Exposing Digital Forgeries Through Specular Highlights on the Eye

Micah K. Johnson and Hany Farid

Department of Computer Science
Dartmouth College
Hanover, NH 03755
{kimo,farid}@cs.dartmouth.edu
www.cs.dartmouth.edu/~{kimo,farid}

Abstract. When creating a digital composite of two people, it is difficult to exactly match the lighting conditions under which each individual was originally photographed. In many situations, the light source in a scene gives rise to a specular highlight on the eyes. We show how the direction to a light source can be estimated from this highlight. Inconsistencies in lighting across an image are then used to reveal traces of digital tampering.

Key words: Digital Tampering, Digital Forensics

1 Introduction

The photograph in Fig. 1 of the host and judges for the popular television show *American Idol* was scheduled for publication when it caught the attention of a photo-editor. Coming on the heels of several scandals that rocked major news organizations, the photo-editor was concerned that the image had been doctored. There was good reason to worry – the image was a composite of several photographs. Shown in Fig. 1 are magnifications of the host's and judge's eyes. The inconsistencies in the shape and position of the specular highlight on the eyes suggest that the people were originally photographed under different lighting conditions. In this work, we show how the location of a specular highlight can be used to determine the direction to the light source. Inconsistencies in the estimates from different eyes, as well as differences in the shape and color of the highlights, can be used to reveal traces of digital tampering.

In related work, the authors of [5] showed how to estimate the light source direction in 2-D. While this approach has the benefit of being applicable to arbitrary objects, it has the drawback that it can only determine the direction to the light source within one degree of ambiguity. In contrast, we estimate the full 3-D light source direction by leveraging a 3-D model of the human eye. Although not specifically developed for a forensic setting, the authors of [7] described a technique for computing an environment map from eyes that embodies the illumination in the scene. While the environment map provides a rich source
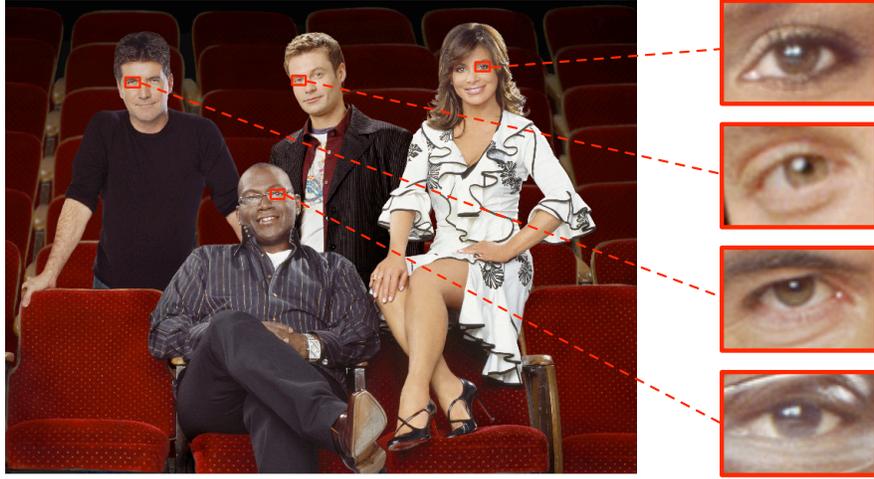
Fig. 1. This photograph of the *American Idol* host and judges is a digital composite of multiple photographs. The inconsistencies in the shape and position of the specular highlight on the eyes suggest that these people were originally photographed under different lighting conditions. Photo courtesy of Fox News and the Associated Press.

of information about the lighting, it has the drawback of requiring a relatively high-resolution image of the eye.

We describe how to estimate the 3-D direction to a light source from specular highlights on the eyes. We show the efficacy of this approach on synthetic and real images and visually plausible forgeries.

2 Methods

The position of a specular highlight is determined by the relative positions of the light source, the reflective surface and the viewer (or camera). In Fig. 2, for example, is a diagram showing the creation of a specular highlight on an eye. In this diagram, the three vectors L, N and R correspond to the direction to the light, the surface normal at the point at which the highlight is formed, and the direction in which the highlight will be seen. For a perfect reflector, the highlight is seen only when the view direction $V = R$. For an imperfect reflector, a specular highlight can be seen for viewing directions V near R, with the strongest highlight seen when $V = R$.

We will first derive an algebraic relationship between the vectors L, N, and V. We then show how the 3-D vectors N and V can be estimated from a single image, from which the direction to the light source L is determined.

The law of reflection states that a light ray reflects off of a surface at an angle of reflection θ_r equal to the angle of incidence θ_i, where these angles are
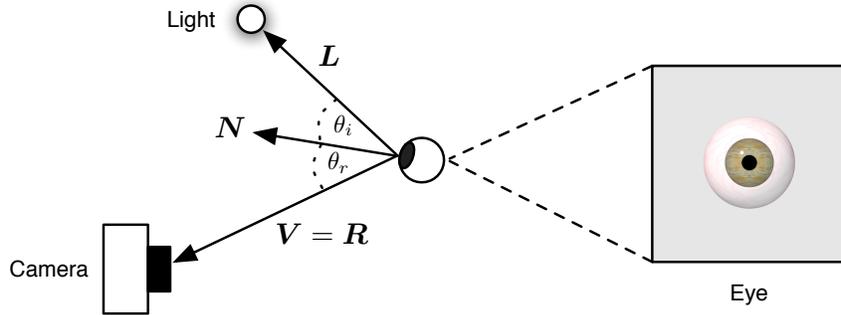
Fig. 2. The formation of a specular highlight on an eye (small white dot on the iris). The position of the highlight is determined by the surface normal N and the relative directions to the light source L and viewer V.

