H. Langweg, M. Meier, B.C. Witt, D. Reinhardt (Hrsg.): Sicherheit 2018,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018 159

Towards Forensic Exploitation of 3-D Lighting Environments in Practice

Julian Seuffert[1], Marc Stamminger[2], Christian Riess[3]

Abstract: The goal of image forensics is to determine authenticity and origin of a digital image or video without an embedded security scheme. Among the existing methods, the probably most well-known physics-based approach is to validate the distribution of incident light on objects of interest. Inconsistent lighting environments are considered as an indication of image splicing. However, one drawback of this approach is that it is quite challenging to use it in practice.

In this work, we propose several practical improvements to this approach. First, we propose a new way of comparing lighting environments. Second, we present a factorization of the overall error into its individual contributions, which shows that the biggest error source are incorrect geometric fits. Third, we propose a confidence score that is trained from the results of an actual implementation. The confidence score allows to define an implementation- and problem-specific threshold for the consistency of two lighting environments.

Keywords: multimedia security; image forensics; physics-based forensics; 3-D lighting estimation

1 Introduction

Visual media plays an important role in our everyday communication. This is partly due to the widespread availability of consumer cameras, and partly due to the ease of distributing visual media, for example over social media. These new possibilities allow to document events in an unprecedented density. However, when a picture or video is taken as evidence that an event actually happened, it is also important to be able to verify its authenticity. Blind image forensics aims to provide technical tools to authenticate visual media without the help of an embedded security scheme. In the recent years, several books and overview papers have been published on image forensics, for example [RTD11, SM13, Fa16].

This work is about so-called physics-based methods in image forensics. The guiding idea of these methods is to validate the physics in the depicted scene for its consistency,

[1] Friedrich-Alexander University Erlangen-Nuernberg, IT-Security Infrastructures Lab, Martensstr. 3, 91058 Erlangen, Germany julian.seuffert@fau.de
[2] Friedrich-Alexander University Erlangen-Nuernberg, Computer Graphics Lab, Cauerstr. 11, 91058 Erlangen, Germany marc.stamminger@fau.de
[3] Friedrich-Alexander University Erlangen-Nuernberg, IT-Security Infrastructures Lab, Martensstr. 3, 91058 Erlangen, Germany christian.riess@fau.de

doi:10.18420/sicherheit2018_12

like the direction and color of the incident light [JF07a, RA10], consistency of specular highlights [JF07b], reflections [OF12], or shadows [Zh09, KOF14]. However, current physics-based methods have the disadvantages that their applicability depends on the visual content in the scene, and that they typically require user interaction.

Several works perform the analysis to human faces only [KF10, Pe16, Pe17]. Faces commonly occur in pictures and videos, and it may be a forensic goal to validate the composition of people in a scene. Operating on faces has the advantage that there exists software to estimate a 3-D model of the face, which then allows to compute the 3-D distribution of incident light.

Nevertheless, using 3-D lighting forensics in practice is still challenging, and oftentimes requires expert knowledge. The goal of this paper is to narrow the gap from the base algorithm in literature towards its use in practice. Specifically, we present an systematic analysis of the algorithmic steps, and we propose a practical, trainable confidence score that adapts to the specific implementation of the algorithm at hand. Using the confidence score, a practitioner not only obtains an assessment whether two lighting environments are identical, but also a probability with which this assessment is true. In detail, the contributions of the paper consist of

- A study of the spherical harmonics model for lighting representation and comparison with an approach that avoids extrapolation over surface normals without observations.
- A factorization of the fitting error into its individual physical contributions, to better understand its impact on the estimation.
- A proposal for a confidence score that describes the reliability of the individual pipeline implementation.

The paper is organized as follows. We present the underlying model and the computation of lighting environments in Sec. 2. Approaches for comparing lighting environment are presented in Sec. 3, the error factorization in Sec. 4, and the proposed confidence score in Sec. 5. Finally, our experiments are presented in Sec. 6.

