Exposing Digital Forgeries From 3-D Lighting Environments

Eric Kee [1], Hany Farid [2]

Department of Computer Science, Dartmouth College
Hanover NH 03755 (USA)
[1] kee@cs.dartmouth.edu
[2] farid@cs.dartmouth.edu

Abstract—When creating a photographic composite, it can be difficult to match lighting conditions. We describe a technique for measuring lighting conditions in an image, and describe its use in detecting photographic composites. Specifically, we describe how to approximate a 3-D lighting environment with a low-dimensional model and how to estimate the model's parameters from a single image. Inconsistencies in the lighting model are then used as evidence of tampering.

I. INTRODUCTION

When creating a photographic composite from multiple images, it is often difficult to exactly match the lighting conditions. In addition, it can be difficult to visually judge inconsistencies in lighting and shadows in a photograph [1]. For example, in an attempt to quell rumors regarding the health of North Korea's leader Kim Jong-Il, the North Korean government released a series of photographs in the Spring of 2008 showing a healthy and active Kim Jong-Il, Fig. 1. Shortly after their release, the *BBC* reported that seemingly incongruous shadows were evidence that the photograph was doctored (the width of the shadows cast by the legs of the men in the front row seems to be inconsistent with a single light source). As a result of this claim, the health of Kim Jong-Il was cast into further doubt.

We have previously shown how inconsistencies in lighting can be estimated and used to detect photo tampering [2], [3], [4]. In [2] it was assumed that the amount of light striking a surface is proportional to the surface normal and the direction to a single light source. With knowledge of 3-D surface normals, the direction to the light source can be estimated. Because 3-D surface normals usually cannot be determined from a single image, only the 2-D surface normals at occluding boundaries were considered. In return, only two of the three components of the light source direction were estimated. In [3], we described how to estimate the 3-D direction to a light source from the light's reflection in the human eye. The required 3-D surface normals were determined by leveraging a 3-D model of the human eye.

In these two earlier works, a simplified lighting model consisting of a single dominant light source was assumed.

Fig. 1. A photograph of Kim Jong-Il released by North Korea to quell rumors of his death. The *BBC* claimed that purported inconsistent shadows were evidence of photo tampering (the width of the shadows cast by the legs are different).

In practice, however, the lighting of a scene can be complex: any number of lights can be placed in any number of positions, creating different lighting environments. In [4] we described how to model such complex lighting environments with a nine-parameter spherical harmonic model. Because 3-D surface normals usually cannot be determined from a single image, we considered the 2-D surface normals at occluding boundaries, from which only five of the nine model parameters could be estimated.

Here we describe how to estimate the full 3-D lighting environment in images of people. In order to extract the required 3-D surface normals, we fit 3-D models to an image of a person's head and automatically align this model to an arbitrary head pose. We describe how to model and estimate lighting environments using this approach and show its efficacy in detecting photographic composites. This 3-D approach removes the ambiguities in the earlier 2-D lighting techniques, and hence allows for a more powerful forensic analysis.

II. METHODS

A. Lighting Environment

We briefly review the modeling of complex lighting environments as previously described in [4]. An arbitrary lighting environment can be expressed as a non-negative function on the sphere, $L(\vec{V})$, specifying the intensity of the incident light along the unit vector direction $\vec{V}$. The irradiance, $E(\vec{N})$, parametrized by the unit length surface normal $\vec{N}$, is written as a convolution of the reflectance function of the surface, $R(\vec{V}, \vec{N})$, with the lighting environment $L(\vec{V})$:

$$E(\vec{N}) \;=\; \int_\Omega L(\vec{V})\, R(\vec{V}, \vec{N})\, d\Omega, \qquad (1)$$

where Ω represents the surface of the sphere. This convolution can be simplified by expressing both the lighting environment and the reflectance function in terms of spherical harmonics to yield:

$$E(\vec{N}) \;=\; \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \hat{r}_n l_{n,m} Y_{n,m}(\vec{N}), \qquad (2)$$

where $\hat{r}_n$ is the reflectance coefficient, and $l_{n,m}$ is the lighting coefficient corresponding to the mth spherical harmonic of order n, $Y_{n,m}(\cdot)$. The key observation in [8] and [9] was that the coefficients $\hat{r}_n$ for a Lambertian reflectance function decay rapidly, and thus the infinite sum in Equation (2) can be well approximated by the first nine terms:

$$E(\vec{N}) \;\approx\; \sum_{n=0}^{2} \sum_{m=-n}^{n} \hat{r}_n l_{n,m} Y_{n,m}(\vec{N}). \qquad (3)$$

