Exposing Digital Image Forgeries by Illumination Color Classification

Tiago José de Carvalho, *Student Member, IEEE*, Christian Riess, *Associate Member, IEEE*, Elli Angelopoulou, *Member, IEEE*, Hélio Pedrini, *Member, IEEE*, and Anderson de Rezende Rocha, *Member, IEEE*

***Abstract*—For decades, photographs have been used to document space-time events and they have often served as evidence in courts. Although photographers are able to create composites of analog pictures, this process is very time consuming and requires expert knowledge. Today, however, powerful digital image editing software makes image modifications straightforward. This undermines our trust in photographs and, in particular, questions pictures as evidence for real-world events. In this paper, we analyze one of the most common forms of photographic manipulation, known as image composition or splicing. We propose a forgery detection method that exploits subtle inconsistencies in the color of the illumination of images. Our approach is machine-learning-based and requires minimal user interaction. The technique is applicable to images containing two or more people and requires no expert interaction for the tampering decision. To achieve this, we incorporate information from physics- and statistical-based illuminant estimators on image regions of similar material. From these illuminant estimates, we extract texture- and edge-based features which are then provided to a machine-learning approach for automatic decision-making. The classification performance using an SVM meta-fusion classifier is promising. It yields detection rates of 86% on a new benchmark dataset consisting of 200 images, and 83% on 50 images that were collected from the Internet.**

***Index Terms*—Color constancy, illuminant color, image forensics, machine learning, spliced image detection, texture and edge descriptors.**

I. Introduction

EVERY day, millions of digital documents are produced by a variety of devices and distributed by newspapers, magazines, websites and television. In all these information channels, images are a powerful tool for communication. Unfortunately, it is not difficult to use computer graphics and image processing techniques to manipulate images. Quoting Russell Frank, a Professor of Journalism Ethics at Penn State University, in 2003 after a Los Angeles Times incident involving a doctored photograph from the Iraqi front: "Whoever said the camera never lies was a liar". How we deal with photographic manipulation raises a host of legal and ethical questions that must be addressed [1]. However, before thinking of taking appropriate actions upon a questionable image, one must be able to detect that an image has been altered.

Fig. 1. How can one assure the authenticity of a photograph? Example of a spliced image involving people.

Image composition (or splicing) is one of the most common image manipulation operations. One such example is shown in Fig. 1, in which the girl on the right is inserted. Although this image shows a harmless manipulation case, several more controversial cases have been reported, e.g., the 2011 Benetton UnHate advertising campaign[1] or the diplomatically delicate case in which an Egyptian state-run newspaper published a manipulated photograph of Egypt's former president, Hosni Mubarak, at the front, rather than the back, of a group of leaders meeting for peace talks[2].

When assessing the authenticity of an image, forensic investigators use all available sources of tampering evidence. Among other telltale signs, illumination inconsistencies are potentially effective for splicing detection: from the viewpoint of a manipulator, proper adjustment of the illumination conditions is hard to achieve when creating a composite image [1].

In this spirit, Riess and Angelopoulou [2] proposed to analyze illuminant color estimates from local image regions. Unfortunately, the interpretation of their resulting so-called *illuminant maps* is left to human experts. As it turns out, this decision is, in practice, often challenging. Moreover, relying on visual assessment can be misleading, as the human visual system is quite inept at judging illumination environments in pictures [3], [4]. Thus, it is preferable to transfer the tampering decision to an objective algorithm.

Manuscript received February 05, 2013; revised May 07, 2013; accepted May 08, 2013. Date of publication June 03, 2013; date of current version June 13, 2013. This work was supported in part by Unicamp, in part by Unicamp Faepex, in part by CNPq, in part by FAPESP, in part by CAPES, in part by IF Sudeste MG, and in part by Microsoft. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Mauro Barni.

T. J. de Carvalho, H. Pedrini, and A. R. Rocha are with the RECOD Laboratory, Institute of Computing, University of Campinas, Campinas, 13083-970, Brazil (e-mail: tjose@ic.unicamp.br; helio@ic.unicamp.br; anderson@ic.unicamp.br).

C. Riess and E. Angelopoulou are with the Pattern Recognition Laboratory, University of Erlangen-Nuremberg, Erlangen, 91054, Germany (e-mail: riess@i5.cs.fau.de; elli@i5.cs.fau.de).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2013.2265677

[1]http://press.benettongroup.com/

[2]http://thelede.blogs.nytimes.com/2010/09/16/doctored-photo-flatters-egyptian-president/

In this work, we make an important step towards minimizing user interaction for an illuminant-based tampering decision-making. We propose a new semiautomatic method that is also significantly more reliable than earlier approaches. Quantitative evaluation shows that the proposed method achieves a detection rate of 86%, while existing illumination-based work is slightly better than guessing. We exploit the fact that local illuminant estimates are most discriminative when comparing objects of the same (or similar) material. Thus, we focus on the automated comparison of human skin, and more specifically faces, to classify the illumination on a pair of faces as either consistent or inconsistent. User interaction is limited to marking bounding boxes around the faces in an image under investigation. In the simplest case, this reduces to specifying two corners (upper left and lower right) of a bounding box.

In summary, the main contributions of this work are:

- Interpretation of the illumination distribution as object texture for feature computation.
- A novel edge-based characterization method for illuminant maps which explores edge attributes related to the illumination process.
- The creation of a benchmark dataset comprised of 100 skillfully created forgeries and 100 original photographs[3]

In Section II, we briefly review related work in color constancy and illumination-based detection of image splicing. In Section III, we present examples of illuminant maps and highlight the challenges in their exploitation. An overview of the proposed methodology, followed by a detailed explanation of all the algorithmic steps is given in Section IV. In Section V, we introduce the proposed benchmark database and present experimental results. Conclusions and potential future work are outlined in Section VI.

II. Related Work

Illumination-based methods for forgery detection are either geometry-based or color-based. Geometry-based methods focus at detecting inconsistencies in light source positions between specific objects in the scene [5]–[11]. Color-based methods search for inconsistencies in the interactions between object color and light color [2], [12], [13].

Two methods have been proposed that use the direction of the incident light for exposing digital forgeries. Johnson and Farid [7] proposed a method which computes a low-dimensional descriptor of the lighting environment in the image plane (i.e., in 2-D). It estimates the illumination direction from the intensity distribution along manually annotated object boundaries of homogeneous color. Kee and Farid [9] extended this approach to exploiting known 3-D surface geometry. In the case of faces, a dense grid of 3-D normals improves the estimate of the illumination direction. To achieve this, a 3-D face model is registered with the 2-D image using manually annotated facial landmarks. Fan *et al.* [10] propose a method for estimating 3-D illumination using shape-from-shading. In contrast to [9], no 3-D model of the object is required. However, this flexibility comes at the expense of a reduced reliability of the algorithm.

Johnson and Farid [8] also proposed spliced image detection by exploiting specular highlights in the eyes. In a subsequent extension, Saboia *et al.* [14] automatically classified these images by extracting additional features, such as the viewer position. The applicability of both approaches, however, is somewhat limited by the fact that people's eyes must be visible and available in high resolution.

Gholap and Bora [12] introduced physics-based illumination cues to image forensics. The authors examined inconsistencies in specularities based on the dichromatic reflectance model. Specularity segmentation on real-world images is challenging [15]. Therefore, the authors require manual annotation of specular highlights. Additionally, specularities have to be present on all regions of interest, which limits the method's applicability in real-world scenarios. To avoid this problem, Wu and Fang [13] assume purely diffuse (i.e., specular-free) reflectance, and train a mixture of Gaussians to select a proper illuminant color estimator. The angular distance between illuminant estimates from selected regions can then be used as an indicator for tampering. Unfortunately, the method requires the manual selection of a "reference block", where the color of the illuminant can be reliably estimated. This is a significant limitation of the method (as our experiments also show).

Riess and Angelopoulou [2] followed a different approach by using a physics-based color constancy algorithm that operates on partially specular pixels. In this approach, the automatic detection of highly specular regions is avoided. The authors propose to segment the image to estimate the illuminant color locally*per segment*. Recoloring each image region according to its local illuminant estimate yields a so-called *illuminant map* . Implausible illuminant color estimates point towards a manipulated region. Unfortunately, the authors do not provide a numerical decision criterion for tampering detection. Thus, an expert is left with the difficult task of visually examining an illuminant map for evidence of tampering. The involved challenges are further discussed in Section III.

In the field of color constancy, descriptors for the illuminant color have been extensively studied. Most research in color constancy focuses on uniformly illuminated scenes containing a single dominant illuminant. For an overview, see e.g., [16]–[18]. However, in order to use the color of the incident illumination as a sign of image tampering, we require multiple, spatially-bound illuminant estimates. So far, limited research has been done in this direction. The work by Bleier *et al.* [19] indicates that many off-the-shelf single-illuminant algorithms do not scale well on smaller image regions. Thus, problem-specific illuminant estimators are required.

