Gradient-Based Illumination Description for Image Forgery Detection

Falko Matern, Christian Riess, *Senior Member, IEEE*, Marc Stamminger

***Abstract*—The goal of blind image forensics is to determine authenticity and origin of an image without using an explicitly embedded security scheme. Most existing forensic methods can roughly be grouped into statistical and physics-based approaches. Statistical methods can oftentimes be fully automated, and achieve impressive results on current state-of-the-art benchmarks. Physics-based methods explain image inconsistencies using an analytic model, and are more robust to common image processing operations such as resizing or recompression.**

In this work, we propose a physics-based forensic descriptor to characterize 2-D lighting environments of objects. The key idea is that the integral over a gradient field of an object indicates the direction of incident light in the image plane. In contrast to prior 2-D lighting methods, the proposed method is remarkably robust to changes in object color and variations in user input, as it operates on the whole object area instead of object contours. Furthermore, we show that the proposed method is unaffected by image resizing or compression, which makes it possible to analyze images that are impossible to analyze with current state-of-the-art statistical methods.

***Index Terms*—Image forensics, lighting direction, image gradient, physics-based methods, manipulation detection**

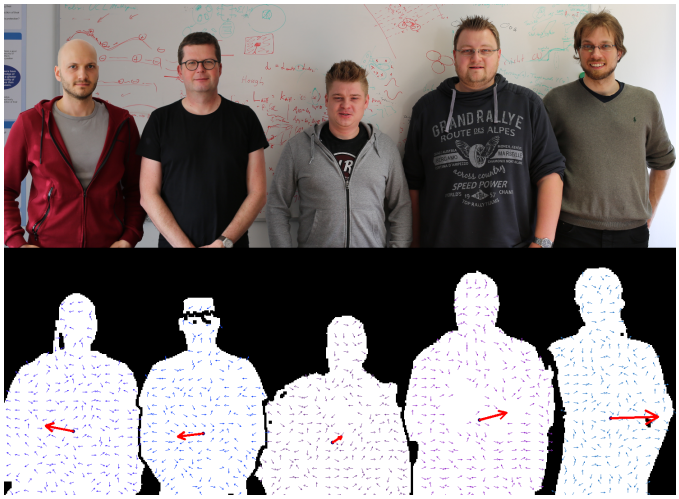


Figure 1. Estimation of the main illumination direction. Image intensity gradients for each masked person are computed and function as input for an estimation of the main light direction. The illumination information can be used as physics-based cue to expose spliced images. The two persons on the left origin from an image with different illumination as indicated by the estimated main light direction.

I. Introduction

IMAGES are now a central element of communication with the wide-spread availability of affordable acquisition devices such as smartphones, and the ease of sharing these images over the internet. At the same time, sophisticated image editing tools make it straightforward to create convincing image manipulations. Such manipulations might not be obvious to an observer [1]. The goal of image forensics is to provide algorithmic tools to detect image manipulations [2], [3].

Most methods for image forensics are statistical. For example, the fixed pattern noise of a camera can be used to associate an image to a unique source device [4]. Other methods analyze for example JPEG compression artifacts [5], [6], traces of resampling [7], or summarized noise statistics [8] based on steganographic descriptors [9]. Recent works perform similar tasks with neural networks, e.g., to condition noise patterns on EXIF entries [10] or to directly search for noise inconsistencies [11], [12]. However, it is still an open challenge to reliably apply statistical methods to recompressed and downsampled data, e.g., from social media or news platforms. In this case, subtle telltales from inter-pixel differences are washed out, with the consequence that detection performance sharply falls off.

Physics-based methods are more resilient to recompression and downsampling. These methods search for physical inconsistencies in images to detect forgeries, like for example in the direction of incident light [13], [14] and the resulting shadows and object shading [15], [16], or the spectral distribution of light [17]. These methods operate on coarse color or intensity distributions of whole image areas, and are hence much less affected by compression and downsampling.

In this work, we propose a physics-based method that validates the consistency of incident light on pairs of objects in the 2-D image plane. This task has been addressed in previous works in the field [13], [18]. However, as we will discuss in the related work in Sec. II, these earlier works make very restrictive assumptions on the reflectance of the objects surface, to the extend that they can only be used by a time-intensive, exact manual annotation of the objects, and on a very small set of scenes that exactly satisfy their photometric assumptions. As a consequence, these earlier methods exhibit excellent performance on quite controlled scenes, but are seldomly applicable to real-world images in the wild.

In contrast, our proposed method is very modest in its photometric assumptions, barely affected by varying surface colors, and remarkably robust to errors in the annotation. The latter property even allows to use a fully automated segmentation method for selecting objects, which considerably simplifies its practical application. The computational efficiency compared to related methods make it particularly well suited to analyze a large number of images.

There also exist methods that estimate a 3-D lighting

distribution from a fitted geometric model [14], [19]–[21], which is also discussed in Sec. II. These methods can only compare objects to which a 3-D model can be fitted with high accuracy, such as faces. In contrast to these methods, the proposed method can be applied to a much wider range of objects.

The proposed method is based on intensity gradients on the surface of an object. In our theoretical derivation, we show that the sum of intensity gradients on a sphere points to the direction of the light source. We empirically show that this condition can be readily relaxed to non-spherical objects, such as persons. Figure 1 shows an example. A spliced image is shown on top, and the mode of the distribution of incident light on each person is shown as red arrows in the bottom, which in this case point for different persons in opposite directions. Furthermore, we empirically show that the distribution of gradients are very valuable forensic cues. For example, the gradient distribution can exhibit light sources that are located in front or behind the camera, which is not possible with previous methods on 2-D lighting distributions. To exploit the full gradient distribution, we encode it in a feature set. These features are a much richer representation than the dominant light direction alone, and allow an automatic classification of the consistency of lighting environments.

In summary, the contributions of this work are:

- We propose a simple, robust, and computationally extremely efficient descriptor of object gradients to estimate the illumination direction in the image plane. The method neither requires a detailed user segmentation nor any 3-D reconstruction of the object geometry.
- We propose a feature set to interpret and classify distributions of image gradients into identical or different lighting environments.
- The method is robust to typical challenges for physics-based methods, such as variations in surface material, local geometry variations and self-shadows.
- The method is robust to strong image compression and downsampling, where it clearly outperforms five other state-of-the-art methods.
- We provide a new dataset for the community to evaluate physics-based splicing detection.

This work is organized as follows. In Sec. II, we review previous works in image forensics on estimating the direction of incident illumination. In Sec. III, we present the theoretical foundation for the estimation of illumination direction from object gradients. Section IV presents the proposed algorithm for manipulation detection. We evaluate the proposed method in Sec. V and conclude this work in Sec. VI.

II. Related Work

Estimating the direction of incident light is an actively researched topic in image forensics. Existing methods can be categorized in estimation of 3-D and 2-D illumination environments.

3-D lighting environments are computed from the intensity distribution on the 3-D geometry of an object. To this end, a 3-D geometry model has to be estimated from 2-D objects under investigation. This model is then used to solve a reflectance equation for the direction of incident light. Existing works propose different ways of addressing the challenge that estimating the 3-D geometry from 2-D images is a severely underconstrained problem. These methods propose either to manually annotate the 3-D surface structure [22], to compute the 3-D geometry with a model-free shape-from shading algorithm [23], or to fit an existing 3-D model to known objects such as faces [14], [19]–[21]. In the latter case, Kee and Farid fit a 3-D morphable face model to persons in the image [19]. The 3-D geometry yields 3-D normal vectors which allow to represent the intensity distribution on the face in a spherical harmonics basis. Forensic comparison consists of comparing these basis coefficients between different faces. Peng *et al.* extended this idea by a more general reflection model, to relax assumptions on face convexity and facial texture [20] and [21]. Further works used a more flexible morphable model to allow for more diverse facial expressions [14].

The success of these approaches critically depends on the quality of the 3-D surface model [24]. On perfect 3-D models, these approaches currently exhibit the best performance for estimating lighting environments. However, in practice, this is difficult to achieve. For faces, there exist high-quality face models and very robust fitting algorithms. However, this limits the forensic analysis to the comparison of faces, while all other scene elements can not be evaluated at the same level of detail. A model-free shape-from-shading approach can extend the analysis beyond faces. However, in practice, it turns out that robust lighting environments can only be computed on rather simple geometries. Manually fitted models or manually annotated surface normals can in principle be fitted to arbitrary objects. However, they considerably increase the effort for the analyst, and the quality of the comparison highly depends on the confidence of the analyst in the annotations.

Conversely, methods that compute so-called 2-D illumination environments mitigate the challenge of estimating 3-D structure from objects. Instead, the direction of incident illumination is estimated as a 2-D projection in the image plane. 2-D models are inherently less expressive than 3-D models, and therefore are clearly outperformed by 3-D models under ideal conditions. However, 2-D environments are in many aspects significantly easier to estimate, and they can be applied to a broader range of objects, which turns out to be very useful in practice. Johnson and Farid pioneered this family of methods [13], [25]. The key idea is that the surface normal of an object contour lies within the image plane, and can therefore be estimated by the local contour gradient. They therefore propose to annotate the occluding contour of an object, and to represent the intensity variations along the contour in a basis of 2-D spherical harmonics. Later works extended this approach by relaxing the requirement that the contour must consist of identical materials [18], [26].

While 2-D methods are overall more generally applicable than 3-D methods, practical experience shows that their application is far from straightforward. For example, the contour line must be carefully annotated to be free from local self-shadowing (e.g., from small creases on clothes, or cast shadow from the head onto the body). In many cases it is

also difficult to find contours that cover a sufficiently large angular range to robustly compute the spherical harmonics model. Additionally, small variations in the manual contour annotations can oftentimes have a large impact on the outcome of the estimation.