2 3-D Lighting Estimation

Johnson, Kee, and Farid [JF07a, KF10] proposed to estimate the distribution of incident light on objects of interest in a scene. Such a distribution is called a lighting environment. Assuming a single, infinitely distant light source (such as the sun in outdoor scenes) and no inter-reflections, then all scene objects must exhibit identical lighting environments. With minor modifications, these assumptions can be used to obtain a forensic test on a given input image, by approximating the rays of the sun or another distant light source as parallel.

Johnson and Farid proposed to compute 2-D lighting environments from monochromatic object contours. Kee and Farid proposed later to estimate more reliable 3-D lighting

environments from the shape of human faces [KF10], using existing software for fitting a 3-D face shape to a 2-D image.

Kee and Farid model a lighting environment with the help of surface normals and their associated pixel intensities [KF10]:

$$e(\vec{n}) = \int_{\Omega} l(\vec{v}) r(\vec{v}, \vec{n}) d\Omega \ . \tag{1}$$

Here, $e(\vec{n})$ is the observed intensity of a pixel with surface normal $\vec{n}$. Ω is the hemisphere of all light directions that fall on a patch with surface normal $\vec{n}$, $l(\vec{v})$ denotes light that falls onto that pixel coming from direction $\vec{v}$, and $r(\vec{v}, \vec{n})$ denotes the reflectance function for that patch. The model becomes particularly convenient when $r(\vec{v}, \vec{n})$ is assumed to be Lambertian (purely diffuse), such that the reflected intensity is the cosine between $\vec{v}$ and $\vec{n}$.

Under the assumption of Lambertian reflectance, the observed intensity $e(\vec{n})$ can be represented by second order spherical harmonics, i.e.,

$$e(\vec{n}) = \sum_{n=0}^{2} \sum_{m=-n}^{n} l_{n,m} Y_{n,m}(\vec{n}) \ , \tag{2}$$

where $Y_{n,m}$ denotes the m-th spherical harmonics basis function of order n, and $l_{n,m}$ is a weighting coefficient. These coefficients can be directly estimated from the observed intensities. Let $i(\vec{x}_i)$ denote the intensity of the i-th pixel $\vec{x}_i$ from a face, and $\vec{n}(\vec{x}_i)$ the surface normal of the face at position $\vec{x}_i$. Then, Kee and Farid propose to estimate the lighting coefficients $\vec{l}$ from the linear equation

$$\mathbf{M} \cdot \vec{l} = \vec{b} \ , \tag{3}$$

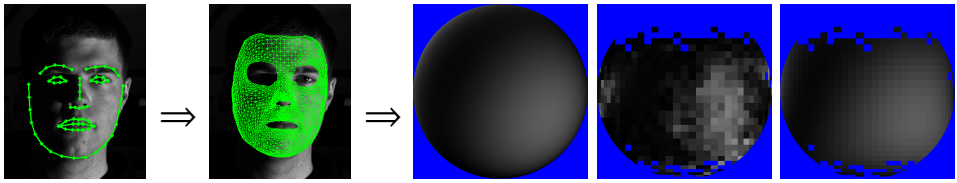
where $\vec{b} = (i(\vec{x}_1), \ldots, i(\vec{x}_N))^{\mathrm{T}}$ are the observed intensities and

$$\mathbf{M} = \begin{pmatrix} \pi Y_{0,0}(\vec{n}(\vec{x}_1)) & \ldots & \frac{\pi}{4} Y_{2,2}(\vec{n}(\vec{x}_1)) \\ \vdots & \ddots & \vdots \\ \pi Y_{0,0}(\vec{n}(\vec{x}_N)) & \ldots & \frac{\pi}{4} Y_{2,2}(\vec{n}(\vec{x}_N)) \end{pmatrix} \ . \tag{4}$$

Equation 3 is solved for $\vec{l}$ via least squares, i.e.,

$$\vec{l} = (\mathbf{M}^{\mathrm{T}}\mathbf{M})^{-1}\mathbf{M}^{\mathrm{T}}\vec{b} \ . \tag{5}$$

The mathematical framework by Kee and Farid is elegant, but several special cases are not explicitly discussed. For example, Peng *et al.* later proposed to automate this pipeline [Pe16], and to add a more complex model for surface reflectance [Pe17].