Ebner [20] presented an early approach to multi-illuminant estimation. Assuming smoothly blending illuminants, the author proposes a diffusion process to recover the illumination distribution. Unfortunately, in practice, this approach oversmooths the illuminant boundaries. Gijsenij *et al.* [21] proposed a pixelwise illuminant estimator. It allows to segment an image into regions illuminated by distinct illuminants. Differently illuminated regions can have crisp transitions, for instance between sunlit and shadow areas. While this is an interesting approach,

[3]The dataset will be available in full two-megapixel resolution upon the acceptance of the paper. For reference, all images in lower resolution can be viewed at: http://www.ic.unicamp.br/ tjose/files/database-tifs-small-resolution.zip.

Fig. 2. Example illuminant map that directly shows an inconsistency.

a single illuminant estimator can always fail. Thus, for forensic purposes, we prefer a scheme that combines the results of multiple illuminant estimators. Earlier, Kawakami *et al.* [22] proposed a physics-based approach that is custom-tailored for discriminating shadow/sunlit regions. However, for our work, we consider the restriction to outdoor images overly limiting.

In this paper, we build upon the ideas by [2] and [13]. We use the relatively rich illumination information provided by both physics-based and statistics-based color constancy methods as in [2], [23]. Decisions with respect to the illuminant color estimators are completely taken away from the user, which differentiates this paper from prior work.

III. Challenges in Exploiting Illuminant Maps

To illustrate the challenges of directly exploiting illuminant estimates, we briefly examine the illuminant maps generated by the method of Riess and Angelopoulou [2]. In this approach, an image is subdivided into regions of similar color (superpixels). An illuminant color is locally estimated using the pixels within each superpixel (for details, see [2] and Section IV-A). Recoloring each superpixel with its local illuminant color estimate yields a so-called *illuminant map*. A human expert can then investigate the input image and the illuminant map to detect inconsistencies.

Fig. 2 shows an example image and its illuminant map, in which an inconsistency can be directly shown: the inserted mandarin orange in the top right exhibits multiple green spots in the illuminant map. All other fruits in the scene show a gradual transition from red to blue. The inserted mandarin orange is the only one that deviates from this pattern.

In practice, however, such analysis is often challenging, as shown in Fig. 3. The top left image is original, while the bottom image is a composite with the right-most girl inserted. Several illuminant estimates are clear outliers, such as the hair of the girl on the left in the bottom image, which is estimated as strongly red illuminated. Thus, from an expert's viewpoint, it is reasonable to discard such regions and to focus on more reliable regions, e.g., the faces. In Fig. 3, however, it is difficult to justify a tampering decision by comparing the color distributions in the facial regions. It is also challenging to argue, based on these illuminant maps, that the right-most girl in the bottom image has been inserted, while, e.g., the right-most boy in the top image is original.

Although other methods operate differently, the involved challenges are similar. For instance, the approach by Gholap and Bora [12] is severely affected by clipping and camera white-balancing, which is almost always applied on images from off-the-shelf cameras. Wu and Fang [13] implicitly create illuminant maps and require comparison to a reference region. However, different choices of reference regions lead to different results, and this makes this method error-prone.

Fig. 3. Example illuminant maps for an original image (top) and a spliced image (bottom). The illuminant maps are created with the IIC-based illuminant estimator (see Section IV-A).

Thus, while illuminant maps are an important intermediate representation, we emphasize that further, automated processing is required to avoid biased or debatable human decisions. Hence, we propose a pattern recognition scheme operating on illuminant maps. The features are designed to capture the shape of the superpixels in conjunction with the color distribution. In this spirit, our goal is to replace the expert-in-the-loop, by only requiring annotations of faces in the image.

Note that, the estimation of the illuminant color is error-prone and affected by the materials in the scene. However, (cf. also Fig. 2), estimates on objects of similar material exhibit a lower relative error. Thus, we limit our detector to skin, and in particular to faces. Pigmentation is the most obvious difference in skin characteristics between different ethnicities. This pigmentation difference depends on many factors as quantity of melanin, amount of UV exposure, genetics, melanosome content and type of pigments found in the skin [24]. However, this intramaterial variation is typically smaller than that of other materials possibly occurring in a scene.

IV. Overview and Algorithmic Details

We classify the illumination for each pair of faces in the image as either consistent or inconsistent. Throughout the paper, we abbreviate illuminant estimation as IE, and illuminant maps as IM. The proposed method consists of five main components:

1) *Dense Local Illuminant Estimation (IE):* The input image is segmented into homogeneous regions. Per illuminant estimator, a new image is created where each region is colored with the extracted illuminant color. This resulting intermediate representation is called illuminant map (IM).
2) *Face Extraction:* This is the only step that may require human interaction. An operator sets a bounding box around

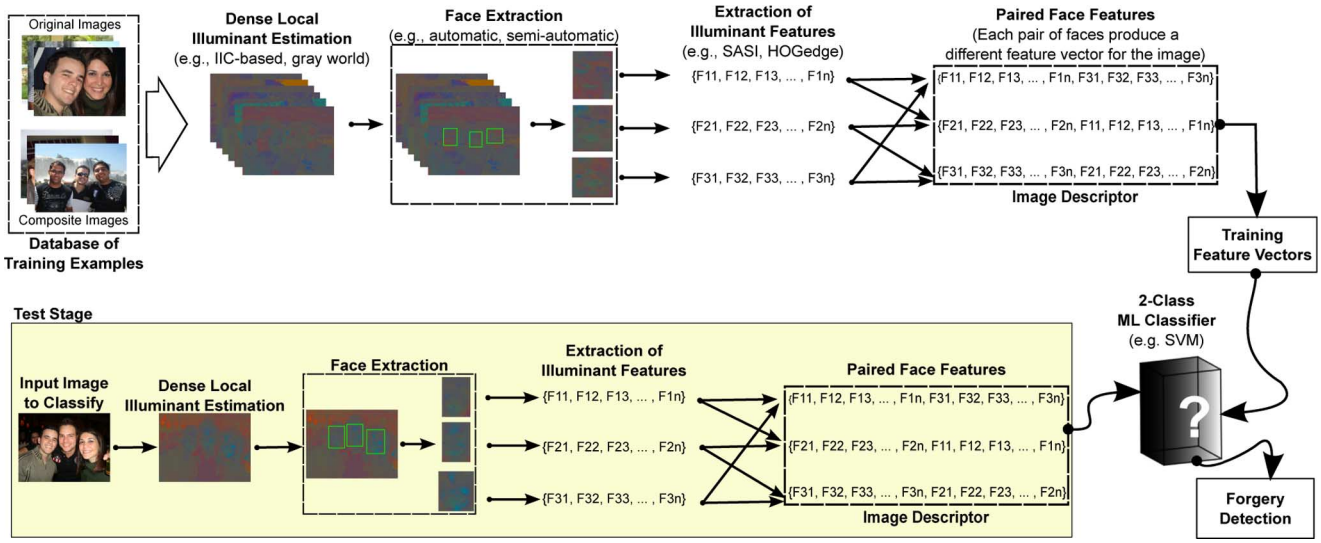


Fig. 4. Overview of the proposed method.

each face (e.g., by clicking on two corners of the bounding box) in the image that should be investigated. Alternatively, an automated face detector can be employed. We then crop every bounding box out of each illuminant map, so that only the illuminant estimates of the face regions remain.

3) *Computation of Illuminant Features:* for all face regions, texture-based and gradient-based features are computed on the IM values. Each one of them encodes complementary information for classification.
4) *Paired Face Features:* Our goal is to assess whether a pair of faces in an image is consistently illuminated. For an image with n_f faces, we construct $\binom{n_f}{2}$ joint feature vectors, consisting of all possible pairs of faces.
5) *Classification:* We use a machine learning approach to automatically classify the feature vectors. We consider an image as a forgery if at least one pair of faces in the image is classified as inconsistently illuminated.

Fig. 4 summarizes these steps. In the remainder of this section, we present the details of these components.

A. Dense Local Illuminant Estimation

To compute a dense set of localized illuminant color estimates, the input image is segmented into superpixels, i.e., regions of approximately constant chromaticity, using the algorithm by Felzenszwalb and Huttenlocher [25]. Per superpixel, the color of the illuminant is estimated. We use two separate illuminant color estimators: the statistical generalized gray world estimates and the physics-based inverse-intensity chromaticity space, as we explain in the next subsection. We obtain, in total, two illuminant maps by recoloring each superpixel with the estimated illuminant chromaticities of each one of the estimators. Both illuminant maps are independently analyzed in the subsequent steps.

