UNIVERSITÀ DEGLI STUDI FIRENZE

UNIVERSITÀ DEGLI STUDI DI PERUGIA

INdAM Istituto Nazionale di Alta Matematica

Università di Firenze, Università di Perugia, INdAM consorziate nel CIAFM

DOTTORATO DI RICERCA
IN MATEMATICA, INFORMATICA, STATISTICA

CURRICULUM IN MATEMATICA
CICLO XXIX

Sede amministrativa Università degli Studi di Firenze
Coordinatore Prof. Graziano Gentili

Image Forensics in the Wild

Settore Scientifico Disciplinare ING-INF/03, SECS-S/01, MAT/06

Dottorando
Massimo Iuliani

Tutore
Prof. Alessandro Piva

Coordinatore
Prof. Graziano Gentili

Graziano Gentili

Anni 2013/2016

Like words together
we can make some sense
Much more than this
way beyond imagination
More than this
beyond the stars

P. Gabriel

Contents

Contents **ii**

List of Figures **iv**

List of Tables **vi**

1 Introduction **1**

2 Image Forensics **4**
2.1 Signal-level vs Scene-level Traces 4
2.2 Forensic Traces and their Capabilities 5

3 Methodologies and Standards for Image Forensics **15**
3.1 Images in the court 15
3.2 Standards and Guidelines 16
3.3 Proposed Methodology 18
3.4 A Case Study 23

4 Image Tampering Detection faces Image Composition **26**
4.1 Introduction 26
4.2 Image Composition Arsenal 27
4.3 Experiments 34
4.4 Conclusions 37

5 Geometric Based Tampering Detection **39**
5.1 Introduction to Pinhole Model 39
5.2 Splicing Detection based on General Perspective Constraints . 42
5.3 Cropping Detection based on Principal Point Estimation 51

6 Social Media Profile Linking **69**
6.1 Introduction to Video Source Identification Based on SPN . . . 71
6.2 Hybrid Sensor Pattern Noise Analysis 73
6.3 PoDIS: a new dataset for video source identification 76
6.4 Experimental validation 77

6.5 Discussion 82
6.6 Future Works: Investigating Stabilised Videos 82

7 Discussion and Open Issues 83

Acknowledgements 86

A Complete Results for Asymmetric Cropping Detection 87

Bibliography 92

List of Figures

2.1 Forensic techniques classification 10

3.1 Methodology scheme 19
3.2 Case study image (face was blurred on purpose). 23
3.3 Wall detail. 24
3.4 Forgery localization map for the case study image 25

4.1 Examples of Composition Techniques 27
4.2 Example of splicing using object transferring techniques 30
4.3 Qualitative Results of Forensic Tools on Image Compositions . . . 35
4.4 A-DJPEG performance for different object transferring approaches 36
4.5 NA-DJPEG performance for different object transferring approaches 37
4.6 Masks applied for Image Composition Techniques 38

5.1 Pinhole Model 40
5.2 Vanishing Points on an Image 41
5.3 Two compared subjects on the reference plane 44
5.4 World reference system 46
5.5 Scheme of height ratio computation 47
5.6 Example of image from dataset 48
5.7 Performance of the method 49
5.8 Detail of one picture in its native version (a), Facebook version (b) and Twitter version (c). 50
5.9 Synthetic data setup 55
5.10 Results on the synthetic data 56
5.11 Images considered in real tests 57
5.12 Graphical visualization of angles obtained from the pairwise intersection of lines concurrent to the same VP 57
5.13 Comparison of results achieved from synthetic and real images . . 58
5.14 Example images produced to test the PP estimation algorithm . . 59
5.15 Examples of lines selected by the user on the real image 60
5.16 Results on real images obtained by varying MVA and lines number 61
5.17 PP in a pristine and cropped image 62
5.18 ROC curves of the cropping detection for synthetic and real data . 64

5.19 ROC curve on synthetic and real data with different cropping percentage . . . 66
5.20 Two examples of cropping detection . . . 68

6.1 Geometric transformations from the full frame to the output format. 73
6.2 Geometric transformations applied to the sensor pattern from the full frame to the image/video outputs on SMPs. . . . 75
6.3 Matching statistics mP_V and mP_I . . . 79
6.4 Matching statistics mP_I obtained using image reference from Facebook LQ and HQ . . . 80
6.5 Performance of profile linking between Facebook and Youtube considering both LQ and HQ Facebook image reference. . . . 81