(a) extract landmarks (b) custom. face model (c) spherical harmonics (d) IBS (e) HIBS

Fig. 1: Computed landmarks (a) and fitted 3-D model (b) for obtaining per pixel a 3-D surface normal. Lighting environments can be computed as spherical harmonics [JF07a, KF10] (c), intensity binned spheres (d), or hybrid intensity binned spheres (e).

3 Comparison of Lighting Environments

Spherical harmonics are the standard representation for lighting environments [JF07a, KF10]. Two lighting environments are compared by expanding the spherical harmonics coefficients $\vec{l}$ of each object of interest to an intensity profile. The correlation between two such intensity profiles is then used to quantify their similarity.

The suitability of spherical harmonics for forensic lighting estimation has to our knowledge not been investigated. We find this surprising, as spherical harmonics consist of low frequencies averaged over potentially unevenly distributed or missing observations, which might lead to wrong results: for example, persons with dense head hair will contribute almost no surface normals pointing upwards. Nevertheless, the spherical harmonics model represents and weighs all angular directions of the hemisphere pointing towards the camera equally. Thus, it is not possible to distinguish the impact of the actual observations on similarity ρ from artifacts from extrapolation of lighting environments. An example spherical harmonics model is shown in Fig. 1c. To compare two spherical harmonics representations, we compute their correlation directly on their SH coefficients as proposed earlier [JF07a].

As an alternative, we propose to consider what we call a "intensity binning sphere" (IBS), where intensities are binned by their surface normals, in steps of 5°. A concrete example IBS is shown in Fig. 1d. In contrast to Fig. 1c, the model does for example not cover surface normals that point upwards, as we do not observe any skin area with that orientation. This idea can be further improved by combining both approaches into a representation that we call "hybrid IBS" (HIBS), shown in Fig. 1e. Here, the spherical harmonics model is cropped to angular bins with a width of 5° that are actually filled with observations. Two IBS or HIBS representations are compared by computing the correlation over the intersection of non-empty bins of both representations.

(a) diff. & spec. (b) purely diffuse (c) global lighting (d) local lighting

Fig. 2: Example renderings from the synthetic dataset. From left to right: diffuse and specular, purely diffuse, purely diffuse with global lighting, purely diffuse with local lighting.

4 Error Factorization

Errors of the individual methods accumulate in the processing chain. To optimize the algorithm, it is important to understand the relative contribution of each factor to the overall error. Based on the physical model in Eqn. 1, we investigate three potential sources of error. First, the face fit might yield slightly incorrect surface normals. Second, the required reflectance model might be more complex than the assumed pure Lambertian model. Third, self-shadows due to occlusions could have an impact on real images.

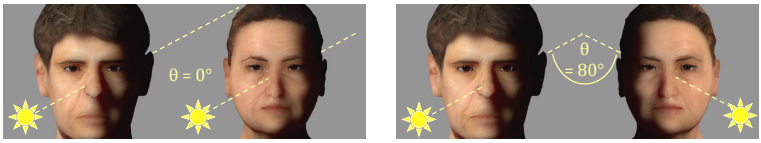
To understand the impact of these three factors, we created a synthetic face dataset consisting of 12 subjects using *MakeHuman* and *Blender*. In *MakeHuman*, subjects are created from a parameters like age or gender. The surface normals from the synthetic face model can be directly taken as ground truth to evaluate the estimation errors in the geometry.

To obtain data for the other two hypotheses, we re-rendered the data with the default amount of specular reflections and the computation of self-shadows. Example faces are shown in Fig. 2. Local lighting denotes illumination of surface patches without self-shadowing. Global lighting includes self-shadowing. The limited photo realism in this rendered data is a minor concern, as only relative performance differences between the variants are considered. The dataset and the code for generating the data is publicly available on our website[4].

