Illuminant-Based Transformed Spaces for Image Forensics

Tiago Carvalho, Fábio A. Faria, Hélio Pedrini, *Senior Member, IEEE*, Ricardo da S. Torres, *Member, IEEE*, and Anderson Rocha, *Senior Member, IEEE*

***Abstract*— In this paper, we explore transformed spaces, represented by image illuminant maps, to propose a methodology for selecting complementary forms of characterizing visual properties for an effective and automated detection of image forgeries. We combine statistical telltales provided by different image descriptors that explore color, shape, and texture features. We focus on detecting image forgeries containing people and present a method for locating the forgery, specifically, the face of a person in an image. Experiments performed on three different open-access data sets show the potential of the proposed method for pinpointing image forgeries containing people. In the two first data sets (DSO-1 and DSI-1), the proposed method achieved a classification accuracy of 94% and 84%, respectively, a remarkable improvement when compared with the state-of-the-art methods. Finally, when evaluating the third data set comprising questioned images downloaded from the Internet, we also present a detailed analysis of target images.**

***Index Terms*— Digital forensics, splicing detection, illuminant maps, image descriptors, machine learning, diversity measures.**

I. Introduction

IN A SOCIETY in which social networks became powerful communication tools and are more ubiquitous than ever, it is now paramount to design and deploy methods that guarantee the authenticity of the broadcast information. Images, for instance, considered one of the most powerful communication media, appear as the most shared documents at these social networks, mainly because current mobile devices allow anyone to capture thousands of images anywhere at anytime. In this context, the development of methods for verifying image authenticity is a real need of the modern society.

Manuscript received July 30, 2015; revised November 6, 2015; accepted November 9, 2015. Date of publication December 18, 2015; date of current version February 1, 2016. This work was supported in part by the Coordination for the Improvement of Higher Education Personnel under Grant 0214-13-2, in part by Microsoft Research, in part by the CAPES DeepEyes Project, in part by the São Paulo Research Foundation under Grant 2010/05647-4, Grant 2010/14910-0, and Grant 2011/22749-8, in part the Brazilian National Research Council under Grant 140916/2012-1, Grant 477662/2013-7, Grant 307113/2012-4, and Grant 304352/2012-8, in part by the Instituto Federal de Educação, Ciência e Tecnologia do Sudeste de Minas Gerais, and in part by the University of Campinas. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gwenaël J. Doërr.

T. Carvalho, H. Pedrini, R. da S. Torres, and A. 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; rtorres@ic.unicamp.br; anderson@ic.unicamp.br).

F. A. Faria is with the GIBIS Laboratory, Federal University of São Paulo, São Paulo 04021-001, Brazil (e-mail: ffaria@unifesp.br).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author.

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.2015.2506548

(a) (b)

Fig. 1. Fake images created through splicing. (a) Conman Dimitri de Angelis (right) alongside US former president Bill Clinton (left). (b) Fake image of Brazil's former president (center) in a purported personal life moment with an investigated gang leader (left).

This verification might be as simple as checking whether an image has been touched up for exhibition improvement purposes (e.g., brightness or contrast) or as complex as detecting if the image has been tampered with aiming at, ultimately, deceiving the viewer.

One of the most common image tampering operations is the splicing or composition. It consists in using parts of two or more different images to construct a new image depicting a moment that never happened in space and time. It is not difficult to find cases in which people use image composition to take business or personal self advantage. Consider the following two cases.

In May 2013, the conman Dimitri de Angelis was sentenced to twelve years in prison for deceiving investors using "photoshoped" photos in which he appeared alongside many different prominent people as, for example, former US president Bill Clinton, as Figure 1(a) portrays. In another case, dating to November 23rd 2012, a fake photo went viral in the Internet purporting Brazil's former president Luiz Inácio Lula da Silva beside Rosemary de Noronha, a suspicious gang leader investigated by the Brazilian Federal Police (see Figure 1(b)). We analyze these two famous cases and many others with details in Section IV-B6.

Unfortunately, these examples are not out of place. Methods for detecting image compositions explore different telltales left behind during the forgery process. Inconsistencies in compression or in the relationship of neighboring pixels are some of the explored image properties. In particular, methods for exploring lighting inconsistencies under different forms [1]–[4] are particularly of interest as a perfect illumination adjustment in a digital composite is very difficult to obtain.

Instead of directly computing physical properties such as the light direction, one could explore image transformed spaces for capturing artifacts and pinpoint possible forgeries. As a motivational example, Pinto *et al.* [5], for instance, represented

input videos as visual rhythms of the Fourier Spectra for highlighting artifacts associated with biometric spoofing in face recognition systems. In this vein, in this paper, we propose to use illuminant maps (IM) [6] as a transformed representation space to highlight different types of inconsistencies present in fake images, not easily detectable in the original image space, and point out possible image forgeries.

In our approach, inconsistencies of color, texture and shape present in a fake image become more pronounced in the transformed image, which is obtained converting an input image to different illuminant maps. More specifically, this work extends upon the method recently proposed by de Carvalho *et al.* [7], in which the authors use texture and edge descriptors to characterize IMs and detect inconsistencies in images pointing out possible tampering operations. In this work, we extensively study different ways to use combinations of different IMs for different color spaces and examine the most appropriate image descriptors and classifiers to better capture visual properties that might lead to forgery detection. We strive for exploring complementary features in order to achieve a very effective classification rate in different setups.

Our main contributions herein are: (1) the use of color descriptors computed upon transformed image spaces (illuminant maps, IMs) and a full study of the effectiveness and complementarity of these image descriptors computed on such transformed spaces; (2) the adoption of a machine learning framework in the proposed approach, for automatically selecting the best combination of all the factors of interest (e.g.,transformation spaces (IMs), color-space representations, descriptors, and classifiers; (3) a quantitative evaluation of the differences among pristine and fake images when represented in different IM spaces; and (4) the introduction of an approach to detecting the most likely doctored part in fake images.

We perform the evaluation of the proposed method in different public benchmarks and also compare it to recent literature techniques that consider, in different levels, the illuminant information in the detection process. The proposed method yields an improvement of 15 percentage points in the classification accuracy and the possibility of providing a confidence degree associated with the classified image when compared to the state-of-the-art [7].

We organized the text into four more sections. Section II describes some of the most recent methods in the literature that consider, in different levels, illuminant information for detection of image splicing. Section III introduces our methodology and its details, whereas Section IV shows the experimental setup used to validate the method, the performed experiments and results for different benchmarks, as well as the comparison with state-of-the-art methods. Finally, Section V presents some conclusions and opportunities for future work.[1]

II. Related Work

The digital age, with all its facilities, also has its nuisances. One of them, empowered by cheap computing devices and powerful image editing software, is photo tampering. With little effort and a proper image manipulation tool (e.g., Adobe Photoshop or Gimp), ordinary people can create masterpieces depicting unbelievably credible photomontages with ease. In addition, the ever-growing quality and power of image editing software have taken image splicing to a whole new level of credibility and difficulty of detection. Such difficulties lead to the need of development of equally sophisticated methods for detecting image telltales left by forgers.

Methods that explore some degree of illumination inconsistencies for detecting image splicing have been the focus of many researchers for over a decade. Basically, they can be divided into two types of approaches: (a) those that look for inconsistencies in light source environment; and (b) the ones that look for inconsistencies in the estimated light source color from the image.

The approaches grounded on inconsistencies of light source environment estimate the environment illumination from an image in the acquisition moment, which involves estimating the light source position or reconstructing a full illumination model from the scene [1]–[3], [8]. On the other hand, approaches grounded on inconsistencies of light source color focus on exploring different kinds and levels of information provided by estimated scene illuminants. Furthermore, this kind of approaches can be subdivided into three groups: the first one explores the specular part in the dichromatic reflectance model; the second one proposes to subdivide the image into small regions, on top of which they compute the illuminant descriptors; and the last one, which can either be seen as an extension of the second group, as it does not contribute to illuminant estimation directly, or as a subgroup by itself, as it focuses on substantial processing on top of IMs.

The first method of group one is represented by Gholap and Bora [9], who pioneered the use of illuminant colors to detect image compositions. They have used the dichromatic reflection model proposed by Tominaga and Wandell [10], which assumes a single light source to estimate illuminant colors from images. Dichromatic planes are estimated through Principal Component Analysis (PCA) from each specular highlight region of an image. Applying a Singular Value Decomposition (SVD) on an RGB matrix extracted from highlighted regions, the authors extract the eigenvectors associated with two significant eigenvalues to construct the dichromatic plane. This plane is then mapped onto a straight line, named *dichromatic line*, in normalized rg-chromaticity space. For distinct objects illuminated by the same light source, the intersection point produced by their dichromatic line intersection represents the illuminant color. If the image has more than one illuminant, it will present more than one intersection point (not expected to happen in pristine images). One problem with this method is the need of well-defined specular highlight regions for estimating the illuminants.

Following Gholap and Bora's work [9], Francis *et al.* [11] also worked with illuminant estimation. They have used the dichromatic reflection model to estimate illuminant colors in human skin highlighted regions. Based on chromaticity coordinates of the estimated illuminant color, the authors

[1]The source code of illuminant maps generation, extracted image descriptors, machine learning framework and databases used at this work can be downloaded on http://dx.doi.org/10.6084/m9.figshare.1593104.

quantify the amount of chromaticity from each person in the image and use it to detect forgeries by matching distributions obtained from people in the same image.

