Handling Multiple Materials for Exposure of Digital Forgeries using 2-D Lighting Environments

Christian Riess · Mathias Unberath · Farzad Naderi · Sven Pfaller · Marc Stamminger · Elli Angelopoulou

Abstract The distribution of incident light is an important physics-based cue for exposing image manipulations. If an image has been composed from multiple sources, it is likely that the illumination environments of the spliced objects differ. Johnson and Farid introduced a proof-of-principle algorithm for a forensic comparison of lighting environments. However, this baseline approach suffers from relatively strict assumptions that limit its practical applicability.

In this work, we address one of the biggest limitations, namely the need to compute a lighting environment from patches of homogeneous material. To compute a lighting environment from multiple-color surfaces, we propose a method that we call "intrinsic contour estimation" (ICE). ICE is able to integrate reflectances from multiple materials into one lighting environment, as long as surfaces of different materials share at least two similar normal vectors. We validate the proposed method in a controlled ground-truth experiment on two datasets, with light from three different directions. These experiments show that using ICE can improve the median estimation error by almost 50%, and the mean error by almost 30%.

C. Riess
Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg
Martensstr. 3, 91058 Erlangen, Germany
Tel.: +49-9131-8527774
Fax: +49-9131-8527270
E-mail: christian.riess@fau.de

Fig. 1 Original or manipulated? The detected direction of incident light (white pointer) differs about 40° (pictures best viewed in color).

1 Introduction

As digital imagery and image processing software become increasingly available there is a need for image forgery detection. Blind image forensics aims at verifying the authenticity and origin of images while not requiring any support from an embedded security scheme. Researchers developed a family of forensic algorithms that either try to detect traces of manipulation in an image, or aim at verifying characteristic scene or image properties to affirm its authenticity. Overviews of existing methods can be found, e. g., in [15] and [5].

Existing methods in image forensics can roughly be categorized into statistical and physics-based. Statistical methods aim to detect manipulations from local bit-level irregularities, while physics-based methods aim to quantify deviations from the interplay of objects with the scene, like the shadow geometry or the direction of incident light. Two key advantages of physics-based algorithms are that they are typically not limited to digital imagery (i. e., they can also be used on analog photographs), and that they are relatively robust to automated counter-forensics methods. The biggest disadvantage of physics-based methods is that they typically require manual user interaction, and as such are not well suited for batch processing. Several physics-based algorithms exploit geometric constraints to detect inconsistencies in cast shadows [19, 12, 11]. Other approaches aim to validate color [16, 3] or motion cues [2].

Another group of physics-based approaches is the validation of lighting environments. Human perception is relatively insensitive to differences in the direction of incident light [13]. Johnson and Farid [9] presented an algorithm to determine the 2-D direction of incident light from the distribution of intensities along object contours. Kee and Farid [10] extended this approach to 3-D by fitting 3-D face geometries to persons under investigation. This approach yields a dense grid of 3-D normals (and hence more robust estimation results), but is also considerably more complex than the 2-D approach. Recently, Peng *et al.* [14] proposed an improvement to the 3-D estimation by Kee and Farid by using a surface reflection model that allows additional flexibility with respect to non-convex local geometries, and non-constant material reflectance. An alternative to this approach was proposed by Fan *et al.* [4]. They re-

placed the estimation of a 3-D surface model with a shade-from-shading algorithm. The attractivity of the 2-D algorithm is that it is relatively lightweight, and widely applicable, in the sense that it does neither require a 3-D model of the objects of interest, nor additional 3-D convexity or simplicity assumptions.

Unfortunately, lighting-based algorithms are oftentimes difficult to apply in practice, which is mainly due to their relatively strong scene assumptions. First, the underlying model only holds for contours of purely diffuse, homogenous materials under direct illumination. Selected contours must exhibit a large variety of normal directions to obtain a numerically stable estimate of the lighting environment. Additionally, all contours used for calculation must be extracted from regions that represent the same material. In practice, it can be challenging to satisfy all three requirements at the same time. For example, hair or highly textured clothes are not admissible under this model. Also, structurally unsmooth regions like for example folds or crumbles in clothes have to be excluded. If applied to real world images, body pose and partial occlusion add to these challenges. Thus, oftentimes it not possible to extract contours represent both a large directional variety and identical materials at the same time.

In this work, we extend the approach by Johnson and Farid [9] by removing the constraint that surface normals have to be selected from the same material. Being able to use a wider variety of surface normals makes the estimation of the lighting environment more robust, or possible in the first place. The proposed method is based on the observation that normals from different materials pointing towards the same direction shall ideally have identical pixel intensities. We propose an straightforward analytic solution to this optimization problem. This work is an extended version of a recent workshop paper [17]. Compared to the previous work, we build a second, more experimental setup, which is considerably more challenging with respect to different subject poses, different materials, and a smaller difference between illuminations. We also considerably expanded the evaluation and discussion of the method.

The idea of the method is illustrated in Fig. 1. On the left, an example image of two spliced persons is shown. The proposed method allows to select contours along *multiple* materials to estimate the direction of incident light. On the right of Fig. 1, an example analysis is shown. The primary illuminants on the persons were estimated to be at angles of 46.8° and 4.6°, respectively. These directions are indicated by the white vectors on the chests. The deviation between both angles indicates inconsistencies of the lighting environments and therefore suggests that the image is spliced.

