# Exposing Photo Manipulation From User-Guided 3-D Lighting Analysis

Tiago Carvalho[a], Hany Farid[b] and Eric Kee[c]

[a]Cemaden, São José dos Campos, SP, Brazil
[b]Dartmouth College, Hanover, NH, USA
[c]Columbia University, New York, NY, USA

## ABSTRACT

We describe a photo forensic technique based on detecting inconsistencies in lighting. This technique explicitly measures the 3-D lighting properties for individual people, objects, or surfaces in a single image. We show that with minimal training, an analyst can accurately specify 3-D shape in a single image from which 3-D lighting can be automatically estimated. A perturbation analysis on the estimated lighting is performed to yield a probabilistic measure of the location of the illuminating light. Inconsistencies in lighting within an image evidence photo tampering.

**Keywords:** Photo Forensics, Lighting

## 1. INTRODUCTION

It has become well established that photo manipulation has eroded our trust in photography. This lack of trust has impacted law enforcement, national security, the media, advertising, commerce, and more. Advances in photo forensics have begun to restore some trust in photography.[1,2]

There are many different and complementary approaches to analyzing a photo for evidence of manipulation. These techniques operate on the general assumption that manipulation will disrupt some statistical, geometric, or physical properties in an image. Manipulation can, therefore, be exposed by quantifying and detecting these perturbations. For example, format-specific analyses exploit artifacts that are introduced when a JPEG image is compressed multiple times.[3–5] Pixel-based analyses detect low-level correlations that are introduced by cloning,[6,7] re-sizing,[8] or non-linear filtering.[9] Sensor-based analyses can detect inconsistencies in chromatic aberrations,[10] color filter array interpolation,[11,12] or sensor noise.[13,14] And, physically-based analyses can detect inconsistencies in reflections,[15] lighting,[16–19] or shadows.[20,21]

Within these different techniques, physically-based methods are particularly attractive because they are applicable in low quality and low resolution images, and can be hard to counter attack since the physical measurements being made are the result of interactions in the 3-D physical world being projected into the 2-D image.

Earlier lighting-based forensic techniques focused on estimating 2-D properties of lighting. This is because estimating the full 3-D lighting requires knowledge of the 3-D structure of the scene which is, of course, generally not readily available. We describe a new 3-D lighting-based technique that overcomes this limitation. This technique leverages the fact that with minimal training, an analyst can often provide fairly reliable estimates of local 3-D scene structure, from which 3-D lighting can be estimated. Estimating the full 3-D lighting properties of a scene makes this forensic technique even more powerful as a determined forger will now be forced to correct for the full 3-D lighting properties – a decidedly more difficult and time-consuming task.

---

Contact: tiagojc@gmail.com, farid@cs.dartmouth.edu, erickee@gmail.com

Figure 1. An analyst specifies 3-D surface normals (red probes on the face) from which the 3-D light direction is estimated (visualized as a probability in azimuth/elevation in the upper right panel). Multiple such constraints are combined to determine if lighting in the entire image is consistent, as in this example (lower right panel).

## 2. METHODS

We seek to estimate 3-D lighting in a single image. We rely on an analyst to specify the required local 3-D shape from which lighting is estimated. We describe an easy to use user-interface for obtaining these shape estimates, how 3-D lighting can be estimated from these estimates, a perturbation analysis that contends with any errors or biases in the user-specified 3-D shape, and a probabilistic technique for combining multiple lighting estimates to determine if they are physically consistent with a single light source.

### 2.1 User-Assisted 3-D Shape Estimation

The projection of a 3-D scene onto a 2-D image sensor results in a basic loss of information. Recovering 3-D shape from a single 2-D image is at best a difficult problem and at worst an under-constrained problem. There is, however, good evidence from the human perception literature that human observers are fairly good at estimating local 3-D shape from a variety of cues including, foreshortening, shading, and familiarity.[22–25] To this end, we ask an analyst to specify the local 3-D shape of surfaces. We have found that with minimal training, this task is relatively easy and accurate.

An analyst estimates the local 3-D shape at different locations on an object by adjusting the orientation of a small 3-D probe. The probe consists of a circular base and a small vector (the stem) orthogonal to the base. An analyst orients a virtual 3-D probe so that when the probe is projected into the image, the stem appears to be orthogonal to the object surface. Shown in Figure 2 is an example of several such probes on a 3-D rendered model of a car.

With the click of a mouse, an analyst can place a probe at any point $\boldsymbol{p}$ in the image. This initial mouse click specifies the location of the probe's base. As the analyst drags their mouse, they control the orientation of the probe by way of the 2-D vector $\boldsymbol{v}$ from the probe's base to the mouse location. This vector is restricted by the interface to have a maximum value of $\delta$ pixels.

