Reflectance Normalization in Illumination-Based Image Manipulation Detection

Christian Riess, Sven Pfaller, and Elli Angelopoulou

Pattern Recognition Lab, Friedrich-Alexander University, Erlangen, Germany
christian.riess@fau.de,
http://www5.cs.fau.de/~riess

Abstract. One approach to detect spliced images is to compare the lighting environment of suspicious objects or persons in the scene. The original method, proposed by Johnson and Farid, requires an investigator to mark occluding contours of multiple objects, from which the distribution of the incident light intensity is estimated. Unfortunately, this method imposes relatively strict constraints on the user and on the scene under investigation.

In this work, we propose a color-normalization approach to relax one important constraint. With our modification, a user is able to select the contours from multiple different materials (instead of having to use a single material). The proposed method will automatically compensate the differences in the reflected intensities. We demonstrate the robustness of the method with a carefully designed ground-truth dataset, consisting of 10 subjects, each of them under 3 controlled lighting environments. With the proposed method, lighting direction as a forensic cue becomes applicable to a much wider range of natural images.

1 Introduction

The goal of blind image forensics is to verify the authenticity and origin of an image without the support from an embedded security scheme. With the increasing availability of digital imagery and image processing software, researchers developed a family of forensic algorithms that either a) detect traces of manipulation in an image or b) verify characteristic scene or imaging properties to determine its authenticity. For an overview of existing methods, please refer to [10, 3].

Several forensic algorithms aim to exploit the physics of geometry in the scene. For instance, to verify the consistency of a scene containing people, Johnson and Farid [7] proposed to exploit the position of specular highlights in the eyes. Zhang *et al.* [12] investigated the shadows of objects on planar surfaces for detecting spliced images. Also based on geometry, Conotter *et al.* [1] proposed an algorithm for verifying ballistic motion in video captures. For a more complete overview on physics-based approaches, please refer, e. g., to [3].

One potentially powerful approach, based on illumination geometry, has been proposed by Johnson and Farid [6]. A user has to annotate the contour of persons

of interest (or objects, respectively[1]). Based on the intensity distribution along the contours, the brightness distribution of the incident light can be estimated as a function of the angle of incidence. This distribution is computed in the image plane, i. e. in two dimensions. It acts as a descriptor for the lighting environment of a person. Thus, if a spliced image contains persons from two different source images, it is likely that their illumination environments also differ. In [8], this approach is extended to three dimensions. However, in three dimensions, this method requires known 3D geometry of the persons under investigation. This leads to another estimation step (for fitting a 3D model), adding complexity and potential sources of error. Recently, Fan *et al.* [2] proposed an alternative to this approach by replacing the estimation of a 3D surface model with a shade-from-shading algorithm.

For humans, assessing lighting environments is a difficult task [9]. Computers can quantify the perceived deviation, or even detect differences that are imperceptible for humans. Additionally, concealing illumination differences in spliced images might force a forger to manually repaint parts of the image, which raises the effort to create a plausible forgery.

In spite these encouraging prospects, these algorithms are not straightforward to apply in practice. The lighting environment can only be estimated from solid, purely diffuse materials. The surface normals of the regions under investigation must exhibit a large variety of directions. Additionally, all marked regions must consist of the same material.

In this paper, we propose a straightforward approach to relax the last constraint, i. e., to be able to estimate lighting environments on mixed materials. We focus on the 2D algorithm by Johnson and Farid [6] to avoid the additional requirement of 3D object geometry. We restate the baseline algorithm in Sec. 2. In Sec. 3, we present the proposed algorithm, which we call *Intrinsic Contour Estimation*. In Sec. 4, we first present our evaluation dataset, and then provide the results of our algorithm. We conclude our work in Sec. 5.

2 Forensic Exploitation of Lighting Environments

We restate the algorithm by Johnson and Farid (for additional details and a full derivation of the equations, please refer to the original work [6]).