Since the constants $\hat{r}_n$ are known for a Lambertian reflectance function, the irradiance of a convex Lambertian surface under arbitrary distant lighting can be well modeled by the nine lighting environment coefficients $l_{n,m}$ up to order $n = 2$.

We assume, as in [4], a linear relationship between image intensity, $I(\cdot)$, and irradiance. As such, image intensity can be written in terms of spherical harmonics by expanding Equation (3):

$$
\begin{aligned}
I(\vec{x}) \;=\;& l_{0,0}\pi Y_{0,0}(\vec{N}(\vec{x})) + l_{1,-1}\tfrac{2\pi}{3} Y_{1,-1}(\vec{N}(\vec{x})) \\
+\;& l_{1,0}\tfrac{2\pi}{3} Y_{1,0}(\vec{N}(\vec{x})) + l_{1,1}\tfrac{2\pi}{3} Y_{1,1}(\vec{N}(\vec{x})) \\
+\;& l_{2,-2}\tfrac{\pi}{4} Y_{2,-2}(\vec{N}(\vec{x})) + l_{2,-1}\tfrac{\pi}{4} Y_{2,-1}(\vec{N}(\vec{x})) \\
+\;& l_{2,0}\tfrac{\pi}{4} Y_{2,0}(\vec{N}(\vec{x})) + l_{2,1}\tfrac{\pi}{4} Y_{2,1}(\vec{N}(\vec{x})) \\
+\;& l_{2,2}\tfrac{\pi}{4} Y_{2,2}(\vec{N}(\vec{x})). \qquad (4)
\end{aligned}
$$

This expression is linear in the nine lighting environment coefficients, $l_{0,0}$ to $l_{2,2}$. Given 3-D surface normals at $p \geq 9$ points on the surface of an object, the lighting environment coefficients can be estimated as the least-squares solution to the following system of linear equations:

$$
\begin{pmatrix}
\pi Y_{0,0}(\vec{N}(\vec{x}_1)) & \cdots & \tfrac{\pi}{4} Y_{2,2}(\vec{N}(\vec{x}_1)) \\
\pi Y_{0,0}(\vec{N}(\vec{x}_2)) & \cdots & \tfrac{\pi}{4} Y_{2,2}(\vec{N}(\vec{x}_2)) \\
\vdots & \ddots & \vdots \\
\pi Y_{0,0}(\vec{N}(\vec{x}_p)) & \cdots & \tfrac{\pi}{4} Y_{2,2}(\vec{N}(\vec{x}_p))
\end{pmatrix}
\begin{pmatrix} l_{0,0} \\ \vdots \\ l_{2,2} \end{pmatrix}
=
\begin{pmatrix} I(\vec{x}_1) \\ I(\vec{x}_2) \\ \vdots \\ I(\vec{x}_p) \end{pmatrix}
$$

$$M\vec{v} \;=\; \vec{b}, \qquad (5)$$

where M is the matrix containing the sampled spherical harmonics, $\vec{v}$ is the vector of unknown lighting environment coefficients, and $\vec{b}$ is the vector of intensities at p points. The least-squares solution to this system is:

$$\vec{v} \;=\; \left(M^T M\right)^{-1} M^T \vec{b}. \qquad (6)$$

This solution requires 3-D surface normals from at least nine points on the object surface. This requirement is easily satisfied from the 3-D models described next.

B. 3-D Model Estimation

In [10], the authors describe a 3-D morphable model for the analysis and synthesis of human faces. The model was derived by collecting a set of 3-D laser scanned faces and projecting them into a lower-dimensional linear subspace. New faces (geometry, texture/color, and expressions) are modeled as linear combinations of the resulting linear basis.