A.1 ROC curves for Synthetic and Real Tests along different single cropping directions . . . 89
A.2 ROC curves for Synthetic and Real Tests along two cropping directions . . . 90
A.3 ROC curves for Synthetic and Real Tests along three cropping directions . . . 91

List of Tables

2.1 Example of a forensic tool steps 9

3.1 Some metadata information included in the Exchangeable Image File Format (Exif) header of an image taken as an example. 20
3.2 Output of forgery detection tools and their fused score 24

5.1 Details of subjects considered in the test 49
5.2 Signal degradation introduced into one image of the dataset by exchange through social networks 50
5.3 AUC for Native, Facebook and Twitter Images 50
5.4 Max STD of estimated PPs between x and y direction 60
5.5 AUC for Perspective based Test on synthetic and real data 63
5.6 Cropping detection on both synthetic and real data 64
5.7 AUC for Characteristic based Test on synthetic and real data . . . 65
5.8 Mean Accuracy for Characteristic based Test on synthetic and real data 65
5.9 AUC for on Facebook Data 67
5.10 Mean Accuracy on Facebook Data 67

6.1 Considered devices with their default resolution settings for image and video acquisition respectively. 77
6.2 Rescaling and cropping parameters for the considered devices from image to video output. 78
6.3 Best achievable accuracy and correspondent PCE threshold. 81

A.1 AUC values for Synthetic and Real Tests along different single cropping directions 88
A.2 AUC values for Synthetic and Real Tests along two cropping directions 88
A.3 AUC values for Synthetic and Real Tests along three cropping directions 88

CHAPTER 1

Introduction

In the last years, visual digital data gained a key role in providing information. Images and videos are acquired and quickly shared between huge amount of users through social media platforms. Statistics [1] show that a relevant portion of the world's total population owns a digital camera and can capture pictures. Furthermore, one-third of the people can go online [2] and upload their pictures on websites and social networks. In this sense, any user is a potential source of information, shared through the visual data he provides.

On the other hand, professional tools for image post processing are available and affordable to both novice and proficient users. Most of them are easy to use and allow any user to create realistic forgeries. This fact poses the problem of relying on digital images and videos as potential source of information, especially when the source is unknown. This issue becomes critical when the visual data is exhibited as a source of potential evidence in legal acts.

In recent years Image Forensics has been proposed as a solution for image authentication problem. This technology concerns the analysis of the traces left by any process occurring in the image lifetime to determine information about its life cycle (e.g., which is its source; which processing it's undergone; if, where and how its content has been modified; . . .).

To date, several tools have been provided by the research community to look to an image at different levels of depth. Some of them analyse the image metadata; others extract pixel level statistics to characterise specific traces (e.g., the sensor noise, compression artefacts, . . .). Finally other tools look at the physical property of the image (light and perspective based) to identify inconsistencies.

The effectiveness and the limits of these tools has been investigated during the years and there's still a big gap to be filled, especially for forensic application. When a specific query on an image is demanded by a legal part, problems may arise, mainly related to the methodology to be applied, the reliability and the interpretation of the tools outputs.

methodology When an image is inspected to provide an evidence in court, the forensic expert has to decide which are the tools to be used for the image analysis and how to apply them. Whatever he does, he's intrinsically applying a methodology.

reliability Tools effectiveness in not the same under every possible environments. That's why tools performance are usually assessed on a certain amount of heterogeneous data (that is usually very limited with respect to the huge amount of image variability). This means that these results cannot be applied indistinctly to any investigated image.

interpretation Most of the tools outputs can be hardly converted the "probability" of the image to be tampered. Furthermore, most of these output cannot be taken as they are and have to be interpreted by the expert, according to the context. The interpretation of the result is much more critical when several tools are applied: the expert, after their comparison, has to merge the achieved results to produce a final record on the investigated image.

These issues are related to technical limits that make the current forensic technologies still unready to work under uncontrolled environments. Some key points can be identified in the following:

- Most of the techniques based on pixel level statistics are fragile against common compression and filtering processing that are usually applied by several social media platforms and smartphone camera software. This means that huge false alarm can possibly occur if the expert applies these techniques in the wild as they are;

- Forgery detection techniques are usually evaluated on unrealistic hoax (synthetically or automatically created to produce huge amount of test data). In most cases it is unknown how the performance changes when facing high quality and sophisticated forgeries;

- Performance are mostly evaluated on limited datasets that are not representative of the wild world of images. Dataset are usually composed by images of the similar category, undergone through the same (simple) chain of processing. Conversely, an image in the wild has possibly undergone to longer or more complex processing chains, making hard to apply the tools as they are;

- Some outputs depend on user interpretation and behaviour meaning that two different users, with the same tool, may achieve different conclusion.