1) Generalized Gray World Estimates: The classical gray world assumption by Buchsbaum [26] states that the average color of a scene is gray. Thus, a deviation of the average of the image intensities from the expected gray color is due to the illuminant. Although this assumption is nowadays considered to be overly simplified [17], it has inspired the further design of statistical descriptors for color constancy. We follow an extension of this idea, the generalized gray world approach by van de Weijer *et al.* [23].

Let $\mathbf{f}(\mathbf{x}) = (f_R(\mathbf{x}), f_G(\mathbf{x}), f_B(\mathbf{x}))^{\mathrm{T}}$ denote the observed RGB color of a pixel at location $\mathbf{x}$. Van de Weijer *et al.* s[23] assume purely diffuse reflection and linear camera response. Then, $\mathbf{f}(\mathbf{x})$ is formed by

$$\mathbf{f}(\mathbf{x}) = \int_{\Omega} e(\lambda, \mathbf{x}) s(\lambda, \mathbf{x}) \mathbf{c}(\lambda) d\lambda, \tag{1}$$

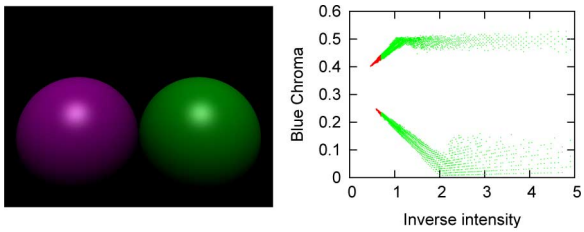
where Ω denotes the spectrum of visible light, λ denotes the wavelength of the light, $e(\lambda, \mathbf{x})$ denotes the spectrum of the illuminant, $s(\lambda, \mathbf{x})$ the surface reflectance of an object, and $\mathbf{c}(\lambda)$ the color sensitivities of the camera (i.e., one function per color channel). Van de Weijer *et al.* [23] extended the original gray world hypothesis through the incorporation of three parameters:

- Derivative order n: the assumption that the average of the illuminants is achromatic can be extended to the absolute value of the sum of the derivatives of the image.
- Minkowski norm p: instead of simply adding intensities or derivatives, respectively, greater robustness can be achieved by computing the p-th Minkowski norm of these values.
- Gaussian smoothing σ: to reduce image noise, one can smooth the image prior to processing with a Gaussian kernel of standard deviation σ.

Putting these three aspects together, van de Weijer *et al.* proposed to estimate the color of the illuminant $\mathbf{e}$ as

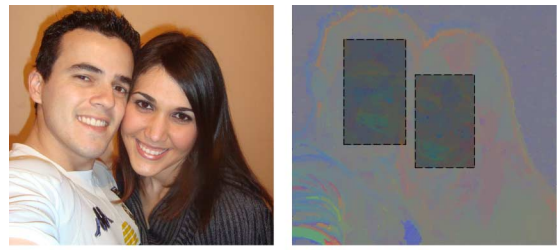
$$k\mathbf{e}^{n,p,\sigma} = \left(\int \left| \frac{\partial^n \mathbf{f}^{\sigma}(\mathbf{x})}{\partial \mathbf{x}^n} \right|^p d\mathbf{x} \right)^{1/p}. \tag{2}$$

Here, the integral is computed over all pixels in the image, where $\mathbf{x}$ denotes a particular position (pixel coordinate). Furthermore, k denotes a scaling factor, $|\cdot|$ the absolute value, ∂ the differential operator, and $\mathbf{f}^{\sigma}(\mathbf{x})$ the observed intensities



(a) (b)

Fig. 5. Illustration of the inverse intensity-chromaticity space (blue color channel). Left: synthetic image (violet and green balls). Right: specular pixels converge towards the blue portion of the illuminant color (recovered at the y-axis intercept). Highly specular pixels are shown in red.



(a) (b)

Fig. 6. Original image and its gray world map. Highlighted regions in the gray world map show a similar appearance. (a) Original. (b) Gray world with highlighted similar parts.

at position $\mathbf{x}$, smoothed with a Gaussian kernel σ. Note that $\mathbf{e}$ can be computed separately for each color channel. Compared to the original gray world algorithm, the derivative operator increases the robustness against homogeneously colored regions of varying sizes. Additionally, the Minkowski norm emphasizes strong derivatives over weaker derivatives, so that specular edges are better exploited [27].

2) Inverse Intensity-Chromaticity Estimates: The second illuminant estimator we consider in this paper is the so-called inverse intensity-chromaticity (IIC) space. It was originally proposed by Tan *et al.* [28]. In contrast to the previous approach, the observed image intensities are assumed to exhibit a mixture of diffuse and specular reflectance. Pure specularities are assumed to consist of only the color of the illuminant. Let (as above) $\mathbf{f}(\mathbf{x}) = (f_R(\mathbf{x}), f_G(\mathbf{x}), f_B(\mathbf{x}))^{\mathrm{T}}$ be a column vector of the observed RGB colors of a pixel. Then, using the same notation as for the generalized gray world model, $\mathbf{f}(\mathbf{x})$ is modelled as

$$\mathbf{f}(\mathbf{x}) = \int_{\Omega} (e(\lambda, \mathbf{x}) s(\lambda, \mathbf{x}) + e(\lambda, \mathbf{x})) \mathbf{c}(\lambda) d\lambda. \tag{3}$$

Let $f_c(\mathbf{x})$ be the intensity and $\chi_c(\mathbf{x})$ be the chromaticity (i.e., normalized RGB-value) of a color channel $c \in \{R, G, B\}$ at position $\mathbf{x}$, respectively. In addition, let γ_c be the chromaticity of the illuminant in channel c. Then, after a somewhat laborious calculation, Tan *et al.* [28] derived a linear relationship between $\mathbf{f}(\mathbf{x})$, $\chi_c(\mathbf{x})$ and γ_c by showing that

$$\chi_c(\mathbf{x}) = m(\mathbf{x}) \frac{1}{\sum_{i \in \{R,G,B\}} f_i(\mathbf{x})} + \gamma_c. \tag{4}$$

Here, $m(\mathbf{x})$ mainly captures geometric influences, i.e., light position, surface orientation and camera position. Although $m(\mathbf{x})$ can not be analytically computed, an approximate solution is feasible. More importantly, the only aspect of interest in illuminant color estimation is the y-intercept γ_c. This can be directly estimated by analyzing the distribution of pixels in IIC space. The IIC space is a per-channel 2-D space, where the horizontal axis is the inverse of the sum of the chromaticities per pixel, $1/\sum_i f_i(\mathbf{x})$, and the vertical axis is the pixel chromaticity for that particular channel. Per color channel c, the pixels within a superpixel are projected onto inverse intensity-chromaticity (IIC) space.

Fig. 5 depicts an exemplary IIC diagram for the blue channel. A synthetic image is rendered (left) and projected onto IIC space (right). Pixels from the green and purple balls form two clusters. The clusters have spikes that point towards the same location at the y-axis. Considering only such spikes from each cluster, the illuminant chromaticity is estimated from the joint y-axis intercept of all spikes in IIC space [28].

In natural images, noise dominates the IIC diagrams. Riess and Angelopoulou [2] proposed to compute these estimates over a large number of small image patches. The final illuminant estimate is computed by a majority vote of these estimates. Prior to the voting, two constraints are imposed on a patch to improve noise resilience. If a patch does not satisfy these constraints, it is excluded from voting.

In practice, these constraints are straightforward to compute. The pixel colors of a patch are projected onto IIC space. Principal component analysis on the distribution of the patch-pixels in IIC space yields two eigenvalues λ_1, λ_2 and their associated eigenvectors $\mathbf{v}_1$ and $\mathbf{v}_2$. Let λ_1 be the larger eigenvalue. Then, $\mathbf{v}_1$ is the principal axis of the pixel distribution in IIC space. In the two-dimensional IIC-space, the principal axis can be interpreted as a line whose slope can be directly computed from $\mathbf{v}_1$. Additionally, λ_1 and λ_2 can be used to compute the eccentricity $\sqrt{1 - \sqrt{\lambda_2}/\sqrt{\lambda_1}}$ as a metric for the shape of the distribution. Both constraints are associated with this eigenanalysis[4]. The first constraint is that the slope must exceed a minimum of 0.003. The second constraint is that the eccentricity has to exceed a minimum of 0.2.

B. Face Extraction

We require bounding boxes around all faces in an image that should be part of the investigation. For obtaining the bounding boxes, we could in principle use an automated algorithm, e.g., the one by Schwartz *et al.* [30]. However, we prefer a human operator for this task for two main reasons: a) this minimizes false detections or missed faces; b) scene context is important when judging the lighting situation. For instance, consider an image where all persons of interest are illuminated by flashlight. The illuminants are expected to agree with one another. Conversely, assume that a person in the foreground is illuminated by flashlight, and a person in the background is illuminated by ambient light. Then, a difference in the color of the illuminants is expected. Such differences are hard to distinguish in a fully-automated manner, but can be easily excluded in manual annotation.