The paper is organized as follows: some basic notation is introduced in Sec. 2. In Sec. 3, we restate the baseline algorithm. In Sec. 4.2, we present the proposed algorithm, which we call *Intrinsic Contour Estimation* (ICE). We captured two sets of ground truth data. The protocol is described in Sec. 5. Quantitative results and a discussion are presented in Sec. 6.

2 Notation of Intensities, Normals and Angles at a Pixel Coordinate

We briefly introduce some mathematical notation that will be used throughout this paper. We denote a pixel position in the image as $\mathbf{x} \in \mathbb{R}^2$. The observed RGB vector at $\mathbf{x}$ is denoted as $\mathbf{p}(\mathbf{x})$. A scalar intensity at $\mathbf{x}$ is denoted as $m(\mathbf{x})$. We assume that any

$Y_{0,0}(\phi)$	$Y_{1,-1}(\phi)$	$Y_{1,1}(\phi)$	$Y_{2,-2}(\phi)$	$Y_{2,2}(\phi)$
$\frac{1}{\sqrt{4\pi}}$	$\sqrt{\frac{3}{4\pi}}\sin(\phi)$	$\sqrt{\frac{3}{4\pi}}\cos(\phi)$	$3\sqrt{\frac{5}{4\pi}}\cos(\phi)\sin(\phi))$	$\frac{3}{2}\sqrt{\frac{5}{12\pi}}(\cos^2(\phi)-\sin^2(\phi))$

Table 1 Spherical Harmonics coefficients used for 2-D estimation of the lighting environment. These functions depend on the normal vector in the image plane, denoted as angle $\phi = v(\mathbf{x})$.

linear operation (e. g., a projection) may be used to reduce a polychromatic signal $\mathbf{p}(\mathbf{x})$ to a monochromatic signal $m(\mathbf{x})$. The presented method makes use of normal vectors of contours that cross pixel $\mathbf{x}$. We denote with $\mathbf{n}(\mathbf{x})$ the 2-D normal of a contour in the image at pixel $\mathbf{x}$. Furthermore, we denote with $v(\mathbf{x})$ the angle of $\mathbf{n}(\mathbf{x})$.

3 Estimation of 2-D Lighting Environments

In the following section we restate the algorithm by Johnson and Farid for estimating a 2-D lighting environment from contours. For additional details please refer to the original work [9].

Consider an object that is illuminated by direct and indirect light sources from different direction. The distribution of these illuminants (i. e., the "lighting environment") can under some assumptions be computed from the intensity distribution on the object and the object's surface normals. Since the picture under examination is a 2-D projection of the true 3-D scene, the scene normals are mostly unknown. The key idea by Johnson and Farid is that at occluding contours, the 3-D scene normals are identical to the observed 2-D image normals (because the 3-D normals lie in the image plane). Thus, a 2-D projection of an object's lighting environment can be directly estimated from the image.

For forensic exploitation, consider that the lighting environment of two objects shall be compared. For each of the two objects, the object contours are manually annotated. Contours can be defined piecewise, but they must be directly illuminated. Thus, contour pixels containing self-shadowing (e.g., from folds in clothes) have to be excluded. The surface normal of each contour pixel is estimated by fitting a 2-D polynomial to the contour in the pixel's neighborhood. The intensity of each point along the contour is extrapolated from pixels in a surrounding neighborhood.

The lighting environment is modeled using a weighted sum of the five second-order basis functions of the 2-D spherical harmonics. These five functions are listed in Tab. 1. Assuming purely diffuse (Lambertian) reflectance of an object of interest, the intensity along the object boundary can be expressed as a linear combination of the basis functions. For a contour consisting of a points $\mathbf{x}_1, \ldots, \mathbf{x}_a$ yielding a normals, the basis functions can be evaluated and stored in a matrix $A \in \mathbb{R}^{a \times 5}$. Then, the unknown weighting factors $\mathbf{h} \in \mathbb{R}^5 \times 1$ must satisfy

$$A\mathbf{h} = \mathbf{m} \ , \tag{1}$$

where $\mathbf{m} = (m(\mathbf{x}_1), \ldots, m(\mathbf{x}_a))^\mathrm{T} \in \mathbb{R}^n$ denotes a vector of intensities along the contour. For color images, this implies that either a single color channel is selected, or
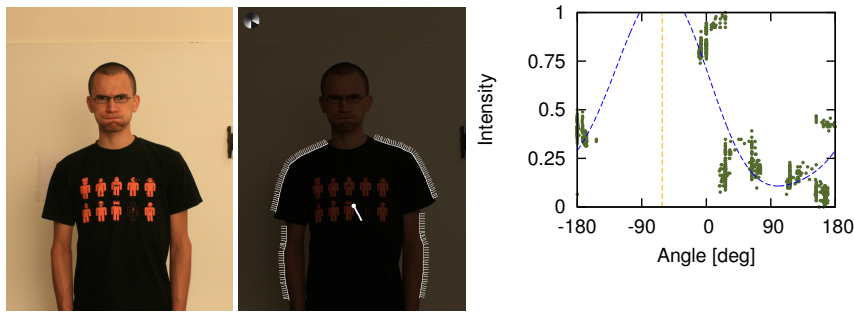
Fig. 2 Illustration of mixed-material contours: the brightness contrast between the black T-shirt and bright skin prevents cross-material estimation of the lighting environment.