The proposed method also estimates a 2-D lighting environment, but it takes a fundamentally different approach than previous works. Instead of relying on relatively few, potentially noisy pixels on the contour, it estimates the projection of light from the full object area, thereby taking all object pixels into consideration. We exploit the fact that the vast majority of gradients on a convex surface points towards the direction of the illuminant. Edges from material boundaries or self-shadows may point into arbitrary directions, but since the number of edge pixels on natural objects is very small compared to the overall surface, these disturbances are robustly suppressed by considering the majority of gradients.

To our knowledge, the theoretical link between image gradients and the position of the light source was discovered by Pentland in 1982 [27] and improved by Lee and Rosenfeld [28]. Both of these works make the very restrictive assumption of observing a perfect sphere under a single light source. Zheng and Chellappa proposed to consider many local spheres [29], and Dosselmann and Yang proposed more efficient estimators, 3-D estimators on spheres [30], [31]. For forensic applications, however, the assumption of observing a sphere under a single light is too restrictive. In this work, we make a leap from these theoretical works into forensic practice, by showing that image gradients can provide very rich information on a full lighting environment on objects with much more general shapes. This requires in particular to not only consider the dominant light direction, but to characterize the gradient distribution as a whole, which we propose to do with a feature set that allows to classify lighting environments in unconstrained scenes.

III. Geometric Link between Object Gradients and Illumination Direction

For the theoretical derivation, it is convenient to assume Lambertian reflection, a linear camera response and an infinitely distant light source. It will be shown in Sec. V that these assumptions are not critical for our application. Lambertian reflectance under an infinitely distant light source models the observed intensity $I(x,y)$ of a surface point at (x,y) as the product

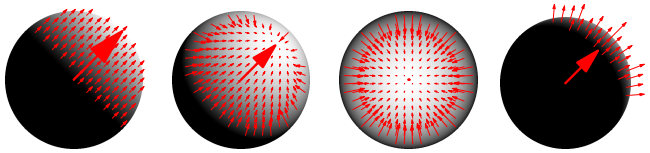
$$I(x,y) = \rho(x,y) \cdot f \cdot \max(0, N(x,y)^{\mathrm{T}} \cdot \boldsymbol{l}) \ , \tag{1}$$

where $\rho(x,y)$, f, $N(x,y)$, and $\boldsymbol{l}$ denote the albedo (i.e., object color) in one color channel, the light flux, the normal vector of the surface, and the direction of incident light onto surface point (x,y), respecively. Coordinates for f and $\boldsymbol{l}$ are omitted, since the light source is assumed to be at infinite distance, and hence constant for all surface points.

With the additional assumption that $\rho(x,y)$ is approximately constant on a small image patch around (x,y), then variations in intensity $I(x,y)$ stem from variations in the surface normal $N(x,y)$ (and thereby, from the surface geometry). This relation can be written via the derivatives $\mathrm{d}I$ and $\mathrm{d}N$ as

$$\mathrm{d}I(x,y) = \xi \cdot (\mathrm{d}N(x,y)^{\mathrm{T}} \cdot \boldsymbol{l}) \ , \tag{2}$$

where we substituted $\xi = \rho(x,y) \cdot f$ as a multiplicative constant consisting of albedo and flux.



(a) Slant: $0°$ (b) Slant: $45°$ (c) Slant: $90°$ (d) Slant: $-45°$

Figure 2. Gradients for a tilt angle of $45°$ and varying slant angles. The average of the gradients is plotted as larger arrow in the center of the sphere.

Generally, the object shape and hence also $\mathrm{d}N$ are unknown. However, we analytically show in Sec. III-A that the gradient average on a sphere points to the light source in the image plane. This assumption will be relaxed to more general objects in Sec. III-B. We also show that we can obtain a weak indicator for the position of the light source outside of the image plane when analyzing the local distribution of the gradients in Sec. III-C. Finally, we show the relation of our approach to contour based lighting estimation approaches in Sec. III-D.

A. Lighting Direction on a Sphere

We introduce the connection between image gradients and the position of the light source on a sphere [27]. To this end, we set the origin of the coordinate system to the center of the sphere, projected onto the image. With appropriate scaling, the image coordinates of the sphere $S(x,y)$ are equal to their normal vectors $N(x,y)$, namely

$$S(x,y) = \begin{pmatrix} x \\ y \\ \sqrt{1-x^2-y^2} \end{pmatrix} = N(x,y) \ . \tag{3}$$

Using the above-stated assumptions of locally constant albedo ρ and light flux f, the image intensity $I(x,y)$ at (x,y) simplifies to

$$\begin{aligned} I(x,y) &= \xi \max(0, N(x,y)^{\mathrm{T}} \cdot \boldsymbol{l}) \\ &= \xi \max(0, x \cdot l_x + y \cdot l_y + \sqrt{1-x^2-y^2} \cdot l_z), \end{aligned} \tag{4}$$

where $\boldsymbol{l} = (l_x, l_y, l_z)^{\mathrm{T}}$. The gradient in the illuminated pixels is

$$\nabla I(x,y) = \begin{pmatrix} \frac{\partial I(x,y)}{\partial x} \\ \frac{\partial I(x,y)}{\partial y} \end{pmatrix} = \xi \begin{pmatrix} l_x - l_z \frac{x}{\sqrt{1-x^2-y^2}} \\ l_y - l_z \frac{y}{\sqrt{1-x^2-y^2}} \end{pmatrix} \ . \tag{5}$$

Then, if $-1 < l_z < 1$, the *average* of all gradients over the illuminated object domain D yields the angle of the light source in the image plane [27], i.e.,

$$\iint_D \nabla I(x,y) \mathrm{d}x\mathrm{d}y = \xi \begin{pmatrix} l_x \\ l_y \end{pmatrix} \ . \tag{6}$$

We interpret the integral over the gradients as the in-plane angle spanned by l_x and l_y, and denote it as *tilt* angle. We define the tilt angle on a circle $[0°, 360°)$ in counter-clockwise direction, where $0°$ points to the right. We will see that the tilt angle can be interpreted as the projection of the dominant lighting direction onto the image plane. We further denote the *slant* angle as the angle between the image plane and the light source, defined by l_z on a unit sphere.

Four example gradient distributions are shown in Fig. 2. In all four cases, the tilt angle is set to $45°$, and the slant angle is set to (from left to right) $0°$, $45°$, $90°$, and $-45°$. The small arrows indicate the local gradient direction, while the large central arrows indicate the average gradient, respectively. In Fig. 2a, all gradients point exactly to the direction of the light source. In Fig. 2b, it can be observed that with increasing slant angle, several gradients change their direction towards the 3-D position of the light source. However, the average gradient points into the exactly same direction as in Fig. 2a, but with reduced magnitude. In Fig. 2c, where the light source is located directly at the camera, it can be observed that all gradients point towards the center, and the average gradient vanishes. Finally, Fig. 2d shows the case that the light source is located behind the object, where the individual gradients diverge, but the average gradient still points towards the direction of the light source.

This example illustrates the implications of Eqn. 6: integrating over the image gradients allows to determine the tilt angle, but not the slant angle. The estimated vector has the maximum magnitude if the light source is located in the image plane, i.e., the slant angle is zero.

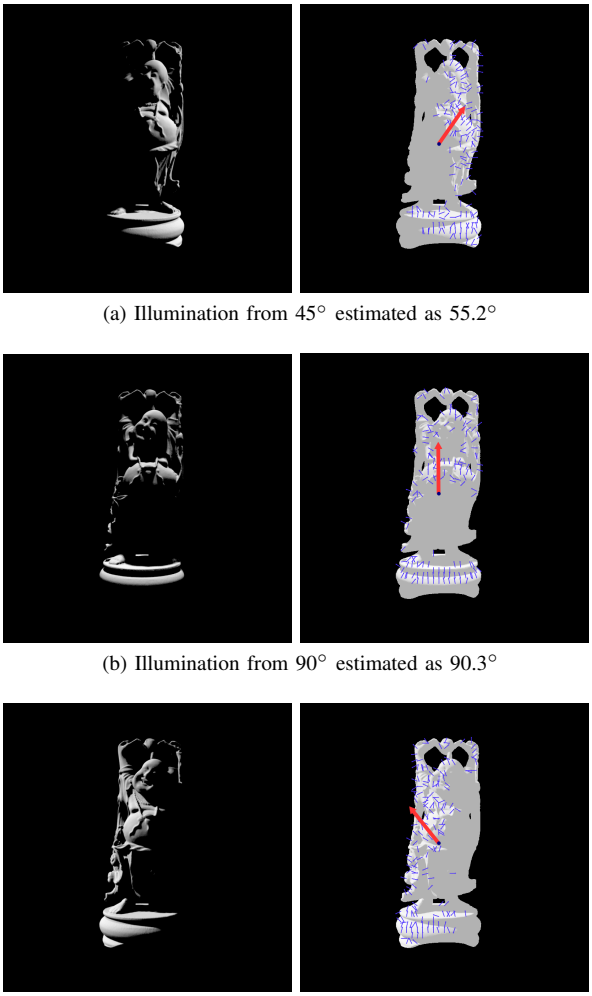
B. Robustness of the Mean Gradient to Variations in Geometry, Texture, and Segmentation

The sphere is an analytically well tractable case, and even the best case for tilt estimation [31]. Natural objects are much less constrained, but Zheng and Chelappa pointed out that mean gradients still point towards the light source as long as the object is locally spherical [29].

To better handle such more general cases, we filter out gradients with a large magnitude, and normalize the remaining vectors before averaging. This simple process removes misleading gradients resulting from inner silhouettes or strong textures and keeps gradients from shading. Details on this process are described in Sec. IV.