We illustrate this setup in Fig. 6. The faces in Fig. 6(a) can be assumed to be exposed to the same illuminant. As Fig. 6(b) shows, the corresponding gray world illuminant map for these two faces also has similar values.

[4]The parameter values were previously investigated by Riess and Angelopoulou [2], [29]. In this paper, we rely on their findings.

C. Texture Description: SASI Algorithm

We use the Statistical Analysis of Structural Information (SASI) descriptor by Carkacioglu and Yarman-Vural [31] to extract texture information from illuminant maps. Recently, Penatti *et al.* [32] pointed out that SASI performs remarkably well. For our application, the most important advantage of SASI is its capability of capturing small granularities and discontinuities in texture patterns. Distinct illuminant colors interact differently with the underlying surfaces, thus generating distinct illumination "texture". This can be a very fine texture, whose subtleties are best captured by SASI.

SASI is a generic descriptor that measures the structural properties of textures. It is based on the autocorrelation of horizontal, vertical and diagonal pixel lines over an image at different scales. Instead of computing the autocorrelation for every possible shift, only a small number of shifts is considered. One autocorrelation is computed using a specific fixed orientation, scale, and shift. Computing the mean and standard deviation of all such pixel values yields two feature dimensions. Repeating this computation for varying orientations, scales and shifts yields a 128-dimensional feature vector. As a final step, this vector is normalized by subtracting its mean value, and dividing it by its standard deviation. For details, please refer to [31].

D. Interpretation of Illuminant Edges: Hogedge Algorithm

Differing illuminant estimates in neighboring segments can lead to discontinuities in the illuminant map. Dissimilar illuminant estimates can occur for a number of reasons: changing geometry, changing material, noise, retouching or changes in the incident light. Thus, one can interpret an illuminant estimate as a low-level descriptor of the underlying image statistics. We observed that the edges, e.g., computed by a Canny edge detector, detect in several cases a combination of the segment borders and isophotes (i.e., areas of similar incident light in the image). When an image is spliced, the statistics of these edges is likely to differ from original images. To characterize such edge discontinuities, we propose a new feature descriptor called *HOGedge*. It is based on the well-known HOG-descriptor, and computes visual dictionaries of gradient intensities in edge points. The full algorithm is described in the remainder of this section. Fig. 7 shows an algorithmic overview of the method. We first extract approximately equally distributed candidate points on the edges of illuminant maps. At these points, HOG descriptors are computed. These descriptors are summarized in a visual words dictionary. Each of these steps is presented in greater detail in the next subsections.

Extraction of Edge Points: Given a face region from an illuminant map, we first extract edge points using the Canny edge detector [33]. This yields a large number of spatially close edge points. To reduce the number of points, we filter the Canny output using the following rule: starting from a seed point, we eliminate all other edge pixels in a region of interest (ROI) centered around the seed point. The edge points that are closest to the ROI (but outside of it) are chosen as seed points for the next iteration. By iterating this process over the entire image, we reduce the number of points but still ensure that every face has a comparable density of points. Fig. 8 depicts an example of the resulting points.

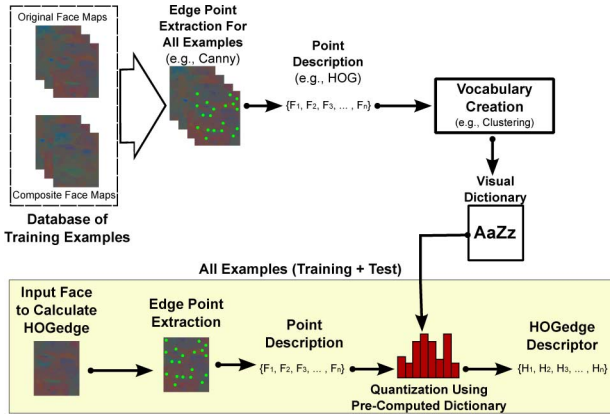


Fig. 7. Overview of the proposed H OGedge algorithm.

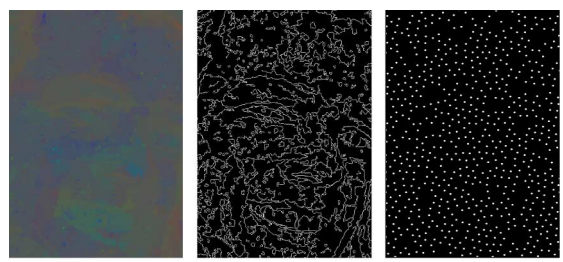


Fig. 8. (a) Gray world IM for the left face in Fig. 6(a). (b) Result of the Canny edge detector when applied on this IM. (c) Final edge points after filtering using a square region. (a) IM derived from gray world. (b) Canny Edges. (c) Filtered Points.

Point Description: We compute Histograms of Oriented Gradients (HOG) [34] to describe the distribution of the selected edge points. HOG is based on normalized local histograms of image gradient orientations in a dense grid. The HOG descriptor is constructed around each of the edge points. The neighborhood of such an edge point is called a cell. Each cell provides a local 1-D histogram of quantized gradient directions using all cell pixels. To construct the feature vector, the histograms of all cells within a spatially larger region are combined and contrast-normalized. We use the HOG output as a feature vector for the subsequent steps.

Visual Vocabulary: The number of extracted HOG vectors varies depending on the size and structure of the face under examination. We use visual dictionaries [35] to obtain feature vectors of fixed length. Visual dictionaries constitute a robust representation, where each face is treated as a set of region descriptors. The spatial location of each region is discarded [36].

To construct our visual dictionary, we subdivide the training data into feature vectors from original and doctored images. Each group is clustered in n clusters using the k-means algorithm [37]. Then, a visual dictionary with $2n$ visual words is constructed, where each word is represented by a cluster center. Thus, the visual dictionary summarizes the most representative feature vectors of the training set. Algorithm 1 shows the pseudocode for the dictionary creation.

Algorithm 1 HOGedge—Visual dictionary creation

Require: V_{TR} (training database examples)n(the number of visual words per class)
Ensure: V_D (visual dictionary containing $2n$ visual words)
  $V_D \leftarrow \emptyset$;
  $V_{NF} \leftarrow \emptyset$;
  $V_{DF} \leftarrow \emptyset$;
  for each face IM $i \in V_{TR}$ **do**
    $V_{EP} \leftarrow$ edge points extracted from i ;
    for each point $j \in V_{EP}$ **do**
      $FV \leftarrow$ apply HOG in image i at position j;
      if i is a doctored face **then**
        $V_{DF} \leftarrow \{V_{DF} \cup FV\}$;
      else
        $V_{NF} \leftarrow \{V_{NF} \cup FV\}$;
    end if
  end for
end for
Cluster V_{DF} using n centers;
Cluster V_{NF} using n centers; $V_D \leftarrow \{$centers of $V_{DF} \cup$ centers of $V_{NF}\}$;
return V_D;

Quantization Using the Precomputed Visual Dictionary: For evaluation, the HOG feature vectors are mapped to the visual dictionary. Each feature vector in an image is represented by the closest word in the dictionary (with respect to the Euclidean distance). A histogram of word counts represents the distribution of HOG feature vectors in a face. Algorithm 2 shows the pseudocode for the application of the visual dictionary on IMs.

Algorithm 2 HOGedge—Face characterization

Require: V_D (visual dictionary precomputed with $2n$ visual words) IM (illuminant map from a face)
Ensure: HFV (HOGedge feature vector)
  $HFV \leftarrow 2n$-dimensional vector, initialized to all zeros;
  $V_{FV} \leftarrow \emptyset$;
  $V_{EP} \leftarrow$ edge points extracted from IM;
  for each point $i \in V_{EP}$ **do**
    $FV \leftarrow$ apply HOG in image IM at position j;
    $V_{FV} \leftarrow \{V_{FV} \cup FV\}$;
  end for
  for each feature vector $i \in V_{FV}$ **do**
    $lower\_distance \leftarrow +\infty$;
    $position \leftarrow -1$;
    for each visual word $j \in V_D$ **do**
      $distance \leftarrow$ Euclidean distance between i and j;
      if $distance < lower\_distance$ **then**
        $lower\_distance \leftarrow distance$;
        $position \leftarrow$ position of j in V_D;
      end if
    end for
    $HFV\,[position] \leftarrow HFV\,[position] + 1$;
  end for
  return HFV;

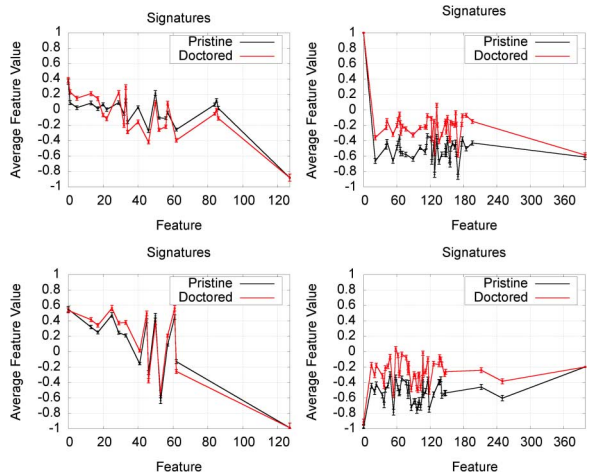


Fig. 9. Average signatures from original and spliced images. The horizontal axis corresponds to different feature dimensions, while the vertical axis represents the average feature value for different combinations of descriptors and illuminant maps. From top to bottom, left to right: SASI-IIC, HOGedge-IIC, SASI-Gray-World, HOGedge-Gray-World.

