Improved 3D Lighting Environment Estimation for Image Forgery Detection

Bo Peng, Wei Wang, Jing Dong and Tieniu Tan
CRIPAC, NLPR, Institute of Automation,
Chinese Academy of Sciences

2015.11.18 Roma Tre University

CRIPAC
智能感知与计算研究中心
Center for Research on Intelligent Perception and Computing

模式识别国家重点实验室
National Laboratory of Pattern Recognition

中国科学院自动化研究所
INSTITUTE OF AUTOMATION CHINESE ACADEMY OF SCIENCES

Outline

- **Introduction & Motivation**
- Methods
  - Reflection Model
  - 3D Face Fitting
  - Lighting Coefficients Estimation
- Experiments & Conclusion
  - Datasets
  - Estimation Accuracy
  - Splicing Detection Efficacy
  - Conclusions

Introduction – Image Forensics

- Pixel based
  - Copy move, resampling, steep edge…
- Format based
  - JPEG quantization, double JPEG…
- Camera based
  - Chromatic aberration, CFA, sensor noise…
- Scene based
  - Illumination color, geometric constraints, lighting direction …

Introduction – Lighting Direction

- An effective kind of forensic method robust for low resolution and low quality images.
- Objects from different images are usually in different lighting conditions.

Previous Work

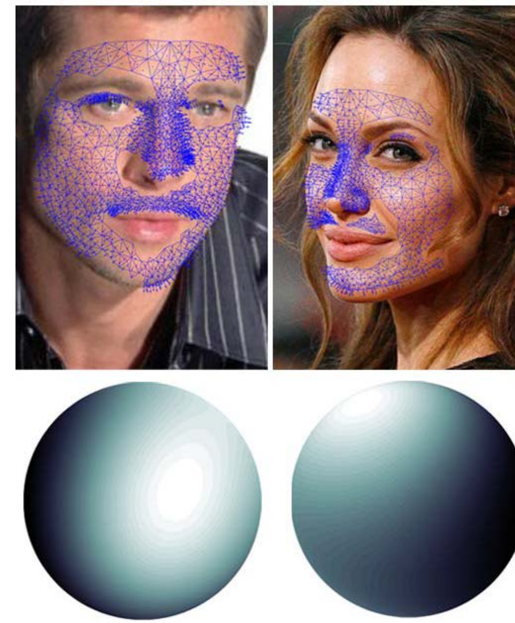
(Johnson and Farid 2005)

2D, single direction

(Johnson and Farid 2007)

2D, complex lighting environment

(Kee and Farid 2010)



3D, complex lighting environment

Motivation

- Previous work's assumptions:
  - Known 3D geometry
  - Distant lighting
  - Lambertian reflection
  - Linear camera response
  - Convex surface
  - Untextured object

Shadows

Facial hair, Pimples...

- Human faces are ***non-convex*** and ***textured*** !
- The relaxation of the two assumptions is more applicable and will lead to improved efficacy.

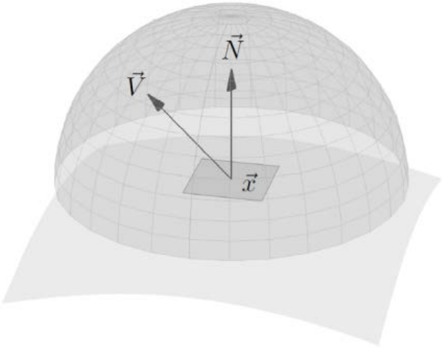
Outline

- Introduction & Motivation
- **Methods**
  - Reflection Model
  - 3D Face Fitting
  - Lighting Coefficients Estimation
- Experiments & Conclusion
  - Datasets
  - Estimation Accuracy
  - Splicing Detection Efficacy
  - Conclusions

Methods – Reflection Model

- Previous model (Kee and Farid 2010):

$$I(\vec{x}) = \rho \int_{\Omega} R(\vec{V}, \vec{N}(\vec{X}))\, L(\vec{V})\, d\vec{V}$$