Assume that an image contain two persons of interest. A user marks the contours of these persons, satisfying several constraints. The 2D contour in the image must correspond to a (true) 3D contour in the scene. Then, surface normals on this contour approximately lie in the image plane. Contours can be piecewise defined, and must everywhere be exposed to the environment light (i. e., regions of locally cast shadows are not admissible). For estimating the direction of the surface normals along these contours, it suffices to fit a 2D polynomial to the contour. The intensity of each point along the contour is extrapolated from the surrounding pixels (for details, confer [6]).

[1] Without loss of generality, we assume in this paper that persons are in the focus of interest.

The lighting environment is modelled as a weighted sum of spherical harmonics. In 2D, using spherical harmonics of up to order 2, only five coefficients need to be estimated. Let

$$\begin{aligned} Y_{0,0}(\boldsymbol{\nu}) &= \tfrac{1}{\sqrt{4\pi}} \\ Y_{1,-1}(\boldsymbol{\nu}) &= \sqrt{\tfrac{3}{4\pi}}y & Y_{1,1}(\boldsymbol{\nu}) &= \sqrt{\tfrac{3}{4\pi}}x \\ Y_{2,-2}(\boldsymbol{\nu}) &= 3\sqrt{\tfrac{5}{4\pi}}xy & Y_{2,2}(\boldsymbol{\nu}) &= \tfrac{3}{2}\sqrt{\tfrac{5}{12\pi}}(x^2-y^2) \end{aligned} \tag{1}$$

denote the five spherical harmonics basis functions $Y_{0,0}$ through $Y_{2,2}$ that are required to model a lighting environment in 2D, depending on a normal vector in the image plane $\boldsymbol{\nu} = (x \;\; y)^{\mathrm{T}}$. If the reflectance of an object material is purely diffuse (Lambertian), an intensity along an object boundary corresponds simply to a linear combination of the basis functions. Thus, for a contour of n points, the basis functions can be evaluated in a matrix $\boldsymbol{M} \in \mathbb{R}^{n\times 5}$. The unknown weighting factors $\boldsymbol{h} \in \mathbb{R}^5$ must then satisfy

$$\boldsymbol{M}\boldsymbol{h} = \boldsymbol{b} \;\; , \tag{2}$$

where $\boldsymbol{b} \in \mathbb{R}^n$ contains the (grayscale) intensities along the contour. Instead of solving Eqn. 2 directly for $\boldsymbol{h}$, an energy function $E(\boldsymbol{h})$ is defined to incorporate a regularization term $\boldsymbol{C} \in \mathbb{R}^{5\times 5}$, $\boldsymbol{C} = \mathrm{diag}(1 \;\; 2 \;\; 2 \;\; 3 \;\; 3)$:

$$E(\boldsymbol{h}) = \|\boldsymbol{M}\boldsymbol{h} - \boldsymbol{b}\|^2 + \lambda_1 \|\boldsymbol{C}\boldsymbol{h}\|^2 \;\; , \tag{3}$$

where the strength of the influence of $\boldsymbol{C}$ is determined by a parameter λ_1. Minimizing Eqn. 3 for $\boldsymbol{h}$ yields

$$\boldsymbol{h} = (\boldsymbol{M}^{\mathrm{T}}\boldsymbol{M} + \lambda_1 \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C})^{-1}\boldsymbol{M}^{\mathrm{T}}\boldsymbol{b} \;\; . \tag{4}$$

Equations 2 and 4 transfer the intensity distribution along the object boundary to a basis of spherical harmonics. In Eqn. 2, differences in contour brightness are assumed to result from differences in the incident illumination. If contours of highly contrasting surface materials, e. g. a black T-shirt and light skin, are both used in Eqn. 2, the computation is severely biased. An example for such a situation is shown in Fig. 1.

3 Reflectance Normalization as a Preprocessing Step

We investigated methods to normalize material brightnesses prior to the estimation of the lighting environment. The separation of object texture and shading is commonly referred to as intrinsic image decomposition. For our task, the shading component is the ideal input for Eqn. 2. We experimented with the recent algorithms by Gehler *et al.* [5] and Shen and Yeo [11]. Although these methods showed encouraging performances on laboratory images, we failed to transfer this performance to real-world images. In particular for large brightness differences, we were not able to obtain satisfying shading components.

