UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di Ingegneria

Corso di Laurea in Ingegneria Informatica

---

# Illuminant map analysis for image splicing detection


*Candidato*

Lorenzo Cioni

*Relatori*

Prof. Alessandro Piva

Prof. Carlo Colombo

*Correlatori*

Dott. Massimo Iuliani

Dott. Marco Fanfani


---


Anno Accademico 2015/2016

*Dedica*

# Contents

# Abstract

Da scrivere

Abstract in italiano

# Introduction

With the rapid development of multimedia and network technology, digital images, as an effective carrier of information, are having a more and more important impact on people's daily lives.

However, the image content can easily be tampered with the increasingly powerful image processing software, which threatens the integrity and authenticity of digital images. Moreover, if the tampered images were used for illegal purposes, there would be no doubt causing extremely bad effects on both the individual and society.

Images manipulated with the purpose of manipulating and deceiving user opinions are present in almost all communication channels including newspapers, magazines, TV shows and, obviously, on Internet. [17]

Image splicing is a fundamental operation used in digital image tampering. It is a copy-and-paste operation of image regions from one image onto the same or another image without performing post-processing such as smoothing [13]. Even without post-processing, the artifacts introduced by image splicing may be almost imperceptible. Therefore, the detection of image splicing is a preliminary but desirable study for image forensics.

Investigating image's lighting is one of the most common approaches for splicing detection. This kind of approach is particularly robust since it's really hard to preserve the consistency of the lighting environment while creating an image composite (i.e. a splicing forgery).

In this scenario, there are mainly two approaches:

1. based on the object-light geometric arrangement

2. based on the image illuminant colors

We focused our attention on the illuminant-based approach, which assumes that a scene is lit by the same light source. More light sources are admitted but only far enough such as to produce a constant brightness across the image. In this condition, pristine images will show a coherent illuminant representation; on the other hand, inconsistencies among illuminant maps will be exploited for splicing detection.

*Illuminant maps* locally describes the lighting in a small region of the image. In the computer vision literature exist many different approaches for determining the illuminant of an image. In particular, such techniques are divided into two main groups: statistical-based and physics-based approaches.

Regarding the first group, we start investigating on the *Grey-World algorithm* [1], which is based on the Grey-World assumption, i.e. the average reflectance in a scene is achromatic. In [6], this algorithm proved to be special instances of the Minkowski-norm. Van de Weijer et al. [21] than proposed an extension of the Gray-World assumption, called *Gray-Edge hypothesis* [21], which assumes that the average of the reflectance differences in a scene is achromatic.

These differences can be determined by taking derivatives of the image. Therefore, the authors present a framework with which many different algorithms can be constructed. We focus our attention on this framework, called *Generalized Grey-World algorithm* (GGE).

For the latter group, it was investigated the method proposed by Riess et al. [16], which extends the *Inverse Intense Chromaticity* (*IIC*) space approach proposed by Tan et al. [19] and tries to model the illuminants considering the dichromatic reflection model [20]. In this case, the illuminant map is evaluated dividing the images into blocks, named *superpixels*, of approximately the same object color, then the illuminant color is evaluated for each block solving the lighting models locally.

Carvalho et al. [2] then presented a method that relies on a combination of the two approaches for the detection of manipulations on images containing

human faces. In addition to maps, a large set of shape and texture descriptors are used together. Note that, from a theoretical viewpoint, it is advantageous to consider only image regions that consist of approximately the same underlying material: for this reason, in [2] the authors focused their analysis on human faces.

This work is structured as follows. Chapter 2 describes the main methods grounded on illumination inconsistencies for detecting image composition and the background theories on which this work is based on. In Chapter 3 our contributions for detecting image splicing is presented. Finally, Chapter 4 collects the experimental results, putting our research in perspective and discussing new research opportunities.

# Chapter 1

# Related work

## 1.1 Image forgery

When dealing with a digital image, it is quite common to wonder if it is original or it has been counterfeited in some way. Images and videos have become the main information carriers in the digital era and they are used to store real world events, but they are very easy to manipulate because of the availability of powerful editing softwares and sophisticated digital cameras.

The contexts where doctored pictures could be involved are very disparate; they could be used in a tabloid or in an advertising poster or included in a journalistic report but also in a court of law where digital (sometimes printed) images are presented as crucial evidences for a trial in order to influence the final judgement. So, especially in the last case, a reliable assessment of image integrity becomes of fundamental importance [26] [4].

*Image forensics* specifically deals with such issues by studying and developing technological tools which generally permit determining, by only analyzing a digital photograph, if that asset has been manipulated or even which could have been the adopted acquisition device [5]. Moreover, if it has been determined that something has been altered, it could be important to understand in which part of the image itself such modification occurred: for instance, if a person or a specific object has been covered, if an area of the image has been cloned, if something (i.e., a face, an object) has been copied from another dif-

ferent image, or, even more, if a mixture of these processes has been carried out.

## 1.2 Image forgery detection techniques

Forgery detection intends to verify the authenticity of images.

Generally speaking, there are two approaches of image forgery detection: *active* and *passive* detection.

The active approaches are mostly concerned with the data hiding techniques, such as *digital watermarks* [15] [24] and digital signatures, wherein prior information is considered essential and integral to the process. Data hiding approaches embed some secondary data into the cover images. Usually, the watermarks are either embedded at the time of the image acquisition through specially equipped devices or later after further processing of the actual image. However, this kind of approach may degrade the quality of the original image [22].

Passive approaches [12] do not require any image pre-processing and do not alter the image content.

These techniques provide a solution to identify image alterations without relying on the insertion of an extrinsic data or digital signatures for the image authentication. Hence, they gain more attention and become a hot research topic in image forensics.