the color values are converted to grayscale. To add some robustness against noise, a regularization term is added to Eqn. 1, leading to the final energy function

$$E(\mathbf{h}) = \|A\mathbf{h} - \mathbf{m}\|^2 + \mu \|C\mathbf{h}\|^2 \ , \tag{2}$$

where $C \in \mathbb{R}^{5\times 5}$, $C = \mathrm{diag}(1\ 2\ 2\ 3\ 3)$, and μ is a user-selectable parameter to guide the strength of the regularizer. Equation 2 has the analytic solution

$$\mathbf{h} = (A^{\mathrm{T}}A + \mu C^{\mathrm{T}}C)^{-1}A^{\mathrm{T}}\mathbf{m} \ . \tag{3}$$

The method fits a 2-D spherical harmonics model to the intensity distribution along the object boundary. All brightness differences along the contour are attributed to differences in the lighting environment. As a consequence, all contour pixels have to be extracted from the same underlying material.

Figure Fig. 2 shows an (extreme) case of what happens if intensities are selected from materials with large brightness differences: the left picture shows a test subject illuminated from top right. Middel picture: To model the lighting environment, contour pixels are selected along the T-shirt and arms. However, since the T-shirt is black, the resulting lighting environment estimates that the light comes from an angle of $-61.5°$, i.e., the bottom right. This is shown in the normal-intensity plot on the right, with the direction of the contour normal on the x-axis ($0°$ points to the right, angular direction in counter-clockwise direction), and the brightness of the respective contour pixel on the y axis. The blue dashed line indicates the estimated lighting environment, the direction of the dominant light at $-61.5°$ is indicated by the orange vertical line.

4 Color Neutralization

If an object consists of just a single color, then most of the normals around an object can be used for the estimation in Eqn. 2. The example in Fig. 2 illustrated that this is generally not the case for multi-colored surfaces. However, for objects consisting

Fig. 3 Example "teabag2" for intrinsic image decomposition from a ground-truth dataset [8]. Left: input image. Middle: shading component. Right: reflectance component.

of multiple materials, an equivalent statement can be made: if we are able to separate shading from reflectance, estimation of Eqn. 2 can be performed only on the shading component.

4.1 The General Problem: Intrinsic Image Decomposition

In computer vision, the problem of separating shading and reflectance is known as intrinsic image decomposition. Figure 3 shows an example image ("teabag2") for intrinsic image decomposition from a publicly available dataset by Grosse *et al.* [8]. The fact that the shading component is completely free from any textures makes it the optimal input for estimating lighting environments.

Thus, we experimented with several algorithms for intrinsic image decomposition, notably the recent methods by Gehler *et al.* [7] and Shen and Yeo [18]. Mathematically, the task of intrinsic image decomposition is to estimate for each observed pixel intensity one scalar shading component and one vector of reflectances, i. e.,

$$\mathbf{p}(\mathbf{x}) = s(\mathbf{x}) \cdot \mathbf{r}(\mathbf{x}) \ . \tag{4}$$

Thus, for example for an RGB image, each pixel has four unknown variables but only three known variables. To obtain a solution, Gehler *et al.* and Shen and Yeo make use of additional assumptions. Most importantly, the set of distinct reflectances in the scene is assumed to be small. Second, shading is assumed to vary smoothly. Third, the scene contains spatially extended areas of constant or very similar reflectance. Besides these conceptual similarities, Gehler *et al.* and Shen and Yeo chose very different paths to algorithmically exploit these constraints. Gehler *et al.* chose a statistical model using a conditional random field. Reflectance constraints are included via the retinex algorithm [1]. Shen and Yeo use weighted red-black wavelets as a sparse reflectance model. The decomposition task is formulated as a L_1-regularized least squares problem where all constraints are correspondingly incorporated. For both methods, global constraints ensure the consistency of the results on 2-D images.

We used the publicly available implementation by Gehler *et al.* and reimplemented the method by Shen and Yeo. Both methods are computationally demanding, which is why we operated on downsampled versions of our images. Both methods have been shown to work very well on the laboratory ground truth data by

Grosse *et al.*. However, we were not able to transfer the success of these methods from the laboratory data to real-world images. Specifically, we were not able to find a good set of parameters across multiple images such that the 2-D shading component does not fall back to a trivial solution.

4.2 A Specialized Solution for Forensics: Intrinsic Contour Estimation

Upon closer examination, our application does not require full recovery of the 2-D shading image. Instead, as a special case, we require only a 1-D shading contour along the user annotations. Thus, we call the proposed method *Intrinsic Contour Estimation*. For each single material, the intensities vary with the direction of the normals. However, having a perfect intrinsic contour implies that for two identical normals, the intensities are also identical. This constraint can be used to neutralize intensity variations that arise from different colors: assume that we observe two different materials with the same surface normal. Then, we seek a multiplicative factor for one material that levels the intensity difference. This multiplier is then applied to all remaining normals of the material, which effectively yields a shading contour. This procedure is not restricted to two materials, but can be applied to any number. Since, typically, multiple normals overlap, we chose the correction factor that minimizes a least squares fit.