  - $L(\vec{V})$: spherical lighting function , distant light assumption
  - ρ: constant albedo, **untextured** assumption
  - $R\left(\vec{V}, \vec{N}(\vec{X})\right) = \max(\cos(\theta), 0)$: Lambertian & **convex** assumption

Methods － Reflection Model

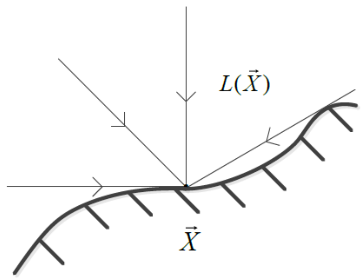
- **Our model**:

RELAXATION !

$$I(\vec{x}) = \int_{\Omega} \rho(\vec{X})G(\vec{X},\vec{V})R(\vec{V},\vec{N}(\vec{X}))L(\vec{V})\,d\vec{V}$$

Texture term

Occlusion term



$\rho(\vec{X})$: spatially varying albedo. Texture!

$G(\vec{X},\ \vec{V})$: spherical mask function indicating the self-occlusion. Non-convex!

Define $A(\vec{X},\vec{V}) = \rho(\vec{X})G(\vec{X},\vec{V})R(\vec{V},\vec{N}(\vec{X}))$ as the transfer function

We have: $I(\vec{x}) = \int_{\Omega} A(\vec{X},\vec{V})L(\vec{V})\,d\vec{V}$

How to get

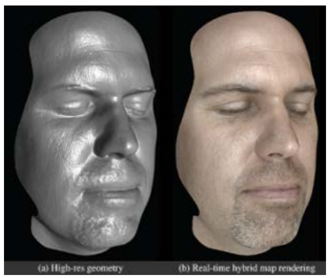
the *TEXTURE* and *OCCLUSION* information?

Outline

- Introduction & Motivation
- **Methods**
  - Reflection Model
  - 3D Face Fitting
  - Lighting Coefficients Estimation
- Experiments & Conclusion
  - Datasets
  - Estimation Accuracy
  - Splicing Detection Efficacy
  - Conclusion

Methods – 3D Face Fitting

- 3D face shape & texture
  - Face scanning: access to involved person (*not practical*)



(Images from the Internet)

  - Face fitting: FaceGen

Two material images to get accurate shape;
Uniform lighting in material images to get accurate texture map;

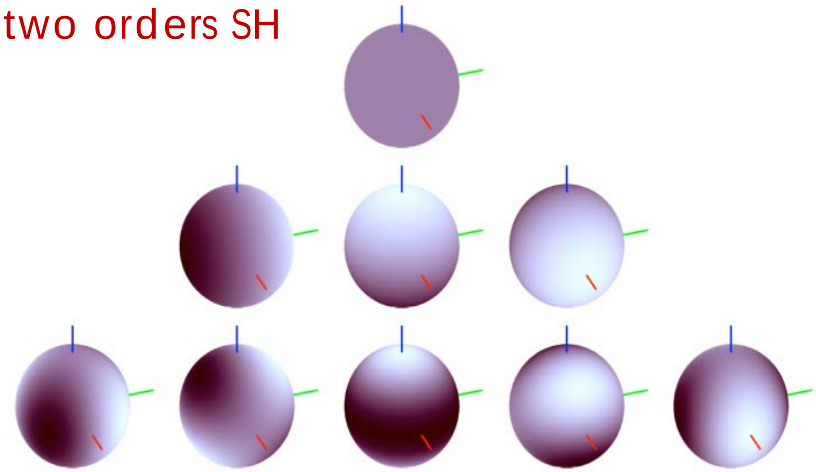
Outline

- Introduction & Motivation
- **Methods**
  - Reflection Model
  - 3D Face Fitting
  - Lighting Coefficients Estimation
- Experiments & Conclusion
  - Datasets
  - Estimation Accuracy
  - Splicing Detection Efficacy
  - Conclusions