The 3-D model parameters can be estimated from a paired frontal and profile view or from only a single frontal view. This estimation requires the manual selection of several fiducial points on the face (11 on the frontal view and 9 on the profile view), from which the 3-D model is then automatically estimated. Shown in Fig. 2(a)-(b), for example, is a frontal and profile view with the selected fiducial points. The estimated model is shown in Fig. 2(d)-(f), which can be seen to be in good agreement with the original head poses shown in panels (a)-(c).

C. 3-D Model Registration

Once estimated, the 3-D model is registered to the face being analyzed. This is done by maximizing an objective function over the camera intrinsic and extrinsic parameters that aligns the 3-D model to the image of the face. Specifically, we seek the rotation matrix R, translation vector $\vec{t}$, focal length
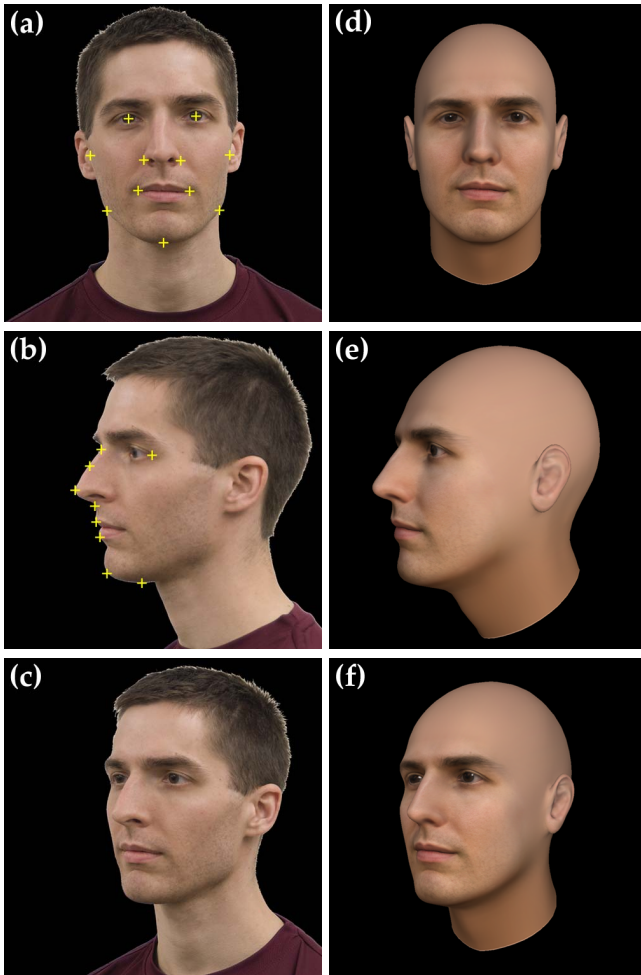
Fig. 2. Shown in panels (a) and (b) are a frontal and profile view used to estimate a 3-D head model (the +'s denote the manually selected fiducial points). The estimated model is shown in panels (d)-(f) with head poses corresponding to panels (a)-(c).

f and camera center (c_x, c_y) that maximizes the correlation between the image $I(\cdot)$ and the rendered model $I_m(\cdot)$:

$$E(R, \vec{t}, f, c_x, c_y) \;\; = \;\; I(x, y) * I_m(x, y), \qquad (7)$$

where $*$ denotes correlation, the spatial coordinates of the rendered model $I_m(\cdot)$ $(x = x_s/s$ and $y = y_s/s)$ are given by:

$$\begin{pmatrix} x_s \\ y_s \\ s \end{pmatrix} \;\; = \;\; \begin{pmatrix} f & 0 & c_x \\ 0 & f & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & \vec{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}, \qquad (8)$$

and (X, Y, Z) are the 3-D coordinates of the face model.

The error function in Equation (7) is maximized as follows. The 3-D model is first manually rotated to approximately align it with the image $I(\cdot)$. At least three corresponding points are selected on the model and image (e.g., the center of each eye and base of the nose), from which the optimal translation is estimated using standard least-squares. With this initial alignment as a starting configuration, a brute force search that maximizes Equation (7) is performed over the three rotation parameters, focal length, and camera center. On each iteration of this search, the translation vector is estimated as described above. In order to reduce the effects of lighting, a high-pass filter is applied to both the image $I(\cdot)$ and rendered model $I_m(\cdot)$ prior to computing the correlation in Equation (7).