We empirically show in this Section and Sec. V that the proposed model indeed holds across a wide range of natural objects and surface geometries. Geometric primitives that are inherently ill-suited for analysis are planar and cylindrical objects. For planar surfaces no change of image intensities is expected, as the surface normal is constant. For cylindrical surfaces all changes of surface normals dN are in one direction only.

To explore the robustness to different geometry and the broader applicability of the method, the Happy Buddha [32] model is rendered with different light positions. Figure 3 shows three example cases of directional illumination without inter-reflections from 45°, 90°, and 135° on the left. On the right, a subsampled set of local gradient vectors is shown in blue, and the mean vector for the dominant lighting direction is shown in red. The model consists of a challenging geometry for the estimation method that extends far beyond the simple sphere of the previous section. It includes self-shadowing and concave surfaces. Additionally, the geometry is not symmetric. As a consequence, the geometry of the model leads to a more perturbed gradient vector field in comparison to the sphere, which also decreases the overall magnitude of the mean vector. The violations of the geometric assumptions introduce an error to the estimation, up to 10.2° in Fig. 3a. However, although the tilt estimation is clearly influenced by this challenging geometry, it still leads to results that are close to the ground truth.



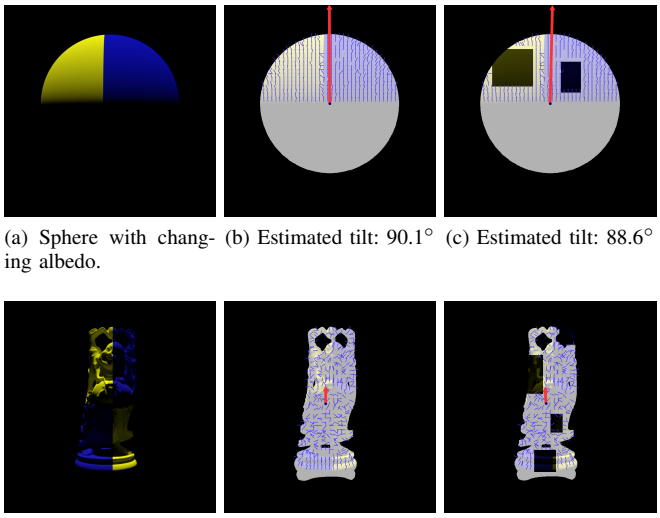
(a) Illumination from 45° estimated as 55.2°

(b) Illumination from 90° estimated as 90.3°

(c) Illumination from 135° estimated as 128.6°

Figure 3. The Buddha model displays challenging geometry, violating assumptions made for the correct tilt estimation. While the geometry influences the estimation and introduces an error, the estimation is still plausible. The left column displays the different illumination situations, the right column the according gradient-based estimations.

Figure 4 illustrates that object gradients are also robust to varying surface color and segmentation errors, two situations that are very challenging for existing 2-D illumination estimation methods. On the top row, the sphere model is used, in the bottom row, the Buddha model is used. Both models are illuminanted at a tilt angle of 90°. The left column shows the



(a) Sphere with changing albedo. (b) Estimated tilt: 90.1° (c) Estimated tilt: 88.6°

(d) Buddah with changing albedo. (e) Estimated tilt: 88.9° (f) Estimated tilt: 93.3°

Figure 4. Illumination estimation for a sphere and Buddha displaying a non-constant albedo and incomplete segmentation (best viewed in color). Both the albedo and the incomplete segmentation only have a minimal effect on the estimation.

input images, which now consist of a bright yellow and a dark blue part. The middle column shows the gradient estimation on these images, which is barely affected by the surface colors. This desirable property stems from the fact that most of the gradients are not subject to a color transition, and hence most vectors indicate the correct direction. Conversely, the (rare) case of objects with a smooth albedo gradient is expected to be a pathological failure case of the proposed method. The right column shows the estimation on the same input images, but with omission of some regions to simulate incomplete or wrong objects segmentations. Although major areas are removed from the image, the estimated illumination direction is barely affected. The robustness against segmentation errors is again an outcome of considering the gradients over the whole object area.

C. Distribution of the Gradient Vector Field

The mean of the gradient vectors can not distinguish between frontal and evenly distributed illumination. In both cases, the magnitude of the mean is close to zero. However, the divergence at the center of the vector field can distinguish this situation, as it indicates whether the overall gradient distribution points inwards or outwards. In our application, a large positive divergence indicates distributed illumination, and a large negative divergence indicates frontal illumination. Mathematically, the divergence is a scalar field that quantifies variations in the vector field,

$$\mathrm{div}\nabla I = \frac{\partial \nabla I_x}{\partial x} + \frac{\partial \nabla I_y}{\partial y} \ , \tag{7}$$

where I_x, I_y denote the x- and y-components of the gradient field, respectively. In our implementation, we split the objects vector field into four quadrants around the object center, and compute the mean gradient for each quadrant. The divergence is computed on these four mean gradients, which reduces to a single scalar value.

An example is shown in Figure 5. The sphere on top is illuminated by four equally distributed illuminants from top, bottom, left, and right. The sphere on bottom is illuminated by a frontal illuminant. In the middle column, the associated gradient fields are shown. In the right column, we split the vector field into four quadrants (indicated by the color coding), and compute the mean gradient vector for each of the quadrants. Their direction is indicated by the larger red arrows. The divergence on top is positive, the divergence on the bottom is negative.

D. Theoretical Connection to Contour-based Estimators

As a sidenote, it might be interesting to see that the proposed integral over the gradient field can be related to previous work on contour-based illumination estimation [13], [18].

Applying Green's theorem, the integrals in Eqn. 6 can be replaced by contour integrals,

$$\begin{aligned} \iint_D \frac{\partial I}{\partial y}(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \oint_C I(x,y)\,\mathrm{d}x \\ \iint_D \frac{\partial I}{\partial x}(x,y)\,\mathrm{d}x\,\mathrm{d}y &= \oint_C I(x,y)\,\mathrm{d}y \ . \end{aligned} \tag{8}$$

The theorem implies that the average intensity along the contour also points into the direction of the tilt angle.

For example, assume without loss of generality that $l_y > 0$, and $l_x = 0$ (e.g., by rotating the coordinate system accordingly). This leads to a sphere where the upper half of the contour is illuminated, while the lower half is not. Accordingly, the image gradient is expected to point towards y-direction. Evaluating the contour integral yields

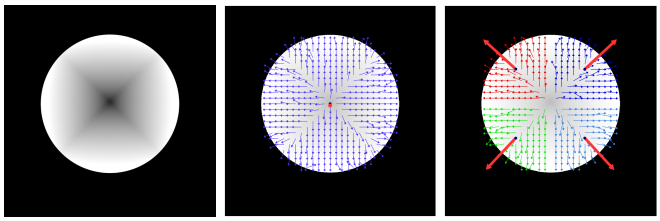
$$\begin{aligned} \oint_C I(x,y)\,\mathrm{d}x &= \int_{-1}^{1} I(x,\sqrt{1-x^2})\,\mathrm{d}x = \frac{\pi}{2}\cdot l_y \\ \oint_C I(x,y)\,\mathrm{d}y &= \int_0^1 I(x,y)\,\mathrm{d}y + \int_1^0 I(x,y)\,\mathrm{d}y = 0 \ , \end{aligned} \tag{9}$$

as expected.

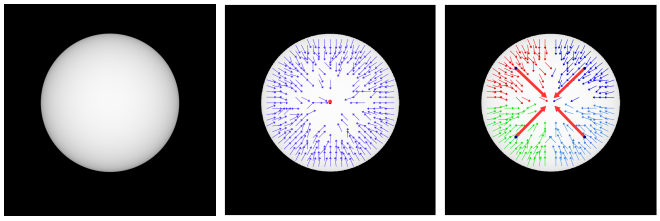
Considering that the gradient indicates the local brightness distribution on an object, it turns out that it is closely related to the contour intensity distribution of earlier works [13]. However, previous methods that use only the contour pixels have to cope with several practical issues [26]: the quality of estimate is very sensitive to the proper selection of the contour, and the relatively low number of contour pixels makes this approach very sensitive to noise. Also, photometric variations due to self-shadows and object materials need to be explicitly addressed. In contrast, using all object pixels as in the proposed method largely removes these limitations.

IV. Proposed Algorithm for Forensic Lighting Analysis

The proposed algorithm operates on individual objects in the image. To this end, it requires a segmentation of the object, which can either be done manually or automatically, e.g., via the Mask R-CNN network [33], [34].



(a) Sphere illuminated from four directions. (b) Absolute of mean near zero. (c) Divergence positive.



(d) Sphere illuminated from front. (e) Absolute of mean near zero. (f) Divergence negative.

Figure 5. Illumination from four directions (top) and frontal illumination (bottom). The gradient mean is in both cases 0. However, the divergence of the gradient field is positive for the four light sources, and negative for frontal lights.

In this section, we present the proposed algorithm. In Sec. IV-A, we describe the gradient computation and filtering. In Sec. IV-B, the features for distinguishing two lighting environments are introduced. The classification of lighting environments via logistic regression is described in Sec. IV-C. Finally, in Sec. IV-D, we add several details on how to fully automate the processing pipeline.

A. Gradient Computation

First, a bilateral filter is applied to the RGB color input image to reduce the influence of very fine texture details and noise. The image is converted to grayscale by averaging the RGB color channels. Consequently, each color channel has the same impact on the gradient vector magnitude. The gradients are computed in x- and y- direction individually with a Sobel operator, resulting in a 2-D vector field over each masked object,

$$\mathrm{d}I(x,y) = (\mathrm{d}I_x(x,y), \mathrm{d}I_y(x,y)) \ . \tag{10}$$

We found that the choice of method to compute the image gradient does not have a significant impact on the results.