Methods – Spherical Harmonics (SH)

- $Y_{n,m}(\vec{V})$: A set of orthogonal basis functions on the spherical surface

First two orders SH



(Johnson and Farid 2007)

$$Y_{0,0} = \frac{1}{\sqrt{4\pi}} \qquad Y_{1,0} = \sqrt{\frac{3}{4\pi}}z$$

$$Y_{1,1} = \sqrt{\frac{3}{4\pi}}x \qquad Y_{1,-1} = \sqrt{\frac{3}{4\pi}}y$$

$$Y_{2,0} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}(3z^2 - 1) \qquad Y_{2,1} = 3\sqrt{\frac{5}{12\pi}}xz$$

$$Y_{2,-1} = 3\sqrt{\frac{5}{12\pi}}yz \qquad Y_{2,2} = \frac{3}{2}\sqrt{\frac{5}{12\pi}}(x^2 - y^2)$$

$$Y_{2,-2} = 3\sqrt{\frac{5}{12\pi}}xy$$

Methods － SH Representation

- Representing $L(\vec{V})$ and $A(\vec{X},\vec{V})$ using SH coefficients

$$L(\vec{V}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} l_{n,m} Y_{n,m}(\vec{V}) \qquad A(\vec{X},\vec{V}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} a_{n,m}(\vec{X}) Y_{n,m}(\vec{V})$$

↓ Lighting Coefficients ↓ Transfer Coefficients

- Image intensity: Integration to inner product

$$I(\vec{x}) = \int_{\Omega} A(\vec{X},\vec{V}) L(\vec{V})\, d\vec{V}$$

$$I(\vec{x}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} l_{n,m}\, a_{n,m}(\vec{X}) = \vec{l}^{\,T} \cdot \vec{a}(\vec{X})$$

Spatial Domain

⇓

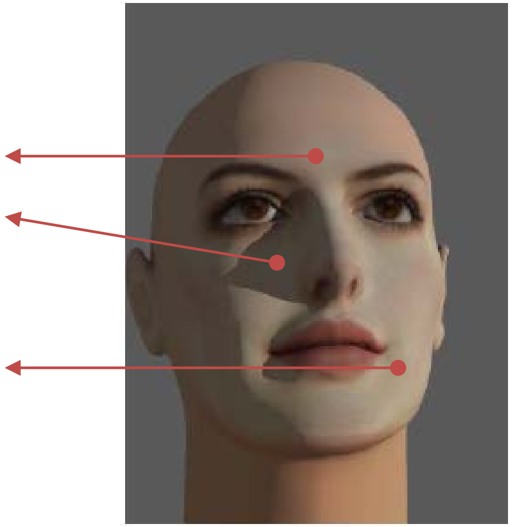
Frequency Domain

Methods - Lighting Coefficients Estimation

- Least Square Error Estimation

$$I(\vec{x}) = \vec{l}^{\,T} \cdot \vec{a}(\vec{X})$$

$$\begin{bmatrix} a_{0,0}(\vec{X}_1) & a_{1,-1}(\vec{X}_1) & \cdots & a_{2,2}(\vec{X}_1) \\ a_{0,0}(\vec{X}_2) & a_{1,-1}(\vec{X}_2) & \cdots & a_{2,2}(\vec{X}_2) \\ \vdots & \vdots & \ddots & \vdots \\ a_{0,0}(\vec{X}_q) & a_{1,-1}(\vec{X}_q) & \cdots & a_{2,2}(\vec{X}_q) \end{bmatrix} \begin{bmatrix} l_{0,0} \\ l_{1,-1} \\ \vdots \\ l_{2,2} \end{bmatrix} = \begin{bmatrix} I(\vec{x}_1) \\ I(\vec{x}_2) \\ \vdots \\ I(\vec{x}_q) \end{bmatrix}$$