Over the past few years, many techniques have been introduced to tamper images or videos. These techniques can be classified into following three general categories:

- *Copy/move forgery*
- *Image retouching*
- *Image splicing*

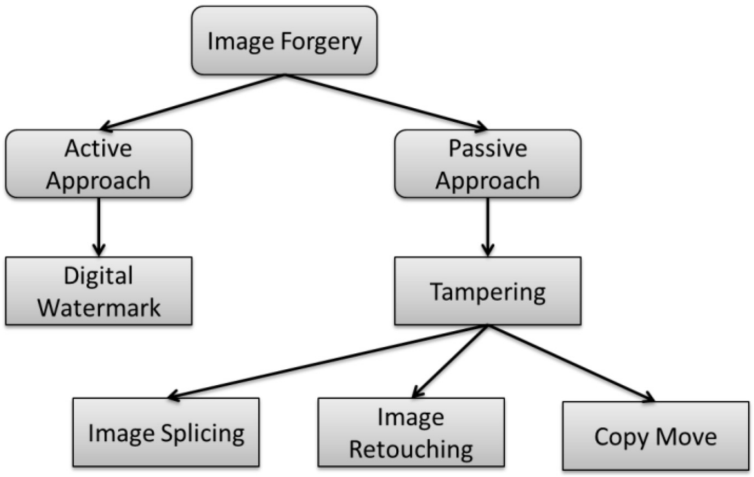


Figure 1.1: Image forgery detection approaches classification

Copy/move forgery is one of the most popular forms of tampering in which some regions are copied from a particular location in an image and thereafter pasted at one or more locations within the same image or a different image of preferably the same scene. Roughly speaking, the detection methods are brute force, involving exhaustive search or block based. These techniques usually rely on the correlation between the original patch and the suspected pasted version.

Image retouching is another class of forensic methods that pertains to a slight change in the image for various aesthetic and commercial purposes, not necessarily conforming to the standards of morality. The retouching is mostly used to enhance or reduce the image features. Usually this type of forgery is realised by changing the colour or texture of the objects, intensify the weather conditions or simply introducing some blur for defusing the objects. Forgery detection, in case of image retouching, involves finding the enhancements, blurring, illumination and colour changing.

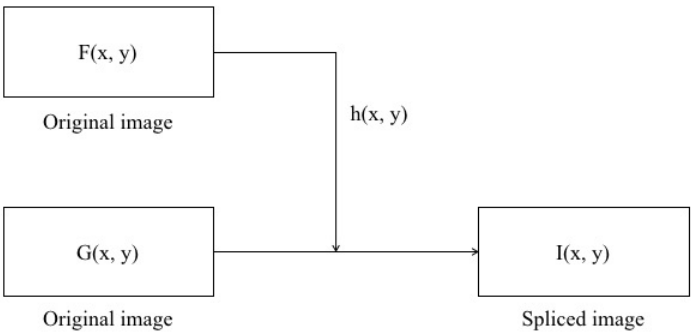


Figure 1.2: Image splicing process

## 1.3 Image splicing

A common form of photographic manipulation is the digital splicing of two or more images into a single composite [14]. Image splicing techniques significantly change the original images. When performed carefully, the border between the spliced regions can be visually imperceptible.

This composition process includes all the necessary operations (such as brightness and contrast adjustment, affine transformations, color changes, etc.) to construct realistic images able to deceive viewer. In this process, normally, we refer to the parts coming from other images as *aliens* and to the image receiving the other parts as *host*.

One common studied case is image composition involving very popular people, which can be employed with many different aims.

### *1.3.1 Some famous cases*

Photography has lost its innocence since the early days of its birth. Image forgery can be traced back to as early as 1840s when Hippolyte Bayrad created the very first fake image **??**, in which he was shown committing a suicide, only a few decades after Niépce created the first photo.

Figure 1.3: The first tampered image

**O.J. Simpson - June 1994** This altered photo of O.J. Simpson appeared on the magazine's cover Time Magazine, soon after his arrest for murder.

In fact, the photograph was altered compared to the original image that appeared on the cover of Newsweek magazine. Time magazine was accused of manipulation of the photography in order to make darker and more menacing the figure of Simpson.

**Iraq - April 2003** This composition of a British soldier in Basra, who keeps pointing toward a civilian Iraqi gesticulating, appeared on the cover of the Los Angeles Times, immediately after the Iraq invasion.


Figure 1.4: The Time Magazine and O.J. Simpson

Brian Walski, a staff photographer for the Los Angeles Times was summarily fired from his publisher because he merged two of his shots in order to improve the composition.

Figure 1.5: An example of image composition

**George W. Bush - March 2004** This image, taken from the election campaign of George W. Bush, outlined a packed audience of soldiers as a backdrop to a child who was flying the American flag. This image was digitally edited, using a crude copy and paste, removing Bush from the podium.

Cases like these show how present image composition is in our daily lives. Unfortunately, it also decreases our trust on images and highlights the need for developing methods for recovering back such confidence.

Figure 1.6: An example of image composition

## 1.4 Methods based on light inconsistencies

Methods for detecting image composition have become actual and powerful tools in the forensic analysis process. Different types of methods have been proposed for detecting image composition. Methods based on inconsistencies in compatibility metrics, JPEG compression features and perspective constraints are just a few examples of inconsistencies explored to detect forgeries.

After studying and analyzing the advantages and drawbacks of different types of methods for detecting image composition, this work herein relies on the research hypothesis that image illumination inconsistencies are strong and powerful evidence of image composition.

This hypothesis has already been used by some researchers in the literature whose work will be detailed in the next sections, and it is specially useful for detecting image composition because, even for expert counterfeiters, a perfect illumination match is extremely hard to achieve. Furthermore, there are some experiments that show how difficult is for humans to perceive image illumination inconsistencies.

We can divide methods that explore illumination inconsistencies into three main groups:

1. methods based on inconsistencies in the **light setting**: this group encloses the approaches that look for inconsistencies in the light position and in models that aim at reconstructing the scene illumination conditions.

2. methods based on inconsistencies in the **shadows**: this group encloses the

approaches that look for inconsistencies in the scene illumination using telltales derived from shadows.

3. methods based on inconsistencies in **light color**: this group of methods encloses the approaches that look for inconsistencies in the color of illuminants present in the scene.