The second group of approaches is represented by the methods proposed by Riess and Angelopoulou [6] and Wu and Fang [12]. Riess and Angelopoulou [6] have used an extension of the Inverse-Intensity Chromaticity Space, originally proposed by Tan *et al.* [13], to estimate illuminants locally from different parts of an image for detecting forgeries. Roughly speaking, Riess and Angelopoulou's [6] method comprises four steps: (1) image segmentation grouping regions of approximately the same object color, for generating blocks named *superpixels*, which are directly illuminated by the investigated light source illuminant and roughly consistent with the physical model of Inverse-Intensity Chromaticity Space; (2) selection of superpixels to be further investigated by the user; (3) estimation of the illuminant color, which is performed twice, one for every superpixel and another one for the selected superpixels; and (4) calculation of the distance from the selected superpixels to the other ones generating a distance map, which is the basis for an expert analysis regarding forgery detection. Differently from Gholap and Bora's work, which estimates illuminants for entire objects, Riess and Angelopoulou's method deals with small parts of objects performing a more robust analysis of the illuminant color. Unfortunately, the high degree of user dependence for the selection of superpixels and for the distance map analysis makes the method strongly susceptible to human errors.

Wu and Fang [12] have proposed a new way to detect forgeries using illuminant colors. Their method divides a color image into overlapping blocks estimating the illuminant color for each block. The authors used the algorithms Gray-World, Gray-Shadow, and Gray-Edge [14] to estimate the illuminant color. In addition, they used a maximum likelihood classifier proposed by Gijsenij and Gevers [15] to select the most appropriate method for representing the illuminant of each block. For forgery detection, the authors choose some blocks as reference and estimate their illuminants. The angular error between reference blocks and a suspicious block is calculated. If this distance exceeds a threshold, a block is classified as manipulated. This method is also strongly dependent on user interaction and perception to choose correct reference blocks. If the reference blocks are incorrectly chosen, for example, the performance of the method is strongly compromised.

Finally, the third and last group of methods is represented by works such as the one proposed by de Carvalho *et al.* [7]. Aiming at reducing the user dependency, de Carvalho *et al.* [7] proposed a different way to use illuminants for detecting forgeries. In a custom-tailored method for detecting image compositions involving people, the authors estimate illuminant maps for the image using statistics-based and physics-based approaches. Texture and shape descriptors are used to characterize these illuminant maps on face regions. The forgery detection is then performed through machine learning techniques such as SVM classifiers. Several important aspects were not explored in their proposed method such as complementary information present in illuminant maps and automatic detection of the fake part in an image classified as fake.

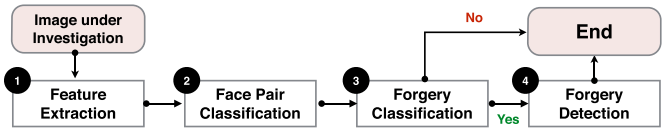


Fig. 2. Overview of the proposed image forgery classification and detection methodology.

In the next section, we perform a broad study of more than 50 different ways of exploring illuminant maps as a transformed space for forgery detection. These description methods are associated with a robust classifier fusion framework to achieve a remarkable improvement when compared with the state of the art [7].

III. Proposed Forgery Detection Methodology

This section describes each step of the proposed image forgery detection methodology.

A. Overview of Forgery Detection

The splicing detection process commonly relies on the expert's experience and background knowledge. This process usually is time consuming and error prone as image splicings are evermore sophisticated, and an aural (e.g., visual) analysis may not be enough to detect forgeries.

Our approach to detecting image splicing, which is specific for pinpointing composites of people, is developed aiming at minimizing the user interaction. The splicing detection task consists in labelling a new image among two pre-defined classes (real and fake) and later pointing out the face with higher probability to be the fake face. In this process, a classification model is created to indicate the class to which a new image belongs.

The image forgery detection methodology comprises four main steps:

1) *Description:* relies on algorithms for estimating IMs, a transformed representation space of the input image, and extracting image visual cues (e.g., color, texture, and shape), encoding the extracted information into feature vectors;
2) *Face Pair Classification:* relies on algorithms that use image feature vectors to learn intra- and inter-class patterns of the images to classify each new image feature vector;
3) *Forgery Classification:* consists in labelling a new image into one of existing known classes (real and fake) based on the previously learned classification model and description techniques;
4) *Forgery Detection:* once knowing that an image is fake, this stage aims at identifying which face is more likely to be fake in the image.

Figure 2 gives an overview of the method whose steps will be refined in the next sections.

B. Description

Image descriptors have been used in many different problems in the literature, such as content-based image

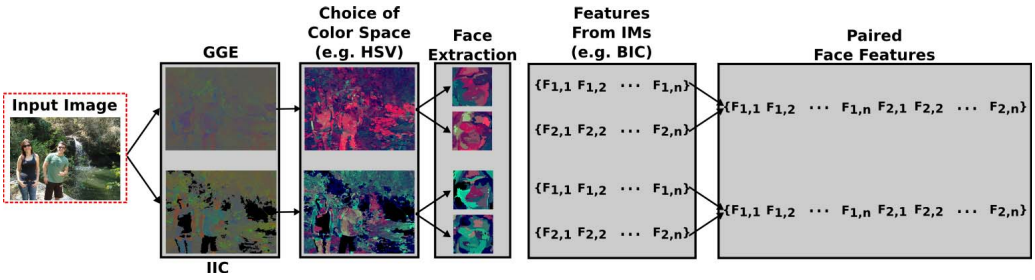


Fig. 3. Image description pipeline. Steps *The Choice of Color Spaces* and *Features From Illuminant Maps* can have many different variants, which allow us to characterize the transformed spaces (IMs) with a wide range of cues and telltales.

retrieval [16], medical image analysis [17], and geographic information systems [18], to name just a few.

The method proposed by de Carvalho *et al.* [7] represents an important step toward an improved and less user dependent approach to image splicing detection. Although effective, the authors explored just a limited range of image descriptors and did not explore many complementary properties in their analysis.

In a real forensic scenario, an improved accuracy in fake detection is much more important than a real-time application. Therefore, in this work, we propose to augment the description complexity of images. The objectives of the more complex augmented description are twofold: first, to boost the classification accuracy; second, to capture more image details and cues in a transformed IM space, which are often imperceptible with a visual analysis. Later on, we present an approach to selecting the most important description techniques.

Our method employs a combination of different algorithms for calculating the IM transformed spaces, the representation of such IMs in different color spaces, and a wide variety of image descriptors to explore different and complementary properties of each method in order to obtain a more robust and effective image representation for detecting forgeries. This description process has a 5-step pipeline, which is depicted in Figure 3 and described as follows.

1) Representation of Images in IM Transformed Space: In general, the literature describes two main classes of algorithms for estimating IMs and thus representing an input image as an illuminant map: statistics-based and physics-based. On one hand, statistics-based IM estimation algorithms rely on hypotheses related to statistics of image pixels (e.g., grayworld illuminant method [14] assumes that, under a white light source, the average pixel colors in a scene is achromatic). On the other hand, physics-based IM estimation methods rely on theoretical formulations of how light interacts with objects (e.g.,considering the dichromatic reflection model). Bearing in mind that both classes of IM estimation methods capture different image illumination information, and taking advantage of the strategy proposed by Riess and Angelopoulou [6], which divides the image into small clusters with similar color (named *superpixels*) to estimate illuminants locally at each *superpixel*, the proposed method herein takes advantage of these different types of information to muster complementary features in the forgery detection process.

For capturing statistical-based information, we transform the images through the generalized grayworld estimates algorithm (GGE) proposed by van de Weijer *et al.* [14]. According [6], [14], this algorithm estimates the illuminant $\mathbf{e}$ from pixels as

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

where $\mathbf{x}$ denotes a pixel coordinate, k is a scale factor, $|\cdot|$ is the absolute value, ∂ the differential operator, $\mathbf{f}^\sigma(\mathbf{x})$ is the observed intensities at position $\mathbf{x}$, smoothed with a Gaussian kernel σ, p is the Minkowski norm, and n is the derivative order.

In turn, for capturing physics-based information, we use a variant of the inverse-intensity chromaticity space (IIC) proposed by Riess and Angelopoulou [6], which modifies the original inverse-intensity chromaticity space estimation proposed by Tan *et al.* [19], to deal with local illuminant estimation. In this modified estimation model, the intensity $f_c(\mathbf{x})$ and the chromaticity $\chi_c(\mathbf{x})$ of a color channel $c \in \{R, G, B\}$ at position $\mathbf{x}$ is represented by

$$\chi_c(\mathbf{x}) = m(\mathbf{x}) \frac{1}{\sum_{i \in \{R,G,B\}} f_i(\mathbf{x})} + \gamma_c. \tag{2}$$

In this equation, γ_c represents the chromaticity of the illuminant in channel c, whereas $m(\mathbf{x})$ mainly captures geometric influences, i.e.,light position, surface orientation and camera position, and is approximate as described in [19].