Probes are displayed to the analyst by constructing them in 3-D, and projecting them into the image. The 3-D probe is constructed in a coordinate system that is local to the object, Figure 3, defined by three mutually orthogonal vectors:

$$\mathbf{b}_1 = \begin{bmatrix} \boldsymbol{p} - \boldsymbol{c} \\ f \end{bmatrix} \qquad \mathbf{b}_2 = \begin{bmatrix} \boldsymbol{v} \\ \frac{1}{f}\boldsymbol{v} \cdot (\boldsymbol{p} - \boldsymbol{c}) \end{bmatrix} \qquad \mathbf{b}_3 = \mathbf{b}_1 \times \mathbf{b}_2 \quad , \tag{1}$$

where $\boldsymbol{p}$ is the location of the probe's base in the image, and $f$ and $\boldsymbol{c}$ are a focal length and principal point (described below). The 3-D probe is constructed by first initializing it into a default orientation in which its stem, a unit vector, is coincident with $\mathbf{b}_1$, and the circular base lies in the plane spanned by $\mathbf{b}_2$ and $\mathbf{b}_3$, Figure 3.
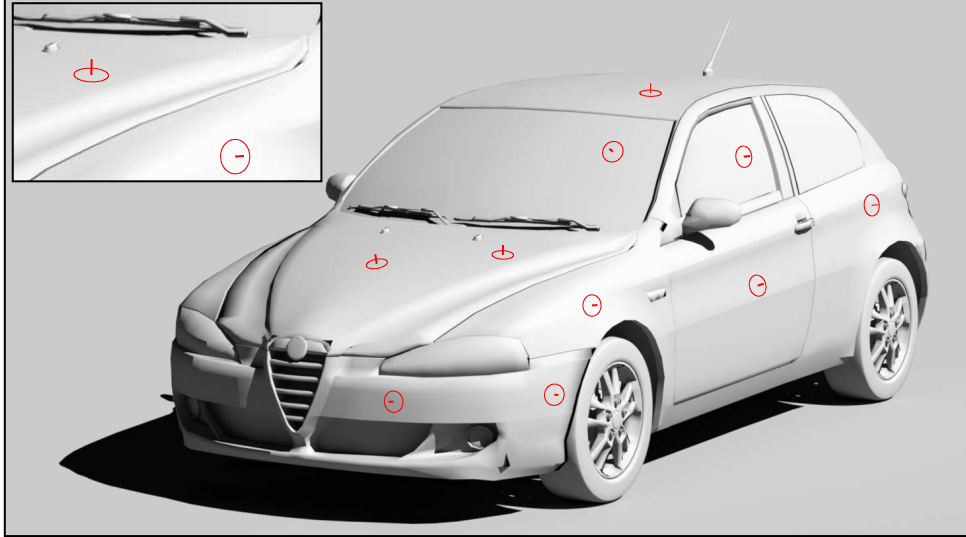
Figure 2. Shown is a rendered 3-D object with user-specified probes that capture the local 3-D structure. Shown in the top left is a magnified view of two probes.

The 3-D probe is then adjusted to correspond with the analyst's desired orientation which is uniquely defined by their 2-D mouse position $\boldsymbol{v}$. The 3-D probe is parameterized by a slant and tilt, Figure 3. The length of the vector $\boldsymbol{v}$ specifies a slant rotation, $\sigma = \sin^{-1}(\|\boldsymbol{v}\|/\delta)$, of the probe around $\mathbf{b}_3$. The tilt, $\tau = \tan^{-1}(v_y/v_x)$, is embodied in the definition of the coordinate system, Equation (1).

The construction of the 3-D probe requires the specification of a focal length $f$ and principal point $\boldsymbol{c}$, Equation (1). There are, however, two imaging systems to be considered. The first is that of the observer relative to the display.[26] This imaging system dictates the appearance of the probe in the image plane. In that case, we assume an orthographic projection with $\boldsymbol{c} = \boldsymbol{0}$.[22, 25] The second imaging system is that of the camera which recorded the image. This imaging system dictates how the surface normal $\mathbf{n}$ is constructed to estimate the lighting (Section 2.2). In this case, if the focal length $f$ and principal point $\boldsymbol{c}$ are unknown then $f$ can be assigned a typical mid-range value and $\boldsymbol{c} = \boldsymbol{0}$.

The slant/tilt convention accounts for linear perspective, and for the analyst's interpretation of the photo.[26–28] A slant of 0 corresponds to a probe that is aligned with the 3-D camera ray, $\mathbf{b}_1$. In this case the probe's stem projects to a single point within the circular base.[27] A slant of $\pi/2$ corresponds to a probe that lies on an occluding boundary in the photo. In this case, the probe projects to a "T"-shape with the stem coincident with the axis $\mathbf{b}_2$, and with the circular base laying in the plane spanned by axes $\mathbf{b}_1$ and $\mathbf{b}_3$. This 3-D geometry is consistent given the analyst's orthographic interpretation of a photo, as derived in.[28]

With user-assisted 3-D surface normals in hand, we proceed with estimating 3-D lighting.