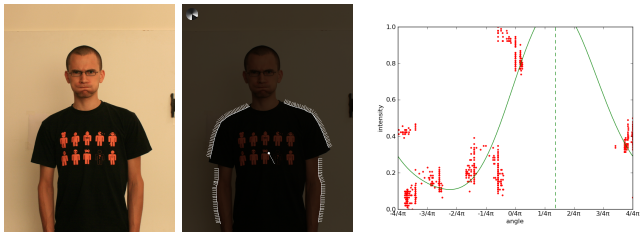

Fig. 1. Illustration of mixed-material contours: the brightness contrast between the black T-shirt and bright skin prevents cross-material estimation of the lighting environment.

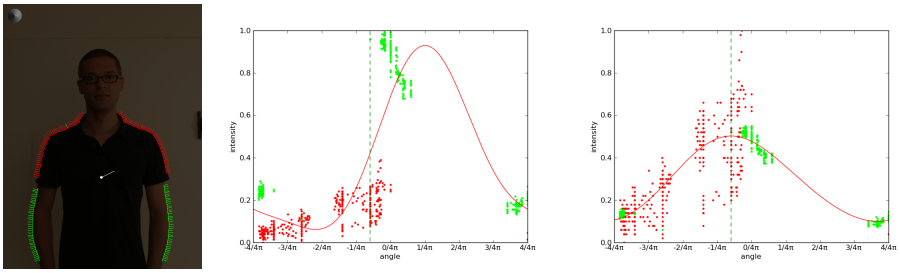

Fig. 2. Idea of intrinsic contour decomposition: contours of different materials (marked in red and blue) can be jointly used by adjusting the reflectance contributions. Left: input image. Middle: uncorrected intensity distribution estimates light to be coming from the right. Right: corrected intensity distributions estimates light to be coming from top.

However, this particular forensic application offers additional constraints compared to what is typically assumed in intrinsic image decomposition: we only need to operate along user-annotated object contours. Thus, the surface normals along our pixels of interest are known. We exploit this fact in a novel algorithm, which we call *intrinsic contour estimation*. Figure 2 illustrates the basic idea. On the left, one of our evaluation subjects is shown. We marked the contour across the black T-shirt, as well as the bright skin. On the right, we plotted the contours of skin regions in green, and shirt regions in red as a function of the normal direction of the contour point. We can reasonably demand that contour points that are pointing in the same direction should exhibit the same brightness. Thus, we search a neutralization factor $\boldsymbol{r}$ that best equalize the intensities of both clusters for points that face in *the same normal direction*. In Fig. 2 (right), this means that multiplication by $\boldsymbol{r}$ should lead to equality of the intensities in the matching blue circles.

The details of the algorithm are outlined below. Assuming purely diffuse (Lambertian) reflectance, let

$$\boldsymbol{p} = \int_{\lambda \in \Omega} \rho(\lambda) e(\lambda) \boldsymbol{c}(\lambda) d\lambda \tag{5}$$

denote the captured color in a pixel $\boldsymbol{p}$, where Ω denotes the visible spectrum of light waves λ, $\rho(\lambda)$ the object color (albedo), $e(\lambda)$ the intensity of the light source, and $\boldsymbol{c}(\lambda)$ a three-component vector of color matching functions of the camera (which ultimately yield the red, green and blue color channels). Assuming that $\boldsymbol{c}(\lambda)$ are linear functions, a change in the material $\rho(\lambda)$ affects the observed colors in p multiplicatively. Thus, to neutralize the distorting effect of different surface materials, we are seeking a multiplicative correction term r_j for each surface material.

Given a contour of k materials, we extend the existing algorithm with a brightness normalizing factor. The colors of the contour pixels are clustered into multiple materials, either automatically using, e. g., the k-means algorithm [4, page 315] or manually by the operator while the contours are marked.