2) Choice of Color Space Model and Face Extraction: Some color space models are more appropriate for extracting meaningful features than others depending on the target application [20]. Additionally, as far as we know, there is no research in digital forensics showing whether or not a specific color space is more suitable for representing image cues when analyzing forgeries, specially when using illuminant maps. Therefore, given that some description techniques are more suitable for specific color spaces, this step converts illuminant maps into different color space representations for further exploration and study.

Differently from de Carvalho *et al.* [7], who have used the conversion of IMs to the YCbCr color space, we propose to augment the number of explored color spaces in order to capture the smallest nuances present in such maps not visible in the original representation of a transformed image to an illuminant map representation. We consider the Lab, HSV, and original normalized RGB color spaces [21]. We have chosen such color spaces, which are popular choices in image processing literature [20]. Once we define a color space,

TABLE I

DIFFERENT DESCRIPTORS USED IN THIS WORK. EACH TABLE ROW REPRESENTS AN IMAGE DESCRIPTOR AND IT IS COMPOSED OF THE COMBINATION (TRIPLET) OF AN IM, A COLOR SPACE (ONTO WHICH IMS HAVE BEEN CONVERTED), AND THE DESCRIPTION TECHNIQUE USED TO EXTRACT THE DESIRED PROPERTY

IM Transf. Space	Color Space	Description Technique	Kind
GGE	Lab	ACC	Color
GGE	RGB	ACC	Color
GGE	YCbCr	ACC	Color
GGE	Lab	BIC	Color
GGE	RGB	BIC	Color
GGE	YCbCr	BIC	Color
GGE	Lab	CCV	Color
GGE	RGB	CCV	Color
GGE	YCbCr	CCV	Color
GGE	HSV	EOAC	Shape
GGE	Lab	EOAC	Shape
GGE	YCbCr	EOAC	Shape
GGE	HSV	LAS	Texture
GGE	Lab	LAS	Texture
GGE	YCbCr	LAS	Texture
GGE	Lab	LCH	Color
GGE	RGB	LCH	Color
GGE	YCbCr	LCH	Color
GGE	HSV	SASI	Texture
GGE	Lab	SASI	Texture
GGE	YCbCr	SASI	Texture
GGE	HSV	SPYTEC	Shape
GGE	Lab	SPYTEC	Shape
GGE	YCbCr	SPYTEC	Shape
GGE	HSV	UNSER	Texture
GGE	Lab	UNSER	Texture
GGE	YCbCr	UNSER	Texture
IIC	Lab	ACC	Color
IIC	RGB	ACC	Color
IIC	YCbCr	ACC	Color
IIC	Lab	BIC	Color
IIC	RGB	BIC	Color
IIC	YCbCr	BIC	Color
IIC	Lab	CCV	Color
IIC	RGB	CCV	Color
IIC	YCbCr	CCV	Color
IIC	HSV	EOAC	Shape
IIC	Lab	EOAC	Shape
IIC	YCbCr	EOAC	Shape
IIC	HSV	LAS	Texture
IIC	Lab	LAS	Texture
IIC	YCbCr	LAS	Texture
IIC	Lab	LCH	Color
IIC	RGB	LCH	Color
IIC	YCbCr	LCH	Color
IIC	HSV	SASI	Texture
IIC	Lab	SASI	Texture
IIC	YCbCr	SASI	Texture
IIC	HSV	SPYTEC	Shape
IIC	Lab	SPYTEC	Shape
IIC	YCbCr	SPYTEC	Shape
IIC	HSV	UNSER	Texture
IIC	Lab	UNSER	Texture
IIC	YCbCr	UNSER	Texture

we extract all faces present in the investigated image using a user-defined manual bounding box.

3) Feature Extraction From IMs: From each extracted face in the previous step, we need to find telltales that allow identification of spliced images. Such information is present in different visual properties (e.g., texture, shape, color, among others) and becomes detectable when we transform suspicious images into an IM representation.[2]

Texture, for instance, allows us to characterize faces whereby illuminants are disposed similarly when comparing two faces. The SASI [22] technique, that was used by de Carvalho *et al.* [7], presented a good performance in their work, therefore, we keep it in our analysis. Furthermore, guided by the excellent results reported in a recent study by Penatti *et al.* [20], we included the LAS [23] technique. Complementarily, we also incorporated the Unser [24] descriptor, which presents a lower complexity and generates compact feature vectors when compared to SASI and LAS.

Differently from texture properties, shape properties present in IMs of fake faces, sometimes, have distinct pixel intensities when compared to shapes present in IMs of faces that originally belong to the analyzed image. In this sense, de Carvalho *et al.* [7] proposed the Hogedge descriptor, which led to a classification accuracy close to 70% in their work. Due to its complexity, in this work, we replace it by two other shape techniques, EOAC [25] and SPYTEC [26]. EOAC is based on shape orientations and correlation between neighboring shapes. These are properties that are potentially useful for forgery detection using IMs transformed space given that neighboring shape in regions of composed faces tend not to be correlated. We selected SPYTEC as it uses the wavelet transform, capturing multi-scale information normally detectable in the frequency domain.

At this point, it is important to highlight that illuminant maps are estimated from a subdivision of the image space (superpixels), which can, to a certain extent, influence the texture and shape descriptors performance. However, such influence has not been deeply or completely investigated in this work and certainly is a future work worth tackling.

According to Riess and Angelopoulou [6], when IMs are analyzed by an expert for detecting forgeries, the main observed feature is color. Thus, in this work we decided to add color description techniques, extracted from the IMs transformed spaces, as an important visual cue to be encoded into the description process. The considered color description techniques are ACC [27], BIC [28], CCV [29], and LCH [30].

ACC is a technique based on color correlograms and encodes image spatial information. This color description technique is very robust when dealing with changes in appearance and shape, information that is indirectly represented in an IM, allowing us to compare faces in different positions. The BIC technique captures border and interior properties in an image and encodes them in a quantized histogram. It presented good performance in the study carried out by Penatti *et al.* [20]. CCV is a well-known segmentation-based color technique often used as a baseline in several analysis. Complementarily, LCH is a simple local color description technique, which encodes color distributions of fixed-size regions of the image. This might be useful when comparing similar regions (e.g.,forehead) extracted from IMs in two different faces.

4) Face Characterization and Paired Face Features: Given that in this work we consider more than one variant of IMs transformed spaces, color space representations, and description techniques, let $\mathcal{D}$ be an image descriptor composed of the triplet (IM Transformed Space, color space representation, and description technique). Assuming all possible combinations of such triplets we consider herein, we have 54 different image descriptors. Table I shows all such combinations.

Finally, to detect a forgery, we need to analyze whether a suspicious part of the image is consistent with other parts from the same image. Specifically, when we try to detect forgeries involving composites of people faces, we need to compare if a suspicious face is consistent with other faces in the image. In the worst case, all faces are suspicious and need to be compared to the others. Thus, instead of analyzing each image face separately, after building $\mathcal{D}$ for each face in the image, we encode the feature vectors of each pair of faces under

[2]Details about image descriptors used in this work are presented in a Supplementary Material along with this paper.

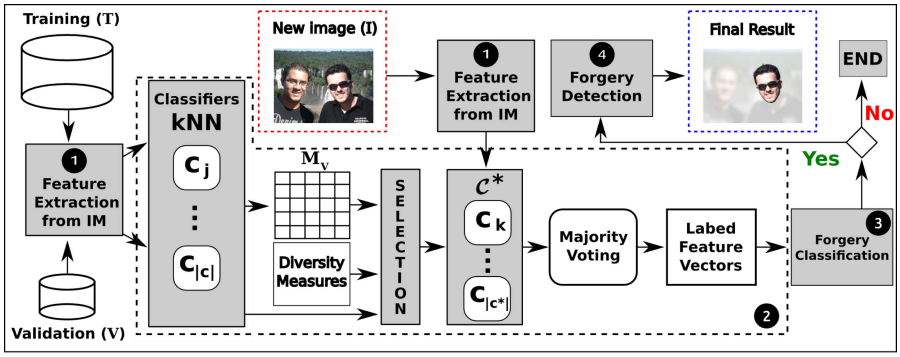


Fig. 4. Proposed framework for image splicing detection.

analysis into one feature vector $\mathcal{P}$, constructed through direct juxtaposition of two feature vectors $\mathcal{D}$.

C. Face Pair Classification

In this section, we show details about the classification step. When using different IMs, color spaces, and description techniques, the obvious question is how to automatically select the most important ones to keep and combine for an improved classification performance. For this purpose, we take advantage of the classifier selection and fusion framework proposed by Faria *et al.* [31].

Let $\mathcal{C}$ be a set of classifiers in which each classifier $c_j \in \mathcal{C}$ $(1 < j \leq |\mathcal{C}|)$ is composed of a tuple comprising a learning method (e.g., Naïve Bayes, k-Nearest Neighbors and Support Vector Machines) and a single image descriptor $\mathcal{D}$.