Once the model has been estimated and registered, 3-D surface normals and corresponding intensities are used to estimate the lighting environment.

III. Results

Shown in the first column of Fig. 3 are nine images of a synthetically generated head (Jen). A 3-D model of Jen was estimated from a single rendered frontal view.[1] This 3-D model was then registered to each of the images shown in Fig. 3, from which the 3-D lighting environment was estimated.[2] Regions around the eyes, nose, and mouth were manually removed as they do not comply with the constant reflectance assumption. Shown in the second and third columns of Fig. 3 are spheres rendered with the estimated and actual lighting environment. In addition to the good qualitative agreement, we computed the difference between the estimated and actual lighting, as described in [4]. The average error is 0.045 with a standard deviation of 0.023. The error is bounded in the range $[0, 1]$, with an error of 0.0 denoting a perfect match. By way of comparison, the average pairwise difference from 1,000 randomly generated lighting environments is 0.5 with a standard deviation of 0.26. A lighting difference less than 0.045 occurs randomly with a probability of only 1%.

Shown in Fig. 6(a) are frontal and side views of Brad Pitt and Angelina Jolie and the 3-D models generated from these images. Shown in the first row of Fig. 6(b) is a composite photo. Shown in the second row are the 3-D models registered onto the image, and shown in the third row are spheres rendered with the estimated lighting environment. Note that the spheres appear very different: the difference between the lighting environments is 0.64. Shown in Fig. 6(c) is an authentic photo. In this case the rendered spheres appear very similar, and the difference between them is 0.01, considerably less than for the composite photo.

Shown in the top panel of Fig. 4 is a portion of the photograph of North Korea's leader Kim Jong-Il (see also Fig. 1). To determine if the shadows are inconsistent, as reported by the *BBC*, we estimated the lighting environment for Kim Jong-Il and the six men surrounding him. Shown in the third panel of Fig. 4 are three views of the 3-D model of Kim Jong-Il, and in the fourth panel is this 3-D model superimposed onto his face. The 3-D model for Kim Jong-Il was estimated from an image of him without sunglasses (not shown here). The 3-D models for the remaining men were estimated directly from the image in Fig. 4. Shown in the second panel of Fig. 4 are the estimated lighting environments

---

[1]3-D models estimated from a single frontal view, or a frontal and profile view produced similar results.

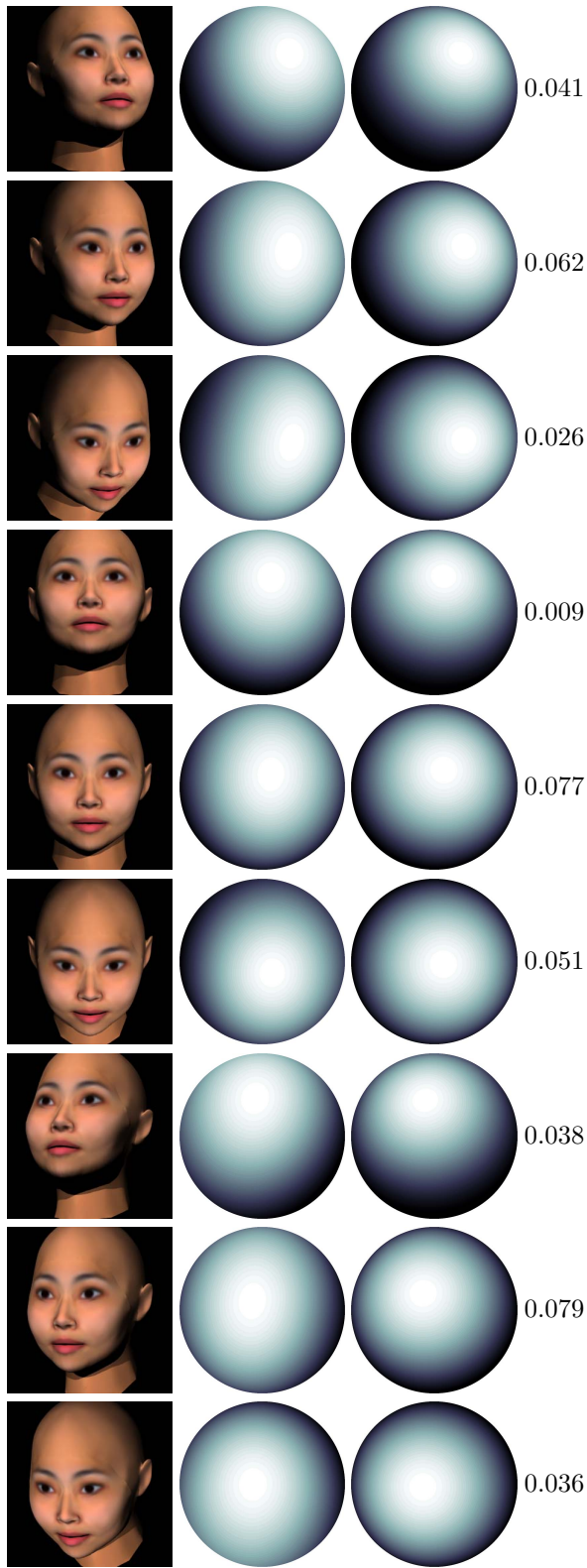[2]Lighting environments were estimated from gray-scale versions of each color image.