In relation to the aforementioned 2-D algorithms on intrinsic image decomposition, the problem of estimating a 1-D contour is far easier. A big difficulty for the general algorithms is the 2-D segmentation of areas of constant reflectance. This is simpler in 1-D, since piecewise constant materials can be thought as lining up on a 1-D string instead of having to fit a 2-D jigsaw puzzle. Furthermore, in the context of forensics, it may happen that suitable object contours are manually annotated anyways. In such a case it is practically also feasible to even provide user-annotated reflectance boundaries.

More technically, the algorithm works as follows. We first identify clusters of contour pixels that are likely to belong to the same material. This can either be done automatically using, e. g., k-means on the contour colors or a combination of color and spatial proximity [6, page 315]. Alternatively, this can be done manually by specifically assigning the annotated contours to clusters. Without loss of generality, assume that a contour points are clustered by reflectance (or color, respectively) into two sets $\mathscr{U}$ and $\mathscr{V}$. Let $\mathbf{x}_i \in \mathscr{U}$ and $\mathbf{x}_j \in \mathscr{V}$ be two points from these clusters. From the perspective of intrinsic image decomposition, the shading component has to be identical if the normal directions $v(\mathbf{x}_i)$, $v(\mathbf{x}_j)$ of these points are identical. For such pairs of points, it is straightforward to analytically find a multiplicative factor to neutralize the brightness difference between the clusters by solving

$$\begin{pmatrix} m(\mathbf{x_i}) \\ m(\mathbf{x_j}) \end{pmatrix}^{\mathrm{T}} \cdot \mathbf{t} = 0 \ . \tag{5}$$

Since it is unlikely to have normals pointing towards exactly the same direction, normals that are pointing to *almost* the same direction are incorporated with a Gaussian

angular distance weight $w(\mathbf{x}_i, \mathbf{x}_j)$,

$$w(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} \exp\left(\frac{(v(\mathbf{x}_i) - v(\mathbf{x}_j))^2}{\sigma^2}\right) & \text{if } |v(\mathbf{x}_i) - v(\mathbf{x}_j)| \leq 2\sigma \\ 0 & \text{otherwise} \end{cases} , \qquad (6)$$

where σ governs the width of the distribution. In our implementation, we empirically set σ to $18.75°$. The threshold in this equation is derived from a Gaussian probability distribution, where less than 5% of the values assume values that are larger in absolute than 2σ.

We rewrite the constraint in Eqn. 5 more generally for arbitrary numbers of similar normals and arbitrary numbers of materials as

$$W\mathbf{t} = \mathbf{0} , \qquad (7)$$

where $W \in \mathbb{R}^{m \times k}$ for m pairs of overlapping normals and k clusters (materials). Each row of W has two entries that are set as follows. Without loss of generality, assume that the example points $\mathbf{x}_i \in \mathcal{U}$, $\mathbf{x}_j \in \mathcal{V}$ from above are the l-th point pair, and that u and v are the cluster indices (counted from 0) of clusters $\mathcal{U}$, $\mathcal{V}$. Then,

$$W_{l,h} = \begin{cases} w(\mathbf{x}_i, \mathbf{x}_j) m(\mathbf{x}_i) & \text{if } h = u, \quad \mathbf{x}_i \in \mathcal{U} \\ w(\mathbf{x}_i, \mathbf{x}_j) m(\mathbf{x}_j) & \text{if } h = v, \quad \mathbf{x}_j \in \mathcal{V} \\ 0 & \text{otherwise} \end{cases} , \qquad (8)$$

The remaining $m-1$ rows of W are filled analogously with data from the $m-1$ other point pairs. To avoid the trivial solution $\mathbf{t} = \mathbf{0}$, we set $t_1 = 1$, which yields the final solution

$$W'\mathbf{t}' = -\mathbf{d} , \qquad (9)$$

where $W = (\mathbf{d} \ W')$ and $\mathbf{t} = \left(\begin{smallmatrix} 1 \\ \mathbf{t}' \end{smallmatrix}\right)$. Equation 9 can be seen as a least-squares problem, and then directly be solved via singular value decomposition (SVD).

As a sidenote, it is also possible to integrate this scheme directly into the solution of the baseline method in Eqn. 3, and jointly estimate material compensation and lighting environment. However, in our implementation the solutions to the integrated approach was somewhat less stable, hence we omitted this approach here.

5 Database

We capture two sets of ground-truth images to quantify the accuracy of the proposed method. In both sets of images, three primary light sources are installed at defined positions in the scene. Additional light sources provide ambient (background) illumination. For each scenario, we capture three images, where in turns exactly one of the three primary light sources is activated. The ground truth direction of the dominant illuminant is the projected angle between the primary light source and the center of the object.
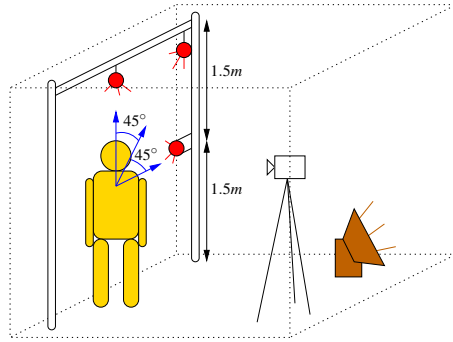
Fig. 4 Experimental setup for the incandescent light dataset. Ambient light is provided by the brown background lamp. Direct illumination (red) on the subjects (yellow) comes from $0°$, $45°$ and $90°$, measured $1.5m$ above the floor.