### *1.4.1 Inconsistencies in the light setting*

Johnson and Farid [10] proposed an approach based on illumination inconsistencies, looking for chromaticity aberrations as an indicator of image forgery. They analyzed the light source direction on different objects in the same image trying to detect traces of tampering. The authors start by imposing different constraints for the problem:

1. All the analyzed objects have *Lambertian surface*[1] [11].

2. The surface reflectance is constant.

3. The object surface is illuminated by an infinitely distant light source.

Using RGB images, the authors assume that the chromaticity deviation is constant (and dependent on each channel wavelength) for all color channels and they create a model, based on image statistical properties, to provide how the ray light should split for each color channel. Given this premise and using the green channel as reference, the authors estimate deviations between the red and green channels and between the blue and green channels for selected parts of the image. Inconsistencies on this split pattern are used as telltales to detect forgeries.

[1] A Lambertian surface for reflection is a surface that appears uniformly bright from all directions of view and reflects the entire incident light. Lambertian reflectance is the property exhibited by an ideal matte or diffusely reflecting surface.

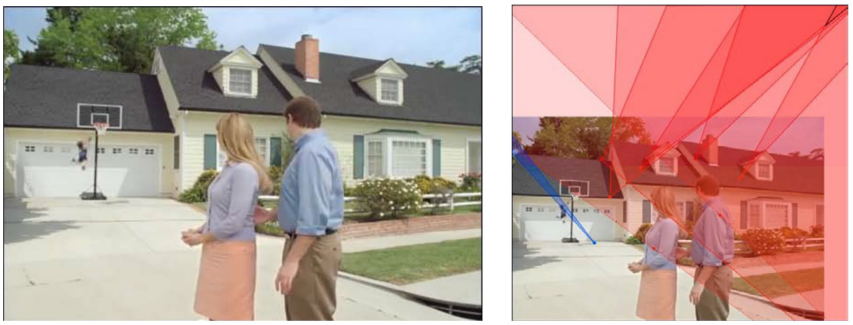

Figure 1.7: Original image (left) and the extracted shadows constraints (right)

A drawback of this method is that chromaticity deviation depends on the camera lens used to take the picture.

### *1.4.2 Inconsistencies in the shadows*

Another set of methods is based on inconsistencies on the shadows in the image.

Zhang and Wang [25] proposed an approach that utilizes the planar homology [18], which models the relationship of shadows in an image for discovering forgeries.

Based on this model, the authors proposed to construct two geometric constraints. The first one is based on the relationship of connecting lines. A connecting line is a line that connects some object point with its shadow. According to planar homology, all of these connecting lines intersect in a vanishing point.

The second constraint is based on the ratio of these connecting lines. In addition, the authors also proposed to explore the changing ratio along the normal direction of the shadow boundaries.

Geometric and shadow photometric constraints together are used to detect image compositions. However, in spite of being a good initial step in forensic shadow analysis, the major drawback of the method is that it only works with images containing casting shadows, a very restricted scenario.

### *1.4.3 Inconsistencies in light color*

The last group of methods investigate the presence, or not, of composition operations in digital images using color inconsistencies.

Gholap and Bora [9] [7] pioneered this approach using the *illuminant colors*. For that, the authors used a *dichromatic reflection model* proposed by Tominaga and Wandell [20], which assumes a single light source to estimate illuminant colors from images.

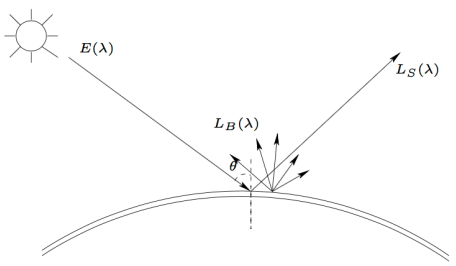


Figure 1.8: The dichromatic reflection model

According to this model, the reflection of any non-homogeneous material may be modelled as additive mixture of two components, diffused reflection and surface reflection as shown in Fig. **??**.

Considering an object point illuminated by a light source, the reflected ray consists of diffuse reflection $L_B(\lambda)$ and surface reflection $L_S(\lambda)$. The reflected light $L(\Theta, \lambda)$ can be written as:

$$L(\Theta, \lambda) = m_S(\Theta)L_S(\lambda) + m_B(\Theta)L_B(\lambda) \tag{1.1}$$

where $m_S(\Theta)$ and $m_B(\Theta)$ are geometrical factors and $\Theta$ the angle of the incident light. This equation can be rewritten in terms of RGB sensors in matrix form.

The two vectors $L_B(\lambda)$ and $L_S(\lambda)$ span the two dimensional plane called *dichromatic plane*.

Dichromatic planes can also be estimated using principal component analysis (PCA) from each specular highlight region of an image. By applying a *Singular Value Decomposition (SVD)* on the RGB matrix extracted from high-

lighted regions, the authors extract the eigenvectors associated with the two most significant eigenvalues to construct the *dichromatic plane*. This plane is then mapped onto a straight line, named dichromatic line, in normalized *r-g-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, which is not expected to happen in pristine (non-forged images). This method represented the first important step toward forgery detection using illuminant colors, but has some limitations such as the need of well defined specular highlight regions for estimating the illuminants.

Following Gholap and Bora's work, Riess and Angelopoulou [16] used an extension of the *Inverse-Intensity Chromaticity Space*, originally proposed by Tan et al. [19], to estimate illuminants locally from different parts of an image for detecting forgeries.

In addition, Wu and Fang [23] 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. To estimate the illuminant color, the authors proposed to use the algorithms Gray-World, Gray-Shadow and Gray-Edge [21].

These approaches will be better explained in the next section due their importance for this proposed method.

## 1.5 Illuminant color estimation

The color of an object observed in an image depends both on its intrinsic color and on the color of light source i.e., the illuminant.

It is important to keep in mind that even the same light source can generate different illuminants. The illuminant formed by the sun, for example, varies in

its appearance during the day and time of year as well as with the weather. We only capture the same illuminant, measuring the sunlight at the same place at the same time.

We can explore illuminants in forensics to check the consistency of similar objects in a scene. If two objects with very similar color stimuli (e.g., human skin) depict inconsistent appearance (different illuminants), it means thatthey might have undergone different illumination conditions hinting at a possible image composition. On the other hand, if we consider a photograph with two people and the color appearance on the faces of such people are consistent, it is likely they have undergone similar lighting conditions.

Riess and Angelopoulou [16] pioneered the approach of using a color-based method that investigates illuminant colors to detect forgeries in forensic scenario. In their work illuminant colors are estimated locally, effectively decomposing the scene in a map of differently illuminated regions. Inconsistencies in such a map suggest possible image tampering.