5 Confidence Score

We seek to transform the correlations from Sec. 3 into a confidence score that indicates whether the underlying lighting environments might be identical. Computation of the confidence score depends on the concrete implementation of the face fitting and lighting environment computation. Thus, we propose to learn it from training data, i.e., from face images that are acquired under known lighting. In our experiments, we further assume a single dominant light source, such that we can determine the angular resolution of the lighting environment estimation. Let us denote identical lighting environments as “consistent”, and different lighting environments as “inconsistent”. By extension, we consider faces under

[4] https://faui1-files.cs.fau.de/public/mmsec/datasets/sfd



(a) consistent lighting (b) inconsistent lighting

Fig. 3: Angle θ between the dominant light sources on two faces.

identical lighting as consistent, and faces under different lighting as inconsistent. If only a single light source illuminates a face, let θ be the angular difference between the dominant light sources of both faces. Then, as shown in Fig. 3, $\theta \neq 0$ indicates inconsistent lighting environments (positive class), and $\theta = 0$ consistent lighting environments (negative class).

We consider the confidence score as a function of the type I and type II error on the respective dataset. The type I error α denotes the relative number of samples being incorrectly classified as inconsistent. The type II error β denotes the relative number of samples being incorrectly classified as consistent. $\kappa_p = 1 - \alpha$ and $\kappa_n = 1 - \beta$ denote the confidence score of labelling the lighting as "inconsistent" and "consistent", respectively.

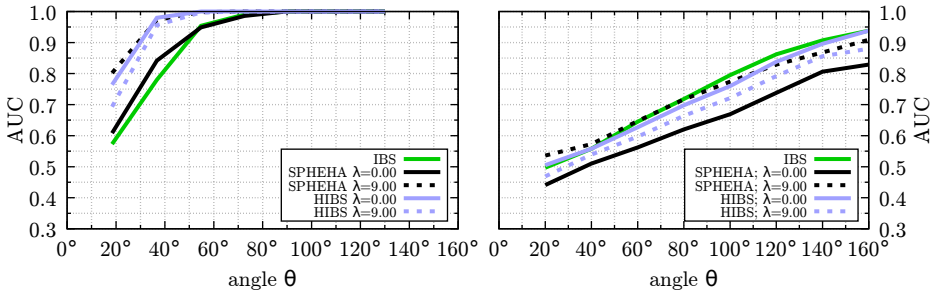
One possible decision template is to choose the class with the highest confidence score κ_n or κ_p. Hence, such a decision exhibits a confidence score function $\kappa_{np}(\rho) = \max(\kappa_n(\rho), \kappa_p(\rho))$ whereas $\kappa_n = 1 - \kappa_p$.

6 Experiments

We first describe the used datasets and experimental protocol in Sec. 6.1 and Sec. 6.2, and then present the experimental results in Sec. 6.3.

6.1 Datasets

We model 12 synthetic subjects, consisting of two males and two females of African, Asian and Caucasian descent and varying age using *MakeHuman v1.1.1*. Rendering is performed with *Blender v2.79*. We refer to this dataset as the Synthetic Face Dataset (SFD). Each subject is illuminated by nine distant point lights with a pitch angle of $\phi = 23°$ and with a yaw angle of $\psi \in \{-80°, -60°, \ldots, +80°\}$. Two light sources exhibit an angular distance of $\theta \in \Theta_{\text{SFD}} = \{0°, 18°, \ldots, 130°\}$. This parametrization allows for a total of 36 possible light source combinations. All combinations of two active lights were rendered with four distinct illumination properties, denoted as "global-spec", "global-lamb", "local-spec", and "local-lamb". "global" denotes global illumination via raytracing and includes cast self-shadows. "local" denotes local lighting, i.e., shading is only determined by the surface orientation, and lighting by the angle between incident light $\vec{v}$ and surface normal $\vec{n}$.