measured with respect to the surface normal N, Fig. 2. Assuming unit-length vectors, the direction of the reflected ray R can be described in terms of the light direction L and the surface normal N:

$$R = L + 2(\cos(\theta_i)N - L)$$
$$= 2\cos(\theta_i)N - L. \tag{1}$$

By assuming a perfect reflector ($V = R$), the above constraint yields:

$$L = 2\cos(\theta_i)N - V$$
$$= 2\left(V^T N\right)N - V. \tag{2}$$

The light direction L can therefore be estimated from the surface normal N and view direction V at a specular highlight. In the following sections, we describe how to estimate these two 3-D vectors from a single image.

Note that the light direction is specified with respect to the eye, and not the camera. In practice, all of these vectors will be placed in a common coordinate system, allowing us to compare light directions across the image.

2.1 Camera Calibration

In order to estimate the surface normal N and view direction V in a common coordinate system, we first need to estimate the projective transform that describes the transformation from world to image coordinates. With only a single image, this calibration is generally an under-constrained problem. In our case, however, the known geometry of the eye can be exploited to estimate this required transform. Throughout, upper-case symbols will denote world coordinates and lower-case will denote camera/image coordinates.

The limbus, the boundary between the sclera (white part of the eye) and the iris (colored part of the eye), can be well modeled as a circle [7]. The image of the limbus, however, will be an ellipse except when the eye is directly facing the

camera. Intuitively, the distortion of the ellipse away from a circle will be related to the pose and position of the eye relative to the camera. We therefore seek the transform that aligns the image of the limbus to a circle.

In general, a projective transform that maps 3-D world coordinates to 2-D image coordinates can be represented, in homogeneous coordinates, as a 3×4 matrix. We assume that points on a limbus are coplanar, and define the world coordinate system such that the limbus lies in the $Z = 0$ plane. With this assumption, the projective transformation reduces to a 3×3 planar projective transform [2], where the world points $\boldsymbol{X}$ and image points $\boldsymbol{x}$ are represented by 2-D homogeneous vectors.

Points on the limbus in our world coordinate system satisfy the following implicit equation of a circle:

$$f(\boldsymbol{X}; \boldsymbol{\alpha}) = (X_1 - C_1)^2 + (X_2 - C_2)^2 - r^2 = 0, \tag{3}$$

where $\boldsymbol{\alpha} = \begin{pmatrix} C_1 & C_2 & r \end{pmatrix}^T$ denotes the circle center and radius.

Consider a collection of points, $\boldsymbol{X_i}$, $i = 1, \ldots, m$, each of which satisfy Equation (3). Under an ideal pinhole camera model, the world point $\boldsymbol{X_i}$ maps to the image point $\boldsymbol{x_i}$ as follows:

$$\boldsymbol{x_i} = H \boldsymbol{X_i}, \tag{4}$$

where H is a 3×3 projective transform matrix.

The estimation of H can be formulated in an orthogonal distance fitting framework. Let $E(\cdot)$ be an error function on the parameter vector $\boldsymbol{\alpha}$ and the unknown projective transform H:

$$E(\boldsymbol{\alpha}, H) = \sum_{i=1}^{m} \min_{\hat{\boldsymbol{X}}} \left\| \boldsymbol{x_i} - H\hat{\boldsymbol{X}} \right\|^2, \tag{5}$$

where $\hat{\boldsymbol{X}}$ is on the circle parametrized by $\boldsymbol{\alpha}$. The error embodies the sum of the squared errors between the data, $\boldsymbol{x_i}$, and the closest point on the model, $\hat{\boldsymbol{X}}$. This error function is minimized using non-linear least squares via the Levenberg-Marquardt iteration [9] (see Appendix A for details).

Once estimated, the projective transform H can be decomposed in terms of intrinsic and extrinsic camera parameters [2]. The intrinsic parameters consist of the camera focal length, camera center, skew and aspect ratio. For simplicity, we will assume that the camera center is the image center, that the skew is 0 and the aspect ratio is 1, leaving only the focal length f. The extrinsic parameters consist of a rotation matrix R and translation vector $\boldsymbol{t}$ that define the transformation between the world and camera coordinate systems. Since the world points lie on a single plane, the projective transform can be decomposed in terms of the intrinsic and extrinsic parameters as:

$$H = \lambda K \begin{pmatrix} \boldsymbol{r_1} & \boldsymbol{r_2} & \boldsymbol{t} \end{pmatrix}, \tag{6}$$

where the 3×3 intrinsic matrix K is:

$$K = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{7}$$

λ is a scale factor, the column vectors $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are the first two columns of the rotation matrix R, and $\boldsymbol{t}$ is the translation vector.