The method can be divided into four steps:

1. Image segmentation. The image is segmented into regions, called *superpixels*, of approximately the same object color. Each superpixel must be:
   (a) Directly illuminated by the same light source.
   (b) Compliant with the *dichromatic reflectance model* [19].
2. Manual user selection of superpixels under investigation.
3. Local illuminant color estimation for each superpixel (both for segmented superpixels and user selected regions).
4. Reference illuminant selection (manual) and *distance map* evaluation. The distance map is the base for an expert analysis for forgery detection.

The final decision of this approach is delegated to an expert and not to the algorithm itself.

Figure 1.9: An example of a generated distance map using Riess and Angelopoulou [16] approach. From left to right: (a) the original image, (b) *illuminant map*, (c) generated distance map

### *1.5.1 Illuminant maps*

#### 1.5.1.1 Generalized Greyworld estimation

The starting point is the *gray world assumption*, proposed by Buchsbaum [1]. In its simplest version, it is assumed that information in the average of each channel of the image is the representative gray level.

Let $f(x) = (\Gamma_R(x), \Gamma_G(x), \Gamma_B(x))^T$ the RGB color of a pixel at position $x$ and $\Gamma_i(x)$ the intensity of that pixel in the $i$-th channel. In their paper, Van de Weijer et al. assumes that the diffuse reflection is diffuse and that the camera response is linear. It is also assumed that the scene is illuminated by a single light source.

The RGB color $f(x)$ can also be rewritten as

$$f(x) = \int_{\omega} e(\beta, x)s(\beta, x)c(\beta)d\beta \tag{1.2}$$

where $\omega$ is the visible light spectrum, $\beta$ is the light wavelenght, $e(\beta, x)$ is the illuminant spectrum, $s(\beta, x)$ is the surface reflectance of an object and $c(\beta)$ is the color sensitivities of the camera for each channel.

As an alternative to the gray-world hypothesis, Van de Weijer et al. [21] proposed the *gray-edge hypothesis*: the average of the reflectance differences in a scene is achromatic.

This idea led to a framework for low-level based illuminant estimation, called

*Generalized Grayworld.*

$$\left( \int \left| \frac{\delta^n f^\sigma(x)}{\delta x^n} \right|^p dx \right)^{\frac{1}{p}} = k e^{n,p,\sigma} \tag{1.3}$$

where $k$ denotes a scaling factor, $|\cdot|$ the norm operand, $\delta$ the differential operator and $f^\sigma(x)$ the pixel intensities at position $x$, smoothed with a Gaussian kernel $\sigma$.

This framework 1.3 produces different estimations for the illuminant color based on three parameters:

1. The order $n$ determines if the method is a gray-world or a gray-edge algorithm. The gray-world methods are based on the RGB values, whereas the gray-edge methods are based on the spatial derivative of order $n$.

2. The Minkowski norm $p$ determines the relative weights of the multiple measurements from which the final illuminant color is estimated.

3. The scale $\sigma$ of the local measurements. For first or higher order estimation, this local scale is combined with the differentiation operation computed with the Gaussian derivative. For zero-order gray-world methods, this local scale is imposed by a Gaussian smoothing operation.

On varying of these parameter, a different method can be used. An overview of the possible methods that derive from Eq. 1.3 is displayed in Table 1.1.

#### 1.5.1.2 Inverse Intensity-Chromaticity estimation

The other considered illuminant estimation method is based on the idea proposed by Tan et al. [19], called *inverse intensity-chromaticity*.

The base for these kind of approaches is the dichromatic reflectance model [9], which states that the amount of light reflected from a point, $x$, of a dielectric, non-uniform material is a linear combination of diffuse reflection and

Table 1.1: Overview of different illuminant estimation methods based on 1.3

| Name | Symbol | Equation | Assumption |
|---|---|---|---|
| Gray-World | $e^{0,1,0}$ | $(\int f(x)dx) = ke$ | The average reflectance in a scene is achromatic |
| max-RGB | $e^{0,\infty,0}$ | $(\int \lvert f(x)\rvert^{\infty}dx)^{\frac{1}{\infty}} = ke$ | The maximum reflectance in a scene is achromatic |
| Shades of Gray | $e^{0,p,0}$ | $(\int \lvert f(x)\rvert^{p}dx)^{\frac{1}{p}} = ke$ | The $p$-th Minkowski norm of scene is achromatic |
| General Gray-World | $e^{0,p,\sigma}$ | $(\int \lvert f^{\sigma}(x)\rvert^{p}dx)^{\frac{1}{p}} = ke$ | The $p$-th Minkowski norm of scene is achromatic after smoothing |
| Gray-Edge | $e^{1,p,\sigma}$ | $(\int \lvert f^{\sigma}_{x}(x)\rvert^{p}dx)^{\frac{1}{p}} = ke$ | The $p$-th Minkowski norm of the image derivative is achromatic |
| Max-Edge | $e^{1,\infty,\sigma}$ | $(\int \lvert f^{\sigma}_{x}(x)\rvert^{\infty}dx)^{\frac{1}{\infty}} = ke$ | The maximum reflectance difference in a scene is achromatic |
| 2nd order Gray-Edge | $e^{2,p,\sigma}$ | $(\int \lvert f^{\sigma}_{xx}(x)\rvert^{p}dx)^{\frac{1}{p}} = ke$ | The $p$-th Minkowski norm of the second order derivative in a scene is achromatic |

specular reflection. Further assumptions assume that the color of the specularities approximates the color of the illuminant, and that the camera response is linear.

Considering a trichromatic camera, the sensor response $I_c(x)$, for each color channel $c \in \{R, G, B\}$ is:

$$I_c(x) = m_S(x)L_S(x) + m_B(x)L_B(x) \tag{1.4}$$

as described in 1.1.