Initially, all classifiers $c_j \in \mathcal{C}$ are trained on the elements of a training set T. Next, the outcome of each classifier on the validation set V, different from T, is computed and stored into a matrix M_V, where $|M_V| = |V| \times |\mathcal{C}|$ and $|V|$ is the number of images in a validation set V and $|C|$ is the number of classifiers. The actual training and validation data samples are known *a priori*.

In the following, M_V is used as input to select a set $\mathcal{C}^* \subset \mathcal{C}$ of classifiers that are good candidates to be combined. In this selection process, five diversity measures (Correlation Coefficient ρ, Double-Fault Measure, Disagreement Measure, Interrater Agreement k, and Q-Statistic [32].[3]) are computed to achieve the degree of agreement/disagreement between all available classifiers in $\mathcal{C}$. Finally, $\mathcal{C}^*$, containing the most promising classifiers and satisfy a defined threshold $\mathcal{T}$, are selected. $\mathcal{T}$ is a threshold defined in terms of the average accuracy among all classifiers using validation set V. For a more detailed description of selection process using diversity measures, the reader is referred to [31].

Given a set of paired feature vectors $\mathcal{P}$, extracted from a new image I, we use each classifier $c_k \in \mathcal{C}^*$ $(1 < k \leq |\mathcal{C}^*|)$ to determine the label (forgery or real) of these feature vectors, producing k outcomes. The k outcomes are used as input of a fusion technique (in this case, majority voting) that takes the final decision regarding the definition of each paired feature vector $\mathcal{P}$ extracted from I.

[3]For a better understanding of all diversity measures used herein, they are described in a Supplementary Material along with this paper.

Figure 4 depicts a fine-grained view of this forgery detection framework. Figure 4-**(2)** shows the entire classifier selection and fusion process.

We should point out that the fusion technique used in the original framework [31] has been exchanged from Support Vector Machines (SVM) [33] to majority voting. When the original framework is used, the SVM technique creates a very specialized model for detecting real images, which increases the number of false negatives. However, in a realistic forensic scenario, we look for decreasing the false negative rate and, in order to achieve it, we adopted a majority voting technique as a more viable alternative. Furthermore, we use these k outcomes to calculate a confidence degree associated with the label. Let k_r be the number of outcomes pointing out an image as real and k_f the number of outcomes pointing it as a fake, where $k = k_r + k_f$. The confidence associated with the image I is given by $\frac{max(k_r,k_f)}{k}$. Tie-breaks (50% confidence) are randomly decided with equal probability.

D. Forgery Classification

Given an image I that contains q people, it is characterized by a set $\mathcal{S} = \{\mathcal{P}_1, \mathcal{P}_2, \cdots, \mathcal{P}_m\}$ being $m = \frac{q \times (q-1)}{2}$ and $q \geq 2$. We adopt a strategy that prioritizes forgery detection. Hence, if any paired feature vector $\mathcal{P} \in \mathcal{S}$ is classified as fake, we classify the image I as fake. Otherwise, we classify it as pristine or non-fake.

E. Forgery Detection

Moving one step forward, we design a specific method for detecting, among all the faces in an image, the one with the highest probability to be the fake face.

Given an image I classified as fake, we now refine the analysis pointing out which part of the image is the result of a composition. This step was overlooked in the work proposed by de Carvalho *et al.* [7]. We cannot employ the same paired feature vectors used in Section III-D (*Forgery Classification*), since we would find the pair with the highest probability instead of the face with highest probability to be fake.

For this task, we take advantage of IMs estimated from different principles (statistical-based and physics-based). The reason is that the aforementioned techniques can produce IMs with different aspects for the same image. In a through analyses of the IMs produced by these two different models,

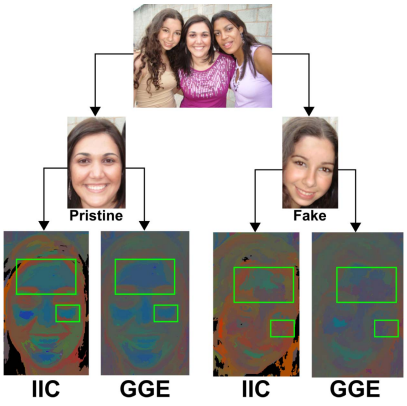


Fig. 5. Differences in ICC and GGE illuminant maps. The highlighted regions exemplify how the difference between ICC and GGE is increased on fake images. In special, the right cheek of the fake face varies from an orange color (in IIC) to a bluish color (in GGE). This is a drastic change wrt. a similar region of a normal face, which varies from an intensity of blue to another. This kind of variation can be captured through the difference between color descriptors, guiding us to a new way of identifying the most probable fake face.

we realized that the appearance in terms of colors in IMs generated for pristine faces are very similar in GGE and IIC. Notwithstanding, when an image contains a fake face, the difference (in terms of color appearance) between GGE and IIC for this fake face is increased. Figure 5 depicts an example of this fact.

Based on this color observation, given an image I already classified as fake (see Section III-D), we propose, for each face f, a descriptor fv_r

$$fv_r = |fc_{IIC} - fc_{GGE}|, \tag{3}$$

where fc_{IIC} and fc_{GGE} are multidimensional feature vectors extracted, using some color image descriptor (e.g., BIC, ACC, etc.), from IIC and GGE maps, respectively.

After training an SVM [33] with a radial basis function (RBF) kernel, we classify the resulting multidimensional feature vector fv_r and collect the two-class output probabilities of this input vector.

IV. Experiments and Results

This section describes the experiments performed in this work to show the effectiveness of the proposed method as well as to compare it with state-of-the-art counterparts. Round #1 intends to show the best k-nearest neighbor (kNN) classifier to be used in the additional rounds of tests. Instead of focusing on a more complex and complicated classifier, we select the simplest one possible for the individual learners in order to show the power of the features we employ as well as the utility of our proposed method for selecting the most appropriate combinations of features, color spaces, and IM transformed space. Round #2 aims at comparing the proposed method to four methods of the literature using DSO-1, a realistic dataset comprising high-resolution pristine and fake images. Round #3 compares the performance of the kNN classifier to more complex learning methods, enforcing our choice for using the simplest method. Round #4 explores the ability of the proposed method to find the actual forged face in an image, while Round #5 shows specific tests

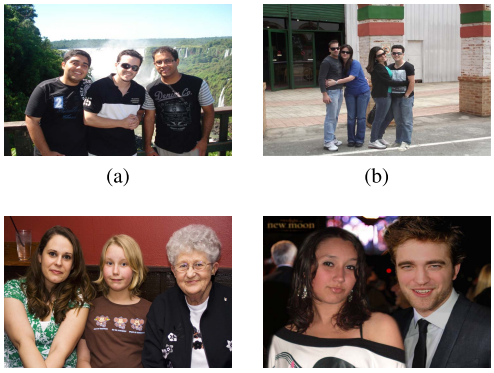


Fig. 6. (a) and (b) Examples of DSO-1 dataset. (c) and (d) Examples of DSI-1 dataset. (a) Pristine. (b) Fake. (c) Pristine. (d) Fake.

with DSI-1, a second dataset comprising original and fake images collected from the Internet. Finally, Round #6 shows a qualitative analysis of famous cases involving questioned images.

A. Datasets and Experimental Setup

For a fair comparison with the state-of-the-art methods, we have used three different public datasets: DSO-1 and DSI-1, both provided by de Carvalho *et al.* [7], and a small set of some famous image composition cases collected from the Internet.[4] DSO-1 dataset comprises 200 indoor and outdoor images, with 100 original and 100 fake images, an image resolution of $2,048 \times 1,536$ pixels. DSI-1 dataset comprises 50 images (25 original and 25 doctored) downloaded from the Internet with different resolutions. In addition, we have used the same users' marks of faces as Carvalho et al. used in their work.[5] Figure 6 (a) and (b) depict examples of the DSO-1 dataset, whereas Figure 6 (c) and (d) depict examples of the DSI-1 dataset.

From Rounds #1 to #5, we have used the 5-fold cross-validation protocol, which allowed us to report results that are directly and easily comparable in the testing scenarios.

Another important point of this work is the form we present the obtained results. We use the average accuracy across the 5-fold cross-validation protocol and its standard deviation. The accuracy rate is calculated as

$$\frac{TP\,(TN+FP)+TN\,(TP+FN)}{2\,(TP+FN)\,(TN+FP)} \times 100 \tag{4}$$

where TP, FN, TN, FP are, respectively, results for true positives, false negatives, true negatives and false positives.

For a direct comparison with the results reported in [7], we also present ROC curves and their AUCs for the most representative methods. Sensitivity (number of true positives or the number of fake images correctly classified) and specificity (number of true negatives or the number of pristine images correctly classified) are also provided for operational points.

For all image descriptors, we have used the standard configuration proposed by Penatti *et al.* [20].

[4]Freely available at http://tinyurl.com/mqrse3s upon acceptance of this paper.

[5]We thank the authors for providing us with all the necessary materials.