Solving: $A\vec{l} = \vec{b}$

$$\vec{l} = (A^T A)^{-1} A^T \vec{b}$$

Two Problems

- How to get the ***correspondence*** between $\vec{x}$ and $\vec{X}$ **?**
- How to get the ***transfer coeff*** $\vec{a}(\vec{X})$ at each point **?**

correspondence

$$\begin{bmatrix} a_{0,0}(\vec{X}_1) & a_{1,-1}(\vec{X}_1) & \cdots & a_{2,2}(\vec{X}_1) \\ a_{0,0}(\vec{X}_2) & a_{1,-1}(\vec{X}_2) & \cdots & a_{2,2}(\vec{X}_2) \\ \vdots & \vdots & \ddots & \vdots \\ a_{0,0}(\vec{X}_q) & a_{1,-1}(\vec{X}_q) & \cdots & a_{2,2}(\vec{X}_q) \end{bmatrix} \begin{bmatrix} l_{0,0} \\ l_{1,-1} \\ \vdots \\ l_{2,2} \end{bmatrix} = \begin{bmatrix} I(\vec{x}_1) \\ I(\vec{x}_2) \\ \vdots \\ I(\vec{x}_q) \end{bmatrix}$$

transfer coeff

Methods – Correspondence

- 3D Face Alignment : we minimize the distance between the detected 2D facial landmarks and the projected 3D ones

Alignment Error:

$$E(R,\vec{t}) = \sum_{i=1}^{N} \|\hat{\vec{x}}_i - K(R\,|\,\vec{t}\,)\vec{X}_i\,\|$$

23 detected facial landmarks and the alignment result

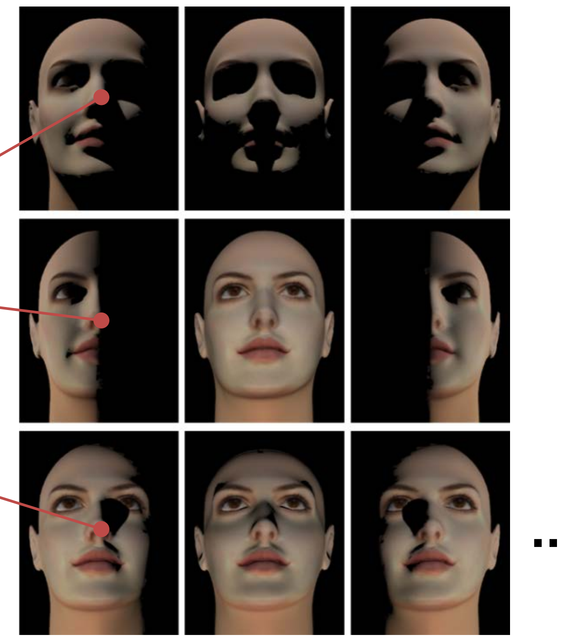
- Solving: Levenberg-Marquardt algorithm

Methods - Transfer Coefficients Fitting

- Render the fitted 3D model under many (42) known distant lighting directions.

$$I(\vec{x}) = \vec{l}^{\,T} \cdot \vec{a}(\vec{X})$$

$$\begin{bmatrix} l_{0,0}^{1} & l_{1,-1}^{1} & \cdots & l_{2,2}^{1} \\ l_{0,0}^{2} & l_{1,-1}^{2} & \cdots & l_{2,2}^{2} \\ \vdots & \vdots & \ddots & \vdots \\ l_{0,0}^{p} & l_{1,-1}^{p} & \cdots & l_{2,2}^{p} \end{bmatrix} \begin{bmatrix} a_{0,0}(\vec{X}) \\ a_{1,-1}(\vec{X}) \\ \vdots \\ a_{2,2}(\vec{X}) \end{bmatrix} = \begin{bmatrix} I^{1}(\vec{x}) \\ I^{2}(\vec{x}) \\ \vdots \\ I^{p}(\vec{x}) \end{bmatrix}$$