Let $\sigma_c$ the image chromaticity, $\Delta_c(x)$ the diffuse chromaticity and $\Gamma_c(x)$ the specular chromaticity defined as follows:

$$\sigma_c(x) = \frac{I_c(x)}{\sum_i I_i(x)} \text{ where } i \in \{R, G, B\} \tag{1.5}$$

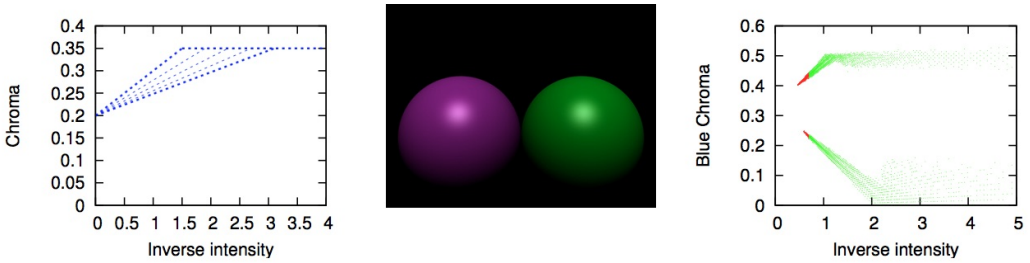


Figure 1.10: Pixel distribution in inverse-intensity chromaticity (IIC) space. Left are as ideal distribution, left on a synthetic image (in the center).

$$\Delta_c(x) = \frac{L_{S,c}(x)}{\sum_i L_{S,i}(x)} \text{ where } i \in \{R, G, B\} \tag{1.6}$$

$$\Gamma_c(x) = \frac{L_{B,c}(x)}{\sum_i L_{B,i}(x)} \text{ where } i \in \{R, G, B\} \tag{1.7}$$

Thus the Eq. 1.4 can be rewritten as

$$I_c(x) = m_S(x)\Delta_c(x) + m_B(x)\Gamma_c(x) \tag{1.8}$$

Tan et al. [19] derived a linear relationship between diffuse, specular and image chromaticities:

$$\sigma_c(x) = m(x)\frac{1}{\sum_i I_i(x)} + \Gamma_c(x) \tag{1.9}$$

where $i \in \{R, G, B\}$ and $m(x)$ is a geometrical factor (light position, surface orientation, camera position, ...), that can be approximated. In the illuminant estimation, the most important aspect is the $y$-intercept $\Gamma_c$.

The 2D space defined by $\frac{1}{\sum_i I_i(x)}$ as domain and $0 \leq \sigma_c \leq 1$ as range is called *inverse-intensity chromaticity (IIC)* space.

An example of the IIC plots for a single channel of a synthetic image are shown in Fig. 1.10. Pixels from the green and purple balls form two clusters. The clusters have spikes that point towards the same location on 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.

Instead of examining the entire pixel distribution, Riess et al. [16] perform the analysis over small connected image regions of roughly uniform object color (*superpixels*). Depending on the outcome of our shape analysis, we can either use this local region to obtain an illuminant estimate, or reject it if it does not seem to fulfill the underlying assumptions of the proposed model. Using local regions allows us to incorporate multiple sampling and voting in the estimation of local illuminants.

## 1.6 Human faces splicing detection

The approach proposed in Chapter 2 is based on two different works, published by Carvalho et al. [2] and Fan et al. [3].

The method proposed by Carvalho et al. [2] aims to detect splicing focusing on human faces, minimizing the user interaction. Faces are previously labeled by a human selecting the box within is contained, than the process will associate a label to each face (i.e. *normal* or *fake*).

The method can be divided into four steps:

1. *Description*: in this step illuminant maps are estimated with the two different approaches, GGE and IIC, and feature vector are generated. Each feature vector is associated with a face pair.

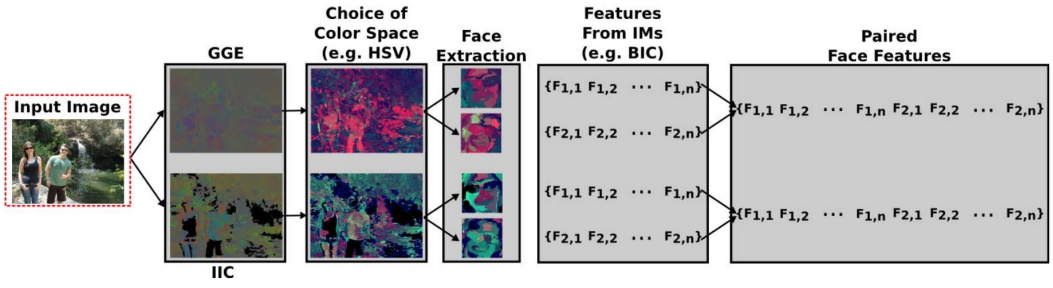


Figure 1.11: Image description pipeline for extracting paired face feature vectors

The Fig. 1.11 shows the image description extraction pipeline. Given an input image, the illuminant maps are estimated and converted to a selected color space (e.g. HSV). So, the faces in the image are extracted (using the defined area). For each face a descriptor is used to generate a feature vector (e.g a color descriptor, texture descriptor or shape descriptors) and finally they are coupled in order to generate paired feature vector simply concatenating the two original vectors.

In this step multiple descriptors and color spaces are used in order to increment the number of final classifiers.

2. *Face Pair Classification*: a set of classification models are trained using the previous step feature vectors. Based on the number of color spaces and descriptors used, a set of KNN classifiers are trained over the paired face feature vectors. The final result is given by a majority voting of all the selected classifiers.

3. *Forgery Classification*: given an image $I$ containing $q$ people (faces), it is characterized by a set $\mathcal{S} = \{\mathcal{P}_1, \ldots, \mathcal{P}_m\}$, where $\mathcal{P}_i$ is the $i$-th paired feature vector and $m = \frac{q(q-1)}{2}, q \geq 2$. If any $\mathcal{P}_i \in \mathcal{S}$ is classified as fake, the image $I$ is classified as fake. Otherwise, the image is considered as pristine.

4. *Forgery Detection*: once knowing that an image is fake, in this stage it is identified which one is more likely to be fake in the image. This is done using a specific SVM classifier.

The Fig. 1.12 shows the above descripted method pipeline.