TABLE II

ACCURACY COMPUTED FOR KNN TECHNIQUE USING DIFFERENT k VALUES AND TYPES OF IMAGE DESCRIPTORS. PERFORMED EXPERIMENTS USING VALIDATION SET AND 5-FOLD CROSS-VALIDATION PROTOCOL HAVE BEEN APPLIED. ALL RESULTS ARE IN %

Descriptors	kNN-1	kNN-3	kNN-5	kNN-7	kNN-9
ACC	72.0	72.8	73.0	72.5	**73.8**
BIC	70.7	71.5	72.7	76.4	**77.2**
CCV	70.9	70.7	**74.0**	72.3	72.5
EOAC	64.8	65.4	**65.5**	65.2	63.9
LAS	67.3	69.1	71.0	**75.0**	72.2
LCH	61.9	**64.0**	62.2	62.1	63.7
SASI	67.9	70.3	**71.6**	69.9	70.1
SPYTEC	63.0	62.4	62.7	**64.5**	**64.5**
UNSER	65.0	66.9	67.0	**67.8**	67.1

B. Experiments

1) Round #1 (Finding the Best kNN Classifier): After characterizing an image with a specific image descriptor, the next step consists of using an appropriate learning method. The method proposed here focuses on using complementary information to describe the IMs. For that, we selected the k-Nearest Neighbor (kNN) classifier [33] instead of more powerful and computational intensive ones such as Support Vector Machines (SVM).

Even with a simple learning method such as kNN, we still need to determine the most appropriate value for parameter k. This round of experiments aims at exploring best k which will be used in the remaining set of experiments.

For this experiment, to describe each paired vector of the face $\mathcal{P}$, we have extracted all image descriptors from IIC in the YCbCr color space. This configuration has been chosen because it was one of the combinations proposed by de Carvalho *et al.* [7] and because the IM produced by IIC was used twice in the metafusion described in their work. We have used DSO-1 with a 5-fold cross-validation protocol from which three folds are used for training, one for validation and one for testing.

Table II shows the results for the entire set of image descriptors we consider herein. kNN-5 yielded the best classification accuracy in three of the image descriptors. As we mentioned before, this work focuses on looking for the best group of features to achieve an improved accuracy. Hence, we decided to choose kNN-5 that is simpler, faster, and with a better accuracy than the alternatives.

2) Round #2 (Performance on DSO-1 Dataset): We now apply the proposed method for classifying an image as fake or real (the actual detection/localization of the forgery will be explored in Section IV-B4). For this experiment, we consider the DSO-1 dataset.

We have used all 54 image descriptors with kNN-5 learning technique resulting in 54 different classifiers. Recall that a classifier is composed of one descriptor and one learning technique. By using the modified combination technique we propose, we select the best combination $|\mathcal{C}^*|$ of different classifiers. We tested different numbers of combinations, $|\mathcal{C}^*| = \{5, 10, 15, \ldots, 54\}$. The best obtained result was an

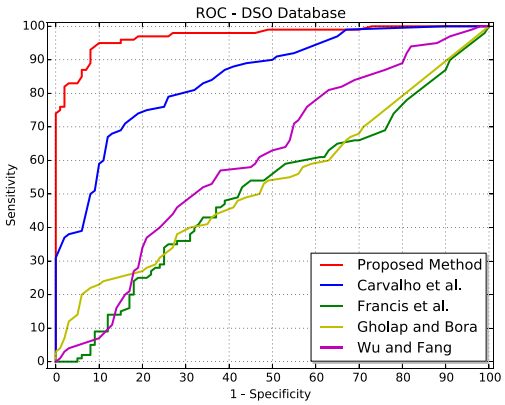
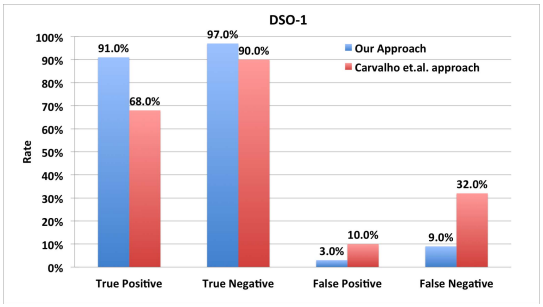


Fig. 7. ROC – Proposed method against state of the art approaches.



DSO-1 dataset

Fig. 8. Comparison between our approach and Carvalho et al.'s method [7] on the DSO-1 dataset.

average accuracy of 94.0% (with an AUC of 97.19% and an operational point with Sensitivity of 91.0% and Specificity of 97.0%) with a standard deviation of 4.5% using all 54 classifiers $\mathcal{C}$ as Figure 7 shows. This result is 15 percentage points better, in terms of accuracy, than the best result reported by de Carvalho *et al.* [7] (the authors report an AUC of 86.3% and operational point with 68.0% Sensitivity and 90.0% of Specificity, with an average classification accuracy of 79.0%). Figure 7 also shows that the proposed method significantly outperforms other methods that consider, in different levels, the illuminant information in the image splicing detection process such as Gholap and Bora [9], Wu and Fang [12], and most recent one proposed by Francis *et a.* [11]. The ROC curve of the proposed method was built using the confidence metric presented in Section III-C.

For a better visualization, Figure 8 depicts a direct comparison between our approach and Carvalho et al.'s method [7], where it is clear the superiority of our results.

Table III shows the results of all tested combinations of $|\mathcal{C}^*|$ on each testing fold and their average and standard deviation. Given that the forensic scenario is more interested in a high classification accuracy than a real-time application (our method takes around three minutes to extract all features from an investigated image), the use of all 54 classifiers is not a major problem. However, the result using only the best subset of them ($|\mathcal{C}^*| = 20$ classifiers) achieves an average accuracy of 90.5% (with a Sensitivity of 84.0% and a Specificity of 97.0%) with a standard deviation of 2.1%, which is a remarkable result compared to the one reported in de Carvalho *et al.* [7].

TABLE III

CLASSIFICATION RESULTS OBTAINED FROM THE METHODOLOGY DESCRIBED IN SECTION III WITH A 5-FOLD CROSS-VALIDATION PROTOCOL FOR DIFFERENT NUMBER OF CLASSIFIERS ($|\mathcal{C}^*|$). ALL RESULTS ARE IN %

DSO-1 dataset											
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**										
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**54 (ALL)**
1	90.0	85.0	92.5	90.0	90.0	95.0	90.0	87.5	87.5	90.0	92.5
2	90.0	87.5	87.5	90.0	90.0	90.0	87.5	90.0	90.0	90.0	90.0
3	95.0	92.5	92.5	92.5	95.0	95.0	95.0	95.0	95.0	95.0	97.5
4	67.5	82.5	95.0	92.5	92.5	95.0	97.5	97.5	95.0	100.0	100.0
5	82.5	80.0	80.0	87.5	85.0	90.0	90.0	90.0	87.5	87.5	90.0
Final (Avg)	85.0	85.5	89.5	90.5	90.5	92.0	92.0	91.0	91.0	92.5	94.0
Std. Dev.	10.7	4.8	6.0	2.1	3.7	2.7	4.1	4.1	3.8	5.0	4.5

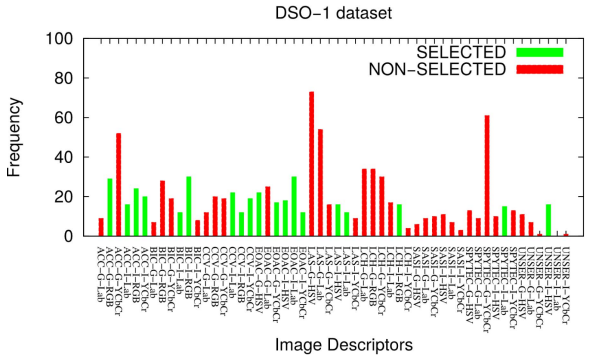


Fig. 9. Classification histograms created during training of the selection process (Section III-C) for the DSO-1 dataset.

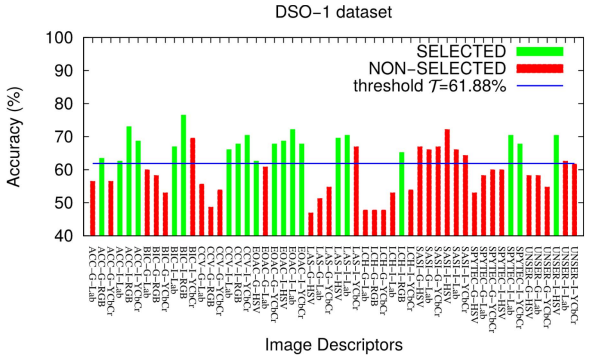


Fig. 10. Classification accuracies of all non-complex classifiers (kNN-5) used in our experiments. The blue line shows the actual threshold $\mathcal{T}$ described in Section III-C used for selecting the most appropriate classification techniques during training. In green, we highlight the 20 classifiers selected for performing the fusion and creating the final classification engine.

The selection process is performed as described in Section III-C and is based on the histogram depicted in Figure 9. The classifier selection approach takes into account both the accuracy performance of classifiers and their correlation.

Figure 10 depicts, in green, the $|\mathcal{C}^*|$ classifiers selected when observing the scenario where just 20 classifiers have been selected. All three kinds of descriptors (texture-, color, and shape-based ones) play a key role in this scenario, reinforcing one of our most important contributions, the use of complementary information to detect image splicing forgeries. Furthermore, this analysis also hints at the importance of representing the transformed IMs spaces with different color space models. Taking the texture image descriptors extracted from IIC maps as an example, Unser [24] descriptor has its best performance when extracted from IIC converted to HSV color space. On the other hand, the EOAC shape descriptor achieves its best performance when extracted from IIC converted to the *Lab* colorspace.