Fig. 3. Shown in the left column are nine views of "Jen". Shown in the second and third columns are spheres rendered with the estimated and actual lighting environments. The numeric values are the difference between the estimated and actual lighting environment (on a scale of [0, 1]).
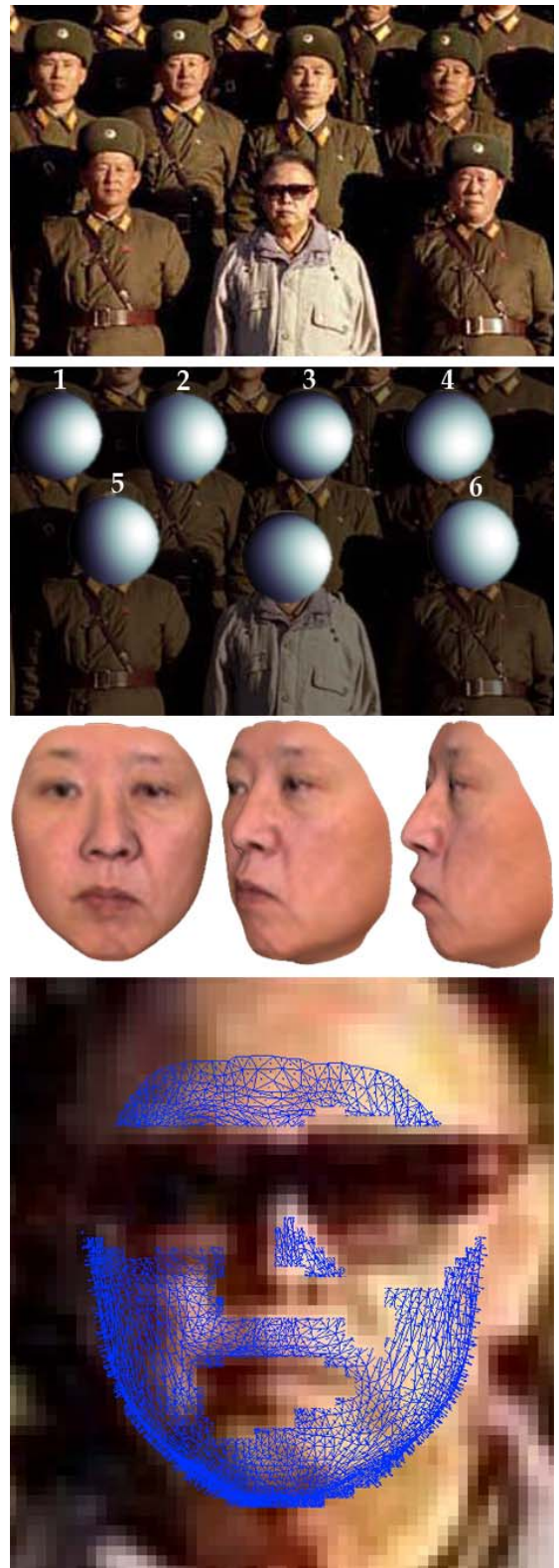