(a) Synthetic Face Dataset (ground truth geometry) (b) Extended Yale Face Dataset B

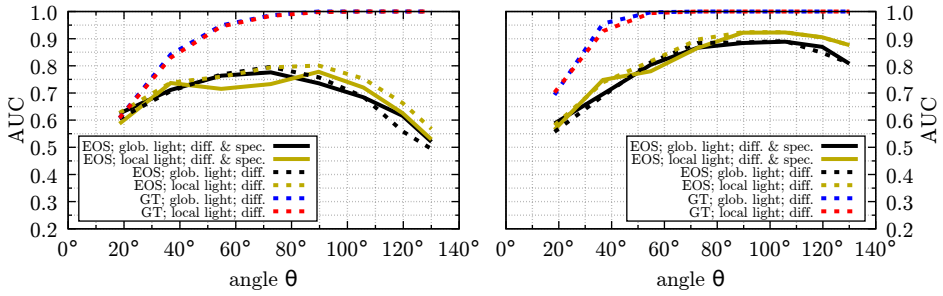
Fig. 4: Detection of inconsistent lighting environments depending on the angular difference between light sources.

"spec" denotes Lambertian and specular reflectance. "lamb" denotes Lambertian reflectance. Additionally, we store the ground truth facial geometry. Only skin pixels are used for further processing, other pixels were ignored. In total, the dataset consists of 432 images. Sample images are shown in Fig. 2 and Fig. 3.

Our experiments on real data are performed on the "Extended Yale Face Database B" [GBK01], consisting of 28 subjects, each illuminated by a light from one out of 64 positions. We use all 1792 frontal view, single light source images. There are in total 2071 possible combinations of light sources. We round θ up to a multiple of 20°. The resulting set of angular distances between lights is $\Theta_{\text{YALE}} = \{0°, 20°, 40°, \ldots, 180°\}$.

6.2 Evaluation Protocol

Evaluation is performed on pairs of randomly chosen (different) subjects, with controlled angular differences between the subjects' lighting environments. Paired lighting environments from the SFD dataset are always rendered with identical options, i.e., "global-spec", "global-lamb", "local-spec", or "local-lamb". Pairs of lighting environments are grouped by their angular distances θ. Each experiment uses N pairs with $\theta = 0°$ and N pairs with $\theta = i \times 20°$ for $i \in \{1, \ldots, 8\}$. For the Extended Yale Face Database B, we use $N_{\text{yale}} = 1700$ and for SFD we chose $N_{\text{SFD}} = 128$. Performances are typically given as area under the curve (AUC). The samples of consistent lighting are identical across experiments. To compute the confidence scores, we use $8 \times N$ different samples of consistent lighting. We set $N_{\text{yale}} = 1500$ and $N_{\text{SFD}} = 64$ due to keep the number of consistent and inconsistent samples balanced.



(a) IBS (b) HIBS $\lambda = 9.0$

Fig. 5: Impact of the reflectance model: local vs. global lighting to model shadows, inclusion or omission of specularities, with estimated ("EOS") or ground truth ("GT") geometry.

6.3 Results

The automated lighting estimation pipeline is set up as follows. Faces are detected using the OpenCV v3.1 library. For each face in the scene, 68 facial landmarks are determined using Dlib [Ki09]. Then, the EOS framework [Hu16] is used to fit the 3-D model. Surface normals and observed intensities are jointly used to estimate the lighting environment (Eqn. 5).

6.3.1 Comparison of Lighting Environment Representations

Figure 4 shows results for comparing the classic spherical harmonics representation (denoted as "SPHEHA") with the intensity binning sphere IBS and the combination of both, HIBS. SPHEHA and HIBS are evaluated once without any regularization (denoted by the regularization weight $\lambda = 0$) and once with a Tikhonov regularizer [JF07a]. In preliminary experiments, we determined that a regularization weight of $\lambda = 9$ worked reasonably well on a range of scenarios, and we continue to use that value.

The results in Fig. 4a are computed on the synthetic dataset with available ground truth geometry. The results in Fig. 4b are computed on the Extended Yale Face Database B. Figure 4a is a best case for all estimators. The AUC is consistently high, and almost perfect for all processing variants at angular differences of about 60°. Figure 4b on real data is the most challenging scenario, which can be seen from the fact that even for lighting environments with an angular distance of 80°, the AUC is still in the range of 0.7.