Solving: $L\vec{a}\,(\vec{X}) = \vec{b}(\vec{x})$

$$\vec{a}\,(\vec{X}) = (L^{T}L)^{-1}L^{T}\vec{b}(\vec{x})$$

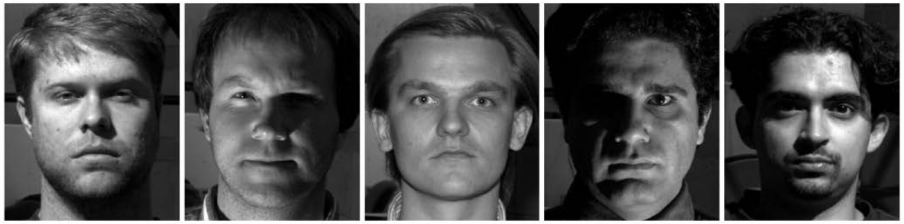
Outline

- Introduction & Motivation
- Methods
  - Reflection Model
  - 3D Face Fitting
  - Lighting Coefficients Estimation
- **Experiments & Conclusion**
  - Datasets
  - Estimation Accuracy
  - Splicing Detection Efficacy
  - Conclusions

Experiments - Datasets

- Synthetic dataset
  - 500 images, random pose, random lighting directions, 1 individual
- Yale B sub-dataset
  - 490 images, 1 frontal pose, 49 lighting directions, 10 individuals

(a) Synthetic dataset



(b) Yale Face Database B

Experiments - Estimation Error Distribution

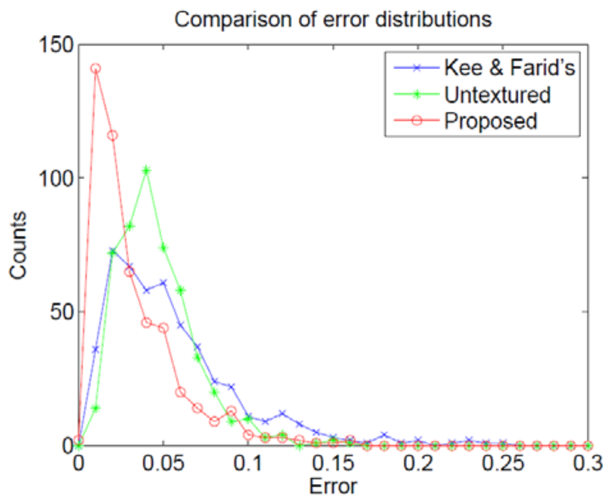
- Distance measurement (Johnson and Farid 2007)

$$D(\vec{l}_1, \vec{l}_2) = \frac{1}{2}(1 - corr(\vec{l}_1, \vec{l}_2))$$

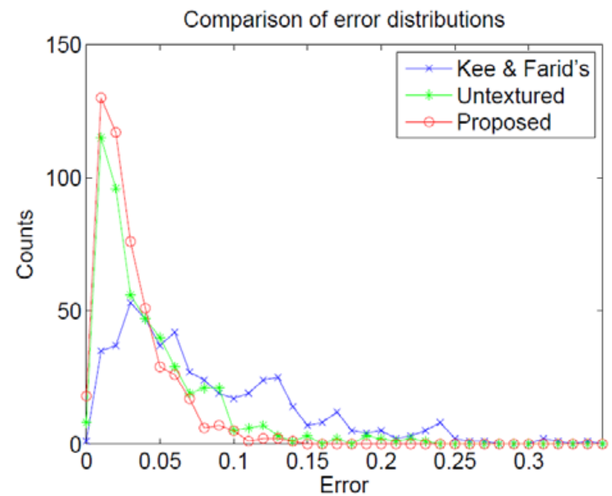
$$corr(\vec{l}_1, \vec{l}_2) = \frac{\vec{l}_1^{\,T} Q \vec{l}_2}{\sqrt{\vec{l}_1^{\,T} Q \vec{l}_1}\sqrt{\vec{l}_2^{\,T} Q \vec{l}_2}}$$

Errata !

- Geometry (occlusion) and texture information can progressively improve estimation accuracy.