## 2.2 3-D Light Estimation

We begin with the standard assumptions that a scene is illuminated by a single distant point light source (e.g., the sun) and that an illuminated surface is Lambertian and of constant reflectance. Under these assumptions, the radiance of a surface patch is given by:

$$r = \mathbf{n} \cdot \mathbf{s} + a, \tag{2}$$

where $\mathbf{n}^T = \begin{pmatrix} n_x & n_y & n_z \end{pmatrix}$ is the 3-D surface normal, $\mathbf{s}^T = \begin{pmatrix} s_x & s_y & s_z \end{pmatrix}$ specifies the direction to the light source (the magnitude of $\mathbf{s}$ is proportional to the light's brightness), and the constant ambient light term $a$ approximates indirect illumination. Note that this expression assumes that the angle between the surface normal and light is less than $90°$.
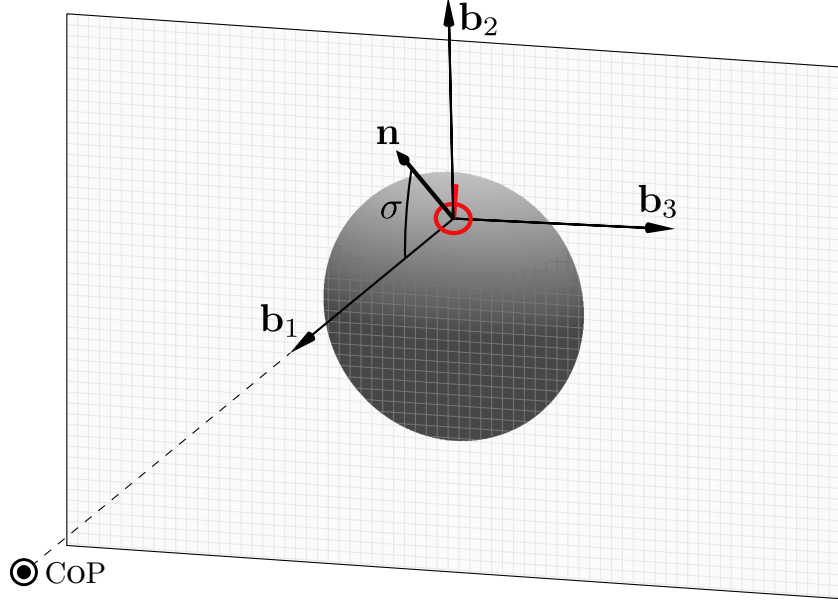
Figure 3. Shown is a shaded sphere in the image plane and a small circular probe corresponding to surface normal $\mathbf{n}$. A local coordinate system is defined by $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$. The slant of $\mathbf{n}$ is specified by a rotation $\sigma$ around $\mathbf{b}_3$, while the tilt $\tau$ is implicitly defined by the axes $\mathbf{b}_2$ and $\mathbf{b}_3$, Equation (1).

The four components of this lighting model (light direction and ambient term) can be estimated from $k \geq 4$ surface patches with known surface normals. The equation for each surface normal and corresponding radiance are packed into the following linear system:

$$
\begin{pmatrix} \mathbf{n}_1^T & 1 \\ \vdots & \vdots \\ \mathbf{n}_k^T & 1 \end{pmatrix} \begin{pmatrix} \mathbf{s} \\ a \end{pmatrix} = \mathbf{r} \tag{3}
$$

$$
\mathbf{Ne} = \mathbf{r} , \tag{4}
$$

where $\mathbf{r}$ is a $k$-vector of observed radiance for each surface patch. The lighting parameters $\mathbf{e}$ can be estimated using standard least squares:

$$
\mathbf{e} = (\mathbf{N}^T\mathbf{N})^{-1}\mathbf{N}^T\mathbf{r}, \tag{5}
$$

where the first three components of $\mathbf{e}$ correspond to the estimated light direction. Because we assume a distant light source, this light estimate can be normalized to be unit sum and visualized in terms of azimuth $\phi = -\text{atan2}\,(s_x/s_z)$, $\phi \in [-\pi, \pi]$ and elevation $\theta = \sin^{-1}(s_y/\|s\|)$, $\theta \in [-\pi/2, \pi/2]$.

In practice there will be errors in the estimated light direction due to errors in the user-specified 3-D surface normals, deviations of the imaging model from our assumptions, sensor noise, etc. To contend with such errors, we perform a perturbation analysis yielding a probabilistic measure of the light direction.

## 2.3 Modeling Uncertainty

For simplicity, we assume that the dominant source of error is the analyst's estimate of the 3-D normals. A model for these errors is generated from a large-scale psychophysical study in which observers were presented with one of twelve different 3-D models, and asked to orient probes, like those used here, to specify the object's shape.[25] The objects were shaded with a simple outdoor lighting environment. Using Amazon's Mechanical Turk a total of $45,241$ probe settings from $560$ observers were collected.