3) Round #3 (Performance of kNN Against More Complex Learning Methods): One of the main goals of the method proposed here is to explore complementary information to characterize images and consequently detect image splicing. As previously highlighted, to better evaluate the performance of different features, the proposed method uses a very simple learning technique (kNN). However, in a real forensic scenario, a fake detection technique needs to be as accurate as possible. Since oftentimes different image descriptors work better with more complex learning methods, in this round of experiments, we investigate the impact of using more complex learning techniques instead of a simple one. We also considered two other learning techniques using the DSO-1 dataset and the same 5-fold cross validation protocol, as in Round #2.

In the first scenario, we replace the simple kNN learning technique by SVM [33] using a polynomial kernel and default parameters. Associated with the 54 image descriptors, we keep the same number of classifiers (here a classifier is a combination of one image descriptor and one learning technique). Table IV shows the results.

The second scenario explores the combination of two previously described learning techniques: kNN and SVM, which result in 108 different classifiers (54 image descriptor $\times$ 2 learning techniques). Table V shows the results.

The proposed framework is able to use complementary information as better classify test samples. However, this round of experiments shows that SVM and kNN, or even fusion of both learning methods, present very similar results, enforcing the choice for a simpler learning technique (kNN).

4) Round #4 (Forgery Detection on DSO-1 Dataset): We now use the methodology proposed in Section III-E to detect the face with the highest probability of being the fake face in an image tagged as fake by a classifier.

Using the same 5-fold cross-validation protocol, we now train an SVM[6] classifier using an RBF kernel. To feed the classifier, we extract feature vectors from each face through the methodology explained in Section III-E. A standard grid-

[6]We have used LibSVM implementation http://www.csie.ntu.edu.tw/cjlin/libsvm/ with its standard configuration (As of June 2015).

TABLE IV

CLASSIFICATION RESULTS OBTAINED FROM THE METHODOLOGY DESCRIBED IN SECTION III WITH A 5-FOLD CROSS-VALIDATION PROTOCOL FOR DIFFERENT NUMBER OF CLASSIFIERS ($|\mathcal{C}^*|$). HERE, THE SIMPLE KNN LEARNING TECHNIQUE USED IN ROUND #2 OF EXPERIMENTS HAS BEEN REPLACED BY SVM, A MORE COMPLEX ONE. ALL RESULTS ARE IN %

DSO-1 dataset											
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**										
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**45**	**50**	**54 (ALL)**
1	85.0	95.0	92.5	92.5	92.5	92.5	95.0	95.0	92.5	90.0	92.5
2	95.0	90.0	90.0	87.5	90.0	90.0	87.5	87.5	87.5	85.0	85.0
3	70.0	87.5	90.0	87.5	90.0	90.0	90.0	92.5	95.0	95.0	95.0
4	90.0	92.5	92.5	92.5	97.5	95.0	97.5	97.5	97.5	97.5	97.5
5	75.0	75.0	75.0	80.0	87.5	85.0	85.0	85.0	82.5	82.5	75.0
Final (Avg)	83.0	88.0	88.0	88.0	91.5	90.5	91.0	91.5	91.0	90.0	89.0
Std. Dev.	9.3	7.0	6.6	4.6	3.4	3.3	4.6	4.6	5.4	5.7	8.2

TABLE V

CLASSIFICATION RESULTS OBTAINED FROM THE METHODOLOGY DESCRIBED IN SECTION III WITH A 5-FOLD CROSS-VALIDATION PROTOCOL FOR DIFFERENT NUMBER OF CLASSIFIERS ($|\mathcal{C}^*|$). HERE, WE USE A COMBINATION BETWEEN KNN AND SVM, TWO LEARNING TECHNIQUES PREVIOUSLY EVALUATED. ALL RESULTS ARE IN %

DSO-1 dataset											
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**										
	10	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**108 (ALL)**
1	92.5	97.5	95.0	95.0	95.0	95.0	95.0	95.0	95.0	95.0	95.0
2	90.0	92.5	92.5	90.0	87.5	90.0	87.5	87.5	87.5	90.0	90.0
3	92.5	92.5	92.5	90.0	95.0	95.0	97.5	95.0	95.0	95.0	95.0
4	95.0	92.5	95.0	95.0	95.0	95.0	95.0	97.5	97.5	97.5	97.5
5	85.0	82.5	85.0	87.5	87.5	90.0	90.0	90.0	85.0	87.5	87.5
Final (Avg)	91.0	91.5	92.0	91.5	92.0	93.0	93.0	93.0	92.0	93.0	93.0
Std. Dev.	3.4	4.9	3.7	3.0	3.7	2.4	3.7	3.7	4.8	3.7	3.7

TABLE VI

ACCURACY FOR EACH COLOR DESCRIPTOR ON FAKE FACE DETECTION APPROACH. ALL RESULTS ARE IN %

Descriptors	**Accuracy (Avg.)**	**Std. Dev.**
ACC	76.0	5.8
BIC	85.0	6.3
CCV	83.0	9.8
LCH	69.0	7.3

search algorithm was used to determine the SVM parameters during the training stage.

In this round of experiments, we assume that I has been classified as fake by the classifier proposed in Section III. Therefore, we just apply the fake face detector over images classified as fake. Once all the faces have been classified, we analyze the probability for the fake class reported by the SVM classifier for each one of them. The face with the highest probability is pointed out as the most probable of being fake.

Table VI shows the detection accuracy using each one of the color descriptors as input descriptor fc for Equation 3, used to calculate descriptors of this round of experiments. The best detection accuracy is obtained when fc is extracted through the BIC descriptor.

5) Round #5 (Performance on DSI-1 Dataset): In this round of experiments, we repeat the setup proposed in the work by de Carvalho *et al.* [7]: we use DSO-1 as training set and DSI-1 as test set. In other words, we perform a cross-dataset validation in which we train our method with images from DSO-1 and test it against images from the Internet (DSI-1). This kind of validation is very useful in forensic scenarios since training and test images come from different sources captured or created under very different conditions (varied scene illumination, compression level, resolution, etc.).

All the parameters used in this round of experiments, such as k of kNN learning technique or the subset of image descriptors, for example, are the same as the ones used in Round #2. More specifically, for each fold, we used the same learning method model generated in Round #2.

As described in Round #2, we classified each one of the 54 ($\mathcal{C}$) classifiers from one image through a kNN-5 and we selected the best combination of them using the modified combination approach. We achieved an average classification accuracy of 83.6% (with an AUC of 91.89% and an operational point with Sensitivity of 75.2% and a Specificity of 92.0%) with a standard deviation of 5.0% using 20 classifiers. This result is around 8 percentage points better, in terms of accuracy, than the result reported in [7] for DSI-1 dataset (the authors report an AUC of 82.6% with the best operational point as 64.0% of Sensitivity and 88.0% of Specificity with a classification accuracy of 76.0%). Figure 11 depicts the two best performing methods, the proposed one and the work by de Carvalho *et al.* [7]. Table VII shows the results of all tested combinations of $|\mathcal{C}^*|$ on each testing fold, as well as their average and standard deviation.

As introduced in Round # 3, we also show a comparison between our results and Carvalho et al.'s results on the DSI-1 dataset as a bar graph (Figure 12).

6) Round #6 (Qualitative Analysis of Famous Cases Involving Questioned Images): In this round of experiments, we perform a qualitative analysis of famous cases involving questioned images. We use the previously trained classification

TABLE VII

ACCURACY COMPUTED THROUGH APPROACH DESCRIBED IN SECTION IV-B5 FOR 5-FOLD CROSS-VALIDATION PROTOCOL IN DIFFERENT NUMBER OF CLASSIFIERS ($|\mathcal{C}^*|$). ALL RESULTS ARE IN %

DSI-1 dataset											
Run	**Number of Classifiers $\|\mathcal{C}^*\|$**										
	5	**10**	**15**	20	**25**	**30**	**35**	**40**	**45**	**50**	**54 (ALL)**
1	88.0	90.0	82.0	92.0	90.0	88.0	86.0	84.0	84.0	84.0	84.0
2	80.0	76.0	78.0	80.0	80.0	80.0	84.0	86.0	88.0	88.0	86.0
3	62.0	80.0	82.0	82.0	82.0	86.0	84.0	78.0	82.0	80.0	80.0
4	76.0	78.0	80.0	80.0	78.0	68.0	72.0	74.0	72.0	78.0	74.0
5	70.0	82.0	78.0	84.0	88.0	84.0	84.0	86.0	84.0	90.0	90.0
Final (Avg)	75.2	81.2	80.0	83.6	83.6	81.2	82.0	81.6	82.0	84.0	82.8
Std. Dev.	9.9	5.4	2.0	5.0	5.2	7.9	5.7	5.4	6.0	5.1	6.1

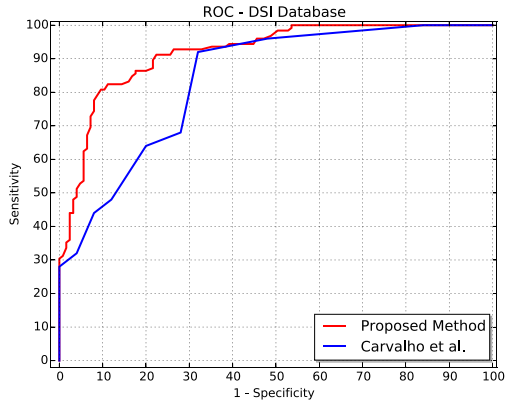
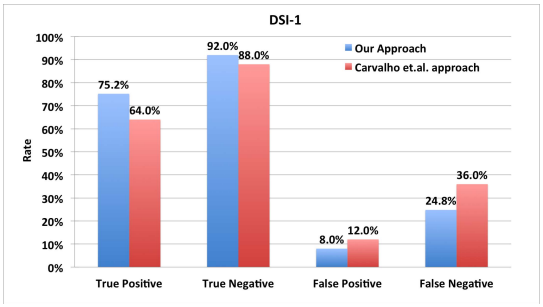


Fig. 11. ROC – Proposed method *vs*. de Carvalho *et al.* [7] for DSI-1 dataset.