For two intensities $p_u(\boldsymbol{\nu})$, $p_v(\boldsymbol{\nu})$ from different clusters u, v with the same normal direction $\boldsymbol{\nu}$, we seek $\boldsymbol{r}$, such that the condition

$$(p_u(\boldsymbol{\nu}) \ - p_v(\boldsymbol{\nu})) \cdot \boldsymbol{r} = 0 \tag{6}$$

is satisfied. For increased numerical robustness, we relax the constraint of identical normals to just *similar* normals. To account for the angular difference, we compute for each pair of points a weighting factor w,

$$w(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) = \begin{cases} \exp\left(\frac{\delta^2}{\sigma^2}\right) & \text{if } \delta \leq 2\sigma \\ 0 & \text{otherwise} \end{cases} \ . \tag{7}$$

Here, $p_u(\boldsymbol{\nu}_1)$ and $p_v(\boldsymbol{\nu}_2)$ denote points from clusters u, v with similar normals $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$, and $\delta = \arccos(\boldsymbol{\nu}_1^{\mathrm{T}} \boldsymbol{\nu}_2)$. In our implementation, we empirically set σ to 18.75°. The threshold for $w = 0$ is derived from a Gaussian filter[2]. Hence, the generalized constraint for $\boldsymbol{r}$ using m pairs of data points on k clusters is

$$\boldsymbol{W}\boldsymbol{r} = \boldsymbol{0} \ , \tag{8}$$

where $W \in \mathbb{R}^{m \times k}$. Let the j-th pair of points $p(\boldsymbol{\nu}_1)$ and $q(\boldsymbol{\nu}_2)$ be from clusters i_1, i_2. Then,

$$W_{j,i_1} = w(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) p(\boldsymbol{\nu}_1) \ , \tag{9}$$

$$W_{j,i_2} = w(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2) q(\boldsymbol{\nu}_2) \ . \tag{10}$$

All other entries of W are set to 0. To avoid the trivial solution $\boldsymbol{r} = \boldsymbol{0}$, we set $r_1 = 1$, which yields the final solution

$$\boldsymbol{W}'\boldsymbol{r}' = -\boldsymbol{l} \ , \tag{11}$$

[2] Integrating the tails of a Gaussian with standard deviation σ outside a range of 2σ yields about 5% of the overall area under the curve

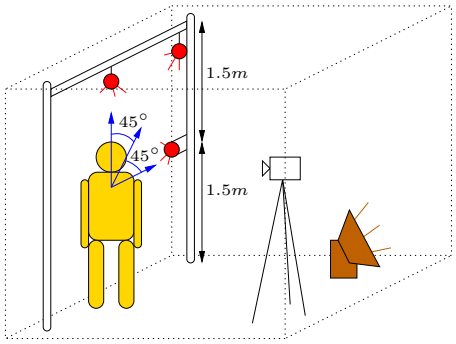

Fig. 3. Experimental setup. Ambient light is provided by the brown background lamp. Direct illumination (red) on the subjects (yellow) comes from 0°, 45° and 90°, measured $1.5m$ above the floor.

where $\boldsymbol{W} = (\boldsymbol{l} \;\; \boldsymbol{W}')$ and $\boldsymbol{r} = \begin{pmatrix} 1 \\ \boldsymbol{r}' \end{pmatrix}$. Equation 11 is solved via singular value decomposition (SVD).

4 Evaluation

To demonstrate the feasibility of the proposed approach, we captured a benchmark dataset where the direction of the dominant illuminant is known[3]. This enables us in the quantitative evaluation to compare the angles of the maximum incident light.

4.1 Data

Figure 3 shows our experimental setup. We captured 30 images consisting of ten subjects under three lighting conditions. In a closed room without windows, one light source was set up for scattered background illumination. Further light sources were fixated at angles of roughly 0°, 45° and 90° to the person in the scene. These lights act as "dominant" light sources, with a distance of only about $1.5m$ from the person. Note that such a close light violates the assumption of the original method [6] of an infinitely distant light source. Indeed, when validating the performance of our re-implementation of the algorithm by Johnson and Farid, we noticed a worse performance than reported in [6], which may partly be due to the violation of this constraint. However, we consider a light source at finite distance to be a reasonable compromise between the theoretical requirements of the algorithm and a practical setup, e.g., for indoor applications.