With a known focal length f, and hence a known matrix K, the world to camera coordinate transform $\hat{H}$ can be estimated directly:

$$\frac{1}{\lambda} K^{-1} H = (\boldsymbol{r_1} \quad \boldsymbol{r_2} \quad \boldsymbol{t})$$
$$\hat{H} = (\boldsymbol{r_1} \quad \boldsymbol{r_2} \quad \boldsymbol{t}), \tag{8}$$

where the scale factor λ is chosen so that $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ are unit vectors. The complete rotation matrix is given by:

$$R = (\boldsymbol{r_1} \quad \boldsymbol{r_2} \quad \boldsymbol{r_1} \times \boldsymbol{r_2}), \tag{9}$$

where $\times$ denotes cross product.

If the focal length is unknown, it can be directly estimated as described in Appendix B.

2.2 View Direction

Recall that the minimization of Equation (5) yields both the transform H and the circle parameters $\boldsymbol{\alpha}$ for the limbus. The unit vector from the center of the limbus to the origin of the camera coordinate system is the view direction, $\boldsymbol{v}$. Let $\boldsymbol{X_c} = (C_1 \quad C_2 \quad 1)$ denote the estimated center of a limbus in world coordinates. In the camera coordinate system, this point is given by:

$$\boldsymbol{x_c} = \hat{H} \boldsymbol{X_c}. \tag{10}$$

The view direction, as a unit vector, in the camera coordinate system is then given by:

$$\boldsymbol{v} = -\frac{\boldsymbol{x_c}}{\|\boldsymbol{x_c}\|}, \tag{11}$$

where the negative sign reverses the vector so that it points from the eye to the camera.

2.3 Surface Normal

The 3-D surface normal $\boldsymbol{N}$ at a specular highlight is estimated from a 3-D model of the human eye [6]. The model consists of a pair of spheres as illustrated in Fig. 3(a). The larger sphere, with radius $r_1 = 11.5$ mm, represents the sclera
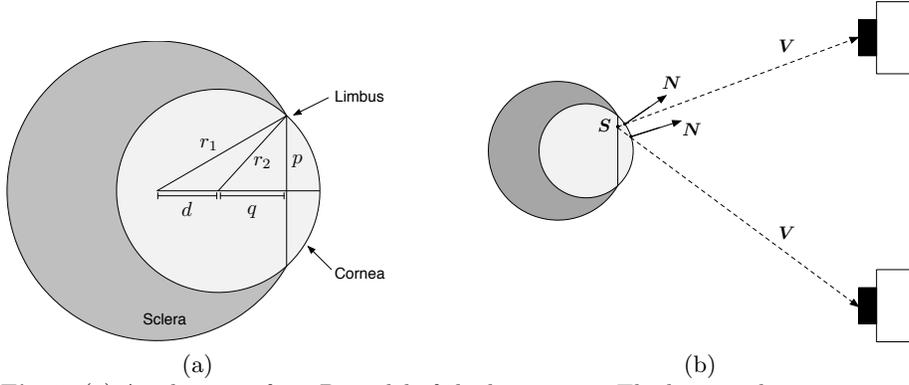
Fig. 3. (a) A side view of a 3-D model of the human eye. The larger sphere represents the sclera and the smaller sphere represents the cornea. The limbus is defined by the intersection of the two spheres. (b) The surface normal at a point S in the plane of the limbus depends on the view direction V.

and the smaller sphere, with radius $r_2 = 7.8$ mm, represents the cornea. The centers of the spheres are displaced by a distance $d = 4.7$ mm. The limbus, a circle with radius $p = 5.8$ mm, is defined by the intersection of the two spheres. The distance between the center of the smaller sphere and the plane containing the limbus is $q = 5.25$ mm. These measurements vary slightly among adults, and the radii of the spheres are approximately 0.1 mm smaller for female eyes [3, 6].

Consider a specular highlight in world coordinates at location $S = (\begin{array}{cc} S_x & S_y \end{array})$, measured with respect to the center of the limbus. The surface normal at S depends on the view direction V. In Fig. 3(b) is a schematic showing this relationship for two different positions of the camera. The surface normal N is determined by intersecting the ray leaving S, along the direction V, with the edge of the sphere. This intersection can be computed by solving a quadratic system for k, the distance between S and the edge of the sphere,

$$(S_x + kV_x)^2 + (S_y + kV_y)^2 + (q + kV_z)^2 = r_2^2$$
$$k^2 + 2(S_x V_x + S_y V_y + qV_z)k + (S_x^2 + S_y^2 + q^2 - r_2^2) = 0, \tag{12}$$

where q and r_2 are specified by the 3-D model of the eye. The view direction $V = (\begin{array}{ccc} V_x & V_y & V_z \end{array})$ in the world coordinate system is given by:

$$V = R^{-1}v, \tag{13}$$

where v is the view direction in camera coordinates, Section 2.2, and R is the estimated rotation between the world and camera coordinate systems, Section 2.1. The surface normal N in the world coordinate system is then given by:

$$N = \begin{pmatrix} S_x + kV_x \\ S_y + kV_y \\ q + kV_z \end{pmatrix}, \tag{14}$$

and in camera coordinates: $n = RN$.