Fig. 4. Shown from top to bottom are, the original Kim Jong-Il photo, the estimated lighting environments for seven men, three views of the 3-D model of Kim Jong-Il, and the 3-D model superimposed onto his face. See also Fig. 5.

person	1	2	3	4	5	6	Kim
1	-	0.019	0.007	0.044	0.021	0.033	0.007
2		-	0.026	0.057	0.009	0.010	0.027
3			-	0.035	0.037	0.032	0.012
4				-	0.073	0.057	0.067
5					-	0.017	0.022
6						-	0.034
Kim							-

Fig. 5. Pairwise lighting differences (on a scale of $[0, 1]$) between each person in Fig. 4.

which can be seen to be in good qualitative agreement with one another. Shown in Fig. 5 are the pairwise differences in lighting between all seven men. Note that all of these differences are on the order of 0.05, similar to the consistent lighting in our earlier simulations, Fig. 3. Had the lighting in this photo been inconsistent we would have found the lighting environment between Kim Jong-Il and those surrounding him to be significantly different – they are not, and we therefore conclude that the lighting in this photo is consistent. We note that each face in this analysis is only on the order of 40×40 pixels in size – even at this low resolution, we are able to accurately estimate the 3-D lighting environment. Note also that it would have been difficult to estimate 2-D normals in this low-resolution image as required in [4].

IV. DISCUSSION

When creating a composite of two or more people, it is often difficult to match the lighting. Lighting environments can be approximated with a nine dimensional model consisting of a linear combination of spherical harmonics. We have shown how to approximate these lighting parameters from a 3-D model of a person's face and head. Lighting inconsistencies across an image are then used as evidence of tampering.

This technique extends earlier 2-D lighting approaches that, due to the lack of 3-D surface normals, were only able to estimate a subset of the full lighting model. This new approach removes the ambiguities in these earlier techniques, and hence allows for a more powerful forensic analysis.

We are extending this technique to automatically fit a 3-D model to an arbitrary head pose (thus obviating the need for a frontal or profile view). We also note that there is no fundamental reason why the proposed technique cannot be extended to arbitrary objects for which a 3-D model can be generated.

ACKNOWLEDGMENTS

REFERENCES

[1] H. Farid and M.J. Bravo, "Image forensic analyses that elude the human visual system," in *SPIE Symposium on Electronic Imaging*, San Jose, CA, 2010.

[2] M.K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *ACM Multimedia and Security Workshop*, New York, NY, 2005.

[3] M.K. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," in *9th International Workshop on Information Hiding*, Saint Malo, France, 2007.

[4] M.K. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 450–461, 2007.

[5] H. Farid, "A survey of image forgery detection," *IEEE Signal Processing Magazine*, vol. 2, no. 26, pp. 16–25, 2009.

[6] C. Riess and E. Angelopulou, "Scene illumination as an indicator of image manipulation," in *12th International Workshop on Information Hiding*, Calgary, Alberta, Canada, 2010.

[7] W. Zhang, X. Cao, J. Zhang, J. Zhu, and P. Wang, "Detecting photographic composites using shadows," in *IEEE International Conference on Multimedia and Expo*, 2009, pp. 1042–1045.

[8] R. Ramamoorthi and P. Hanrahan, "On the relationship between radiance and irradiance: determining the illumination from images of a convex Lambertian object," *Journal of the Optical Society of America A*, vol. 18, pp. 2448–2559, 2001.