On SFD, regularized spherical harmonics and HIBS perform very well. On the real data, IBS performs best together with regularized spherical harmonics, but the differences between the approaches are overall less pronounced. The AUC of the spherical harmonics representation is slightly higher, but IBS and HIBS yield quite good results for the confidence score computation below, which is why we believe that both approaches are worth to consider.

$\mathbf{E}\{\kappa_{np}\}$	**IBS**	**SPHEHA** $\lambda = 0$	**SPHEHA** $\lambda = 9$	**HIBS** $\lambda = 0$	**HIBS** $\lambda = 9$
SFD	0.926	0.898	0.954	**0.974**	0.948
ExtYaleB	0.696	0.622	0.687	**0.698**	0.670

Tab. 1: Expected decision confidence score

6.3.2 Error Factorization

Fig. 5 shows a comparison between local and global lighting, diffuse and diffuse+specular reflectance, and estimated geometry ("EOS") versus ground truth geometry ("GT") on the synthetic dataset. The plots consist of two clusters. Using ground truth geometry outperforms all other variants by a large margin. At the same time, the differences between all other variants are minor. Figure 5 shows that accuracy of surface normals has by far the biggest impact on the estimation error, and that limitations in the computational model are of secondary concern.

6.3.3 Confidence Score

Figure 6 shows the confidence scores on the SFD dataset with known geometry and on the Yale dataset. Left, confidence scores for consistent lighting are shown. Right, confidences for inconsistent lighting are shown. On the SFD dataset, HIBS exhibits the steepest transition between consistent and inconsistent lighting. On the Yale dataset, confidence scores are mostly lower. The expected confidence score $E\{\kappa_{np}\}$ incorporates both the confidence score value $\kappa_{np}(\rho)$ and the correlation value density $p(\rho)$,

$$E\{\kappa_{np}\} = \sum_{\rho=-1.0}^{1.0} p(\rho) \cdot \kappa_{np}(\rho) \quad . \tag{6}$$

From a user perspective, higher values $E\{\kappa_{np}\}$ can suggest a more reliable decision. The expectation values in Tab. 1 show that HIBS with $\lambda = 0$ performs best on both datasets.

7 Conclusions

We present three components that support the use of 3-D lighting environments for physics-based detection of image manipulations. First, we propose to include for the similarity computation of lighting environments only angular ranges that are backed up by actual observations (as opposed to extrapolated intensities). Second, we present a method to factorize and analyze the error contributions of the face fitting and correlation computation. It turns out that the impact of violations to the physical model due to specularities and