E. Face Pair

To compare two faces, we combine the same descriptors for each of the two faces. For instance, we can concatenate the SASI-descriptors that were computed on gray world. The idea is that a feature concatenation from two faces is different when one of the faces is an original and one is spliced. For an image containing n_f faces ($n_f \geq 2$), the number of face pairs is $(n_f(n_f - 1))/2$.

The SASI and HOGedge descriptors capture two different properties of the face regions. From a signal processing point of view, both descriptors are *signatures* with different behavior. Fig. 9 shows a very high-level visualization of the distinct information that is captured by these two descriptors. For one of the folds of our experiments (see Section V-C), we computed the mean value and standard deviation per feature dimension. For a less cluttered plot, we only visualize the feature dimensions with the largest difference in the mean values for this fold. This experiment empirically demonstrates two points. Firstly, SASI and HOGedge, in combination with the IIC-based and gray world illuminant maps create features that discriminate well between original and tampered images, in at least some dimensions. Secondly, the dimensions, where these features have distinct value, vary between the four combinations of the feature vectors. We exploit this property during classification by fusing the output of the classification on both feature sets, as described in the next section.

F. Classification

We classify the illumination for each pair of faces in an image as either consistent or inconsistent. Assuming all selected faces are illuminated by the same light source, we tag an image as manipulated if one pair is classified as inconsistent. Individual feature vectors, i.e., SASI or HOGedge features on either gray world or IIC-based illuminant maps, are classified using a support vector machine (SVM) classifier with a radial basis function (RBF) kernel.

The information provided by the SASI features is complementary to the information from the HOGedge features. Thus, we use a machine learning-based fusion technique for improving the detection performance. Inspired by the work of Ludwig *et al.* [38], we use a late fusion technique named SVM-Meta Fusion. We classify each combination of illuminant map and feature type independently (i.e., SASI-Gray-World, SASI-IIC, HOGedge-Gray-World and HOGedge-IIC) using a two-class SVM classifier to obtain the distance between the image's feature vectors and the classifier decision boundary. SVM-Meta Fusion then merges the marginal distances provided by all m individual classifiers to build a new feature vector. Another SVM classifier (i.e., on meta level) classifies the combined feature vector.

V. Experiments

To validate our approach, we performed six rounds of experiments using two different databases of images involving people. We show results using classical ROC curves where *sensitivity* represents the number of composite images correctly classified and *specificity* represents the number of original images (non-manipulated) correctly classified.

A. Evaluation Data

To quantitatively evaluate the proposed algorithm, and to compare it to related work, we considered two datasets. One consists of images that we captured ourselves, while the second one contains images collected from the internet. Additionally, we validated the quality of the forgeries using a human study on the first dataset. Human performance can be seen as a baseline for our experiments.

1) DSO-1: This is our first dataset and it was created by ourselves. It is composed of 200 indoor and outdoor images with an image resolution of $2{,}048 \times 1{,}536$ pixels. Out of this set of images, 100 are original, i.e., have no adjustments whatsoever, and 100 are forged. The forgeries were created by adding one or more individuals in a source image that already contained one or more persons. When necessary, we complemented an image splicing operation with postprocessing operations (such as color and brightness adjustments) in order to increase photorealism.

2) DSI-1: This is our second dataset and it is composed of 50 images (25 original and 25 doctored) downloaded from different websites in the Internet with different resolutions[5]. Fig. 10 depicts some example images from our databases.

B. Human Performance in Spliced Image Detection

To demonstrate the quality of DSO-1 and the difficulty in discriminating original and tampered images, we performed an experiment where we asked humans to mark images as tampered or original. To accomplish this task, we have used Amazon Mechanical Turk[6]. Note that on Mechanical Turk categorization experiments, each batch is evaluated only by experienced users which generally leads to a higher confidence in the outcome of the task. In our experiment, we setup five identical categorization experiments, where each one of them is called a batch. Within a batch, all DSO-1 images have been evaluated. For each image, two users were asked to tag the image as original or manipulated. Each image was assessed by ten different users, each user expended on average 47 seconds to tag an image. The final accuracy, averaged over all experiments, was 64.6%. However, for spliced images, the users achieved only an average accuracy of 38.3%, while human accuracy on the original images was 90.9%. The kappa-value, which measures the degree of agreement between an arbitrary number of raters in deciding the class of a sample, based on the Fleiss [39] model, is 0.11. Despite being subjective, this kappa-value, according to the Landis and Koch [40] scale, suggests a slight degree of agreement between users, which further supports our conjecture about the difficulty of forgery detection in DSO-1 images.

Fig. 10. Original (left) and spliced images (right) from both databases. (a) DSO-1 Original image. (b) DSO-1 Spliced image. (c) DSI-1 Original image. (d) DSI-1 Spliced image.

C. Performance of Forgery Detection Using Semiautomatic Face Annotation in DSO-1

We compare five variants of the method proposed in this paper. Throughout this section, we manually annotated the faces using corner clicking (see Section V-D). In the classification stage, we use a five-fold cross validation protocol, an SVM classifier with an RBF kernel, and classical grid search for adjusting parameters in training samples [37]. Due to the different number of faces per image, the number of feature vectors for the original and the spliced images is not exactly equal. To address this issue during training, we weighted feature vectors from original and composite images. Let w_o and w_c denote the number of feature vectors from original and composite images, respectively. To obtain a proportional class weighting, we set the weight of features from original images to $w_c/(w_o + w_c)$ and the weight of features from composite images to $w_o/(w_o + w_c)$.

We compared the five variants SASI-IIC, SASI-Gray-World, HOGedge-IIC, HOGedge-Gray-World and Metafusion. Compound names, such as SASI-IIC, indicate the data source (in this

[5]Original images were downloaded from Flickr (http://www.flickr.com) and doctored images were collected from different websites such as Worth 1000 (http://www.worth1000.com/), Benetton Group 2011 (http://press.benetton-group.com/), Planet Hiltron (http://www.facebook.com/pages/Planet-Hiltron/150175044998030), etc.

[6]https://www.mturk.com/mturk/welcome

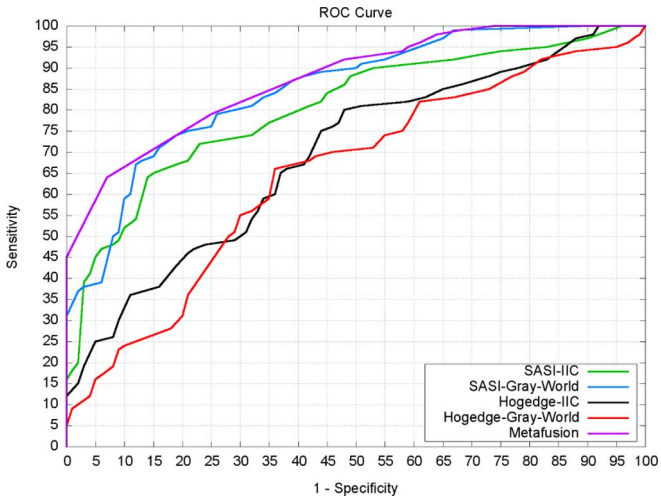


Fig. 11. Comparison of different variants of the algorithm using semiautomatic (corner clicking) annotated faces.