### *1.6.1 Method drawbacks*

The main drawback of this approach relies in the manual selection of faces by a human agent. The definition of a face in the image by the operator is made

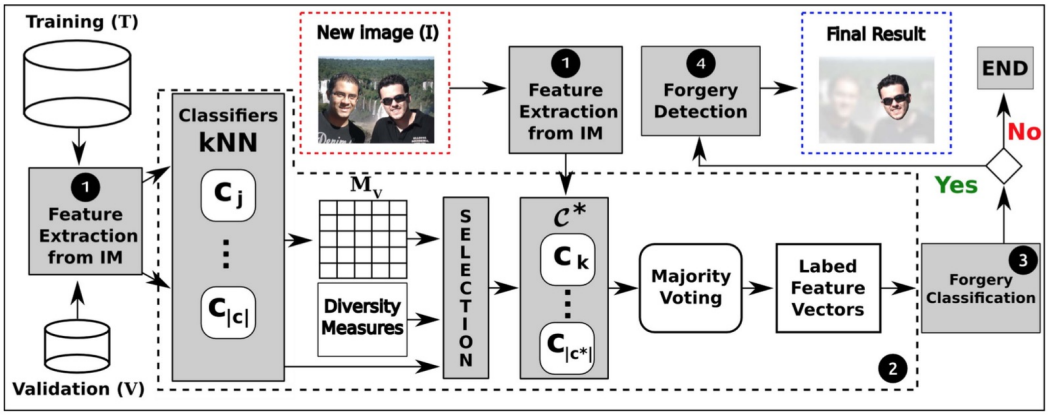


Figure 1.12: Image splicing detection over human faces

by identification of the bounding box in which it is fully contained.

This manual operation does not allow a blind approach to the problem. Because the faces are considered in pairs, this method can be applied only in the case when there are at least two faces, so you need to know in advance the content of the images to assess the applicability of this algorithm.

Working with pairs of faces, another problem is the choice of which face has been manipulated among those considered.

In addiction, this method is based on the observation that, given two faces (a pristine face and a spliced one), the difference between the two forged faces illuminant maps tend to be higher. This intuition, however, proved to be not applicable in all contexts, so cannot be considered as a starting point for the final evaluation.

## 1.7 Region splicing detection

The other considered approach is based on the work of Fan et al. [3]. This method relies on the Van de Weijer et al. [21] Generalized Gray-World framework 1.3. These algorithm perform better on a scene when it is rich of colors, but in a splicing detection scenario we are interest in a specific region of the image.

This approach splits image in vertical and horizontal bands (rectangular

regions), assuming that each band contains sufficient colors for a correct illuminant estimation. Five different algorithm are used, deduced from Eq. 1.3: Grey-World, Max-RGB, Shades of Grey, first- order Grey-Edge and second-order Grey-Edge, in order to have as many illuminant estimates for each band.

This algorithm can be divided into two steps:

1. Subsampling of the image horizontally and vertically
2. Illuminant estimation for each band using the 5 different algorithms and spliced region location.

In the first step the image is sampled into two band categories: horizontal and vertical bands. The height and the width of each band is configured *a priori* based on the minimal height and minimal width separately among the objects of interest. An object of interest is encompassed by a virtual rectangle with its height and width decided by the object itself.

The second step can be also divided into five steps:

1. For each direction and for each algorithm, a reference illuminant is estimated for a single band.
2. For each direction and for each algorithm, a reference illuminant for each direction is evaluated as the median of all references of that direction. At this stage there are two reference illuminats (one for the vertical and one for the horizontal direction) for each algorithm.
3. A detection map is created, with the same dimensions of the image.
4. For each algorithm, every band estimate are compared with the reference illuminant of that band direction and algorithm with the Euclidean distance. If the distance exceed a fixed threshold, the band is considered fake and all the pixel values in the detection map are increased by one unit.
5. The forgery is located using the resulting thresholded detection map.

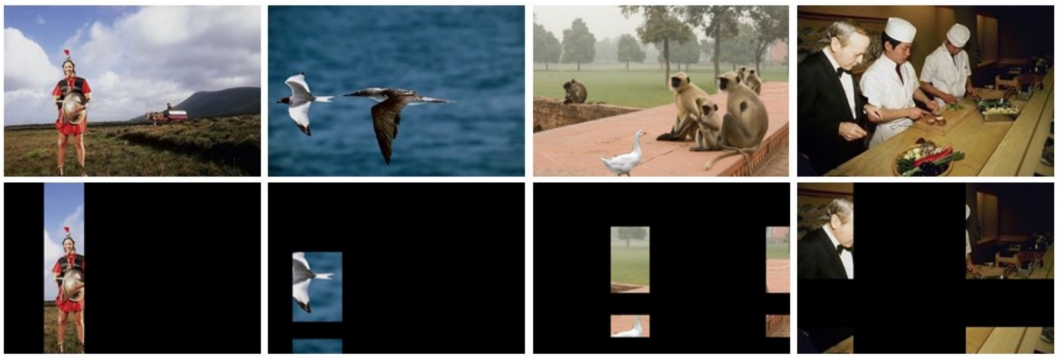

Figure 1.13: Image regions splicing detection

The main advantage of this method is its blind approach: the algorithm is not based on the image content (as it was in the previous case) and no human interaction is needed. The only parameters are the bands minimal dimensions and the threshold. Due these facts it is very simple to implement, but it is also afflicted by false alarm caused by a not easy to tune threshold (which is the key the final step).

### *1.7.1 Method drawbacks*

Due its simplicity, this approach has some drawbacks. First, because the classification is based on a simple distance between a reference value, a key point is the identification of optimal thresholds.

These thresholds can be calculated experimentally, but their accuracy will depend on the dataset with which they were computed.

For how it's built, moreover, this method tends to have a low detection rate compared to a high false positive rate.

# Chapter 2

# Proposed approach

In the previous chapter, a review of two splicing detection methods based on illuminant colors analysis has been presented. However, their effectiveness still needed to be improved for real forensic applications.

The approach proposed in this chapter has been developed to correct some drawbacks and mainly to achieve an improved accuracy over the two approaches presented in Chapter 1.

## 2.1 Overview

Most of the times, the splicing detection process relies on the expert's experience and background knowledge. This process usually is time consuming and error prone once that image splicing is more and more sophisticated and an aural (e.g., visual) analysis may not be enough to detect forgeries.

This approach to detecting image splicing is developed aiming at minimizing the user interaction.