2.4 Light Direction

Consider a specular highlight $\boldsymbol{x_s}$ specified in image coordinates and the estimated projective transform H from world to image coordinates. The inverse transform H^{-1} maps the coordinates of the specular highlight into world coordinates:

$$\boldsymbol{X_s} = H^{-1}\boldsymbol{x_s} \tag{15}$$

The center $\boldsymbol{C}$ and radius r of the limbus in the world coordinate system determine the coordinates of the specular highlight, $\boldsymbol{S}$, with respect to the model:

$$\boldsymbol{S} = \frac{p}{r}\left(\boldsymbol{X_s} - \boldsymbol{C}\right), \tag{16}$$

where p is specified by the 3-D model of the eye. The position of the specular highlight $\boldsymbol{S}$ is then used to determine the surface normal $\boldsymbol{N}$, as described in the previous section. Combined with the estimate of the view direction $\boldsymbol{V}$, Section 2.2, the light source direction $\boldsymbol{L}$ can be estimated from Equation (2). In order to compare light source estimates in the image, the light source estimate is converted to camera coordinates: $\boldsymbol{l} = R\boldsymbol{L}$

3 Results

We tested our technique for estimating the 3-D light source direction on both synthetically generated and real images. In all of these results the direction to the light source was estimated from specular highlights in both eyes. This required a slight modification to the minimization in Equation (5) which is described in Appendix A. The view direction, surface normal and light direction were then estimated separately for each eye.

3.1 Synthetic Images

Synthetic images of eyes were rendered using the `pbrt` environment [8]. The shape of the eyes conformed to the 3-D model described in Section 2.3 and the eyes were placed in one of 12 different locations. For each location, the eyes were rotated by a unique amount relative to the camera. The eyes were illuminated with two light sources: a fixed light directly in line with the camera, and a second light placed in one of four different positions. The twelve locations and four light directions gave rise to 48 images, Fig. 4. Each image was rendered at a resolution of 1200×1600 pixels, with the cornea occupying less than 0.1% of the entire image. Shown in Fig. 4 are several examples of the rendered eyes, along with a schematic of the imaging geometry.

   The limbus and position of the specular highlight(s) were automatically extracted from the rendered image. For each highlight, the projective transform H, the view direction $\boldsymbol{v}$ and surface normal $\boldsymbol{n}$ were estimated, from which the
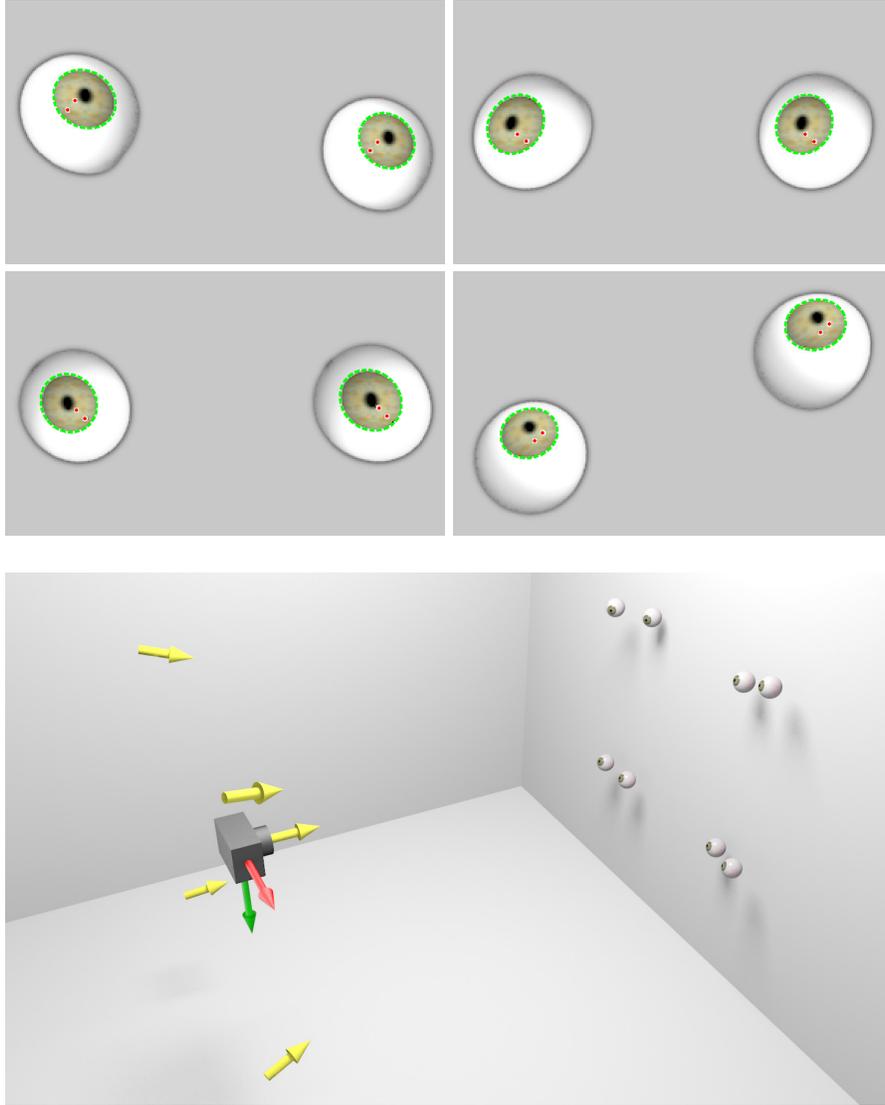
Fig. 4. Synthetically generated eyes. Each of the upper panels corresponds to different positions and orientations of the eyes and locations of the light sources. The ellipse fit to each limbus is shown in dashed green, and the red dots denote the positions of the specular highlights. Shown below is a schematic of the imaging geometry: the position of the lights, camera and a subset of the eye positions.
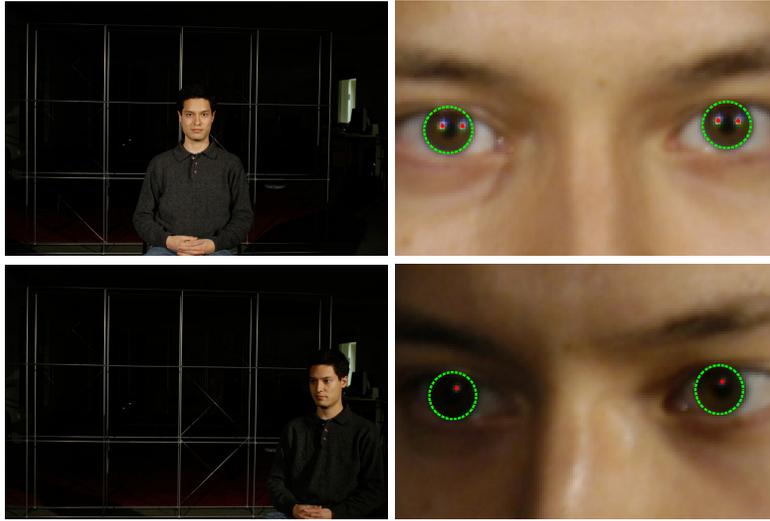
Fig. 5. A subject at different locations and orientations relative to the camera and two light sources. Shown to the right are magnified views of the eyes. The ellipse fit to each limbus is shown in dashed green and the red dots denote the positions of the specular highlights. See also Table 1.