Occluding contours, variations in albedo (e.g., from different clothing) or self-shadowing lead to gradients that are not related to the light source. To lower their influence, we assume that such gradients are generally larger than shading gradients. As a consequence, we filter out gradients with a magnitude above an adaptive threshold t. To obtain t, we compute the mean μ and standard deviation σ of the magnitude of all non-zero gradient vectors of an object, i.e.,

$$\mu = \frac{1}{m} \sum_{x,y} \|\mathrm{d}I(x,y)\| \tag{11}$$

$$\sigma = \sqrt{\frac{1}{m} \sum_{x,y} (\|\mathrm{d}I(x,y)\| - \mu)^2} \ , \tag{12}$$

where m is the number of pixels in the processed area. The threshold t is then set to

$$t = (\mu + \sigma) \ . \tag{13}$$

B. Lighting Representation and Dissimilarity Features

The mean of the normalized gradient vectors, i.e., the dominant lighting direction, is computed as

$$\overline{dI} = \left(\frac{1}{N} \sum_{x,y} d\hat{I}_x(x,y), \frac{1}{N} \sum_{x,y} d\hat{I}_y(x,y) \right)^{\mathrm{T}} \ , \tag{14}$$

where $d\hat{I}(x,y)$ denotes the magnitude-normalized gradient vectors, computed as

$$d\hat{I}(x,y) = \left(\frac{dI_x(x,y)}{\|dI(x,y)\|}, \frac{dI_y(x,y)}{\|dI(x,y)\|} \right)^{\mathrm{T}} \ . \tag{15}$$

The normalization leads to the effect that only the direction of the vector is taken into account [31]. It contributes to a high robustness to changes in albedo, and mitigates the need for intrinsic image decomposition for albedo neutralization. The mean in Eqn. 14 is used as the estimate for the dominant illuminant direction in the image plane, referred to as tilt angle in Sec. III.

Using Eqn. 14, we can compute the dominant lighting directions $\boldsymbol{a}$ and $\boldsymbol{b}$ for two objects A and B. To quantify differences in illumination, we use the cosine dissimilarity L_D,

$$L_D(\boldsymbol{a},\boldsymbol{b}) = 1 - \frac{(c(\boldsymbol{a},\boldsymbol{b}) + 1.0)}{2} \ , \tag{16}$$

where

$$c(\boldsymbol{a},\boldsymbol{b}) = \frac{\boldsymbol{a}^{\mathrm{T}} \cdot \boldsymbol{b}}{\|\boldsymbol{a}\| \cdot \|\boldsymbol{b}\|} \tag{17}$$

is the cosine of the angle between the two illuminant directions. This measure ranges from 0 to 1, with value 0 for identical directions, and 1 for opposite directions.

To compute the divergence, we divide the gradient field into four quadrants around the center of each object, and compute the divergence of the four quadrants according to Eqn. 7, which is a single scalar. The differences in the divergences $\mathrm{div}A$ and $\mathrm{div}B$ for objects A and B is computed as

$$D_D(A,B) = |\mathrm{div}A - \mathrm{div}B| \tag{18}$$

Additionally, we calculate the mean lighting directions in each tile using Eqn. 16 and average their pairwise distances, i.e.,

$$T_D(A,B) = \frac{1}{K} \sum_i L_D(A_i, B_i) \ , \tag{19}$$

where A_i, B_i denote the quadrants of A and B, and $K = 4$ due to the use of four quadrants.

To further characterize the gradient vector field, a histogram of the gradient vectors is computed. The gradient directions are discretized into 72 bins, which corresponds to an angular resolution of 5°. The bins are collected in a histogram, and the sum of the histogram entries is normalized to 1 to accomodate for different object sizes of A and B. Let $\boldsymbol{h}(A)$ and $\boldsymbol{h}(B)$

denote the histograms of objects A and B. We compare these histograms via zero-normalized cross-correlation, i.e.,

$$H_D(\boldsymbol{h}(A), \boldsymbol{h}(B)) = \frac{(\boldsymbol{h}(A) - \bar{h}(A))^{\mathrm{T}}(\boldsymbol{h}(B) - \bar{h}(B))}{\|\boldsymbol{h}(A) - \bar{h}(A)\| \cdot \|\boldsymbol{h}(B) - \bar{h}(B)\|} \ , \tag{20}$$

where $\bar{h}(A) = \bar{h}(B) = 1/72$ denote the means of the normalized histograms $\boldsymbol{h}(A)$ and $\boldsymbol{h}(B)$, respectively.

C. Fake-Score Regression

The proposed features, namely the differences in tilt L_D, the divergence of tiled estimates D_D, individual differences of the tiled estimates T_D, and the correlation between the gradient histograms H_D are used to quantify the dissimilarity of two lighting environments.

The features are combined in a logistic regression model to generate a single distance between 0 and 1, which can be seen as a "fake-score". The logistic function is defined as

$$f_l(t) = \frac{1}{1 + e^{-t}} \ , \tag{21}$$

where t is a linear combination of the features L_D, D_D, T_D and H_D,

$$t = (\beta_0 + \beta_1 \cdot L_D + \beta_2 \cdot D_D + \beta_3 \cdot T_D + \beta_4 \cdot H_D) \ . \tag{22}$$

The parameters $(\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ can be fitted from training data via a logistic regression regularized by a L_2 penalty.

D. Automated Forgery Detection

Some forensic scenarios require batch processing of large amounts of data, which enforces the use of a fully automated algorithm. Many physics-based algorithms are quite difficult to automate, including previous 2-D lighting estimators. For example, previous work on 2-D lighting environments is highly sensitive to the careful selection of an occluding contour [26]. Conversely, the proposed method is considerably more robust, and additionally makes it possible to fully automate the method.

We use the popular Mask R-CNN neural network to segment the image [33], [34], which provides a coarse pixel-wise segmentation of object instances. Objects with an area below 1% of the image pixels are excluded from the analysis, since these consist of too few pixels. The considered object classes can be restricted in advance depending on forensic relevance and availability given by the segmentation method. After filtering all detected object instances based on image area and object category, the fake-score is computed between all pairs of the remaining instances, and divided by the number of pairs.

V. Results

We first evaluate the accuracy of only estimating the main illumination direction in Sec. V-A. Further, in Sec. V-B, the full proposed feature set is applied to classification of lighting on pre-segmented data. Three related methods estimating lighting based on objects are evaluated for comparison. Two of the comparative methods are estimating 2-D light environments based on object contours: the method as proposed by Johnson and Farid [13] and the work by Riess *et al.* [26] extending this by estimating an intrinsic decomposition for the contour. The third method estimates 3-D light environments based on shape-from-shading, similar to the method as proposed by Fan *et al.* [23]: to estimate the illumination we employ the SIRFS method by Barron and Malik [35] and compare the resulting 3-D spherical harmonic coefficients as proposed by Johnson and Farid [13]. We will refer to these methods as Contour [13], ICE [26] and SIRFS [35], respectively. All methods use the same segmentation masks. Contours are generated by applying a Canny edge detector. A comparison of the runtime of these methods is presented in Sec. V-C. In Sec. V-D, we report results for the most general case on several datasets: the fully-automated detection of manipulations including the object segmentation. The performance of the proposed method is compared to two additional state-of-the-art methods. Limitations of the algorithm are discussed in Sec. V-F.

(a) Example objects of the ALOI dataset.

(b) Illumination conditions with single light source L1-L5.

(c) Illumination conditions with multiple light sources L6-L8.

Figure 6. Example images of the ALOI dataset. [36]

A. Estimation of the Main Illumination Direction

We first show that the dominant lighting direction from Eqn. 14 can be estimated with high accuracy from a wide range of real-world objects. To this end, we use the "Amsterdam Library of Object Images" (ALOI) dataset [36]. This laboratory dataset consists of 1000 objects captured under different lighting directions. Figure 6a shows several example images from the dataset. The exact experimental setup of the dataset can be found in [36]. We select five different illumination conditions with a single light source. The light sources are at angles of 30° (L1), 60° (L2), 90° (L3), 120° (L4) and 150° (L5). An example object illuminated from these five directions is shown in Figure 6b. The light sources are slightly in front of the objects as can be seen in the reflections of the object.

Table I
MEDIAN ERROR OF ESTIMATIONS FOR FIVE LIGHT DIRECTIONS AND ALL OBJECTS GIVEN BY THE ALOI DATASET.

	L1	L2	L3	L4	L5	All
Contour	34.69°	27.73°	10.55°	22.27°	34.92°	25.08°
ICE	33.63°	22.03°	**9.84°**	19.81°	32.23°	22.85°
SIRFS	**21.49°**	34.30°	31.83°	29.16°	**20.49°**	26.08°
Proposed	26.65°	**21.54°**	14.14°	**17.93°**	22.05°	**20.55°**

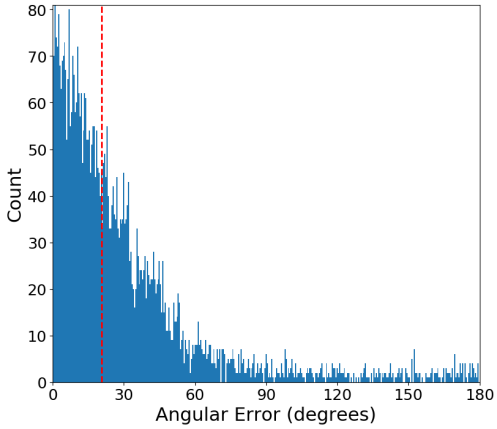
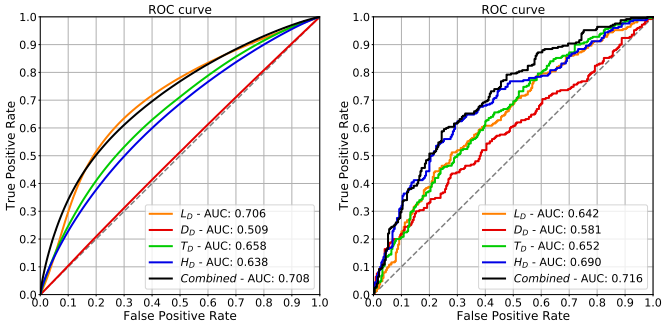


Figure 7. Histogram of angular errors for all samples evaluated with the proposed method. Red line shows the median error of 20.55°.