The two methods, presented in the previous chapter [2] and [3], are now being used in synergy with each other, going to analyze each image at the same time looking for potential signs of forgery.

Starting from an image we want to analyze, the method will output a set of results.

- A classification **label** indicating whether an image is believed to be original

or counterfeit.

- A classification **score** indicating the confidence of the method output.
- A **detection map** highlighting the detected spliced regions.

The proposed approach minimizes human interaction being fully automated. However, not both the modules can operate in any circumstance. The face splicing detection module will work only if there is a number of faces greater than or equal to two.

## 2.2 Face splicing detection module

## 2.3 Region splicing detection module

The second form of the algorithm is to implement the method proposed by Fan et al. [3] with some changes in order to try to correct some of its major drawbacks presented in Section 1.7.

The splicing detection task performed by our approach consists in labelling a new image among two pre-defined classes (real and fake) and later pointing 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.

In summary, this module consists of the 6 main steps:

- **Image segementation**: relies on vertical and horizontal image segmentations. The outputs of this stage are two set of directional image bands.
- **Band illuminant estimation**: consists in estimating the illuminant color for each segmented band using 5 different GGE algorithms.
- **Reference illuminant estimation**: consists in estimating the illuminant reference value for each direction.

- **Feature vector evaluation**: relies on encoding the singular band illuminant information into a feature vector for further classification. The feature vector elements are the differences between the current illuminant color and the reference one.

- **Band classification**: consists in labelling each image band into one of the know classes (real or fake) based on the previously learned classification model.

- **Detection map**: using the classification output of the previous step, a detection map is build. The higher the value of this map, the higher the resulting classification score for a single pixel.

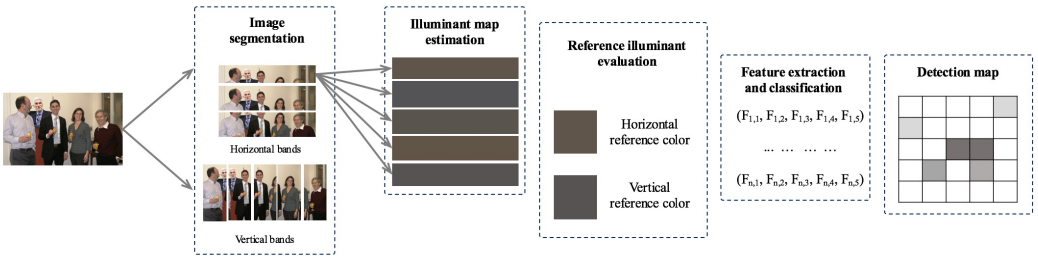


Figure 2.1: Image regional splicing detection module pipeline

The Fig 2.1 summarize the module pipeline.

### *2.3.1 Image segmentation*

In the first step of the process, the input image is segmented in order to obtain two image bands categories: horizontal and vertical bands. This kind is segmentation method is chosen because of its simplicity.

First, a band width $B_w$ and a band height $B_h$ is set. Due to the fact that the segmentation has to produce overlapping bands, we set a delta factor of $\frac{1}{4}$. As a result we get overlapping stripes for a total of 25% of their area.

The choice of the band size is a crucial phase of the algorithm: a band too tight would fail to capture the information necessary to classify our object of interest as falsified, unlike a band too wide would capture instead too much additional information.

The choice of the overlapped area percentage makes possible a more detailed evaluation. In this way it is possible to classify the same region of the image more than once, increasing the expressive power of our final classifier, the detection map.

In summary, let $I$ be the input image. After the segmentation process we obtain a set $B$ of bands containing all vertical and horizontal bands:

$$B = \{V_1, \ldots, V_n, H_1, \ldots, H_m\}$$

The dimension of $B$ is given by the sum of the number of vertical ($n$) and horizontal ($m$) bands.

### *2.3.2 Band illuminant estimation*

The resulting image bands are now processed in order to evaluate the illuminant color using different techniques.

For this step, the Generalized Grayworld [21] algorithms are used, as presented in Chapter 1. For each band, the illumination estimation is accomplished by using one of algorithms composed of Grey-World, Max-RGB, Shades of Grey, first-order Grey-Edge and second-order Grey-Edge. Thus we have 5 illuminant estimates for each band. Table 1.1 in Chapter 1 summarize this algorithm parameters.

In this way we obtain 5 different illuminant estimation for each single band.

$$\forall b \in B \qquad R_a(b) = GGE_a(b) \qquad a \in \mathcal{A}$$

where $\mathcal{A}$ is the set of the previously mentioned algorithms and $GGE$ is the

algorithm implementation.

### 2.3.3 Reference illuminant estimation

After estimating the illuminant of each horizontal/vertical band, it is evaluated a reference illuminant color for each used algorithms.

$$\forall a \in \mathcal{A} \qquad RV_a = median(R_a(b)) \quad \forall\, b \text{ vertical}$$

$$\forall a \in \mathcal{A} \qquad RH_a = median(R_a(b)) \quad \forall\, b \text{ horizontal}$$

where $RV_a$ and $RH_a$ are the reference illuminant colors for the $a$ algorithm for vertical and horizontal direction respectively. As these two values are calculated using the median, will be identical to one of the values of a band of that same direction.

### 2.3.4 Feature vector estimation

Given the two reference color for each of the two directions, a feature band for each single image band can be built.

Assuming a single light source in the image, all the evaluated illuminant will point at the same color. Based on this assumption, a feature vector that capture how the gang present of singularities in the illuminant is built.

Let $b \in B$ a single band lying on $d \in \{vertical, horizontal\}$ direction. The feature vector for $b$ will be:

$$f_b = \{m_1, m_2, m_3, m_4, m_5\} \tag{2.1}$$

where

$$m_i = dist(R_a(b) - RC)$$

where $RC = RV_a$ or $RC = RH_a$ in case of vertical and horizontal band respectively and $dist$ is the Euclidean distance function between two RGB values.

### 2.3.5 Band classification

Given a feature vector, a machine learning approach is used to automatically classify the band. In a prior step, a classification model is trained using a set of sample data. A Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel is used.

$$\forall\, b \in B \quad Label(b) = SVM(b)$$

### 2.3.6 Detection map

In order to collect all the classification outputs, a detection map is build and updated after each evaluation. If the result of the classification is *positive* (*fake*), all the pixel values that belong to the band portion in the image will be increased by one unit.