This usually happens when a user is required to select features in the scene or to interpret image content;

- When a tool ”provides” some kind of evidence, scientific results have to be transmitted to the legal part. The lack of communication between these parts poses a problem of choosing the best investigation methodology, of proving the pertinence of the applied method, of presenting the achieved results.

This work, starting from the available technologies, is focused in making strides on some of above limitations. All the debated topics aim to improve the application of forensic technologies in real case scenario, where images from unknown source are investigated as a potential source of evidence. This thesis addresses the following issues: i) Provide a methodology to investigate a digital image with several tools and provide results to be presented in court; ii) Improve the available forensic tools to solve current limitations; iii) Assess the accuracy variability of available technologies under specific conditions iv) Develop new applications for the available technologies that take advantage of side information available on the web.

This document is organised as follow: in Chapter 2, Image Forensic technologies and main traces are summarised. A new forensic scale is defined to provide more than a binary answer (the image is tampered or authentic) and forensic traces are linked to the kind of evidence they are able to provide. Chapter 3 summarises the best practices and standard available for digital image investigation and introduces a new methodology for the forensic analysis of images (and multimedia contents). Chapter 4 addresses the problem of forgery detection against sophisticated image alteration, obtained by means of advanced techniques. Image Compositions tools are surveyed and both qualitative and quantitative tests are performed. In Chapter 5, focusing on geometric-based features, we provide two contributes: i) a generalisation of a perspective-based technique for tampering detection and ii) the reliability assessment of a cropping detection technique based on principal point estimation. In Chapter 6, basing on the promising trace of the sensor pattern noise, we introduce an intelligence application to link social media profiles where images and/or videos are captured with the same device. Finally, in Chapter 7 we summarise the achieved results, their limitations and the open issues for future works.

CHAPTER 2

Image Forensics

Image Forensics has been proposed as a solution for authenticating the contents of digital images [3, 4, 5]. This technology is based on the observation that each phase of the image history—from the acquisition process, through its storage in a compressed format, to any editing operation—leaves distinctive traces on the data, as a sort of digital fingerprint [6]. It is then possible to determine whether a digital image is authentic or modified, by detecting the presence, the absence or the incongruence of such traces intrinsically tied to the digital content itself. In the literature different classifications of traces and tools have been proposed, each fitting different purposes.

In this chapter, starting from the available classifications, we highlight main issues related to the applicability of these techniques in the wild. Furthermore, we introduce a new general classification scale, called the FD (short for Forgery Detection scale), based in the concept of image nativity. This scale aims to extend the concept of distinguishing between pristine and tampered images. The most relevant techniques are surveyed and their capabilities, both in terms of applicability (i.e. when we can use them) and assessment (i.e. the level that can be achieved in the FD scale) are investigated. Specifically, Section 2.1 introduces the main distinction between signal-level and scene-level traces, that exhibit few relevant differences in terms of performance evaluation and robustness when applied in the wild; in Section 2.2 we define the FD scale, a classification of the most relevant forensic traces is reported and their capabilities are investigated.

2.1 Signal-level vs Scene-level Traces

When a forensic tool provides an output, it's important to clarify its reliability in the specific context. Few general issues can be highlighted by distinguishing between signal-level and scene-level traces: the former include invisible footprints introduced in the signal statistics at a pixel level, e.g., demosaicing artifacts [7], sensor noise [8] or compression artifacts [9, 10]; the latter are based on a physical interpretation of the depicted scene, e.g., inconsistencies in shadows [11], lights [12, 13], or in perspective and geometry of objects [14, 15]). Signal-level traces are typically detected automatically on most kind of image contents, allowing to test the algorithms on a huge amount of heterogeneous data; on the other hand, any pixel processing possibly alter previous traces present on the image. Indeed, they often exhibit lower effectiveness when the investigated content has been subjected to an unknown chain of processes (e.g., filtering, resizing, compression) that partially or completely spoiled the searched traces [16]. Scene-level traces are typically different: they usually have stronger requirements on scene constraints (e.g. the presence of Lambertian convex surfaces for lighting estimation [17], or some objects with specific geometric shape), but have the advantage of being robust to common image processing operations, thus appearing suitable even for low resolution images, or when the content has undergone multiple and/or strong compressions. Conversely, a critical point for this techniques is their performance evaluation. This is mainly due to the fact that such algorithms are usually tested on small datasets only, since they cannot exclude some human intervention, e.g. for image feature selection or analysis supervision. This distinction is useful to understand two general open issues: i) assess and improve the reliability of tools based on signal-level traces when they are applied on images under uncontrolled environments; ii) automatise tools based on scene-level traces to remove the human-in-the-loop.