From this data, we construct a probability distribution for the actual light slant and tilt conditioned on the estimated slant and tilt. We then model the uncertainty in the analyst's estimated light position by randomly
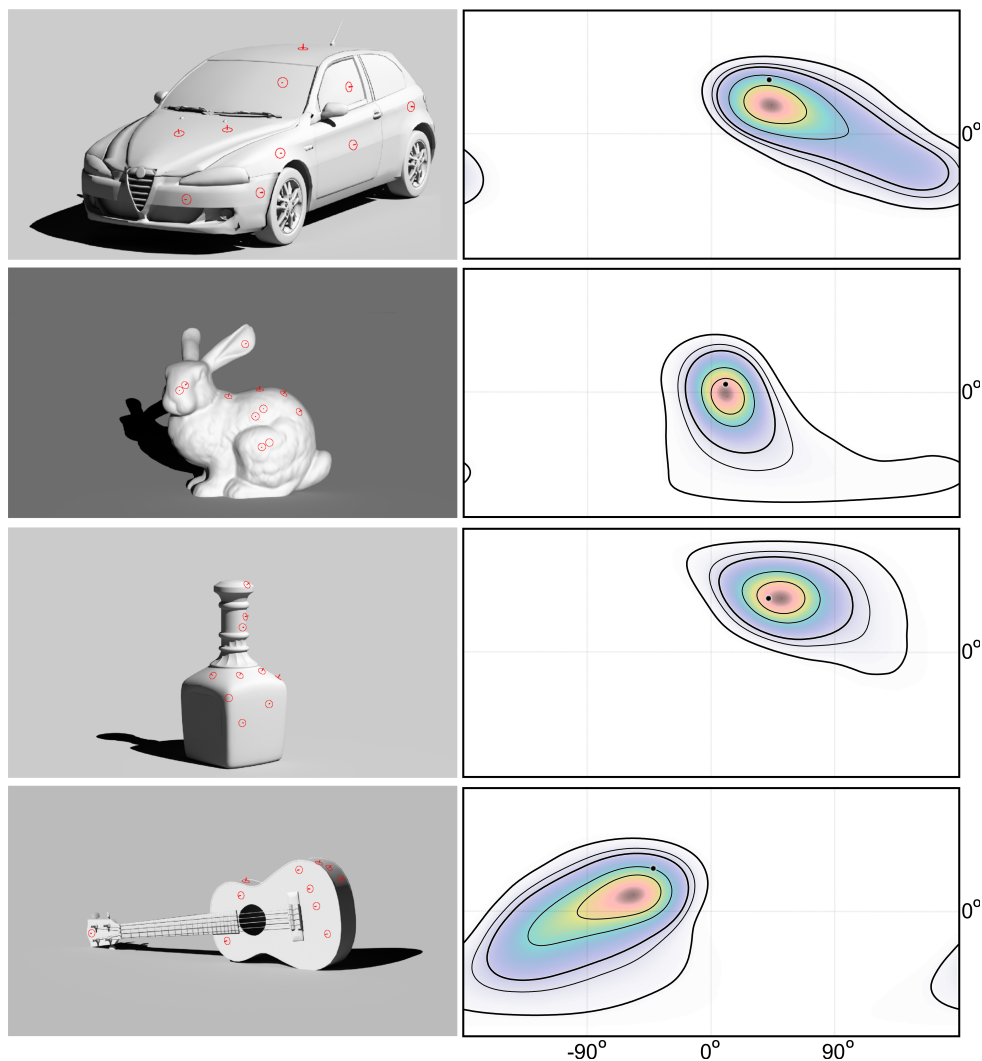
Figure 4. Shown on the left are four representative objects with user-specified probes. Shown on the right is the estimated light position specified as confidence intervals in an azimuth/elevation space. The contours correspond to probabilities 30% (smallest), 60%, 90%, 95%, and 99% (largest). The small black dot corresponds to the actual light position.

drawing from this conditional probability distribution and re-estimating the light position. Each estimated light position contributes a small Gaussian density that is projected into azimuth/elevation space. These densities are accumulated across $20,000$ random draws, producing a kernel-density estimation of the uncertainty in the analyst's estimate of lighting.

The probe setting data is used to construct a probability distribution over the ground truth slant and tilt conditioned on the analyst's slant and tilt. This slant/tilt distribution is then used to build a distribution over the 3-D light direction. Specifically, for each of the analyst's probes, a ground truth slant/tilt is randomly drawn. The light direction is then estimated, and contributes a small Gaussian density (over the 3-D light direction) that is projected into azimuth/elevation space. This density estimation procedure is repeated for $20,000$ random perturbations of the analyst's probes. The result is a kernel-density estimate of the distribution over light directions, given the analyst's estimate of the object geometry.