[9] R. Basri and D.W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[10] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *SIGGRAPH, Computer Graphics Proceedings*, Los Angeles, 1999, pp. 187–194.
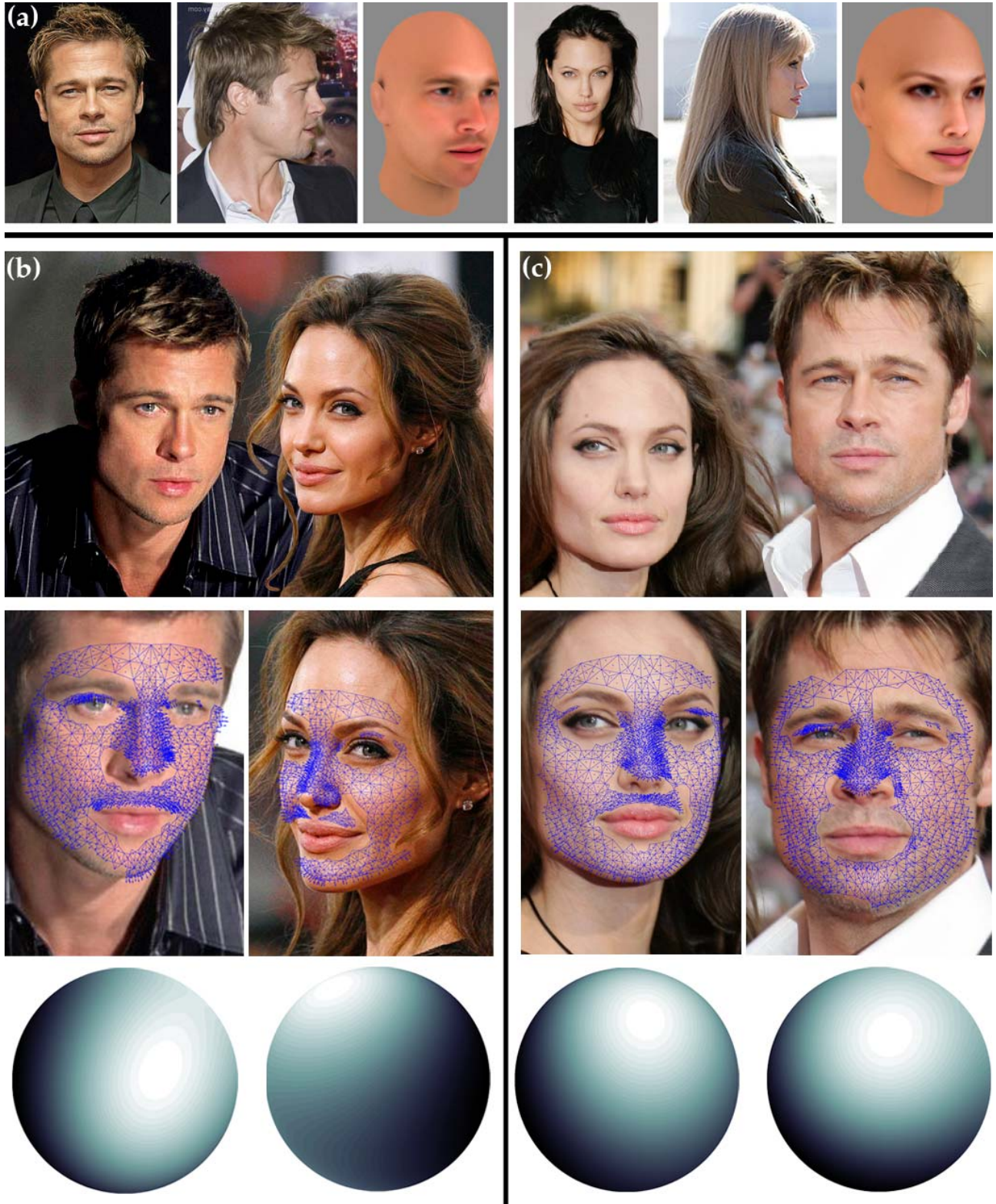
Fig. 6. (a) Frontal and profile views of Brad Pitt and Angelina Jolie and the estimated 3-D models: (b) a composite photo, the estimated 3-D model superimposed onto each face, and a sphere rendered with each estimated lighting environment: and (c) an authentic photo. Note the differences in the rendered sphere for the composite case and the similarity in the authentic case.