DSI-1 dataset

Fig. 12. Comparison between our approach and the one proposed by Carvalho et al. over the DSI-1 dataset.

models of Section IV-B.2 in a cross-dataset evaluation protocol. We classify the suspicious image using the model built for each training set (fold) and the final class is calculated based on average confidence of all folds. Thus, given five folds, we obtain the maximum confidence between fake average confidences and pristine average confidences.

1) *Brazil's Former President:* On November 23, 2012 Brazilian Federal Police started an operation named "Safe Harbor", which dismantled an undercover gang on federal agencies for fraudulent technical advices. One of the gang's leaders was Rosemary Novoa de Noronha.[7] For some 15-minute fame, at the same time, people started to broadcast on the Internet, images purporting Brazil's former president Luiz Inácio Lula da Silva alongside Rosemary de Noronha, in apparently daily activities. Shortly after, another image in exactly the same scenario started to be broadcasted, however, this time, without de Noronha in the scene. We analyzed both images (see Figures 13(a-b)), using our proposed method. Figure 13(a) has been classified as pristine with 59.25% confidence, whereas Figure 13(b) has been classified as fake with 57.4% confidence.

(a) (b)

Fig. 13. Questioned images involving Brazil's former president. (a) Original image, taken by photographer Ricardo Stucker, and (b) The fake one, whereby Rosemary Novoa de Noronha's face (left side) is composed with the image. (a) Pristine. (b) Fake.

2) *The Situation Room Image:* Another recent forgery that quickly went viral on the Internet was based on an image depicting the Situation Room[8] when the Operation Neptune's Spear, a mission against Osama bin Laden, was taking place. The original image depicts U.S. President Barack Obama along with members of his national security team during the operation Neptune's Spear on May 1, 2011. Shortly after the release of the original image, several fake images depicting the same scene had been disseminated in the Internet. One of the most famous among them depicts Italian soccer player Mario Balotelli in the center of image. We analyzed both images, the original (officially broadcasted by the White House) and the fake one. Figures 14(a-b) show both images.

Even though the image containing the player Mario Balotelli has undergone several compression stages,[9] which could compromise the forgery detection

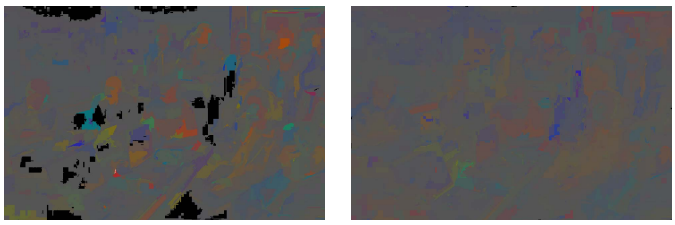
[7]Veja Magazine, *Operação Porto Seguro*, http://veja.abril.com.br/tema/operacao-porto-seguro (As of 2015-06-01).

[8]Original image from http://upload.wikimedia.org/wikipedia/commons/a/ac/Obama_and_Biden_await_updates_on_bin_Laden.jpg (As of Jun. 2015).

[9]de Carvalho *et al.* [7] have hinted that successive JPEG compressions may compromise the performance of IM estimation methods.

(a) (b)

Fig. 14. The situation room images. (a) depicts the original image released by American government; (b) depicts one among many fake images broadcasted in the Internet. (a) Pristine. (b) Fake.



(a) (b)

Fig. 15. IMs extracted from Figure 14(b). Successive JPEG compressions applied on the image make it almost impossible to detect a forgery by a visual analysis of IMs as proposed by Riess and Angelopoulou's method [6]. (a) IIC. (b) GGE.

method, using different characterization methods and exploring complementary features, our method classifies such image as fake with 56.49% confidence. The original one is tagged as pristine with 64.07% confidence.

Figures 15(a-b) depict IIC and GGE transformed maps, respectively, produced by the fake image containing the Italian player Mario Balotelli. Just performing a visual analysis on these transformed images, as proposed by Riess and Angelopoulou's method [6], is almost impossible to detect any pattern capable of indicating a forgery by itself. However, once that our method explores complementary statistical information on texture, shape and color, it was able to detect the forgery.

3) *Dimitri de Angeli's Case:* In March 2013, Dimitri de Angelis was found guilty and sentenced to serve 12 years in jail for swindling investors in more than 8 million dollars. To garner the investor's confidence, de Angelis produced several images, side by side with celebrities, using Adobe Photoshop. We analyzed one of de Angeli's case where the conman is alongside former U.S. president Bill Clinton.

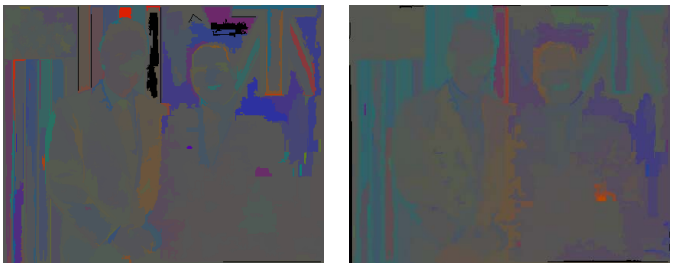
Unfortunately, in this case, the analyzed image was misclassified as pristine with a confidence of 74.81%. This happened because this image has a very low resolution and has undergone strong JPEG compression harming the IMs estimation, as depicted in Figures 17(a-b), which is the first step of our method.

V. Conclusions and Future Work

Image composition involving people is one of the most common tasks nowadays. The reasons vary from simple jokes with colleagues to harmful montages defaming or impersonating third parties. Independently of the reasons, it is paramount to design and deploy appropriate solutions to detect

Dimitri de Angeli's and Bill Clinton

Fig. 16. Dimitri de Angelis used Adobe Photoshop to falsify images side by side with celebrities.



(a) (b)

Fig. 17. IMs extracted from Figure 16. Successive JPEG compressions applied on the image, allied with a very low resolution, compromised the first step of the proposed method and leading our method to misclassify the image. (a) IIC. (b) GGE.

such activities. The complexity of such forgeries are also uphill. A few years ago, a montage involving people normally depicted a person innocently put side by side with another one. Nowadays, complex scenes involving politicians, celebrities and child pornography are in place.

Unfortunately, although technology is capable of helping us solve such problems, most of the available solutions still rely on experts' knowledge and background to perform well. Taking a different path, in this work, we investigated how to use multiple types of information to formulate an approach able to decrease user interaction and increase the classification accuracy on image splicing detection.

In our work, we analyze illuminant maps, as a possible image transformed space that capture, to some degree, the lighting information in a scene and that emphasize telltales left behind during the forgery process. To capture such properties, we explored image descriptors that analyze color, texture and shape cues. The color descriptors identify if similar parts of the object are colored in the IM in a similar way. The texture descriptors characterize the distribution of colors through IMs in a given region. Finally, shape descriptors encompass properties related to the object borders in such IMs. In a previous work, de Carvalho *et al.* [7] investigated only two descriptors when analyzing an image converted into an illuminant map. In this work, we presented an improved approach to detecting composites of people that explore complementary information for characterizing images. However, instead of just stockpiling a huge number of image descriptors, we need to effectively find the most appropriate ones for the task. For that, we proposed an automatic way of selecting and combining the

best image descriptors with their appropriate color spaces and IMs. The final classifier is fast and effective for determining whether an image is real or fake.

This work also introduced two important contributions for the forensic community. First, a confidence metric associated with each classified image. Second, we proposed a method for effectively pointing out the region of an image that was forged. For a fair validation scenario, we considered three different benchmarks and strict validation protocols, including a cross-dataset one. The automatic forgery classification, in addition to the actual forgery localization, represent an invaluable asset for forensic analysts with a 94% classification rate, in the best scenario, a remarkable 72% error reduction when compared to the state-of-the-art method proposed by de Carvalho *et al.* [7].