Multiple distributions are constructed by an analyst, each from a particular object or feature in the image. A confidence region is computed in each distribution, identifying a potentially non-contiguous area in which the light source must lie. The physical consistency of an image is determined by intersecting these confidence
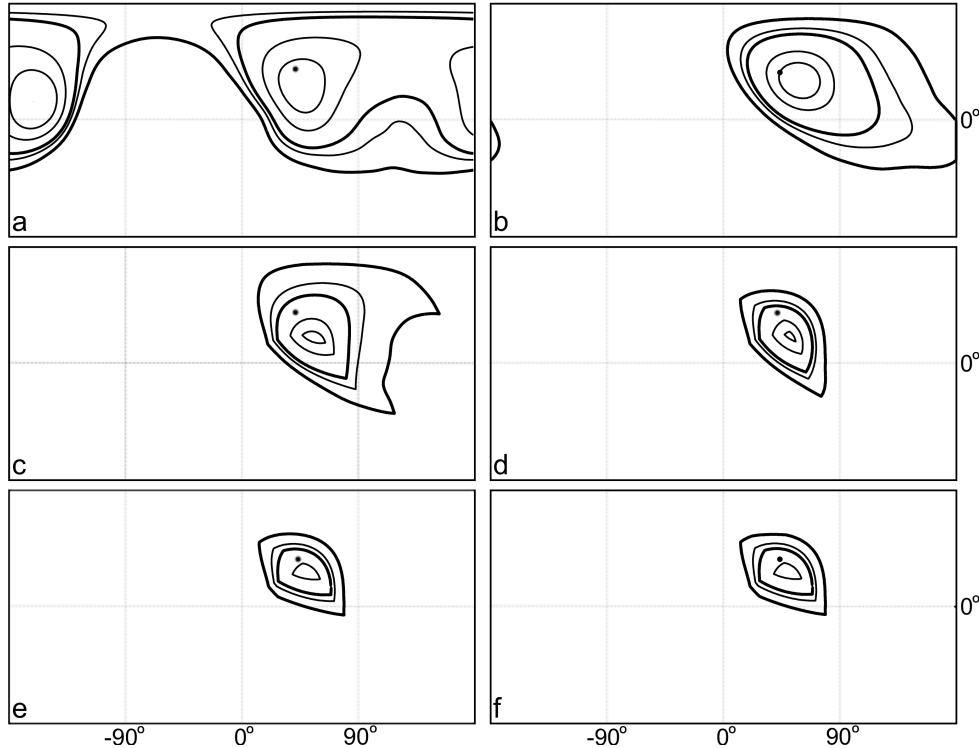
Figure 5. Shown are (left to right, top to bottom) the confidence intervals for the lighting estimate from 1 through 6 objects in the same scene. This interval shrinks as more objects are added. Confidence intervals are shown at 30% (smallest), 60%, 90% (bold), 95%, and 99% (largest and bold). The small black dot corresponds to the actual light position.

regions. Forgery is detected if the intersection* is empty, up to a specified confidence threshold.

## 3. RESULTS

We rendered ten objects under six different lighting conditions. Sixteen previously untrained users were each instructed to place probes on ten objects. Shown in the left column of Figure 4 are four representative objects with the user-selected probes. Shown in the right column of this same figure are the estimated light positions specified as confidence intervals. The small black dot in each figure corresponds to the actual light position. On average, users were able to estimate the azimuth and elevation with an average accuracy of 11.1 and 20.6 degrees with a standard deviation of 9.4 and 13.3 degrees, respectively. On average, a user placed 12 probes on an object in 2.4 minutes.

In a real-world analysis, an analyst will be able to specify the 3-D shape of multiple objects which can then be combined to yield an increasingly more specific estimate of lighting. Shown in Figure 5, for example, are the results of sequentially intersecting the confidence regions from six rendered objects (Figure 4) in the same scene. From left to right and top to bottom, the confidence intervals shrink. Of course, the smaller this confidence region, the more effective this technique will be in detecting inconsistent lighting.

Shown in Figure 1 is the result from a real-world image. Shown in the top row is the original image. Shown in the middle row are sample probes from one person's face. Shown next to this panel is the estimated light source. An analyst also selected probes on the other person. Shown in the bottom panel of the same figure is the intersection of the estimated light sources from each of these. These results reveal that the lighting in this image is physically consistent. Shown in Figure 6 is a forgery based on the image in Figure 1. Shown in the lower panel is the estimated light source for the added soccer ball (labeled B) which is clearly inconsistent with the estimated light source from Figure 1 (labeled A).