direction to the light source l was determined. The angular error between the estimated l and actual l_0 light directions is computed as:

$$\phi = \cos^{-1}\left(l^T l_0\right). \tag{17}$$

where the vectors are normalized to be unit length.

With a known focal length, the average angular error in estimating the light source direction was 2.8° with a standard deviation of 1.3° and a maximum error of 6.8°. With an unknown focal length, the average error was 2.8° with a standard deviation of 1.3° and a maximum error of 6.3°.

3.2 Real Images

To further test the efficacy of our technique, we photographed a subject under controlled lighting. A camera and two lights were arranged along a wall, and the subject was positioned 250 cm in front of the camera and at the same elevation. The first light L_1 was positioned 130 cm to the left of and 60 cm above the camera. The second light L_2 was positioned 260 cm to the right and 80 cm above the camera. The subject was placed in five different locations and orientations relative to the camera and lights, Fig. 5. A six mega-pixel Nikon D100 camera with a 35 mm lens was set to capture in the highest quality JPEG format.

For each image, an ellipse was manually fit to the limbus of each eye. In these images, the limbus did not form a sharp boundary – the boundary spanned

image	left eye		right eye		left eye		right eye	
	L_1	L_2	L_1	L_2	L_1	L_2	L_1	L_2
1	5.8	7.6	3.8	1.6	5.8	7.7	3.9	1.7
2	–	8.7	–	0.8	–	10.4	–	18.1
3	9.3	–	11.0	–	17.6	–	10.1	–
4	12.5	16.4	7.5	7.3	10.4	13.6	7.4	5.6
5	14.0	–	13.8	–	17.4	–	16.5	–

Table 1. Angular errors (degrees) in estimating the light direction for the images shown in Fig. 5. On the left are the errors for a known focal length, and on the right are the errors for an unknown focal length. A '–' indicates that the specular highlight for that light was not visible on the cornea.

roughly 3 pixels. As such, we fit the ellipses to the better defined inner outline [4], Fig. 5. The radius of each limbus was approximately 9 pixels, and the cornea occupied 0.004% of the entire image.

Each specular highlight was localized by specifying a bounding rectangular area around each highlight and computing the centroid of the selection. The weighting function for the centroid computation was chosen to be the squared (normalized) pixel intensity.

The location to the light source(s) was estimated for each pair of eyes assuming a known and unknown focal length. The angular errors, Equation (17), for each image are given in Table 1. Note that in some cases an estimate for one of the light sources was not possible when the highlight was not visible on the cornea. With a known focal length, the average angular error was 8.6°, and with an unknown focal length, the average angular error was 10.5°.

There are several reasons for the increase in error over the synthetic images. First, the average size of the cornea in our real images is much smaller than the size of the cornea in the synthetic images, 256 pixels^2 versus over 1000 pixels^2. Second, the limbus in an adult human eye is slightly elliptical, being 1 mm wider than it is tall [3], while our model assumes a circular limbus.

Shown in Fig. 1 is a photograph of the host and judges of the television show *American Idol*, and shown in Fig. 6 are the results of estimating the direction to the light source for each person. These estimates are rendered as Gaussian blobs ($\sigma = 15°$) on a hemisphere. The final estimate is depicted as a sum of Gaussians, one for each specular highlight. Note that the estimates in the two right-most plots are visually consistent with one another, but are significantly different from the two left-most estimates.