5.1 Incandescent Light Dataset

Figure 4 shows the experimental setup for the first dataset. The data was captured in a closed room without windows. 10 different subjects were asked to stand with the back to the wall. We use three incandescent lights that act as dominant light sources. The first light was mounted at about breast height exactly to the right (seen from the camera) of the subject, which we denote as a projected angle of $0°$. The second light was mounted top right (seen from the camera), forming a projected angle of about $45°$. The third light was mounted on top of the subject, with a projected angle of about $90°$. A fourth, stronger light source was located behind the camera pointing to the backside wall to provide a floor of scattered environment light. We captured and manually annotated a total of 30 images. Example pictures of this dataset are shown in Fig. 5. The distance of the primary lights to the subjects are about $1.5\,\mathrm{m}$. This is a more challenging scenario than when using direct sunlight, since the method's assumption of parallel rays originating from an "infinitely distant" light source is violated. For the present setup geometry, the error varies between $0°$ and about $9.4°$, with the highest errors occuring for normals a) located at maximum distance to the line connecting the light source and the center of illumination and b) that point in a direction orthogonal to this line. All other normals exhibit a lower error, down to 0. The details of this calculation are in Appendix A. Nevertheless, since the angle between dominant light sources is $45°$ and $90°$, respectively, we found that this deviation is still manageable.

5.2 Flash Dataset

Figure 4 shows the experimental setup for the second dataset[1]. The setup somewhat deviates from the first setup. First, we used camera-synchronized flash lights as dominant light sources. Second, we used a room with windows to provide natural light as

---

[1] Both datasets, are available from our lab's web page `http://www5.cs.fau.de/`.

Fig. 5 Example images from the first dataset. The rightmost subject is shown under the three different illumination directions from 0°, 45°, and 90°.
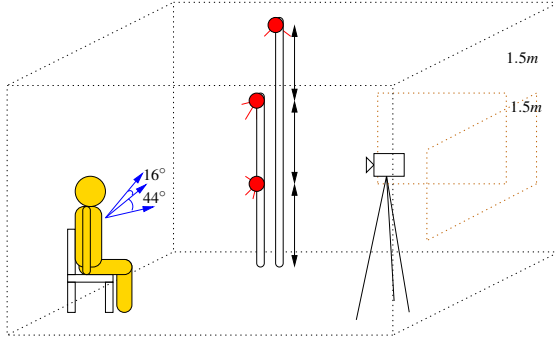

Fig. 6 Experimental setup for the flash dataset. Ambient light is provided by windows in the back of the room (images were captured on an overcast day). Direct illumination (red) on the subjects (yellow) comes from 0°, 44° and 60°, measured 0.95 m above the floor.

background light. The room has a side length of more than 4 m, and the sky was overcast, to prevent the background illumination to dominate the scene. The flash lights were moved slightly in front of the subject to reduce effects from self-shadowing. The projected (2-D) angles of the light sources were at 0°, 44° and 60°. Thus, the angular distances of the light sources were shorter, but the absolute distances in meters between light source and subject were between 1.57 m for the low light and 2.30 m for the uppermost light, which leads to a worst-case angular uncertainty of 9.03° and 6.20°, respectively. We captured 9 scenes under all three different dominant illuminations. To enforce a challenging variability in the the different materials, we used a single, sitting subject that we purposefully dressed in different clothes, with variations of bright and dark garments in various angular directions. We avoided very dark (black) garments to increase the signal-to-noise ratio in the pixels. Example pictures from this dataset are shown in Fig. 7. Note that in this setup, the cast shadow in the background might also indicate the lighting direction. However, we did not use any of such information for this analysis.

6 Evaluation

For evaluation, we use the images from both experimental setups. From each image, we selected contours using a single material and using multiple materials.

Fig. 7 Example images from the second dataset. The three pictures on the right show the three different illumination directions from 0°, 44°, and 60°.

Fig. 8 Images used to demonstrate the dependency on normal direction coverage. The white lines represent the contours and normals, respectively.

The angle of the dominant direction of the incident light is estimated by applying the method in Sec. 3 for single-material contours, or by applying ICE (Sec. 4) and then the single-material method.

There is always more than one possible choice for finding an annotation of single-material contours (see Sec. 6.4 for general comments on the annotation). To be as fair as possible, we always used the a-posteriori best contour. From a practical viewpoint, this might imply that the single-material results are somewhat too optimistic. However, by doing so, we enforce that there is no better choice if operating on only single materials. Thus, all improvements of multi-material contours over single-material contours can be attributed to the benefit of the proposed method.

In Sec. 6.1, we first illustrate that estimating a lighting environment from normals with a limited angular range is unstable. In Sec. 6.2, we report results on the first dataset. In Sec. 6.3, we report results on the second dataset. We add a brief discussion on the results in Sec. 6.4.