Table I lists the median absolute angular error for the estimation of the dominant lighting direction per ground truth illuminant for each method. The input images for SIRFS were scaled by factor 0.25 to achieve a feasible runtime. The computational time of the methods is further evaluated in Sec. V-C. For the SIRFS method the dominant light direction in the image plane is inferred from the corresponding spherical harmonic coefficients. The median error of the proposed method for the different light directions varies between 14° and 26°. The median error for all light directions is 20°. The other methods perform slightly worse with median errors for all light directions between 22° and 26°.

Figure 7 shows a histogram of the error distribution over all samples for the proposed method. The median error is indicated by the red line. The main contributing factor to large errors are objects that violate one or more of the theoretical assumptions in this very diverse database: some objects display the degenerate case of cylindrical, planar or even mostly concave geometry. Additionally, some objects show strong specular reflection which is in contrast to the assumed Lambertian reflectance model. These deviations from our assumptions can in some cases lead to a large error. In scenarios where a human operator guides the analysis, most of the degenerate cases can be easily removed, which yields a robust, easy-to-use physics-based estimator. In fully automated scenarios, the overall good results indicate that different illumination environments can be distinguished for a large variety of objects.

A practical example for comparing the dominant illumination directions is shown in Fig. 1 on the title page. The top part of the image shows a composite image, where the two persons on the left are spliced into the image. The three persons on the right are illuminated by a window to the right. The two persons on the left were illuminated from the left and do not fit the scene illumination, as indicated by the estimation. In contrast to many objects in the ALOI dataset, the persons in Fig. 1 fit the theoretical assumptions quite well: clothing exhibits mostly diffuse reflectance and the body surface is predominantly convex, which allows an analyst to have overall high confidence in the estimate.



(a) ALOI (b) COCO

Figure 8. ROC curves for the classification of lighting environments with individual features and full feature set on the ALOI and COCO datasets.

B. Classification of Lighting Environments

The full feature vector for lighting classification is evaluated on pre-segmented real objects. For splicing detection, we make the common assumption that objects from different images exhibit different lighting environments [13]. Thus, the classification task is to determine whether two objects stem from the same image or from two different images.

Prior work [21], [26] shows that related methods, while strong in theory, are hard to apply in a realistic setting, needing precise annotations and being vulnerable to less constraint prerequisites in natural scenes. We present results for laboratory data and for natural scenes. We evaluate the classification of lighting environments including single and multiple light sources in a controlled laboratory setting using the ALOI dataset. Here, especially the methods estimating lighting based on object contours benefit from the given high quality object segmentation. To further evaluate the performance in a realistic setting, we use the COCO dataset [37] to emulate splices.

1) Laboratory: To evaluate the ability of distinguishing lighting environments in a defined laboratory setting we again use the ALOI dataset and add three additional light settings with multiple light sources. An example object subject to the light settings L6 to L8 is shown in Fig. 6c. We generate samples by pairing each object in one light setting with all other objects in another light setting. For each pair of light settings (e.g. L1L1, L1L2, etc.) we can create $k = n \cdot (n-1)/2$ samples, where $n = 999$ is the number of objects. We choose multiple subsets and create all possible combinations of light settings. The subsets include all light settings L1 to L8, just

Table II
AUC OF ROC CURVES FOR BINARY CLASSIFICATION INTO SAME OR DIFFERENT LIGHTING ENVIRONMENTS. THE PERFORMANCE OF THE RELATED METHODS DROPS SIGNIFICANTLY ON THE MORE CHALLENGING NATURAL SCENES. THE PROPOSED METHOD PERFORMS CONSISTENTLY WELL ON THE NATURAL SCENES AND OUTPERFORMS THE RELATED METHODS.

	Laboratory (ALOI)			**Natural Scenes (COCO)**				
	All	**Single**	**Multi**	**Person**	**Animal**	**Furniture**	**Vehicle**	**Mixed**
Contour	0.728	0.766	0.756	0.589	0.654	0.567	0.609	0.539
ICE	**0.740**	**0.776**	0.776	0.598	0.641	0.592	0.588	0.518
SIRFS	0.738	0.763	**0.793**	0.580	0.665	0.581	0.559	0.524
Proposed	0.708	0.738	0.740	**0.716**	**0.735**	**0.633**	**0.657**	**0.677**

the settings with single light sources L1 to L5 and just multiple light sources L6 to L8.

First, we evaluate each of the four proposed distance features individually, and with a combined logistic regression model as described in Sec. IV-C. The individual features and the combined "fake-score" are evaluated via thresholding to create a receiver operating characteristic (ROC) curve. We fit the logistic regression model using 100 of the 999 objects and evaluate on the remaining. The ROC curves for binary classification of pairs including all eight light settings are shown in Fig. 8a. The proposed method achieves an area under the curve (AUC) value of 0.708 with the combined classifier. The distance between the estimated tilt angle L_D alone obtains an AUC value of 0.706. Consistent with the mostly directional light in the dataset, the comparison of divergence D_D alone with an AUC of 0.509 is not suited to distinguish the light settings. Table II shows AUC values for the different subsets and the comparison methods. Again, the input images for SIRFS were scaled by factor 0.25. The experiment shows that all methods are able to distinguish the lighting environments with single and multiple light sources in a laboratory setting. The proposed method achieves AUC values of 0.708-0.740. In this artificial setting with high quality segmentations the comparative methods leveraging spherical harmonics as features play their strength and slightly exceed the proposed method with AUC values of 0.728-0.793.

2) Natural Scenes: To evaluate the performance in a realistic setting, we use the publicly available COCO dataset [37]. This dataset contains annotated objects with associated segmentation masks. Given specific object classes, we select images containing at least two objects, each covering at least 1% of the image area. We fit the logistic regression model to the object category "person". Persons can be considered relevant objects for a forensic analysis, and they fit the assumptions for the proposed gradient-based illuminant estimator reasonably well, i.e., their surface geometry is mostly convex. We select either two persons from within the same image or two different images together with their respective segmentation masks. Two persons from the same image are assumed to be exposed to identical illumination. Two persons from different images are assumed to be exposed to differing lighting environments. While this assumption might not always hold, we still adopt it in order to perform the forensic task of splicing detection, where we would like to know whether an object has been inserted into an image. Thus, the reported performances are obtained with a high degree of label ambiguity, i. e., under relatively adversarial conditions.

The logistic regression model is fitted to 8000 person pairs from the MS COCO 2014 training data where 50% stem from the same source image and 50% from different source images.

For evaluation we use 500 randomly chosen pairs from different source images contained in the validation data. Again, we evaluate each of the four proposed distance features individually, and with the combined logistic regression model. The results of this experiment are shown in Fig. 8b. The combined classifier achieves the best result, with an AUC of 0.716. Among the individual metrics, the best performing distance is H_D, i.e., between histograms. The distances between tiled estimates T_D, tilt angle L_D and the divergence D_D are slightly worse, but still obtain AUC values of 0.581-0.652.

We repeat the experiment for objects other than persons and compare the results with the related methods. The results are obtained on 500 object pairs of the respective category, using the same regression parameters and the same evaluation protocol as before. Table II shows the AUC values for classifying samples of the object groups "person", "animal", "furniture", and "vehicle". Additionally, we create a "mixed" dataset by randomly choosing pairs out of all the aforementioned object classes. As shown in Tab. II, the proposed method achieves AUC values of 0.633-0.735. While these results are overall comparable to the "person" experiment, the performance varies depending on how well the objects fit the initial assumptions for the estimation: furniture, and particularly the planar, metallic surfaces of vehicles significantly deviate from the physical model of the method. The comparative methods achieving AUC values of 0.518-0.665 are outperformed by the proposed method for all categories. All methods perform best for the object group "animal". We assume the lighting environments are easier to classify as a lot of animal images are taken in outdoor settings with distinctive illumination.

Table II also shows that the performance of the related methods degrades considerably when transitioning from the laboratory ALOI data to the real data. This reveals the strong impact of natural scenes and segmentation quality on the performance of the related methods. While the related methods perform well given the laboratory ALOI data, their performances drop in many cases almost to guessing chance on the COCO samples from natural scenes. The proposed method performs overall consistently for both experiments. Especially for the object category "person" and "mixed" samples of ob-

Table III
TIMINGS FOR EVALUATING 2500 SAMPLES WITH GIVEN SEGMENTATION AT A RESOLUTION OF 640 PIXEL IN THE LARGER DIMENSION.

	Total time	**Time per sample**
Contour	15.87 min	0.38 s
ICE	56.65 min	1.36 s
SIRFS	1655.39 min	39.73 s
Proposed	**0.89 min**	**0.02 s**

jects from multiple classes, the proposed methods outperforms the others by a large margin.

C. Computational Time

When the proposed method is used to automatically process images, the by far most costly operation is the image segmentation. The core algorithm is highly efficient to compute, as it works with a fixed resolution of 640 pixels in the larger dimension and mainly consists of a few image filtering operations to obtain the image gradients.