Shown in Fig. 7 is a composite where the father's face has been replaced with a different face. Two specular highlights are visible on each of the children's eyes. The light direction was estimated from each specularity and for each eye. Across the children's eyes, the average pair-wise difference in orientation for the first specularity was 8.5° with a maximum difference of 11.6°. The average difference for the second specularity was 9.4° with a maximum difference of 13.9°. By comparison, the average difference in orientation between the father's specularities to those of the children was 40.3°. We did not estimate the light
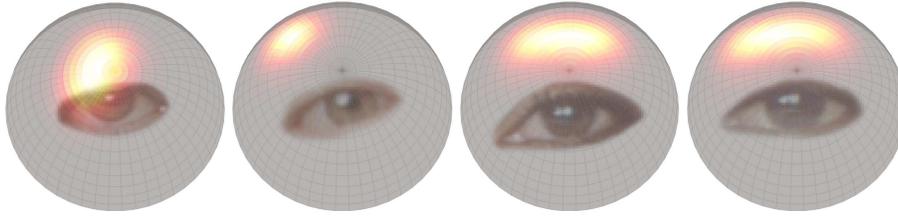
Fig. 6. The estimated light source direction for each person in Fig. 1 is depicted as Gaussian blobs on a hemisphere, each centered about the estimated 3-D direction. Superimposed on each hemisphere is an image of one of the eyes from which the estimates were made. Note that the inconsistencies in the light source direction suggest that the photograph is a composite of at least three photographs.

direction for the woman because we have found that glasses distort the shape and location of the specularity on the eye.

4 Discussion

When creating a composite of two or more people it is often difficult to match the lighting conditions under which each person was originally photographed. Specular highlights that appear on the eye are a powerful cue as to the shape, color and location of the light source(s). Inconsistencies in these properties of the light can be used as evidence of tampering. We have described how to measure the 3-D direction to a light source from the position of the highlight on the eye. While we have not specifically focused on it, the shape and color of a highlight are relatively easy to quantify and measure and should also prove helpful in exposing digital forgeries.

Since specular highlights tend to be relatively small on the eye, it is possible to manipulate them to conceal traces of tampering. To do so, the shape, color and location of the highlight would have to be constructed so as to be globally consistent with the lighting in other parts of the image. Inconsistencies in this lighting may be detectable using the technique described in [5]. Also working in our favor is that even small artifacts on the eyes are visually salient. Nevertheless, as with all forensic tools, it is still possible to circumvent this technique.

We expect this technique, in conjunction with a growing body of forensic tools, to be effective in exposing digital forgeries.
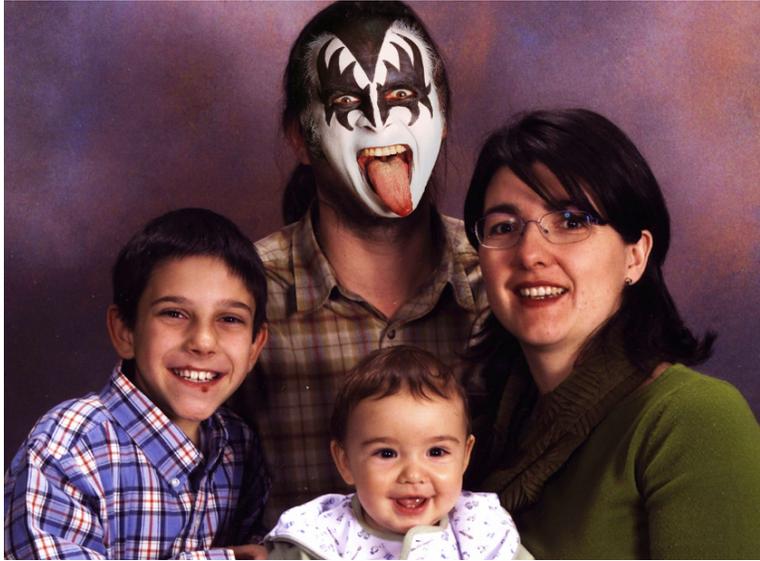
5 Acknowledgments

Fig. 7. A composite where rock star Gene Simmons' face has been inserted into a family portrait.

are those of the author and do not represent the official position or policies of the United States Department of Justice).

References

1. Sung Joon Ahn. *Least Squares Orthogonal Distance Fitting of Curves and Surfaces in Space*, volume 3151 of *Lecture Notes in Computer Science*. Springer, 2004.
2. Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
3. Michael J. Hogan, Jorge A. Alvarado, and Joan Esperson Weddell. *Histology of the Human Eye*. W.B Saunders Company, 1971.
4. D. Robert Iskander. A parametric approach to measuring limbus corneae from digital images. *IEEE Transactions on Biomedical Engineering*, 53(6):1134–1140, June 2006.
5. Micah K. Johnson and Hany Farid. Exposing digital forgeries by detecting inconsistencies in lighting. In *ACM Multimedia and Security Workshop*, New York, NY, 2005.
6. Aaron Lefohn, Richard Caruso, Erik Reinhard, Brian Budge, and Peter Shirley. An ocularist's approach to human iris synthesis. *IEEE Computer Graphics and Applications*, 23(6):70–75, 2003.
7. Ko Nishino and Shree K. Nayar. Eyes for relighting. *ACM Transactions on Graphics*, 23(3):704–711, 2004.
8. Matt Pharr and Greg Humphreys. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann, 2004.
9. Andrzej Ruszczyński. *Nonlinear Optimization*. Princeton University Press, 2006.