6.1 Estimation of Light Direction from a Limited Angular Range

We first assess the effect of a limited angular range of normals on the estimation of the dominant lighting direction. We created a generic test case from the three images shown in Figure 8. Here, the same light was used (coming from the right, at an angle of 44°), but the blue sweater was moved to different sides of the body. We exclusively extract contours from the sweater to track the behavior of the lighting direction estimates. From left to right, the estimated angular directions are 73.8°, 84.7°, and 33.0°, respectively. Thus, the differences to the true light direction are +29.8°, +40.7°, and −11°. In order to better understand where this variation comes from, we show in Figure 9 (top and bottom left) plots of the data that is used for the estimation. Along the x-axis is the direction of the normal in degrees, i. e., 0 denotes normals pointing to the right, 90° denotes normals pointing upwards. The y-axis represents the pixel intensities per normal. Additionally, a smooth solid line is plotted which shows the intensity distribution as a result of fitting the data to a lighting environment. It can be seen that a sparse angular range by tendency draws the fitted curve towards the direction where the observations happen. Thus, if only normals are selected that point towards the direction of the light source, a perfect estimate can be obtained "by luck". If only normals are selected that point away from the light source, the result can become arbitrarily bad. This shows that if only normals from a limited angular range are available, there is a large room for estimation bias.

Conversely, in Fig. 9 (bottom right), we plotted the normals from all three images into a single diagram. The angular range is now much more widely covered, which implies tighter boundaries of the estimated lighting environment. Indeed, when we estimated the direction light, we obtained on this rich vector a lighting direction of 43.2°, i. e., the error to the ground truth is below one degree.

6.2 Evaluation on the Incandescent Light Dataset

The first set of results is computed on the incandescent light dataset. We compare the single-material estimation by Johnson and Farid [9] (denoted as "Original") to the proposed ICE multi-material estimation (denoted as "ICE"). Quantitative results are presented in Tab. 2. In the first row, we used the original method only on single-material contours. In the second row, we used the original method on multi-material contours. In the third row, we applied ICE to multi-material contours. We compute the median error, the mean error, and the number of cases where a method is able to distinguish two different lighting environments. Since all light sources are in a distance of 45°, we counted the cases where the absolute error was less than 22.5°.

Quite expected, the original method breaks on multi-material contours. ICE slightly improves the mean error, and is able to solve one additional case within 22.5°. However, the original method still achieves the best mean value. While it can be seen that ICE gently integrates multiple materials, it is not apparent that there is a significant benefit with respect to the final outcome of the forensic analysis. Upon analyzing these results, we found that for this dataset, the original method actually greatly benefits from estimation bias illustrated in the previous Section: although the angular
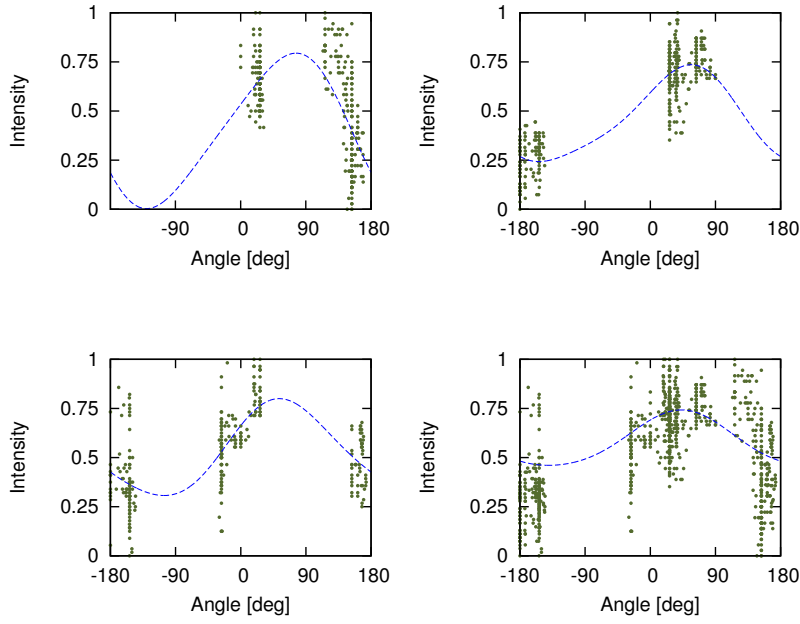
Fig. 9 Contour intensity as a function of the normal orientation. Top: intensity plots from the contours at the left and middle of Fig. 8. Bottom left: intensity plot from the contours at the right of Fig. 8. Bottom right: intensity plot where all three parts have been combined together.

	Median	Mean	Within 22.5°
Original (single-material contour)	**10.7**	13.6	25/30 (83%)
Original (multi-material contour)	40.2	56.5	10/30 (33%)
ICE (multi-material contour)	12.6	**13.0**	**26/30 (86%)**

Table 2 Results on the incandescent light dataset. See text for details.

range of each single material is limited, there is very often a single good contour where the normals point towards the light source, i. e., many single-contour results are "lucky". The next Section shows that this is not generally the case.

6.3 Evaluation on the Flash Dataset

The second set of results is computed on the flash dataset. This dataset is designed to be more comprehensive with respect to body poses, occlusions, and varying materials at varying locations. We use the same evaluation protocol as in the previous section, with one notable exception. Since the light sources were located at $0°$, $44°$, and $60°$, the angles between the light sources are not equidistant anymore. Thus, we counted correct attribution of a lighting environment for the first and last light if the absolute

	Median	Mean	Correctly attributed
Original (single-material contour)	15.1	16.2	15/27 (56%)
ICE (multi-material contour)	**7.8**	**11.6**	**19/27 (70%)**

Table 3 Results on the flash dataset. See text for details.

error is within $22°$ and $8°$, respectively, and for the middle light if the estimate is located in the interval within. Note that these are much tighter boundaries than in the previous scenario, which also explains the lower percentage of correct attributions.