(a) Synthetic dataset



(b) Yale Face Database B

Experiments - Different Individuals

- Our method constantly outperforms previous method for all individuals and is more stable.

TABLE I. ESTIMATION ACCURACY ON THE YALE FACE DATABASE B

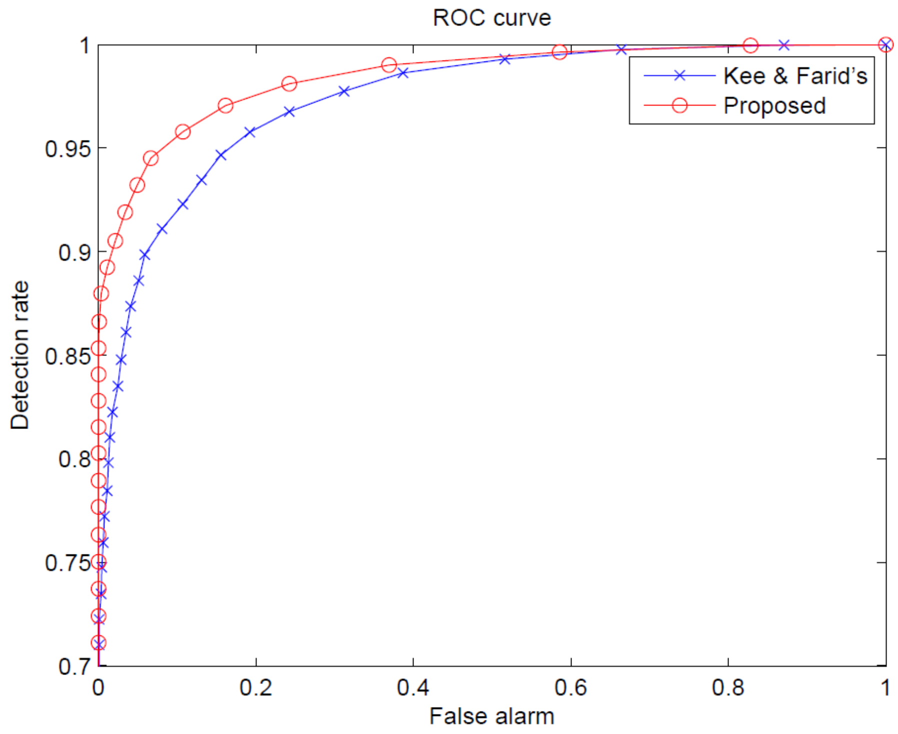
	ID 1	ID 2	ID 3	ID 4	ID 5
Kee & Farid's	0.121	0.045	0.068	0.083	0.100
Proposed	**0.040**	**0.028**	**0.022**	**0.014**	**0.040**

	ID 6	ID 7	ID 8	ID 9	ID 10
Kee & Farid's	0.061	0.074	0.091	0.127	0.040
Proposed	**0.021**	**0.020**	**0.051**	**0.035**	**0.023**

ID1 & ID9 have relatively heavy facial hair. Previous method does not incorporate facial texture.

Experiments – Splicing Detection Efficacy

- All possible pairs in YaleB are "virtually" spliced together. Images taken under the same lighting direction are treated as "pristine". Those taken under different lightings are treated as "spliced".
- At a false alarm rate of around 1%, the detection rate of Kee & Farid's is 78.5%, while ours achieves 89.2%, achieving **an improvement of more than 10% !**



Experiments – A Splicing Example

- Using the threshold at 1% false alarm rate, our method can detect more subtle inconsistency.

Conclusions

- The relaxation of the convexity and constant reflectance assumptions are more applicable to human faces, and it can get improved forgery detection efficacy
- The more information we have, the more reliable forensic determination we will get. (in this case, the non-convex shape & facial texture)
- More assumption relaxations, e.g. distant light and Lambertian reflection, may further benefit the lighting based forensic techniques.

Thanks! *Q & A*

E-mail: bo.peng@nlpr.ia.ac.cn