We compare the wall-clock time needed to evaluate the 2500 COCO samples with given segmentation for the results shown in Tab. II. All methods process the images in the given resolution of 640 pixels in the larger dimension. We use an Intel Core i7-5820K PC with 32GB RAM. For all methods we use the unoptimized research implementations available and run 10 processes in parallel. The amount of intermediate harddrive output slightly varies between methods. The timings include all I/O operations and the computation of the core algorithms, i.e., lighting estimation, feature computation, and classification. The required total time for each method and the average time per sample are shown in Tab. III. The runtime of the proposed method is significantly shorter compared to the others. The two methods estimating 2-D lighting from object contours are about one to two magnitudes slower than the proposed method. By far the most costly method is SIRFS which solves for shape, reflectance and shading and estimates the 3-D illumination. It takes more than 1800 times as long to process the data.

D. Fully Automated Splicing Detection

We evaluate the fully automated pipeline, including automated instance segmentation on several datasets. To evaluate the performance on data that is shared over the internet, we also demonstrate the performance on resized and recompressed datasets. For the proposed method we use the same regression parameters as fitted for the COCO data, described in Sec. V-B. The performance of the proposed method is compared to five other methods. Three of the methods are the already described closely related works based on illumination estimation we refer to as Contour [13], ICE [26] and SIRFS [35]. Additionally, we compare to two other state-of-the-art methods: the statistical "SpliceBuster" by Cozzolino *et al.* [8], and a statistical deep-learning approach by Huh *et al.* [10]. For both methods we use the publicly available implementations [38], [39]. Both methods generate a heat-map as output. For the task of manipulation detection, we take the maximum output of the response map as score indicating a forgery. In the discussion, we refer to these methods as Cozzolino2015 [38] and Huh2018 [10], respectively.

Table IV
ROC CURVE AUC VALUES FOR CLASSIFYING THE DSO-1 [17] DATASET AND THE TRAINING CORPUS OF THE IEEE IFS-TC IMAGE FORENSICS CHALLENGE [40]. THE PERCENTAGE VALUE INDICATES THE FRACTION OF SUCCESSFULLY PROCESSED IMAGES.

	DSO-1 [17]	**DSO-1 [17] 960px, JPEG70**	**IFS-TC [40]**
Huh2018	0.718 (100%)	0.504 (100%)	**0.669** (100%)
Cozzolino2015	**0.803** (100%)	0.530 (100%)	0.473 (97%)
Contour	0.503 (97%)	0.513 (96%)	0.569 (35%)
ICE	0.501 (97%)	0.519 (96%)	0.570 (35%)
SIRFS	0.538 (97%)	0.553 (96%)	0.569 (35%)
Proposed	0.555 (97%)	**0.565** (96%)	0.538 (35%)

1) DSO-1 Dataset: The publicly available *DSO-1* dataset contains 200 images displaying multiple persons [17]. 100 images contain one or more spliced in persons. Post-processing such as color and brightness adjustment were applied to increase photorealism.

We perform two experiments on this dataset. First, we use the dataset as-is. Second, we downsample the dataset to 960 pixels in the larger image dimension and compress the images with JPEG level 70. The AUCs for both experiments are shown in Tab. IV. The percentages in the table indicate the fraction of images that the methods were able to process. The statistical methods could be applied on all images. The physics-based methods excluded 3% of the images, as the segmentation did not yield at least two persons with a minimum size of 1% of the image area.

Our proposed method achieves an AUC of only 0.555, which indicates a performance close to guessing. The other illumination-based methods exhibit a comparable performance. However, as pointed out by Peng *et al.* [21], this is a known issue of this dataset. The dataset is not well suited for analysis of lighting environments, as many spliced images are captured with frontal flash, which leads to relatively similar illumination environments. Additionally, persons in many non-tampered images are exposed to varying illumination. This is why also the the method by Peng *et al.* only achieves slightly better results, although it specifically operates on high quality 3-D geometry on faces for illumination estimation [21]. The statistical methods Huh2018 and Cozzolino2015 are better suited here, with an AUC of up to 0.803.

However, the situation changes drastically after resizing and compressing the dataset. In this case, the performance of the statistical methods dramatically drops, while the physics-based methods are not affected by this operation. The proposed

method performs best in this case, with an AUC of 0.565.

2) IEEE IFS-TC Challenge: We use the publicly available training corpus of the IEEE IFS-TC Image Forensic Challenge as a second dataset [40]. It consists of 1050 pristine and 450 manipulated images. Various manipulation techniques are applied, including inpainting, copy-move and splicing. The results are shown in Tab. IV. Huh2018 performs best, achieving an AUC value of 0.669. The performance of the other methods is relatively weak. The physics-based methods are mainly limited by the fact that many manipulations are unrelated to splicing. Additionally, only a small part of the dataset contains salient objects. We configured the Mask R-CNN segmentation [34] to retrieve all available object categories, to maximize the set of images for automated analysis. However, also this broad selection returned only in 35% of the cases suitable pairs of objects, as most manipulations in this dataset are on non-salient regions.

3) OpenImages Splices (OIS): To address the limitations of the other datasets, we propose a new image dataset to evaluate splicing. It consists of 450 images with two well visible persons each. In 150 of these images, one person is inserted from a different image. The source images of the dataset stem from the publicly available OpenImages V4 dataset [41], which contains about nine million images annotated with image-level labels and bounding-boxes.

The untampered images of the proposed dataset are directly taken from the original URLs provided by the OpenImages dataset and scaled to 1280 pixels in the larger image dimension. As the images in the dataset might themselves be preprocessed, we only consider the splicing of persons for manipulation detection.

The tampered images are created by selecting target and donor images using the provided image labels and Mask R-CNN for segmentation [33], [34]. The target image is chosen to show exactly one well-visible person in foreground. The donor image is chosen with the aim to find a person with reasonable semantic consistency to the target image. There was no strict focus regarding the illumination situation. Thus, it is not guaranteed that the illumination is inconsistent in spliced images, and vice versa, but it is at least highly likely that illumination conditions differ due to the randomness in pairing a splicing donor and target. Target and donor image are scaled to 1280 pixels in the larger image dimension. The segmentation of Mask R-CNN is manually refined using GrabCut [42]. Care was taken that both persons do not completely occlude each other upon splicing. The quality of the splices is partly limited by the segmentation. The spliced persons are scaled and placed manually to fit the target image, and might be slightly blurred or copied with feathered edges, but no additional post-processing is applied.

We believe that this dataset presents an interesting benchmark for instance-level forensic methods, as the image provenance from the web (e.g., Flickr) is a plausible use case, and the detection of spliced persons is a semantically meaningful goal. The proposed dataset will be referred to as OpenImages Splices (OIS). Example spliced images from the dataset are shown in Figure 9.

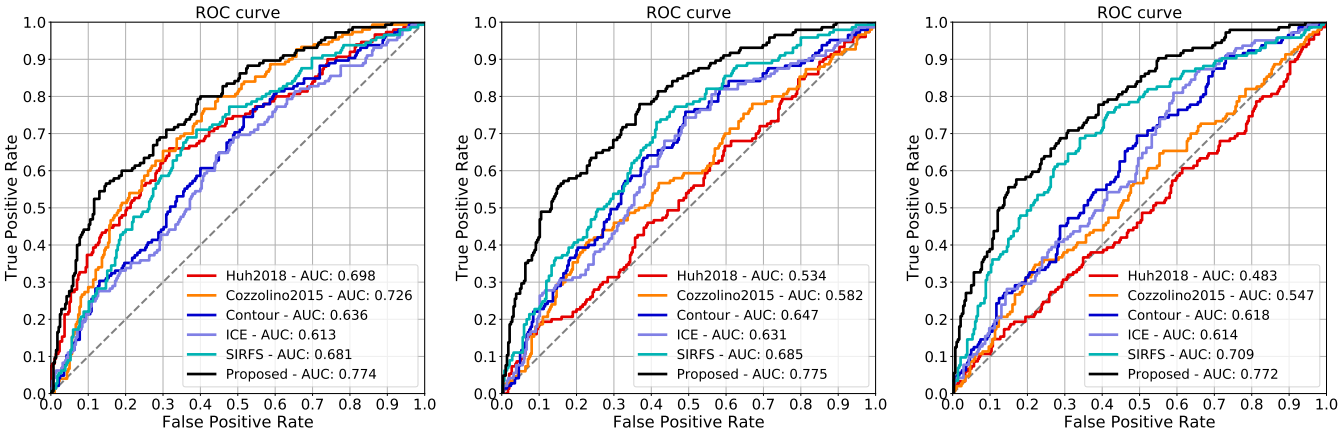
For evaluation, the same protocol is applied as for the previous datasets. We evaluate the full dataset, and again evaluate robustness to a downsampling to 960 pixels in the larger dimension and compression with JPEG quality 70. To further demonstrate the robustness of the proposed method, we evaluate a third variant with resizing to 600 pixels in the larger dimension and compression to JPEG quality 30.

The results to these experiments are shown in Tab. V, and the associated ROC curves are shown in Fig. 10. In all variants,

(a) True positives. Fake-scores (top down): 0.810, 0.794, 0.789

(b) False negatives. Fake-scores (top down): 0.196, 0.253, 0.257

Figure 9. Example splices of the proposed OIS dataset. The upper three rows show the successful application of the proposed method with a high fake-score indicating image splicing. The lower three rows show failure cases of the proposed method with a low fake-score. The right side shows the estimated 2-D light directions as indicated by the red arrows.