Results for the original method on the best found single-material contours, and for the proposed method ICE on multi-material contours are shown in Tab. 3. Here, the benefit of multi-material contours becomes apparent. The median error of ICE is about half of the median error of the original method. The mean error is almost 30% lower. Conversely, the number of correctly attributed lighting situations is considerably higher.

6.4 Discussion

These controlled experiments show that ICE allows to integrate contours of different materials. Although incorporating ICE means that the processing pipeline for finding the dominant illuminant is extended by another step, the (unavoidable) estimation errors do not increase, even in a close-to best-case scenarios for the single-material method, like when using the first dataset.

The biggest impact of ICE is that it allows to use contours over a much wider angular range. The two main advantages are a) the increased robustness (and hence reduced estimation bias) of the baseline method, and b) the general applicability of the method for unconstrained images, where it is all to often extremely difficult to find a good set of single-material contours. Thus, in principle, the proposed method is applicable to the same scenarios as the method by Johnson and Farid [9]. However, our method makes the estimation numerically more robust, with the added benefit that there are somewhat less stringent color constraints on the scene. Thus, the advantage over the previous method is not that it allows to process a completely new class of forgeries, but that its performance declines more gently over a large number of increasingly difficult (in a numerical sense) examples.

Still, the choice of good contours can be tricky at times. To our experience, a reasonably good strategy for extracting contour information is to annotate not exactly the object contour, but instead a line that is parallel to the actual object contour but two or three pixels inside the object, to avoid noise from edge interpolation. Folds or crumbles in the clothes also show up as noise, and should be excluded from the contour, as well as self-shadows. Generally, hair, beard, and metallic (or generally clearly non-Lambertian) surfaces should be avoided as well. A single contour segment does not need to be long, as few pixels (seven, in our implementation) suffice for computing the normal. The overall number of pixels in a contour may be low, but it is crucial that there are contour pixels with similar normals across different materials. If the

number of pixels across clusters is massively inbalanced, it may help (although we did not explicitly test that) to resample the large clusters. For reference and comparison, we provide our contours together with the dataset for download from our web page.

The theoretical model for estimating lighting environments relies on parallel light rays, which is only the case for an infinitely distant light source (or, approximately, the sun). This implies that in theory the method is only applicable to objects exposed to direct sunlight, which brings practical complications for a quantitative performance evaluation. Computer-generated scenes are also far from straightforward to set up for simulating all effects of real images, including image noise, crumbles in clothing, or even physically-real material reflectances. Thus, we decided for capturing real-data in an indoor environment, although near-light sources propagate in a cone beam. Our calculation in Appendix A shows that the introduced model error is not negligible, but indeed manageable, which is also confirmed by the experiments.

All in all, using ICE as color normalization for the classical lighting estimator greatly increases the applicability and robustness of illumination tools for forensic analysis. In future work, it may be worth investigating other objective functions for optimization than the least-squares approach, and to quantify the impact of assuming Lambertian reflectance for various materials.

7 Conclusions

Estimating lighting environments for a physics-based forensic image analysis requires to compute a representative set of "good" normals, i.e., normals that match the underlying physical model. In practice, it turns out that normal selection is very challenging, and even impossible at times.

In this work, we address one of the biggest limitations in the physical model of 2-D lighting estimation: the requirement that all such normals have to be located on the same material. We propose a method that compensates different materials, which we call *Intrinsic Contour Estimation* (ICE). We exploit the fact that normals with the same orientation, but different underlying materials, have to provide the same intensity contribution to the estimation of the lighting environment. We captured two quantitative ground-truth datasets to evaluate the efficacy of ICE. It turns out that lighting estimation can greatly benefit from material neutralization, a) for reducing estimation bias due to a limited angular range of single-material normals, b) for increasing the overall accuracy of estimating the direction of the dominant light source.

A Intrinsic Dataset Error due to Indoor Illumination

We perform an error calculation to illustrate the effect of a light source at finite distance, as used for capturing both datasets. A light source at a finite distance effectively casts a cone of light. Assuming that object normals are selected from a light cone of 50 cm diameter around the object center, only the central

ray of the cone is with exactly the same angle incident to the object as if the rays were parallel. The maximum angular deviation of a ray incident to the object is

$$\phi_{\text{max,incand}} \tan^{-1}(25\,\text{cm}/150\,\text{cm}) = 9.46° \quad , \tag{10}$$

which occurs at the outer boundary of the cone. Under the method's assumption of Lambertian reflectance, this error propagates into the cosine between the ray direction and the surface normal. The derivative (and hence variation) of the cosine is maximum at $\pi/2$, with $\sin(\pi/2) = 1$. Distributing the error symmetrically around $\pi/2$ leads to a upper bound of the intrinsic dataset error of

$$e_{\text{orth,incand}} = 2 \cdot \cos(\pi/2 - (9.46°/2)) = 9.44° \quad . \tag{11}$$

This worst case bound occurs if a normal is located at the cone-beam boundary and is directed orthogonally to the light source. If a normal lies at the cone beam boundary but points towards the light source, the intrinsic dataset error is negligible, i. e.,

$$e_{\text{towards,incand}} = 1 - (2 \cdot \cos(9.46°/2)) = 0.39° \quad . \tag{12}$$

Similarly, the closer a normal is to the center of the cone-beam, the smaller is its estimation error. In other words, normals that are orthogonal to the light source *and* located at the cone boundary exhibit the same intensities as if the light source-contour angle were $90° + 9.44° \approx 100°$ degrees under an infinitely distance light source. All normals that do not adhere to both of these conditions exhibit a lower error, with minima at the center of the cone-beam or whenever normals are parallel to the central ray of the light source.