In the next section we extend the concept of image authenticity considering that each tool provides different kind of evidence, i.e. proving that the image is not native, or detect a forged image, or localise a forgery, or identify the tool exploited to make the forgery.

2.2 Forensic Traces and their Capabilities

According to recent surveys [3][18][19], there are two main fronts to digital image forensics: source identification and forgery detection[1] Source identification tries to link an image to the device that took it. Acquisition traces are extracted from the investigated image and then compared with a dataset of possible fingerprints specific for each class/brand/model of devices. Matching fingerprints are exploited to identify the image source device. This topic will be examined in depth is Section 6. In this section we focus on forgery detection, i.e. on determining whether and how a target image has been altered. Forgery detection mostly works twofold: either looking for patterns where

[1] Recently, some other applications are also addressed by the research community, e.g., image phylogeny and reverse engineering. This topics are out of the scope of this thesis and are not addressed in this thesis.

there shouldn't be, or looking for the lack of patterns where there should be. Let us consider, for example, an object spliced from an image into another, and resized to be in the same scale as the target picture. The resizing operation inserts a pattern of correlation between neighboring pixels where there shouldn't be. Concurrently, several image traces can be possibly disrupted by the splicing operation. The most simplistic view would pose that there are only two outcomes for forgery detection: the image has been altered, or no evidence of alteration is found. However, this classification might not be sufficient. Simply compressing an image might be considered an alteration, even though it is a commonplace operation, making this classification useless. Different forensic techniques work on different assumptions of what traces could be present on the image and what it can be inferred from them, e.g., the location or the nature of the forgery. However, there is no standard in the literature, for classifying and comparing techniques based on their outcomes.

2.2.1 The Forgery Detection Scale

Here, we propose a new general classification scale called the FD (short for Forgery Detection scale). This scale is based in the concept of an image being native or not: a native image is an image that was captured by a device and then outputted to the user "as-is". Conceptually, this is easy to define, but technically there might be some complications: different devices process the image differently. These problems will be discussed in detail further in this section.

The FD scale ranks forensic techniques based on the type of evidence they can possibly provide about an image's history. The first possible outcome is the negative case, when it is not possible to discover information supporting that the image has undergone any form of alteration with respect to its original form. This could happen because the image is really native, or because the analyzed traces do not show forgeries, but it makes no difference: it is not possible to say that an image is truly native, only that there is no evidence supporting it to be altered. This outcome falls outside our scale in practice, but can be called FD0 for simplicity. The following are the different levels of our Forensics Detection scale:

FD0 No evidence can be found that the image is not native.

FD1 The image has undergone some form of alteration from its native state, but the nature and location of it is unknown.

FD2 The image has undergone some form of alteration from its native state and the location of the alteration can be determined, but its nature is not known.

FD3 The image has undergone some form of alteration from its native state, the location of the alteration can be determined and the nature of it is known.

FD4 All the conclusions of the previous item, and a particular processing tool or technique can be linked to the forgery.

These should be referred as FD (short for forgery detection) scale, with values FD0-FD4. It could be argued that the ultimate form of forgery detection would go beyond identifying the used technique, by locating a forger or even estimating the historic of the image's alterations [20]. It is a valid point, but they are not common in digital image forensics; then we do not consider them for the moment. FD scale is backwards inclusive for FD>0, meaning that if FD4 can be guaranteed so can FD3, FD2 and FD1. The following subsections provide in-depth explanation of the different levels of FD scale and further considerations.

FD1: Nativity

The FD1 level differs from the negative case FD0 because it is possible to determine that the image is not native. This is not so simple to assess, as most modern cameras have a processing pipeline comprised of several operations (demosaicing, white balance, etc), changing the image before the user has access to it. Furthermore, demosaicing is such a fundamental operation in modern cameras that it makes little sense talking about images without it. For the sake of generality, we propose that any form of pre-processing on the image up until a single in-camera compression can be accepted without breaching the image nativity. A forensic technique achieves FD1 when it is able to find evidence of alteration after capture. Techniques that analyze an image's EXIF information are an example of FD1: they can detect an inconsistency in the metadata proving an image is not native, but nothing can be said about location or nature of the alteration.