---

*By intersecting confidence regions, rather than multiplying probabilities, every constraint must be satisfied for the lighting to be consistent.
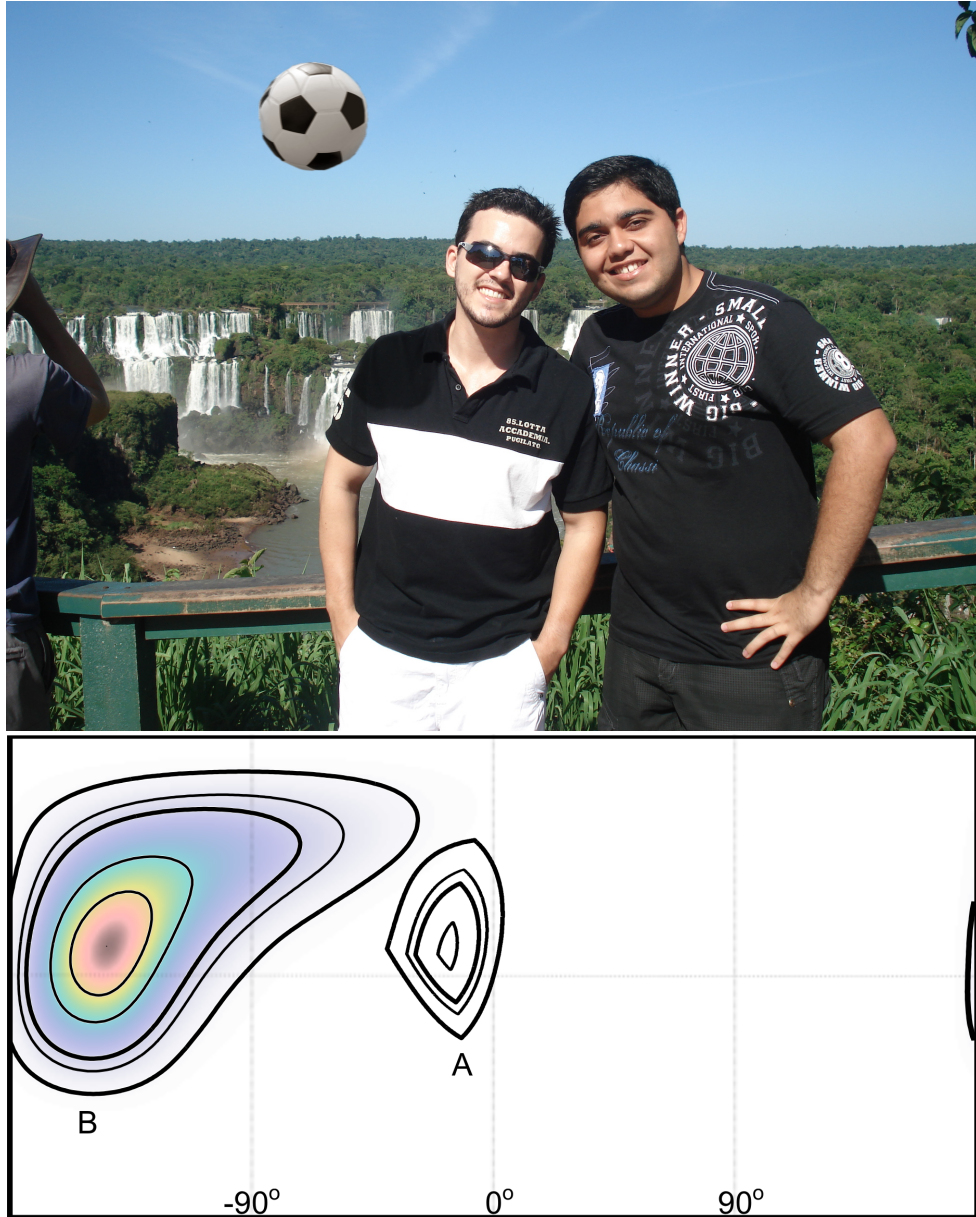
Figure 6. A forgery in which the soccer ball was inserted. The 3-D lighting estimated from the ball (B) is detected as being inconsistent with the rest of the scene (A). Confidence intervals are shown at 30% (smallest), 60%, 90% (bold), 95%, and 99% (largest and bold).
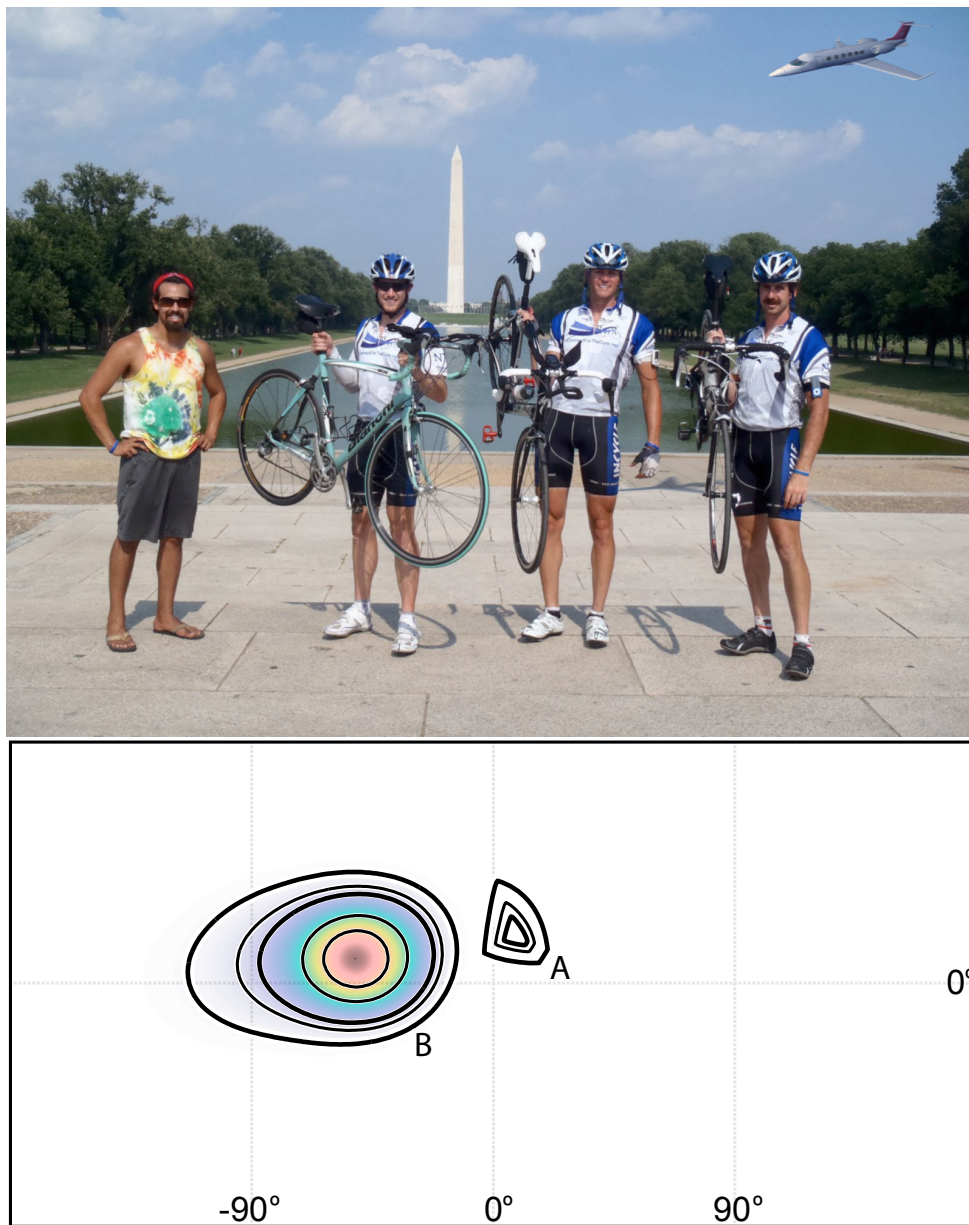
Figure 7. A forgery in which the jet was inserted. The 3-D lighting estimated from the jet (B) is detected as being inconsistent with the rest of the scene (A). Confidence intervals are shown at 30% (smallest), 60%, 90% (bold), 95%, and 99% (largest and bold). [Original image copyright 2010, Connect for the Cure, http://connectforthecure.blogspot.com]