4.2 Experiments

In our evaluation, we compared three methods. First, denoted as "Original", we used the method by Johnson and Farid [6] (cf. Sec. 2). Second, denoted as

[3] The dataset is publicly available at `http://www5.cs.fau.de/`

	Single-colored contour		
	Median	Mean	Within 22.5°
Original	10.7	13.6	25/30 (83%)
Gehler	9.1	12.5	**26/30 (86%)**
ICE	10.9	14.1	24/30 (80%)

	Multi-colored contour		
	Median	Mean	Within 22.5°
Original	40.2	56.5	10/30 (33%)
Gehler	33.0	50.7	13/30 (43%)
ICE	12.6	13.0	**26/30 (86%)**

Table 1. Median and mean angular error on the lighting environment database, and the number images for which the estimation error of the dominant light direction was less than 22° degrees. In the left columns, the best single-colored contour per image is used, in the right columns, mixed-color contours are used.

"Gehler", we used the input images to compute intrinsic images with the method by Gehler *et al.* [5], which, in theory, should isolate the shading component. We then applied the method by Johnson and Farid on the shading component. Finally, denoted as "ICE" (intrinsic contour estimation), we evaluated the proposed algorithm. Table 1 shows the results of this evaluation. In the first three rows, we report results when selecting only the best, single-material contour of our database subjects. Per method, the median and mean angular error of the estimated dominant light direction are shown. We also (somewhat arbitrarily) set a binary threshold of 22.5° (one eighth of a circle), and counted for how many test cases the dominant light direction was estimated within 22.5° degrees of the ground truth.

For single-colored contours, all methods achieve similar performance, with a median angular deviation between 9.1° (Gehler) and 10.9° (ICE). Note that, in theory, Gehler's intrinsic image decomposition should not change a single-colored contour at all. Yet, it affects to some degree the estimated shading image, which leads to a slightly better result for this method.

The situation is quite different in case that we select multi-colored contours. While (expectedly) the method by Johnson and Farid can not deal with this situation, it turns out that also the Gehler's intrinsic image decomposition is not really able to produce a shading image that yields good results. Only the proposed intrinsic contour estimation (ICE) is able to maintain the performance of the single-colored contour case.

5 Conclusion

Johnson and Farid proposed a pioneering method for exploiting inconsistencies in the lighting direction for forensic applications. Unfortunately, the original method imposes a number of relatively strict constraints to the user. Particularly, in order for the method to be applicable, it is necessary that the objects under examination exhibit a wide angular range of a same-material occluding contour. In practice, such a contour oftentimes does not exist.

In this work, we propose a slight extension to the original method, which relaxes the requirement on the contour material. We use a ground-truth dataset

containing people, where illuminants were located at 0°, 45° and 90° in the scene. On this ground-truth data, we demonstrate that the proposed intrinsic contour estimation method can reliably compensate multi-material contours.

Acknowledgment

The authors gratefully acknowledge support of the Erlangen Graduate School of Heterogeneous Image Systems (HBS) by the German National Science Foundation (DFG).

References

1. 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 (Feb 2012)
2. 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 (Aug 2012)
3. Farid, H.: Digital Image Forensics. `http://www.cs.dartmouth.edu/farid/downloads/tutorials/digitalimageforensics.pdf` (Jun 2011)
4. Forsyth, D.A., Ponce, J.: Computer Vision — A Modern Approach. Pearson Education Inc. (2003)
5. 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 (Dec 2011)
6. Johnson, M., Farid, H.: Exposing Digital Forgeries in Complex Lighting Environments. IEEE Transactions on Information Forensics and Security 2(3), 450–461 (Sep 2007)
7. Johnson, M., Farid, H.: Exposing Digital Forgeries through Specular Highlights on the Eye. In: Proceedings of the 9th International Workshop on Information Hiding (IH 2007). vol. Lecture Notes in Computer Science 4567, pp. 311–325. Saint Malo, France (2007)
8. 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 (Dec 2010)
9. Ostrovsky, Y., Cavanagh, P., Sinha, P.: Perceiving Illumination Inconsistencies in Scenes. Perception 34(11), 1301–1314 (Nov 2005)
10. Redi, J., Taktak, W., Dugelay, J.L.: Digital Image Forensics: A Booklet for Beginners. Multimedia Tools and Applications 51(1), 133–162 (Jan 2011)
11. 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 (Jun 2011)
12. 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 (Sep 2010)