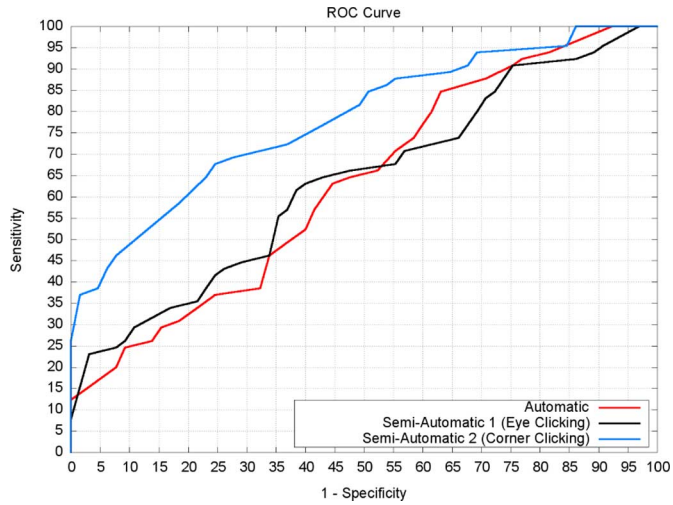


Fig. 12. Experiments showing the differences for automatic and semiautomatic face detection.

case: IIC-based illuminant maps) and the subsequent feature extraction method (in this case: SASI). The single components are configured as follows:

- **IIC:** IIC-based illuminant maps are computed as described in [2].
- **Gray-World:** Gray world illuminant maps are computed by setting $n = 1$, $p = 1$, and $\sigma = 3$ in (2).
- **SASI:** The SASI descriptor is calculated over the Y channel from the YC_bC_r color space. All remaining parameters are chosen as presented in [32][7].
- **HOGedge:** Edge detection is performed on the Y channel of the YC_bC_r color space, with a Canny low threshold of 0 and a high threshold of 10. The square region for edge point filtering was set to 32×32 pixels. Furthermore, we used 8-pixel cells without normalization in HOG. If applied on IIC-based illuminant maps, we computed 100 visual words for both the original and the tampered images (i.e., the dictionary consisted of 200 visual words). On gray world illuminant maps, the size of the visual word dictionary was set to 75 for each class, leading to a dictionary of 150 visual words.
- **Metafusion:** We implemented a late fusion as explained in Section IV-F. As input, it uses SASI-IIC, SASI-Gray-World, and HOGedge-IIC. We excluded HOGedge-Gray-World from the input methods, as its weaker performance leads to a slightly worse combined classification rate (see below).

Fig. 11 depicts a ROC curve of the performance of each method using the corner clicking face localization. The area under the curve (AUC) is computed to obtain a single numerical measure for each result.

From the evaluated variants, Metafusion performs best, resulting in an AUC of 86.3%. In particular for high specificity (i.e., few false alarms), the method has a much higher sensitivity compared to the other variants. Thus, when the detection threshold is set to a high specificity, and a photograph is classified as composite, Metafusion provides to an expert high confidence that the image is indeed manipulated.

[7]We gratefully thank the authors for the source code.

Note also that Metafusion clearly outperforms human assessment in the baseline Mechanical Turk experiment (see Section V-B). Part of this improvement comes from the fact that Metafusion achieves, on spliced images alone, an average accuracy of 67%, while human performance was only 38.3%.

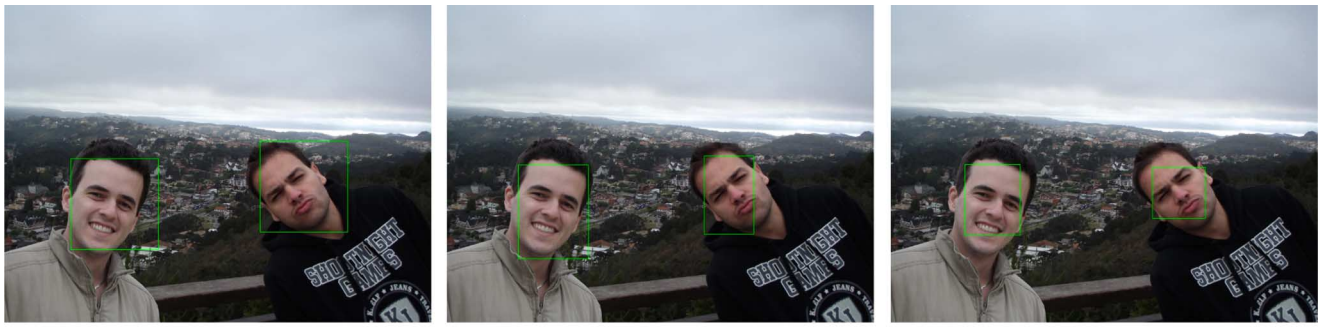
The second best variant is SASI-Gray-World, with an AUC of 84.0%. In particular for a specificity below 80.0%, the sensitivity is comparable to Metafusion. SASI-IIC achieved an AUC of 79.4%, followed by HOGedge-IIC with an AUC of 69.9% and HOGedge-Gray-World with an AUC of 64.7%. The weak performance of HOGedge-Gray-World comes from the fact that illuminant color estimates from the gray world algorithm vary more smoothly than IIC-based estimates. Thus, the differences in the illuminant map gradients (as extracted by the HOGedge descriptor) are generally smaller.

D. Fully Automated Versus Semiautomatic Face Detection

In order to test the impact of automated face detection, we reevaluated the best performing variant, Metafusion, on three versions of automation in face detection and annotation.

- **Automatic Detection:**we used the PLS-based face detector [30] to detect faces in the images. In our experiments, the PLS detector successfully located all present faces in only 65% of our images. We then performed a 3-fold cross validation on this 65% of the images. For training the classifier, we used the manually annotated bounding boxes. In the test images, we used the bounding boxes found by the automated detector.
- **Semiautomatic Detection 1 (Eye Clicking):** an expert does not necessarily have to mark a bounding box. In this variant, the expert clicks on the eye positions. The Euclidean distance between the eyes is the used to construct a bounding box for the face area. For classifier training and testing we use the same setup and images as in the automatic detection.
- **Semiautomatic Detection 2 (Corner Clicking):** in this variant, we applied the same marking procedure as in the previous experiment and the same classifier training/testing procedure as in automatic detection.

Fig. 12 shows the results of this experiment. The semiautomatic detection using corner clicking resulted in an AUC



(a) (b) (c)

Fig. 13. Different types of face location. Automatic and semiautomatic locations select a considerable part of the background, whereas manual location is restricted to face regions. (a) Automatic. (b) Semiautomatic (eye clicking). (c) Semiautomatic (corner clicking).

of 78.0%, while the semiautomatic using eye clicking and the fully-automatic approaches yielded an AUC of 63.5% and AUC of 63.0%, respectively. Thus, as it can also be seen in Figs. 13(a)–13(c), proper face location is important for improved performance.

Although automatic face detection algorithms have improved over the years, we find user-selected faces more reliable for a forensic setup mainly because automatic face detection algorithms are not accurate in bounding box detection (location and size). In our experiments, automatic and eye clicking detection have generated an average bounding box size which was 38.4% and 24.7% larger than corner clicking detection, respectively. Thus, such bounding boxes include part of the background in a region that should contain just face information. The precision of bounding box location in automatic detection and eye clicking has also been worse than semiautomatic using corner clicking. Note, however, that the selection of faces under similar illumination conditions is a minor interaction that requires no particular knowledge in image processing or image forensics.

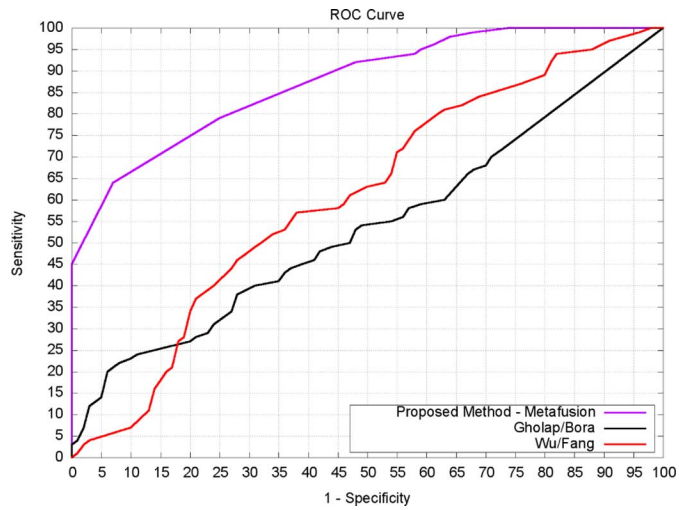


Fig. 14. Comparative results between our method and state-of-the-art approaches performed using DSO-1.

E. Comparison With State-of-the-art Methods

For experimental comparison, we implemented the methods by Gholap and Bora [12] and Wu and Fang [13]. Note that neither of these works includes a quantitative performance analysis. Thus, to our knowledge, this is the first direct comparison of illuminant color-based forensic algorithms.

For the algorithm by Gholap and Bora [12], three partially specular regions per image were manually annotated. For manipulated images, it is guaranteed that at least one of the regions belongs to the tampered part of the image, and one region to the original part. Fully saturated pixels were excluded from the computation, as they have presumably been clipped by the camera. Camera gamma was approximately inverted by assuming a value of 2.2. The maximum distance of the dichromatic lines per image were computed. The threshold for discriminating original and tampered images was set via five-fold cross-validation, yielding a detection rate of 55.5% on DSO-1.