And lastly, shown in Figure 7 is another result from a real-world image in which the jet was digitally inserted into an authentic photo. An analyst selected probes on the face and body of each man and on the bicycle wheel of the second cyclists from the left. Shown in the bottom panel of the same figure is the intersection of the estimated light sources from each of these. Despite the somewhat unusual collection of characters, these results reveal that the lighting in this image (minus the jet) is physically consistent. An analyst also selected 47 probes on the jet. Shown in the bottom panel of Figure 7 is the estimated light source for the jet (labeled B) which is clearly inconsistent with the rest of the scene (labeled A).

## 4. DISCUSSION

We have described a user-guided forensic technique for estimating 3-D lighting. With minimal training, an analyst can quickly and accurately specify the required 3-D surface normals needed to estimate lighting from a single image. Unlike previous lighting-based techniques, this approach is able to estimate the full 3-D lighting properties of a scene. Estimating the full 3-D lighting makes this forensic technique more powerful than previous 2-D methods as it removes the inherent ambiguity present when only analyzing a subset of the full lighting model. Although we have focused here on estimating light direction only, this technique could easily be used to estimate more complex 3-D lighting environments.[17]

More generally, other forensic techniques might benefit from user-guided assistance when dealing with otherwise poorly-defined estimation problems such as that described here.

## REFERENCES

[1] Farid, H., "A survey of image forgery detection," *IEEE Signal Processing Magazine* **2**(26), 16–25 (2009).

[2] Rocha, A., Scheirer, W., Boult, T. E., and Goldenstein, S., "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Computing Surveys (CSUR)* **43**(4), 26:1–26:42 (2011).

[3] Popescu, A. and Farid, H., "Statistical tools for digital forensics," in [*Information Hiding*], Fridrich, J., ed., *Lecture Notes in Computer Science* **3200**, 395–407, Springer Berlin / Heidelberg (2005).