(a) OIS (b) OIS 960px, JPEG70 (c) OIS 600px, JPEG30

Figure 10. ROC curves for the classification of different versions of the proposed OIS dataset in accordance with the AUC values reported in Table V.

Table V
ROC CURVE AUC VALUES FOR CLASSIFYING THE PROPOSED OIS DATASET. THE PERCENTAGE VALUE INDICATES THE FRACTION OF SUCCESSFULLY PROCESSED IMAGES.

	OIS	**OIS 960px, JPEG70**	**OIS 600px, JPEG30**
Huh2018	0.698 (100%)	0.534 (100%)	0.483 (100%)
Cozzolino2015	0.726 (100%)	0.582 (100%)	0.547 (100%)
Contour	0.636 (96%)	0.647 (96%)	0.618 (95%)
ICE	0.613 (96%)	0.631 (96%)	0.614 (95%)
SIRFS	0.681 (96%)	0.685 (96%)	0.709 (95%)
Proposed	**0.774** (96%)	**0.775** (96%)	**0.772** (95%)

the proposed method outperforms all other methods. On the full-resolution images, the statistical methods Huh2018 and Cozzolino2015 obtain an AUC of up to 0.726. We consider this a strong performance, considering that the donor and target images have already been resampled during dataset creation. The proposed method performs only slightly better with an AUC of 0.774. The contour-based methods only achieve AUC values of up to 0.636. Their main weakness here is their sensitivity to the selected contour. Even the much more computationally expensive shape-from-shading approach SIRFS achieves an AUC of only 0.681.

When downsampling and recompressing the final images, as it oftentimes happens when sharing data over social media or news web pages, the performance of Huh2018 and Cozzolino2015 drops significantly to AUC values between 0.483 and 0.582. Again, as in the previous experiments, the proposed method and the other methods based on lighting estimation are mostly unaffected even by very strong downsampling and compression. The proposed method outperforms all other methods in the given scenarios.

Figure 9 shows true positive examples (splices assigned a high fake-score) and false negative examples (splices assigned a low fake-score) for the proposed method. The estimated 2-D light directions are shown on the right. The examples indicate that the method is subject to the underlying assumption that a splice will display differences in illumination which is in the majority of cases true, but not always fulfilled by the proposed dataset.

E. Automated Object Segmentation

With images playing a major role in communication nowadays, it becomes increasingly important to analyze a large number of images in a forensic context. Section V-D shows the feasibility of an automated physics-based analysis, given recent methods for instance-wise image segmentation and the proposed estimator, robust to segmentation errors.

Nevertheless, the segmentation of object instances and its quality can impact the performance of the proposed processing pipeline. Unrelated to the specific method used for object segmentation, the considered object categories and minimum image area per object have to be chosen in advance. Only objects detected by the segmentation method and meeting the chosen criteria are part of the analysis. Spliced objects beyond this scope will be missed. Generally, the method requires for comparison a minimum of two objects that are not entirely flat. In the proposed pipeline, the Mask-RCNN method is used for segmentation. Detailed benchmarks regarding the detection and segmentation performance of the method can be found in [33]. The segmentation method can be exchanged for better suited methods, when available.

The segmentation quality of Mask-RCNN is mostly sufficient for the proposed pipeline. Some typical examples are shown in Fig. 9 on the right side, with segmentations highlighted in white. Figure 11 shows two typical segmentation errors. In the top image additional image parts are segmented as part of an object instance. In this case, a fish is considered as part of a person. The lighting estimation stays plausible, but from a forensic perspective it might not be desired to include this additional object in the analysis. The bottom

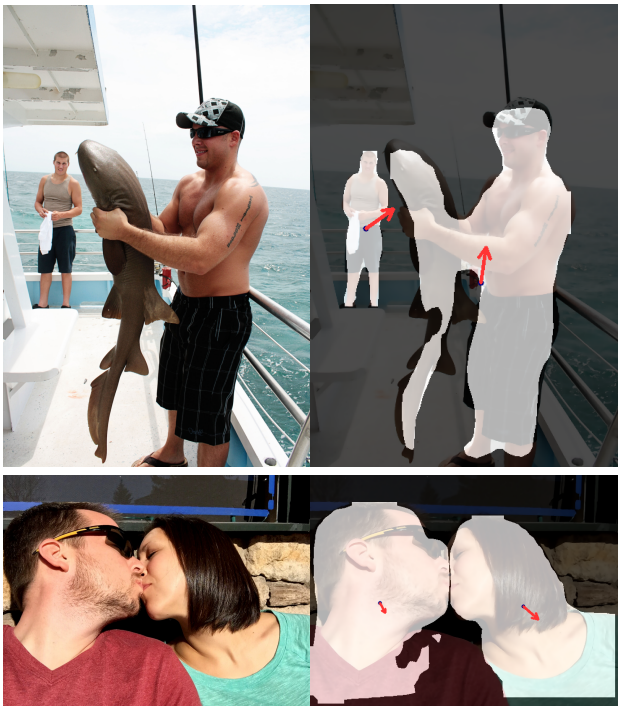


Figure 11. Examples of pristine images with challenging segmentation result. The segmentation result is highlighted in white and the estimated main light direction is shown as red arrow. In the top image the fish is included unintentionally. In the bottom image the segmentation is incomplete.

image displays a segmentation result missing a large part of one person. Prior work estimating 2-D lighting environments based on contours is highly sensitive to such errors [26]. The performance of the proposed estimator is more robust to such errors. However, the performance may also degrade when the assumptions from Sec. III are significantly violated by missing segmentations.

F. Limitations

The comparison of incident illumination is subject to inherent limitations and failure cases. The underlying assumption is that a spliced image displays differences in the lighting environment under direct illumination, while a pristine image exhibits identical lighting environments. However, the DSO-1 and IFS-TC datasets, for example, contain several practically relevant examples where this assumption does not hold. For example, a pristine image may show two faces, where each face is illuminated by a different local light source, presumably a floor lamp. Another example shows two faces under one light source, but one face shadows the other, such that the lighting environments differ. More generally, if one of the objects under analysis are in shadow, and the other is exposed to direct light, the analysis result will be wrong. This special case could be potentially filtered out with a dedicated shadow segmentation method in future work.

The method requires a minimum of two salient objects for comparison. It is therefore not suited to detect manipulations such as inpainting. The analysis of datasets covering various manipulation techniques without constraints regarding the purpose of the manipulation, such as IFS-TC, are especially challenging for the proposed and related methods, based on checking the consistency of features between specific objects.

As a final caveat, if several objects are inserted from the same donor image, the illumination between these inserted objects can be consistent. In this case, inconsistencies can only occur when comparing an inserted object with a background object.

VI. Conclusions

In this work, we propose a highly robust physics-based approach for comparing 2-D lighting environments. This even allows to use the method in a fully automated pipeline. As such, we propose this method as an interesting tradeoff: it is robust and widely applicable on a broad spectrum of images, which is a common limitation of lighting-based methods. This comes at the expense of a somewhat lower descriptive power compared to much more constrained lighting approaches.

The method is based on four features that are computed on the gradients of an object. Our theoretical derivation shows that the direction of a single light source can be recovered exactly on a sphere. On less constrained objects, the distribution of the gradients across the object is still highly informative, although the theoretical conditions are not strictly met. Our derivation also shows that the divergence of the vector field allows to distinguish between frontal lights outside of the 2-D image plane and multiple lights within the image plane. Compared to previous works on 2-D lighting environments, the descriptor draws its remarkable performance from the fact that it uses the whole object area instead of a single, potentially noisy, contour.

We demonstrate the performance of the gradient-based descriptors in several steps. We first show that the primary light direction can be reliably estimated on a single object using the large and diverse ALOI object dataset. We also show that the proposed features can be used to distinguish light environments with single and multiple light sources. A simple logistic regression on the proposed features can be deployed to distinguish between objects from identical or different images, which is demonstrated on the COCO dataset. In comparison to related lighting methods, the proposed method is highly efficient to compute and robustly applicable to images of natural scenes, even when the segmentation quality is limited.

Finally, we show on three datasets that a fully automated segmentation and classification achieves an AUC of up to 0.774 on spliced persons. One of the strongest properties of the proposed method is its remarkable robustness to strong downsampling or strong JPEG compression, which are regularly applied, e.g., in social media and on news web sites. In this scenario, the performance of statistical approaches quickly drops to guessing chance. However, the proposed descriptor remains unaffected, even for downsampling to less than 50% of the image size and JPEG compression quality 30. This robustness enables the use of the proposed descriptor on images that can not be analyzed by statistical approaches.

Acknowledgment

This material is based on research sponsored by the Air Force Research Laboratory and the Defense Advanced Re-

search Projects Agency under agreement number FA8750-16-2-0204. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon.

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory and the Defense Advanced Research Projects Agency or the U.S. Government.

REFERENCES

[1] S. J. Nightingale, K. A. Wade, and D. G. Watson, "Can people identify original and manipulated photos of real-world scenes?" *Cognitive Research: Principles and Implications*, vol. 2, no. 1, p. 30, Jul. 2017. [Online]. Available: https://doi.org/10.1186/s41235-017-0067-2

[2] H. Farid, *Photo Forensics*. The MIT Press, 2016.

[3] J. Redi, W. Taktak, and J.-L. Dugelay, "Digital Image Forensics: A Booklet for Beginners," *Multimedia Tools and Applications*, vol. 51, no. 1, pp. 133–162, Jan. 2011.

[4] J. Lukáš, J. Fridrich, and M. Goljan, "Digital Camera Identification From Sensor Pattern Noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.

[5] T. Bianchi, A. Piva, and F. Perez-Gonzalez, "Near Optimal Detection of Quantized Signals and Application to JPEG Forensics," in *IEEE International Workshop on Information Forensics and Security*, Nov. 2013, pp. 168–173.