Performing the same calculation for the flash dataset yields $\phi_{\text{max,flash}} = 9.04°$ for the closest ("0°") light source, and $\phi_{\text{max,flash}} = 6.20°$ for the most distant ("60°") light source. The maximum errors $e_{\text{orth,flash}}$ are therefore $9.03°$ and $6.20°$, respectively. The minimum errors $e_{\text{towards,flash}}$ are $0.36°$ and $0.17°$, respectively.

References

1. Brian V. Funt Mark S. Drew, M.B.: Recovering shading from color images. In: European Conference on Computer Vision, pp. 124–132 (2005)
2. Conotter, V., O'Brien, J.F., Farid, H.: Exposing Digital Forgeries in Ballistic Motion. IEEE Transactions on Information Forensics and Security **7**(1), 283–296 (2012)
3. de Carvalho, T.J., Riess, C., Angelopoulou, E., Pedrini, H., Rocha, A.: Exposing Digital Image Forgeries by Illumination Color Classification. IEEE Transactions on Information Forensics and Security **8**(7), 1182–1194 (2013)
4. Fan, W., Wang, K., Cayre, F., Xiong, Z.: 3D Lighting-Based Image Forgery Detection using Shape-from-Shading. In: Proceedings of the 20th European Signal Processing Conference (EUSIPCO-2012), pp. 1777–1781. Bucarest, Romania (2012)
5. Farid, H.: Digital Image Forensics. `http://www.cs.dartmouth.edu/farid/downloads/tutorials/digitalimageforensics.pdf` (2011)
6. Forsyth, D.A., Ponce, J.: Computer Vision — A Modern Approach. Pearson Education Inc. (2003)
7. Gehler, P.V., Rother, C., Kiefel, M., Zhang, L., Schölkopf, B.: Recovering Intrinsic Images with a Global Sparsity Prior on Reflectance. In: Advances in Neural Information Processing Systems (NIPS 2011), vol. 24, pp. 765–773. Granada, Spain (2011)
8. Grosse, R., Johnson, M., Adelson, E., Freeman, W.: Ground Truth Dataset and Baseline Evaluations for Intrinsic Image Algorithms. In: Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV 2009), pp. 2335–2342. Kyoto, Japan (2009)
9. Johnson, M., Farid, H.: Exposing Digital Forgeries in Complex Lighting Environments. IEEE Transactions on Information Forensics and Security **2**(3), 450–461 (2007)
10. Kee, E., Farid, H.: Exposing Digital Forgeries from 3-D Lighting Environments. In: Proceedings of the 2nd IEEE International Workshop on Information Forensics and Security (WIFS 2010). Seattle, WA, USA (2010)
11. Kee, E., O'Brien, J., Farid, H.: Exposing Photo Manipulation with Inconsistent Shadows. ACM Transactions on Graphics **32**(4), 28:1–12 (2013)

12. Kee, E., O'Brien, J., Farid, H.: Exposing Photo Manipulation from Shading and Shadows. ACM Transactions on Graphics **33**(5), 165:1–21 (2014)
13. Ostrovsky, Y., Cavanagh, P., Sinha, P.: Perceiving Illumination Inconsistencies in Scenes. Perception **34**(11), 1301–1314 (2005)
14. Peng, B., Wang, W., Dong, J., Tan, T.: Improved 3D lighting environment estimation for image forgery detection. In: Proceedings of the 7th IEEE International Workshop on Information Forensics and Security (WIFS 2015). Rome, Italy (2015)
15. Redi, J., Taktak, W., Dugelay, J.L.: Digital Image Forensics: A Booklet for Beginners. Multimedia Tools and Applications **51**(1), 133–162 (2011)
16. Riess, C., Angelopoulou, E.: Scene Illumination as an Indicator of Image Manipulation. In: Proceedings of the 12th International Conference on Information Hiding (IH 2010), vol. Lecture Notes in Computer Science 6387, pp. 66–80. Calgary, AB, Canada (2010)
17. Riess, C., Pfaller, S., Angelopoulou, E.: Reflectance Normalization in Illumination-Based Image Manipulation Detection. In: International Workshop on Recent Advances in Digital Security: Biometrics and Forensics, pp. 3–10 (2015)
18. Shen, L., Yeo, C.: Intrinsic Images Decomposition Using a Local and Global Sparse Representation of Reflectance. In: Proceedings of the 24th IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2011), pp. 697–704. Colorado Springs, CO, USA (2011)
19. Zhang, W., Cao, X., Qu, Y., Hou, Y., Zhang, C.: Detecting and Extracting the Photo Composites Using Planar Homography and Graph Cut. IEEE Transactions on Information Forensics and Security **5**(3), 544–555 (2010)