[4] Mahdian, B. and Saic, S., "Detecting double compressed JPEG images," in [*3rd International Conference on Crime Detection and Prevention (ICDP)*], 1–6 (2009).

[5] Kee, E., Johnson, M. K., and Farid, H., "Digital image authentication from JPEG headers," *IEEE Transactions on Information Forensics and Security* **6**, 1066–1075 (September 2011).

[6] Fridrich, J., Soukal, D., and Lukas, J., "Detection of copy move forgery in digital images," in [*In Proceedings of Digital Forensic Research Workshop (DFRWS)*], (2003).

[7] Pan, X. and Lyu, S., "Region duplication detection using image feature matching," *IEEE Transactions on Information Forensics and Security* **5**(4), 857 –867 (2010).

[8] Popescu, A. C. and Farid, H., "Exposing digital forgeries by detecting traces of re-sampling," *IEEE Transactions on Signal Processing* **53**(2), 758–767 (2005).

[9] Lin, Z., Wang, R., Tang, X., and Shum, H.-Y., "Detecting doctored images using camera response normality and consistency," *IEEE Conference on Computer Vision and Pattern Recognition* **1**, 1087–1092 (2005).

[10] Johnson, M. K. and Farid, H., "Exposing digital forgeries through chromatic aberration," in [*Proceedings of the 8th workshop on multimedia and security*], MM&Sec'06, 48–55, ACM (2006).

[11] Popescu, A. C. and Farid, H., "Exposing digital forgeries in color filter array interpolated images," *IEEE Transactions on Signal Processing* **53**(10), 3948–3959 (2005).

[12] Kirchner, M., "Efficient estimation of CFA pattern configuration in digital camera images," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], **7541** (2010).

[13] Chen, M., Fridrich, J., Goljan, M., and Lukas, J., "Determining image origin and integrity using sensor noise," *IEEE Transactions on Information Forensics and Security (TIFS)* **3**, 74–90 (March 2008).

[14] Fridrich, J., "Digital image forensic using sensor noise," *IEEE Signal Processing Magazine* **26**(2), 26–37 (2009).

[15] O'Brien, J. F. and Farid, H., "Exposing photo manipulation with inconsistent reflections," *ACM Transactions on Graphics* **31**, 4:1–11 (Jan. 2012). Presented at SIGGRAPH 2012.

[16] Johnson, M. K. and Farid, H., "Exposing digital forgeries by detecting inconsistencies in lighting," in [*Proceedings of the 7th workshop on Multimedia and security*], *MM&Sec '05*, 1–10, ACM (2005).

[17] Johnson, M. K. and Farid, H., "Exposing digital forgeries in complex lighting environments," *IEEE Transactions on Information Forensics and Security* **3**(2), 450–461 (2007).

[18] Kee, E. and Farid, H., "Exposing digital forgeries from 3-D lighting environments," in [*IEEE International Workshop on Information Forensics and Security (WIFS)*], 1–6 (2010).

[19] de Carvalho, T. J., Riess, C., Angelopoulou, E., Pedrini, H., and de Rezende Rocha, A., "Exposing digital image forgeries by illumination color classification," *IEEE Transactions on Information Forensics and Security* **8**(7), 1182–1194 (2013).

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

[21] Kee, E., O'Brien, J., and Farid, H., "Exposing photo manipulation with inconsistent shadows," *ACM Trans. Graph.* **32**(3), 28:1–28:12 (2013).

[22] Koenderink, J. J., Van Doorn, A., and Kappers, A., "Surface perception in pictures," *Attention, Perception, & Psychophysics* **52**, 487–496 (1992). 10.3758/BF03206710.

[23] Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., and Todd, J. T., "Ambiguity and the 'mental eye' in pictorial relief," *Perception* **30**(4), 431–448 (2001).

[24] Todd, J. T., Koenderink, J. J., van Doorn, A. J., and Kappers, A. M. L., "Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces.," *Journal of Experimental Psychology: Human Perception and Performance* **22**(3), 695–706 (1996).

[25] Cole, F., Sanik, K., DeCarlo, D., Finkelstein, A., Funkhouser, T., Rusinkiewicz, S., and Singh, M., "How well do line drawings depict shape?," in [*ACM Transactions on Graphics (Proc. SIGGRAPH)*], **28** (Aug. 2009).

[26] Cooper, E. A., Piazza, E. A., and Banks, M. S., "The perceptual basis of common photographic practice," *Journal of Vision* **12**(5) (2012).

[27] Koenderink, J. J., van Doorn, A. J., de Ridder, H., and Oomes, S., "Visual rays are parallel," *Perception* **39**(9), 1163–1171 (2010).

[28] Kee, E., O'Brien, J., and Farid, H., "Exposing photo manipulation from shading and shadows," *ACM Trans. Graph.* (in press 2014).