In the implementation of the method by Wu and Fang, the Weibull distribution is computed in order to perform image classification prior to illuminant estimation. The training of the image classifier was performed on the ground truth dataset by Ciurea and Funt [41] as proposed in the work [13]. As the resolution of this dataset is relatively low, we performed the training on a central part of the images containing 180×240 pixels (excluding the ground-truth area). To provide images of the same resolution for illuminant classification, we manually annotated the face regions in DSO-1 with bounding boxes of fixed size ratio. Setting this ratio to 3:4, each face was then rescaled to a size of 180×240 pixels. As the selection of suitable reference regions is not well-defined (and also highly image-dependent), we directly compare the illuminant estimates of the faces in the scene. Here, the best result was obtained with three-fold cross-validation, yielding a detection rate of 57%. We performed five-fold cross-validation, as in the previous experiments. The results drop to 53% detection rate, which suggests that this algorithm is not very stable with respect to the selection of the data.

To reduce any bias that could be introduced from training on the dataset by Ciurea and Funt, we repeated the image classifier training on the reprocessed ground truth dataset by Gehler [42]. During training, care was taken to exclude the ground truth information from the data. Repeating the remaining classification yielded a best result of 54.5% on two-fold cross-validation, or 53.5% for five-fold cross-validation.

Fig. 14 shows the ROC curves for both methods. The results of our method clearly outperform the state-of-the-art. However,

these results also underline the challenge in exploiting illuminant color as a forensic cue on real-world images. Thus, we hope our database will have a significant impact in the development of new illuminant-based forgery detection algorithms.

F. Detection After Additional Image Processing

We also evaluated the robustness of our method against different processing operations. The results are computed on DSO-1. Apart from the additional preprocessing steps, the evaluation protocol was identical to the one described in Section V-C. In a first experiment, we examined the impact of JPEG compression. Using libJPEG, the images were recompressed at the JPEG quality levels 70, 80 and 90. The detection rates were 63.5%, 64% and 69%, respectively. Using imagemagick, we conducted a second experiment adding per image a random amount of Gaussian noise, with an attenuated value varying between 1% and 5%. On average, we obtained an accuracy of 59%. Finally, again using imagemagick, we randomly varied the brightness and/or contrast of the image by either $+5\%$ or -5% . These brightness/contrast manipulations resulted in an accuracy of 61.5%.

These results are expected. For instance, the performance deterioration after strong JPEG compression introduces blocking artifacts in the segmentation of the illuminant maps. One could consider compensating for the JPEG artifacts with a deblocking algorithm. Still, JPEG compression is known to be a challenging scenario in several classes of forensic algorithms [43]–[45]

One could also consider optimizing the machine-learning part of the algorithm. However, here, we did not fine-tune the algorithm for such operations, as postprocessing can be addressed by specialized detectors, such as the work by Bayram *et al.* for brightness and contrast changes [46], combined with one of the recent JPEG-specific algorithms (e.g., [47]).

G. Performance of Forgery Detection Using a Cross-Database Approach

To evaluate the generalization of the algorithm with respect to the training data, we followed an experimental design similar to the one proposed by Rocha *et al.* [48]. We performed a cross-database experiment, using DSO-1 as training set and the 50 images of DSI-1 (internet images) as test set. We used the pretrained Metafusion classifier from the best performing fold in Section V-C without further modification. Fig. 15 shows the ROC curve for this experiment. The results of this experiment are similar to the best ROC curve in Section V-C, with an AUC of 82.6%. This indicates that the proposed method offers a degree of generalization to images from different sources and to faces of varying sizes.

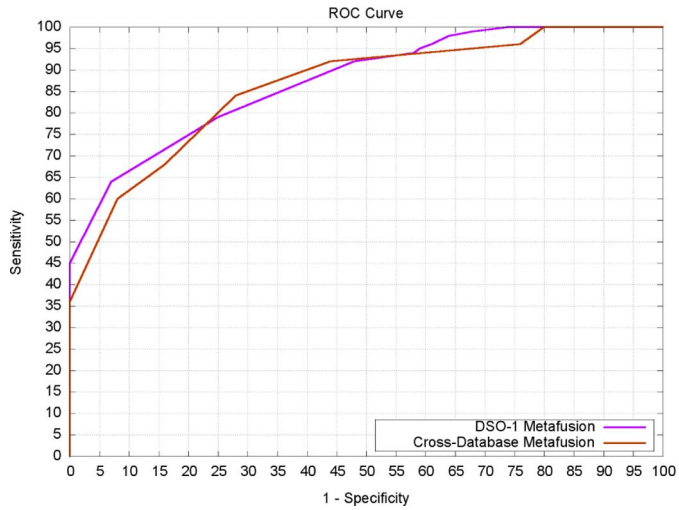


Fig. 15. ROC curve provided by cross-database experiment.

VI. Conclusions and Future Work

In this work, we presented a new method for detecting forged images of people using the illuminant color. We estimate the illuminant color using a statistical gray edge method and a physics-based method which exploits the inverse intensity-chromaticity color space. We treat these illuminant maps as texture maps. We also extract information on the distribution of edges on these maps. In order to describe the edge information, we propose a new algorithm based on edge-points and the HOG descriptor, called HOGedge. We combine these complementary cues (texture- and edge-baed) using machine learning late fusion. Our results are encouraging, yielding an AUC of over 86% correct classification. Good results are also achieved over internet images and under cross-database training/testing.

Although the proposed method is custom-tailored to detect splicing on images containing faces, there is no principal hindrance in applying it to other, problem-specific materials in the scene.

The proposed method requires only a minimum amount of human interaction and provides a crisp statement on the authenticity of the image. Additionally, it is a significant advancement in the exploitation of illuminant color as a forensic cue. Prior color-based work either assumes complex user interaction or imposes very limiting assumptions.

Although promising as forensic evidence, methods that operate on illuminant color are inherently prone to estimation errors. Thus, we expect that further improvements can be achieved when more advanced illuminant color estimators become available. For instance, while we were developing this work, Bianco and Schettini [49] proposed a machine-learning based illuminant estimator particularly for faces. An incorporation of this method is subject of future work.

Reasonably effective skin detection methods have been presented in the computer vision literature in the past years. Incorporating such techniques can further expand the applicability of our method. Such an improvement could be employed, for instance, in detecting pornography compositions which, according to forensic practitioners, have become increasingly common nowadays.

References

[1] A. Rocha, W. Scheirer, T. E. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Comput. Surveys*, vol. 43, pp. 1–42, 2011.

[2] C. Riess and E. Angelopoulou, "Scene illumination as an indicator of image manipulation," *Inf. Hiding*, vol. 6387, pp. 66–80, 2010.

[3] H. Farid and M. J. Bravo, "Image forensic analyses that elude the human visual system," in *Proc. Symp. Electron. Imaging (SPIE)*, 2010, pp. 1–10.