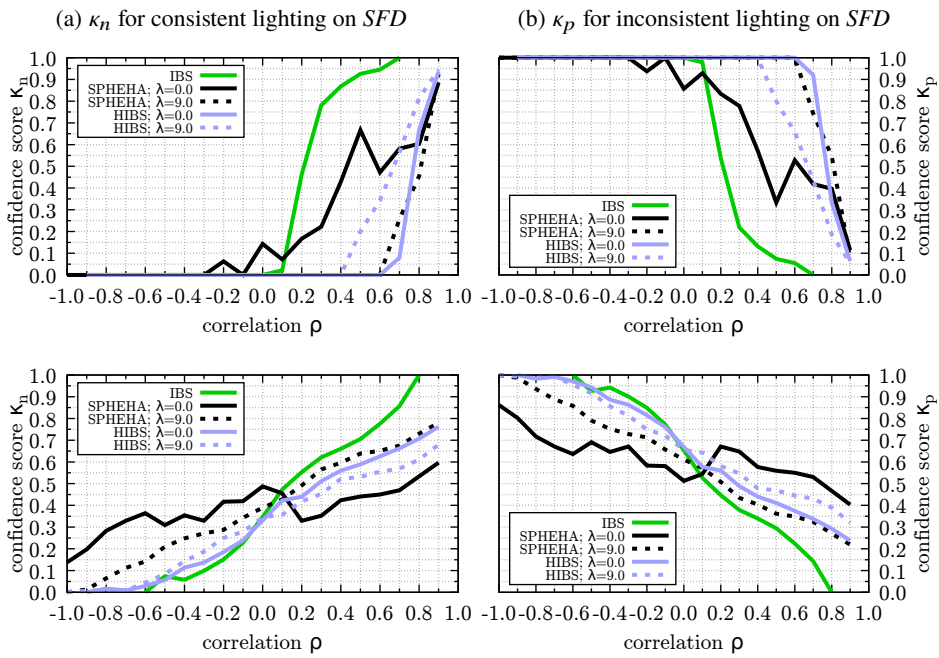


Fig. 6: Confidence scores with known geometry on synthetic data (SFD) and on real-world data (Extended Yale Face Dataset B).

self-shadows is minor compared to geometric fitting errors. This indicates that a high-quality face fit has by far the biggest impact on the overall accuracy. Third, for practical use, we propose a lighting environment confidence score that is learned from the actual data, specifically for the available implementation of the processing pipeline.

Acknowledgements

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

[Fa16] Farid, Hany: Photo Forensics. MIT Press, 2016.

[GBK01] Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J.: From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6):643–660, June 2001.

[Hu16] Huber, P.; Hu, G.; Tena, R.; Mortazavian, P.; Koppen, W.; Christmas, W.; Rätsch, M.; Kittler, J.: A Multiresolution 3D Morphable Face Model and Fitting Framework. In: International Conference on Computer Vision Theory and Applications. pp. 79–86, February 2016.

[JF07a] Johnson, M.; Farid, H.: Exposing Digital Forgeries in Complex Lighting Environments. IEEE Transactions on Information Forensics and Security, 2(3):450–461, September 2007.

[JF07b] Johnson, Micah; Farid, Hany: Exposing Digital Forgeries through Specular Highlights on the Eye. In: Proceedings of the 9th International Workshop on Information Hiding. Saint Malo, France, pp. 311–325, September 2007.

[KF10] Kee, Eric; Farid, Hany: Exposing Digital Forgeries from 3-D Lighting Environments. In: Proceedings of the 2nd IEEE International Workshop on Information Forensics and Security. Seattle, WA, USA, December 2010.

[Ki09] King, Davis E.: Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research, 10:1755–1758, July 2009.

[KOF14] Kee, Eric; O'Brien, James F.; Farid, Hany: Exposing Photo Manipulation from Shading and Shadows. ACM Transactions on Graphics, 33(5):165:1–21, August 2014.

[OF12] O'Brien, James F.; Farid, Hany: Exposing Photo Manipulation with Inconsistent Reflections. ACM Transactions on Graphics, 31(1):1–11, January 2012.

[Pe16] Peng, Bo; Wang, Wei; Dong, Jing; Tan, Tieniu: Automatic Detection of 3-D Lighting Inconsistencies via a Facial Landmark based Morphable Model. In: IEEE International Conference on Image Processing. pp. 3932–3936, September 2016.

[Pe17] Peng, Bo; Wang, Wei; Dong, Jing; Tan, Tieniu: Optimized 3D Lighting Environment Estimation for Image Forgery Detection. IEEE Transactions on Information Forensics and Security, 12(2):479–494, February 2017.

[RA10] Riess, C.; Angelopoulou, E.: Scene Illumination as an Indicator of Image Manipulation. In: Proceedings of the 12th International Conference on Information Hiding. Calgary, AB, Canada, pp. 66–80, June 2010.

[RTD11] Redi, Judith; Taktak, Wiem; Dugelay, Jean-Luc: Digital Image Forensics: A Booklet for Beginners. Multimedia Tools and Applications, 51(1):133–162, January 2011.

[SM13] Sencar, Husrev Taha; Memon, Nasir, eds. Digital Image Forensics. Springer, 2013.

[Zh09] Zhang, Wei; Cao, Xiaochun; Zhang, Jiawan; Zhu, Jigui; Wang, Ping: Detecting Photographic Composites using Shadows. In: Proceedings of the IEEE International Conference on Multimedia and Expo. Cancun, Mexico, pp. 1042–1045, June 2009.