Finally, note that although our method employs illuminant maps, it is not known whether differences in the physical illuminant color are being detected. To reach that conclusion, images would need to be collected in which the lighting color and environment are controlled. The detector would then be tested using pairs of images in which the illuminant color differs by a known amount. Such experiments are complex, and we leave them for future work.

Future developments of this work, in addition to the design of controlled experiments for quantitatively measuring the contribution of the illuminant differences in faces of different people in a composition, may include the investigation of how lighting changes affect the entire framework. Given that our method compares skin material, it is feasible to investigate how different skin tones influence the method.

REFERENCES

[1] E. Kee, J. F. O'brien, and H. Farid, "Exposing photo manipulation with inconsistent shadows," *ACM Trans. Graph.*, vol. 32, no. 3, pp. 28:1–28:12, Jul. 2013.

[2] E. Kee and H. Farid, "Exposing digital forgeries from 3-D lighting environments," in *Proc. IEEE Int. Workshop Inf. Forensics Secur.*, Dec. 2010, pp. 1–6.

[3] M. K. Johnson and H. Farid, "Exposing digital forgeries by detecting inconsistencies in lighting," in *Proc. ACM 7th Workshop Multimedia Secur.*, New York, NY, USA, 2005, pp. 1–10.

[4] P. Saboia, T. Carvalho, and A. Rocha, "Eye specular highlights telltales for digital forensics: A machine learning approach," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 1937–1940.

[5] A. Pinto, W. Robson Schwartz, H. Pedrini, and A. De Rezende Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.

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

[7] T. J. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. Rocha, "Exposing digital image forgeries by illumination color classification," *IEEE Trans. Inf. Forensics Security*, vol. 8, no. 7, pp. 1182–1194, Jul. 2013.

[8] W. Fan, K. Wang, F. Cayre, and Z. Xiong, "3D lighting-based image forgery detection using shape-from-shading," in *Proc. 20th Eur. Signal Process. Conf.*, Aug. 2012, pp. 1777–1781.

[9] S. Gholap and P. K. Bora, "Illuminant colour based image forensics," in *Proc. IEEE Region 10 Conf.*, Nov. 2008, pp. 1–5.

[10] S. Tominaga and B. A. Wandell, "Standard surface-reflectance model and illuminant estimation," *J. Opt. Soc. Amer. A*, vol. 6, no. 4, pp. 576–584, Apr. 1989.

[11] K. Francis, S. Gholap, and P. K. Bora, "Illuminant colour based image forensics using mismatch in human skin highlights," in *Proc. Nat. Conf. Commun.*, Feb./Mar. 2014, pp. 1–6.

[12] X. Wu and Z. Fang, "Image splicing detection using illuminant color inconsistency," in *Proc. 3rd Int. Conf. Multimedia Inf. Netw. Secur.*, Nov. 2011, pp. 600–603.

[13] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Amer. A*, vol. 21, no. 3, pp. 321–334, 2004.

[14] 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.

[15] A. Gijsenij and T. Gevers, "Color constancy using natural image statistics and scene semantics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 687–698, Apr. 2011.

[16] P. A. S. Kimura, J. M. B. Cavalcanti, P. C. Saraiva, R. da Silva Torres, and M. A. Gonçalves, "Evaluating retrieval effectiveness of descriptors for searching in large image databases," *J. Inf. Data Manage.*, vol. 2, no. 3, pp. 305–320, 2011.

[17] A. Rocha, T. Carvalho, H. F. Jelinek, S. 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.

[18] J. A. dos Santos, P.-H. Gosselin, S. Philipp-Foliguet, R. da Silva Torres, and A. Xavier Falcao, "Interactive multiscale classification of high-resolution remote sensing images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 4, pp. 2020–2034, Aug. 2013.

[19] R. T. Tan, K. Nishino, and K. Ikeuchi, "Color constancy through inverse-intensity chromaticity space," *J. Opt. Soc. Amer. A*, vol. 21, no. 3, pp. 321–334, 2004.

[20] O. A. B. Penatti, E. Valle, and R. da Silva Torres, "Comparative study of global color and texture descriptors for Web image retrieval," *J. Vis. Commun. Image Represent.*, vol. 23, no. 2, pp. 359–380, 2012.

[21] M. H. Asmare, V. S. Asirvadam, and L. Iznita, "Color space selection for color image enhancement applications," in *Proc. Int. Conf. Signal Acquisition Process.*, Apr. 2009, pp. 208–212.

[22] A. Çarkacıoğlu and F. T. Yarman-Vural, "SASI: A generic texture descriptor for image retrieval," *Pattern Recognit.*, vol. 36, no. 11, pp. 2615–2633, 2003.

[23] B. Tao and B. W. Dickinson, "Texture recognition and image retrieval using gradient indexing," *J. Vis. Commun. Image Represent.*, vol. 11, no. 3, pp. 327–342, 2000.

[24] M. Unser, "Sum and difference histograms for texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 1, pp. 118–125, Jan. 1986.

[25] F. Mahmoudi, J. Shanbehzadeh, A.-M. Eftekhari-Moghadam, and H. Soltanian-Zadeh, "Image retrieval based on shape similarity by edge orientation autocorrelogram," *Pattern Recognit.*, vol. 36, no. 8, pp. 1725–1736, 2003.

[26] D.-H. Lee and H.-J. Kim, "A fast content-based indexing and retrieval technique by the shape information in large image database," *J. Syst. Softw.*, vol. 56, no. 2, pp. 165–182, Mar. 2001.

[27] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1997, pp. 762–768.

[28] R. O. Stehling, M. A. Nascimento, and A. Falcão, "A compact and efficient image retrieval approach based on border/interior pixel classification," in *Proc. ACM 11th Int. Conf. Inf. Knowl. Manage.*, 2002, pp. 102–109.

[29] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," in *Proc. 4th ACM Int. Conf. Multimedia*, 1996, pp. 65–73.

[30] M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vis.*, vol. 7, no. 1, pp. 11–32, 1991.

[31] F. A. Faria, J. A. dos Santos, A. Rocha, and R. da Silva Torres, "A framework for selection and fusion of pattern classifiers in multimedia recognition," *Pattern Recognit. Lett.*, vol. 39, no. 4, pp. 52–64, Apr. 2014.

[32] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Mach. Learn.*, vol. 51, no. 2, pp. 181–207, 2003.

[33] C. Bishop, *Pattern Recognition and Machine Learning* (Information Science and Statistics). Secaucus, NJ, USA: Springer-Verlag, 2006.

Tiago Carvalho received the B.Sc. (computer science) degree from the Federal University of Juiz de Fora, Brazil, in 2008, and the M.Sc. and Ph.D. (computer science) degrees from the University of Campinas, Brazil, in 2010 and 2014, respectively. As part of his Ph.D., he developed digital forensics methods for detecting image splicing detection. He is currently an Assistant Professor with the So Paulo Federal Institute of Education, Science and Technology Technologist. He has served as a Member of the Technical Committee and Reviewer for several conferences and journals. His main interests include digital image forensics, image processing, pattern analysis, machine learning, general computer vision, and problems related with environment weather monitoring. In 2011, he was awarded the Best Image Processing Master's Dissertation from SIBGRAPI'2011 for his work on diabetic retinopathy anomalies detection in eye funds images. In 2014, he received the Best Ph.D. Thesis from SIBGRAPI'2014 for his digital forensics research. Finally, in 2015, his Ph.D. Thesis was given second place by the Brazilian Computer Society Award.

Fábio A. Faria received the B.Sc. degree in computer science from So Paulo State University, in 2007, and the M.Sc. and Ph.D. degrees in computer science from the Institute of Computing, University of Campinas, Brazil, in 2010 and 2014, respectively. He is currently an Assistant Professor with the Institute of Science and Technology, Federal University of São Paulo. His research interests include machine learning, image processing, information fusion, and data mining.

Hélio Pedrini (SM'15) received the B.Sc. degree in computer science and the M.Sc. degree in electrical engineering from the University of Campinas, Brazil, and the Ph.D. degree in electrical and computer engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently a Professor with the Institute of Computing, University of Campinas. His research interests include image processing, computer vision, pattern recognition, machine learning, computer graphics, and computational geometry. He is a member of the Brazilian Computer Society. He has served as a member of technical committees and reviewer for several conferences and journals.

Ricardo da S. Torres received the B.Sc. degree in computer engineering and the Ph.D. degree in computer science from the University of Campinas (Unicamp), Brazil, in 2000 and 2004, respectively. He has been the Director of the Institute of Computing, Unicamp, since 2013, where he is currently a Full Professor of Computer Science. He is the Cofounder and a member of the RECOD Laboratory, where he has been developing multidisciplinary research involving image analysis, content-based image retrieval, databases, digital libraries, and geographic information systems. He has authored or coauthored over 100 articles in refereed journal and conferences and serves as a PC member for several international and national conferences.

Anderson Rocha received the B.Sc. (computer science) degree from the Federal University of Lavras, Brazil, in 2003, and the M.S. and Ph.D. (computer science) degrees from the University of Campinas (Unicamp), Brazil, in 2006 and 2009, respectively. He is currently an Associate Professor with the Institute of Computing, Unicamp. His main interests include digital forensics, reasoning for complex data, and machine intelligence.