[4] Y. Ostrovsky, P. Cavanagh, and P. Sinha, "Perceiving illumination inconsistencies in scenes," *Perception*, vol. 34, no. 11, pp. 1301–1314, 2005.
[5] H. Farid, A 3-D lighting and shadow analysis of the JFK Zapruder film (Frame 317), Dartmouth College, Tech. Rep. TR2010–677, 2010.
[6] M. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proc. ACM Workshop on Multimedia and Security*, New York, NY, USA, 2005, pp. 1–10.
[7] M. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," *IEEE Trans. Inf. Forensics Security*, vol. 3, no. 2, pp. 450–461, Jun. 2007.
[8] M. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," in *Proc. Int. Workshop on Inform. Hiding*, 2007, pp. 311–325.
[9] E. Kee and H. Farid, "Exposing digital forgeries from 3-D lighting environments," in *Proc. IEEE Int. Workshop on Inform. Forensics and Security (WIFS)*, Dec. 2010, pp. 1–6.
[10] W. Fan, K. Wang, F. Cayre, and Z. Xiong, "3D lighting-based image forgery detection using shape-from-shading," in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*, Aug. 2012, pp. 1777–1781.
[11] J. F. O'Brien and H. Farid, "Exposing photo manipulation with inconsistent reflections," *ACM Trans. Graphics*, vol. 31, no. 1, pp. 1–11, Jan. 2012.
[12] S. Gholap and P. K. Bora, "Illuminant colour based image forensics," in *Proc. IEEE Region 10 Conf.*, 2008, pp. 1–5.
[13] X. Wu and Z. Fang, "Image splicing detection using illuminant color inconsistency," in *Proc. IEEE Int. Conf. Multimedia Inform. Networking and Security*, Nov. 2011, pp. 600–603.
[14] P. Saboia, T. Carvalho, and A. Rocha, "Eye specular highlights telltales for digital forensics: A machine learning approach," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2011, pp. 1937–1940.
[15] C. Riess and E. Angelopoulou, "Physics-based illuminant color estimation as an image semantics clue," in *Proc. IEEE Int. Conf. Image Processing*, Nov. 2009, pp. 689–692.
[16] K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms–Part I: Methodology and Experiments With Synthesized Data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 972–983, Sep. 2002.
[17] K. Barnard, L. Martin, A. Coath, and B. Funt, "A comparison of computational color constancy algorithms – Part II: Experiments With Image Data," *IEEE Trans. Image Process.*, vol. 11, no. 9, pp. 985–996, Sep. 2002.
[18] A. Gijsenij, T. Gevers, and J. van de Weijer, "Computational color constancy: Survey and experiments," *IEEE Trans. Image Process.*, vol. 20, no. 9, pp. 2475–2489, Sep. 2011.
[19] M. Bleier, C. Riess, S. Beigpour, E. Eibenberger, E. Angelopoulou, T. Tröger, and A. Kaup, "Color constancy and non-uniform illumination: Can existing algorithms work?," in *Proc. IEEE Color and Photometry in Comput. Vision Workshop*, 2011, pp. 774–781.
[20] M. Ebner, "Color constancy using local color shifts," in *Proc. Eur. Conf. Comput. Vision*, 2004, pp. 276–287.
[21] A. Gijsenij, R. Lu, and T. Gevers, "Color constancy for multiple light sources," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 697–707, Feb. 2012.
[22] R. Kawakami, K. Ikeuchi, and R. T. Tan, "Consistent surface color for texturing large objects in outdoor scenes," in *Proc. IEEE Int. Conf. Comput. Vision*, 2005, pp. 1200–1207.
[23] J. van de Weijer, T. Gevers, and A. Gijsenij, "Edge-based color constancy," *IEEE Trans. Image Process.*, vol. 16, no. 9, pp. 2207–2214, Sep. 2007.
[24] T. Igarashi, K. Nishino, and S. K. Nayar, "The appearance of human skin: A survey," *Found. Trends Comput. Graph. Vis.*, vol. 3, no. 1, pp. 1–95, 2007.
[25] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, 2004.
[26] G. Buchsbaum, "A spatial processor model for color perception," *J. Franklin Inst.*, vol. 310, no. 1, pp. 1–26, Jul. 1980.
[27] A. Gijsenij, T. Gevers, and J. van de Weijer, "Improving color constancy by photometric edge weighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 5, pp. 918–929, May 2012.
[28] R. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Amer. A*, vol. 21, pp. 321–334, 2004.
[29] C. Riess, E. Eibenberger, and E. Angelopoulou, "Illuminant color estimation for real-world mixed-illuminant scenes," in *Proc. IEEE Color and Photometry in Comput. Vision Workshop*, Barcelona, Spain, Nov. 2011.
[30] W. R. Schwartz, A. Kembhavi, D. Harwood, and L. S. Davis, "Human detection using partial least squares analysis," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2009, pp. 24–31.
[31] A. Carkacioglu and F. T. Yarman-Vural, "Sasi: A generic texture descriptor for image retrieval," *Pattern Recognit.*, vol. 36, no. 11, pp. 2615–2633, 2003.
[32] O. A. B. Penatti, E. Valle, and R. S. Torres, "Comparative study of global color and texture descriptors for web image retrieval," *J. Visual Commun. Image Representat.*, vol. 23, no. 2, pp. 359–380, 2012.
[33] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Jun. 1986.
[34] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, 2005, pp. 886–893.
[35] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Proc. Workshop on Statistical Learning in Comput. Vision*, 2004, pp. 1–8.
[36] J. Winn, A. Criminisi, and T. Minka, "Object categorization by learned universal visual dictionary," in *Proc. IEEE Int. Conf. Comput. Vision (ICCV)*, 2005, pp. 1800–1807.
[37] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc, 2006.
[38] O. Ludwig, D. Delgado, V. Goncalves, and U. Nunes, "Trainable classifier-fusion schemes: An application to pedestrian detection," in *Proc. IEEE Int. Conf. Intell. Transportation Syst.*, 2009, pp. 1–6.
[39] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971.
[40] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
[41] F. Ciurea and B. Funt, "A large image database for color constancy research," in *Proc. IS&T/SID Eleventh Color Imaging Conf.: Color Sci. and Eng. Syst., Technologies, Applicat. (CIC 2003)*, Scottsdale, AZ, USA, Nov. 2003, pp. 160–164.
[42] L. Shi and B. Funt, Re-processed Version of the Gehler Color Constancy Dataset of 568 Images, Jan. 2011 [Online]. Available: http://www.cs.sfu.ca/colour/data/shi_gehler/
[43] A. C. Popescu and H. Farid, "Statistical tools for digital forensics," in *Proc. Inf. Hiding Conf.*, Jun. 2005, pp. 395–407.
[44] M. Kirchner, "Linear row and column predictors for the analysis of resized images," in *Proc. ACM SIGMM Multimedia Security Workshop*, Sep. 2010, pp. 13–18.
[45] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Trans. Inf. Forensics Security*, vol. 1, no. 2, pp. 205–214, Jun. 2006.
[46] S. Bayram, I. Avcibas, B. Sankur, and N. Memon, "Image manipulation detection with binary similarity measures," in *Proc. Eur. Signal Processing Conf. (EUSIPCO)*, 2005, vol. I, pp. 752–755.
[47] T. Bianchi and A. Piva, "Detection of non-aligned double JPEG compression based on integer periodicity maps," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 2, pp. 842–848, Apr. 2012.
[48] A. Rocha, T. Carvalho, H. Jelinek, S. K. Goldenstein, and J. Wainer, "Points of interest and visual dictionaries for automatic retinal lesion detection," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2244–2253, Aug. 2012.
[49] S. Bianco and R. Schettini, "Color constancy using faces," in *Proc. IEEE Comput. Vision and Pattern Recognition*, Providence, RI, USA, Jun. 2012.

Tiago José de Carvalho (S'12) received the B.Sc. degree (computer science) from Federal University of Juiz de Fora (UFJF), Brazil, in 2008. He received the M.Sc. degree (computer science) from University of Campinas (Unicamp), Brazil, in 2010. Currently, he is working toward the Ph.D. degree at the Institute of Computing, Unicamp, Brazil.

His main interests include digital forensics, pattern analysis, data mining, machine learning, computer vision, and image processing.

Christian Riess (S'10–A'12) received the Diploma degree in computer science in 2007 and the doctoral degree in 2013, both from the University of Erlangen-Nuremberg, Germany.

From 2007 to 2010, he was working on an industry project with Giesecke+Devrient on optical inspection. He is currently doing his postdoc at the Radiological Sciences Laboratory at Stanford University, Stanford, CA, USA. His research interests include all aspects of image processing, in particular with applications in image forensics, medical imaging, optical inspection, and computer vision.

Elli Angelopoulou (S'89–M'90) received the Ph.D. degree in computer science from the Johns Hopkins University in 1997.

She did her postdoc at the General Robotics, Automation, Sensing and Perception (GRASP) Laboratory at the University of Pennsylvania. She then became an assistant professor at Stevens Institute of Technology. She is currently an associate research professor at the University of Erlangen-Nuremberg. Her research focuses on multispectral imaging, skin reflectance, reflectance analysis in support of shape recovery, image forensics, image retrieval, and reflectance analysis in medical imaging (e.g., capsule endoscopy).

Dr. Angelopoulou has over 50 publications, multiple patents, and has received numerous grants, including an NSF CAREER award. She has served on the program committees of ICCV, CVPR, and ECCV and is an associate editor of *Machine Vision and Applications* (MVA) and the *Journal of Intelligent Service Robotics* (JISR).

Hélio Pedrini (S'99–M'00) received the Ph.D. degree in electrical and computer engineering from Rensselaer Polytechnic Institute, Troy, NY, USA. He received the M.Sc. degree in electrical engineering and the B.Sc. degree in computer science, both from the University of Campinas, Brazil.

He is currently a professor with the Institute of Computing at the University of Campinas, Brazil. His research interests include image processing, computer vision, pattern recognition, computer graphics, and computational geometry.

Anderson de Rezende Rocha (S'05–M'10) received the CS B.Sc. degree from Federal University of Lavras (UFLA), Brazil, in 2003. He received the Computer Science M.S. and Ph.D. degrees from University of Campinas (Unicamp), Brazil, in 2006 and 2009, respectively.

Currently, he is an assistant professor in the Institute of Computing, Unicamp. As of 2011, Prof. Rocha is a Microsoft Research Faculty Fellow and an elected member of the Brazilian Academy of Sciences. He is also an elected member of the IEEE Information Forensics and Security Technical Committee (IFS-TC). His interests include digital image and video forensics, machine intelligence, and general computer vision.