Appendix A

In this appendix we describe the minimization of the error function:

$$E(\boldsymbol{\alpha}, H) = \sum_{i=1}^{m} \min_{\hat{\boldsymbol{X}}} \left\| \boldsymbol{x_i} - H\hat{\boldsymbol{X}} \right\|^2, \tag{18}$$

which yields the perspective transform H and circle parameters $\boldsymbol{\alpha}$. For notational convenience, we express this error function as $E(\boldsymbol{u})$ where $\boldsymbol{u} = (\boldsymbol{\alpha} \quad \boldsymbol{h})$, and where the vector $\boldsymbol{h}$ contains the nine elements of the 3×3 matrix H.

This error function is minimized in nested iterations as described in [1]. The inner iteration computes the closest point $\hat{\boldsymbol{X}}$ on the model for each image point $\boldsymbol{x_i}$, where the model is specified by the current state of $\boldsymbol{u}$. The outer iteration then updates the parameter vector $\boldsymbol{u}$ according to the results of the inner model fitting. This process is repeated and terminates when the norm of the update to $\boldsymbol{u}$ is below a specified threshold.

Closest point: For a given point $\boldsymbol{x_i}$ in image coordinates, we seek the closest point $\hat{\boldsymbol{X}}$ on the model. The point $\hat{\boldsymbol{X}}$ that satisfies this condition must, of course, be on the model, $f(\hat{\boldsymbol{X}}; \boldsymbol{\alpha}) = 0$. Recall that $f(\cdot)$ is the equation of a circle, Equation (3). In order to contend with the scale ambiguity inherent to homogeneous coordinates, this model takes on a slightly different form:

$$f(\boldsymbol{X}; \boldsymbol{\alpha}) = (X_1/X_3 - C_1)^2 + (X_2/X_3 - C_2)^2 - r^2 \tag{19}$$

For $\hat{\boldsymbol{X}}$ to be the closest point, it must satisfy two additional criteria. First, the vector between the image point $\boldsymbol{x_i}$ and the model point $H\hat{\boldsymbol{X}}$ (expressed in image coordinates) must be parallel to the gradient of the model in image coordinates, $H^{-T}\nabla f$, yielding the following constraint:

$$\boldsymbol{z}^T((\boldsymbol{x_i} - H\hat{\boldsymbol{X}}) \times H^{-T}\nabla f) = 0, \tag{20}$$

where $\boldsymbol{z}^T = (0 \quad 0 \quad 1)$ restricts this constraint to the image plane. Second, the model point $H\hat{\boldsymbol{X}}$ must lie in the image plane (recall that the homogeneous points $\boldsymbol{x_i}$ lie in the plane $z = 1$):

$$\boldsymbol{z}^T(\boldsymbol{x_i} - H\hat{\boldsymbol{X}}) = 0. \tag{21}$$

These three constraints form a system of non-linear equations that can be solved using the Gauss-Newton method, where the vector-valued function to be minimized is:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{x_i}, \boldsymbol{X}) = \begin{pmatrix} f(\boldsymbol{X}, \boldsymbol{\alpha}) \\ \boldsymbol{z}^T((\boldsymbol{x_i} - H\boldsymbol{X}) \times H^{-T}\nabla f) \\ \boldsymbol{z}^T(\boldsymbol{x_i} - H\boldsymbol{X}) \end{pmatrix}. \tag{22}$$

In practice, the image point $\boldsymbol{x_i}$ is expressed in terms of world coordinates $\boldsymbol{x_i} = H\boldsymbol{X_i}$. This error function is given by:

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{X_i}, \boldsymbol{X}) = \begin{pmatrix} f(\boldsymbol{X}, \boldsymbol{\alpha}) \\ \boldsymbol{z}^T(H(\boldsymbol{X_i} - \boldsymbol{X}) \times H^{-T}\nabla f) \\ \boldsymbol{z}^T H(\boldsymbol{X_i} - \boldsymbol{X}) \end{pmatrix}. \tag{23}$$

This inner iteration is initialized with H equal to the identity matrix, and $\boldsymbol{\alpha}$, the circle parameters, equal to a bounding circle fit to the image data.

Parameter update: Once the inner iteration completes and the closest points $\hat{\boldsymbol{X}}$ have been computed for each image point $\boldsymbol{x_i}$, the parameter vector $\boldsymbol{u}$ can be updated. The outer iteration uses a Levenberg-Marquardt minimization, which requires the derivative of $\boldsymbol{x_i} - H\boldsymbol{X}$ with respect to $\boldsymbol{u}$, evaluated at the closest point $\hat{\boldsymbol{X}}$:

$$\frac{\partial}{\partial \boldsymbol{u}}(\boldsymbol{x_i} - H\boldsymbol{X})\Big|_{\boldsymbol{X}=\hat{\boldsymbol{X}}} = -\left(\frac{\partial H}{\partial \boldsymbol{u}}\right)[\boldsymbol{X}]\Big|_{\boldsymbol{X}=\hat{\boldsymbol{X}}} - H\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{u}}\Big|_{\boldsymbol{X}=\hat{\boldsymbol{X}}}, \tag{24}$$

where $[\boldsymbol{X}]$ is a block-diagonal matrix with $\boldsymbol{X}$ on the diagonal. The derivative $\partial H/\partial \boldsymbol{u}$ is computed by simply differentiating the matrix H with respect to each of its components h_i. The derivative $\partial X/\partial \boldsymbol{u}$ is computed by implicitly differentiating $\boldsymbol{g}(\cdot)$ with respect to $\boldsymbol{u}$:

$$\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}} + \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{X}}\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{u}} + \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{X_i}}\frac{\partial \boldsymbol{X_i}}{\partial \boldsymbol{u}} = 0, \tag{25}$$

and solving for $\partial X/\partial \boldsymbol{u}$:

$$\frac{\partial \boldsymbol{X}}{\partial \boldsymbol{u}} = -\left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{X}}\right)^{-1}\left(\frac{\partial \boldsymbol{g}}{\partial \boldsymbol{u}} + \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{X_i}}\frac{\partial \boldsymbol{X_i}}{\partial \boldsymbol{u}}\right). \tag{26}$$