At the end of the process, all the splicing region pixels will have greater values than the others. At this stage a color map is displayed to give a visual feedback for locating the splicing image parts.

# Chapter 3

# Experiments and results

## 3.1 Evaluation datasets

### *3.1.1 Colorchecker*

The ColorChecker dataset is a collection of images for evaluating Color Constancy algorithms built as additional material to [8]. It consists in 568 RGB colored images of different scenes, both indoor and outdoor taken under different illuminations. In each scene a Gretag MacBeth Color Checker Chart was placed such that it was illuminated by the main scene illuminant and thus its color could be retrieved. The data is available in Canon RAW format free of any correction.

Figure 3.1: Example of an image of the ColorChecker dataset

This dataset has been mainly used for experimenting on pristine data: due

its characteristics is very varied and lends itself well to image analysis based on color.

### 3.1.2 DSO-1

The DSO-1 dataset[1] is composed of 200 indoor and outdoor images with image resolution of 2048 × 1536 pixels. Out of this set of images, 100 are original, i. e., have no adjustments whatsoever, and 100 are forged.

Figure 3.2: Example of an image of the DSO-1 dataset

The forgeries were created by adding one or more individuals in a source image that already contained one or more people. When necessary, we complemented an image splicing operation with post-processing operations (such as color and brightness adjustments) in order to increase photorealism.

### 3.1.3 DSI-1

The DSI-1 dataset is composed of 50 images (25 original and 25 doctored) downloaded from different websites in the Internet with different resolutions. Original

[1]Public available for download at https://recodbr.wordpress.com/code-n-data

images were downloaded from Flickr and doctored images were collected from different websites such as Worth 1000, Benetton Group 2011, Planet Hiltron, etc.

Figure 3.3: Example of an image of the DSI-1 dataset

### *3.1.4 NIMBLE*

## 3.2 Test cases

## 3.3 Performance

# Conclusions

# Bibliography


[1] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310(1):1 – 26, 1980.

[2] T. Carvalho, F. A. Faria, H. Pedrini, R. d. S. Torres, and A. Rocha. Illuminant-based transformed spaces for image forensics. *IEEE Transactions on Information Forensics and Security*, 11(4):720–733, 2016.

[3] Y. Fan, P. Carré, and C. Fernandez-Maloigne. Image splicing detection with local illumination estimation. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2940–2944. IEEE, 2015.

[4] H. Farid. Digital doctoring: can we trust photographs? 2009.

[5] H. Farid. Image forgery detection. *IEEE Signal processing magazine*, 26(2):16–25, 2009.

[6] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[7] K. Francis, S. Gholap, and P. Bora. Illuminant colour based image forensics using mismatch in human skin highlights. In *Communications (NCC), 2014 Twentieth National Conference on*, pages 1–6. IEEE, 2014.

[8] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[9] S. Gholap and P. Bora. Illuminant colour based image forensics. In *TENCON 2008-2008 IEEE Region 10 Conference*, pages 1–5. IEEE, 2008.

[10] M. K. Johnson and H. Farid. Exposing digital forgeries by detecting inconsistencies in lighting. In *Proceedings of the 7th Workshop on Multimedia and Security*, MM&Sec '05, pages 1–10, New York, NY, USA, 2005. ACM.

[11] B. Mazin, J. Delon, and Y. Gousseau. Estimation of illuminants from projections on the planckian locus. *IEEE Transactions on Image Processing*, 24(6):1944–1955, 2015.

[12] T.-T. Ng, S.-F. Chang, C.-Y. Lin, and Q. Sun. Passive-blind image forensics. *Multimedia Security Technologies for Digital Rights*, 15:383–412, 2006.

[13] T.-T. Ng, S.-F. Chang, and Q. Sun. Blind detection of photomontage using higher order statistics. In *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*, volume 5, pages V–V. IEEE, 2004.

[14] T. Qazi, K. Hayat, S. U. Khan, S. A. Madani, I. A. Khan, J. Kolodziej, H. Li, W. Lin, K. C. Yow, and C.-Z. Xu. Survey on blind image forgery detection. *IET Image Processing*, 7(7):660–670, 2013.

[15] C. Rey and J.-L. Dugelay. A survey of watermarking algorithms for image authentication. *EURASIP Journal on Advances in Signal Processing*, 2002(6):218932, 2002.

[16] C. Riess and E. Angelopoulou. Scene illumination as an indicator of image manipulation. In *International Workshop on Information Hiding*, pages 66–80. Springer, 2010.

[17] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein. Vision of the unseen: Current trends and challenges in digital image and video forensics. *ACM Computing Surveys (CSUR)*, 43(4):26, 2011.

[18] C. Springer. Geometry and analysis of projective spacesfreeman. *New York*, 1964.

[19] R. T. Tan, K. Nishino, and K. Ikeuchi. Color constancy through inverse-intensity chromaticity space. *JOSA A*, 21(3):321–334, 2004.

[20] S. Tominaga and B. A. Wandell. Standard surface-reflectance model and illuminant estimation. *JOSA A*, 6(4):576–584, 1989.

[21] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007.

[22] T. Van Lanh, K.-S. Chong, S. Emmanuel, and M. S. Kankanhalli. A survey on digital camera image forensic methods. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 16–19. IEEE, 2007.

[23] X. Wu and Z. Fang. Image splicing detection using illuminant color inconsistency. In *Multimedia Information Networking and Security (MINES), 2011 Third International Conference on*, pages 600–603. IEEE, 2011.

[24] S. Ye, E.-C. Chang, and Q. Sun. Watermarking based image authentication using feature amplification. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 610–613. IEEE, 2005.

[25] W. Zhang, X. Cao, J. Zhang, J. Zhu, and P. Wang. Detecting photographic composites using shadows. In *Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on*, pages 1042–1045. IEEE, 2009.

[26] B. B. Zhu, M. D. Swanson, and A. H. Tewfik. When seeing isn't believing [multimedia authentication technologies]. *IEEE Signal Processing Magazine*, 21(2):40–49, 2004.

# Ringraziamenti