[6] T. H. Thai, R. Cogranne, F. Retraint, and T.-N.-C. Doan, "JPEG Quantization Step Estimation and Its Applications to Digital Image Forensics," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 123–133, Jan. 2017.

[7] M. Kirchner, "Linear Row and Column Predictors for the Analysis of Resized Images," in *ACM SIGMM Multimedia & Security Workshop*, Sep. 2010, pp. 13–18.

[8] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: A new blind image splicing detector," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, Nov. 2015, pp. 1–6.

[9] J. Fridrich and J. Kodovský, "Rich Models for Steganalysis of Digital Images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 868–882, Jun. 2012.

[10] M. Huh, A. Liu, A. Owens, and A. A. Efros, "Fighting Fake News: Image Splice Detection via Learned Self-Consistency," *European Conference on Computer Vision (ECCV)*, 2018.

[11] O. Mayer and M. C. Stamm, "Learned Forensic Source Similarity for Unknown Camera Models," in *International Conference on Acoustics, Speech and Signal Processing*, Apr. 2018.

[12] D. Cozzolino and L. Verdoliva, "Noiseprint: a CNN-based camera model fingerprint," University of Naples, arXiv preprint, 2018, arXiv:1808.08396.

[13] M. K. Johnson and H. Farid, "Exposing Digital Forgeries in Complex Lighting Environments," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 450–461, Sep. 2007. [Online]. Available: http://dx.doi.org/10.1109/TIFS.2007.903848

[14] B. Peng, W. Wang, J. Dong, and T. Tan, "Automatic detection of 3D lighting inconsistencies via a facial landmark based morphable model," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 3932–3936.

[15] J. Zheng, X. Song, J. Ren, and T. Zhu, "Exposing photo manipulation using geometry and shadows," in *Sixth International Conference on Digital Image Processing (ICDIP 2014)*, C. M. Falco, C.-C. Chang, and X. Jiang, Eds. SPIE-Intl Soc Optical Eng, Apr. 2014. [Online]. Available: http://dx.doi.org/10.1117/12.2064531

[16] E. Kee, J. F. O'Brien, and H. Farid, "Exposing Photo Manipulation from Shading and Shadows," *ACM Transactions on Graphics*, vol. 33, no. 5, pp. 1–21, Sep. 2014. [Online]. Available: http://dx.doi.org/10.1145/2629646

[17] T. J. Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. d. R. Rocha, "Exposing Digital Image Forgeries by Illumination Color Classification," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 7, pp. 1182–1194, Jul. 2013.

[18] C. Riess, S. Pfaller, and E. Angelopoulou, *Reflectance Normalization in Illumination-Based Image Manipulation Detection*. Cham: Springer International Publishing, 2015, pp. 3–10.

[19] E. Kee and H. Farid, "Exposing digital forgeries from 3-D lighting environments," in *2010 IEEE International Workshop on Information Forensics and Security*. Institute of Electrical and Electronics Engineers (IEEE), Dec. 2010. [Online]. Available: http://dx.doi.org/10.1109/WIFS.2010.5711437

[20] B. Peng, W. Wang, J. Dong, and T. Tan, "Improved 3D lighting environment estimation for image forgery detection," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, Nov. 2015, pp. 1–6.

[21] ——, "Optimized 3D Lighting Environment Estimation for Image Forgery Detection," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 2, pp. 479–494, Feb. 2017.

[22] T. Carvalho, H. Farid, and E. Kee, "Exposing photo manipulation from user-guided 3D lighting analysis," in *Media Watermarking, Security, and Forensics 2015*, A. M. Alattar, N. D. Memon, and C. D. Heitzenrater, Eds. SPIE-Intl Soc Optical Eng, Mar. 2015. [Online]. Available: http://dx.doi.org/10.1117/12.2075544

[23] W. Fan, K. Wang, F. Cayre, and Z. Xiong, "3D lighting-based image forgery detection using shape-from-shading," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug. 2012, pp. 1777–1781.

[24] J. Seuffert, M. Stamminger, and C. Riess, "Towards Forensic Exploitation of 3-D Lighting Environments in Practice," in *Sicherheit*, Apr. 2018, pp. 159–169.

[25] M. K. Johnson and H. Farid, "Exposing Digital Forgeries by Detecting Inconsistencies in Lighting," in *Proceedings of the 7th Workshop on Multimedia and Security*. New York, NY, USA: ACM, 2005, pp. 1–10. [Online]. Available: http://doi.acm.org/10.1145/1073170.1073171

[26] C. Riess, M. Unberath, F. Naderi, S. Pfaller, M. Stamminger, and E. Angelopoulou, "Handling Multiple Materials for Exposure of Digital Forgeries using 2-D Lighting Environments," *Multimedia Tools and Applications*, 2016. [Online]. Available: https://www5.informatik.uni-erlangen.de/Forschung/Publikationen/2016/Riess16-HMM.pdf

[27] A. P. Pentland, "Finding the illuminant direction," *J. Opt. Soc. Am.*, vol. 72, no. 4, pp. 448–455, Apr. 1982. [Online]. Available: http://www.osapublishing.org/abstract.cfm?URI=josa-72-4-448

[28] C.-H. Lee and A. Rosenfeld, "Improved Methods of Estimating Shape from Shading Using the Light Source Coordinate System," *Artif. Intell.*, vol. 26, no. 2, pp. 125–143, May 1985. [Online]. Available: http://dx.doi.org/10.1016/0004-3702(85)90026-8

[29] Q. Zheng and R. Chellappa, "Estimation of Lambertian Reflectance Map," in *1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers, 1990.*, vol. 2, Nov. 1990, pp. 805–.

[30] R. Dosselmann and X. D. Yang, "Determining Light Direction in Spheres using Average Gradient," University of Regina, Tech. Rep., 2009.

[31] ——, "Improved method of finding the illuminant direction of a sphere," *Journal of Electronic Imaging*, vol. 22, no. 1, pp. 013 035–013 035, 2013. [Online]. Available: http://dx.doi.org/10.1117/1.JEI.22.1.013035

[32] "The Stanford 3D Scanning Repository." [Online]. Available: http://graphics.stanford.edu/data/3Dscanrep/

[33] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 2980–2988.

[34] W. Abdulla, "Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," https://github.com/matterport/Mask_RCNN, 2017.

[35] J. T. Barron and J. Malik, "Shape, Illumination, and Reflectance from Shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 8, pp. 1670–1687, Aug 2015.

[36] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam Library of Object Images," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 103–112, 2005. [Online]. Available: https://ivi.fnwi.uva.nl/isis/publications/2005/GeusebroekIJCV2005

[37] "COCO - Common Object in Context." [Online]. Available: cocodataset.org

[38] "GRIP - Splicebuster: A new blind image splicing detector." [Online]. Available: http://www.grip.unina.it/research/83-image-forensics/100-splicebuster.html

[39] "GitHub - minyoungg/selfconsistency: Code for the paper: Fighting Fake News: Image Splice Detection via Learned Self-Consistency." [Online]. Available: https://github.com/minyoungg/selfconsistency

[40] "IEEE IFS-TC Image Forensics Challenge - Image Corpus." [Online]. Available: http://ifc.recod.ic.unicamp.br/fc.website/index.py?sec=5

[41] I. Krasin, T. Duerig, N. Alldrin, V. Ferrari, S. Abu-El-Haija, A. Kuznetsova, H. Rom, J. Uijlings, S. Popov, S. Kamali, M. Malloci, J. Pont-Tuset, A. Veit, S. Belongie, V. Gomes, A. Gupta,

C. Sun, G. Chechik, D. Cai, Z. Feng, D. Narayanan, and K. Murphy, "OpenImages: A public dataset for large-scale multi-label and multi-class image classification." *Dataset available from https://storage.googleapis.com/openimages/web/index.html*, 2017.

[42] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": Interactive Foreground Extraction Using Iterated Graph Cuts," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. New York, NY, USA: ACM, 2004, pp. 309–314. [Online]. Available: http://doi.acm.org/10.1145/1186562.1015720

Falko Matern received the M.Sc. degree in computer science from the Friedrich-Alexander University Erlangen-Nuremberg (FAU), Erlangen, Germany, in 2016, where he is currently pursuing the Ph.D. degree in computer science. His current research interests include image and video forensics, image processing and computer vision.

Christian Riess received the Ph.D. degree in computer science from the Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany, in 2012. From 2013 to 2015, he was a Postdoc at the Radiological Sciences Laboratory, Stanford University, Stanford, CA, USA. Since 2015, he is the head of the Phase-Contrast X-ray Group at the Pattern Recognition Laboratory at FAU. Since 2016, he is senior researcher and head of the Multimedia Security Group at the IT Infrastructures Lab at FAU. He is currently a member of the IEEE Information Forensics and Security Technical Committee. His research interests include all aspects of image processing and imaging, particularly with applications in image and video forensics, X-ray phase contrast, color image processing, and computer vision.

Marc Stamminger is a full professor for Visual Computing at the Friedrich-Alexander University Erlangen-Nuremberg. His research covers various areas of visual computing, in particular real-time rendering and visualization, 3d reconstruction of static and dynamic objects, and virtual and augmented reality, as well as image forensics. He is associate editor of the Journal on Computer Graphics Tools, was co-chair of the program committee of Eurographics 2009 and the Eurographics Symposium on Rendering in 2010, and member of the program committee of all major computer graphics conferences, such as Siggraph, Siggraph Asia, Eurographics, ACM I3D, or ACM High Performance Graphics. Marc Stamminger is member of the Eurographics Executive Committee and currently serves as the Eurographics Professional Board chair.