The individual derivatives in this expression are determined by straight-forward differentiation of each function with respect to its unknowns. The derivatives for all m image points, $\boldsymbol{x_1}$ to $\boldsymbol{x_m}$, are then stacked into a $3m \times 12$ Jacobian matrix, where 12 corresponds to the total number of unknowns (9 elements of H and 3 circle parameters $\boldsymbol{\alpha}$). This Jacobian matrix is used by the Levenberg-Marquardt minimization to compute the update to the parameter vector $\boldsymbol{u}$.

Constraints: The minimization described above can be extended to handle two circles by creating a block-diagonal Jacobian matrix from the Jacobian matrices of the individual eyes. In addition, constraint equations can be added to the error function $E(\boldsymbol{u})$, Equation (18), to ensure that the transform H for both eyes is the same and that the radii of the circles are equal to 5.8 mm. The error function for both eyes with constraints is then given by:

$$\begin{aligned} \hat{E}(\boldsymbol{u_1}, \boldsymbol{u_2}) = &\, E(\boldsymbol{u_1}) + E(\boldsymbol{u_2}) \\ &+ w\left(\|\boldsymbol{h_1} - \boldsymbol{h_2}\|^2 + (\det(H_1) - 1)^2 + (\det(H_2) - 1)^2\right. \\ &\left. + (r_1 - 5.8)^2 + (r_2 - 5.8)^2\right), \end{aligned} \tag{27}$$

where w is a scalar weighting factor. The Jacobian of this system is:

$$J(\boldsymbol{u_1}, \boldsymbol{u_2}) = \begin{pmatrix} J_1(\boldsymbol{u_1}) & \\ & J_2(\boldsymbol{u_2}) \\ \hat{J}_1(\boldsymbol{u_1}) & \hat{J}_2(\boldsymbol{u_2}) \end{pmatrix}, \tag{28}$$

where J_1 and J_2 are the Jacobian matrices from the individual eyes, and $\hat{J}_1$ and $\hat{J}_2$ are the Jacobians of the constraint equations with respect to $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$. The transforms H_1 and H_2 are initially set to the identity matrix, and the circle parameters were chosen to enclose the limbus of each eye.

Appendix B

In this appendix we describe how to decompose the projective transform H in Equation (6) in the case when the focal length f is unknown.

The transform H has eight unknowns: the focal length f, the scale factor λ, the three rotation angles θ_x, θ_y and θ_z for the rotation matrix R, and the three coordinates of the translation vector $\boldsymbol{t}$. By multiplying the matrices on the right-hand side of Equation (6), H can be expressed in terms of these unknowns:

$$H = \lambda \begin{pmatrix} fc_yc_z & fc_ys_z & ft_x \\ f(s_xs_yc_z - c_xs_z) & f(s_xs_ys_z + c_xc_z) & ft_y \\ c_xs_yc_z + s_xs_z & c_xs_ys_z - s_xc_z & t_z \end{pmatrix}, \tag{29}$$

where $c_x = \cos(\theta_x)$, $s_x = \sin(\theta_x)$, etc, and where the rotation matrix follows the "x-y-z" convention.

Consider the upper-left 2×2 sub-matrix of H rewritten in terms of the four unknowns θ_x, θ_y, θ_z, and $\hat{f} = \lambda f$. These unknowns are estimated by minimizing the following error function using non-linear least-squares:

$$E(\theta_x, \theta_y, \theta_z, \hat{f}) = (\hat{f}c_yc_z - h_1)^2 + (\hat{f}c_ys_z - h_2)^2 + (\hat{f}(s_xs_yc_z - c_xs_z) - h_4)^2$$
$$+ (\hat{f}(s_xs_ys_z + c_xc_z) - h_5)^2, \tag{30}$$

where h_i corresponds to the ith entry of H. A Gauss-Newton iterative approach is employed to minimize $E(\cdot)$. In practice, we have found that $\theta_z = \tan^{-1}(h_2/h_1)$, $f = 1$ and random values for θ_x and θ_y provide good starting conditions for this minimization. These estimated parameters then yield two possible estimates of the focal length:

$$f_1 = \frac{\hat{f}(c_xs_yc_z + s_xs_z)}{h_7} \quad \text{and} \quad f_2 = \frac{\hat{f}(c_xs_ys_z - s_xc_z)}{h_8}. \tag{31}$$

These two estimates are combined using the following weighted average:

$$f = \frac{h_7^2 f_1 + h_8^2 f_2}{h_7^2 + h_8^2}. \tag{32}$$

Note that the focal length f is undefined for $h_7 = h_8 = 0$. In addition, this estimation is vulnerable to numeric instabilities for values of h_7 and h_8 near zero. As such, the weighting was chosen to favor larger values of h_7 and h_8.