FD2: Location

The FD2 level is obtained when the general location of the alteration in the image is known. It is possible that a region of an image has been erased by a series of copy-pasting operations, and then retouched with smoothing brushes. In this case the boundaries of the forgery might not be as clear. If a technique is able to obtain any form of specificity on the altered region, FD2 is achieved. This is the case when analyzing traces such as PRNU (Photo Response Non-Uniformity, Section 2.2.2), CFA or ELA (Error Level Analysis, Section 2.2.2), that are locally structured on the image. If evidence of any global alteration on the image is found, then the location of the forgery is the whole image. Similarly, operations that remove parts of the image such as seam carving and cropping can be detected but the actual altered area is not present in the analyzed image anymore. It is argued that the forgery location can be considered to be all image, reaching FD2.

FD3: Nature

The nature of the forgery can be subjective, because it is not possible to predict all ways in which an image can be altered. The most commonly studied forms of forgery such as splicing, copy-pasting and erasing, are just a subset of possibilities. As was discussed on the introduction, image composition techniques are able to alter the shape, texture and orientation of objects, and even merge them together. For simplicity, any meaningful information in addition to location of the processing that can be used to assist the forensics analyst can be considered FD3. For instance, identifying that a spliced object has been rotated and scaled awards an FD3 level on the scale. Even identifying that an object is just spliced is worth an FD3 on the scale because the image is not native (FD1), its location on the target image is evident (FD2), and the nature of the alteration is known (FD3).

FD4: Technique

The highest level on our scale, FD4, is achieved when the analyst finds evidences that can link the forgery to a particular technique or tool. A splicing can be done by simply cutting a region from an image an pasting over another, but there are also sophisticated ways to blend them, such as Alpha Matting or Seamless Cloning. A forensic technique that is able to, after obtaining FD3, provide further insight into the technique or tool used to perform the forgery achieves FD4.

Accuracy and Confidence

The FD scale describes the scope of the information that a forensic technique can achieve. **It does not evaluate the accuracy of techniques, their confidence or applicability**. If a technique provides an output map of irregular pixels based on a general trace such as PRNU, it is going to be FD2. Two different techniques that produce the same type of maps based on PRNU, but one has better results are still both FD2. A forensic technique that outputs the same type of probability map per pixel, but is looking for traces left by a particular processing like local gaussian filtering would be an FD3 on our scale. Visually both maps could be similar, but they are providing a very different type of information.

Forensic approaches that heavily rely on feature descriptors or machine learning to identify a specific type of forgery may at first seem to fall outside our scale. For instance, a technique that evaluates the image as a whole to identify if it has been spliced, but without providing a location seems to fall in

FD3 but not in FD2. Our argument is that such general techniques are usually trained on dataset composed only by one class of tampered images, exactly the one that the method is looking for. Detecting splicing in this sense is only detecting non-nativity (FD1) as it is unsure how the technique responds to other types of forgery.

2.2.2 The forensics arsenal

The current state-of-the-art on digital image forensics provides an arsenal of tools and techniques for forensics analysts. In this section we investigate the most relevant approaches and their capabilities, both in terms of applicability (i.e. when we can use them) and assessment (i.e. the level that can be achieved in the FD scale). In a general way, it can be noted that there is a trade off between the generality and the FD level that a technique is able to reach. This is intuitive, because the higher the level on the scale, the more specific the assessments are. FD1 can be simplified as a boolean statement (the image is either native or not). From FD2 onwards, there is a large set of possible answers (all different combinations of pixels in the image). To identify the nature of the forgery in the image (FD3), a technique must be looking for more specific features or traces. An image forensic tool is usually designed considering three steps:

1. Some **traces** in the image - possibly introduced by the forgery process - are considered;
2. Some image statistics are determined based on the considered trace, resulting in **features**, which are usually numeric in nature;
3. A **decision** is taken about the image. This can be done using simple thresholds on the calculated feature or on sophisticated machine learning techniques.

Table 2.1: The steps of the tool by Carvalho et. al. [12].

Layer	**Example**
Trace	Illuminant or light source of the image.
Feature	Estimated illuminant colors and light intensity on object edges.
Decision	SVM classification.

In Table 2.1 we show a practical example of previous steps for the technique developed by Carvalho et. al. [12] to detect splicing. The used **trace** is the

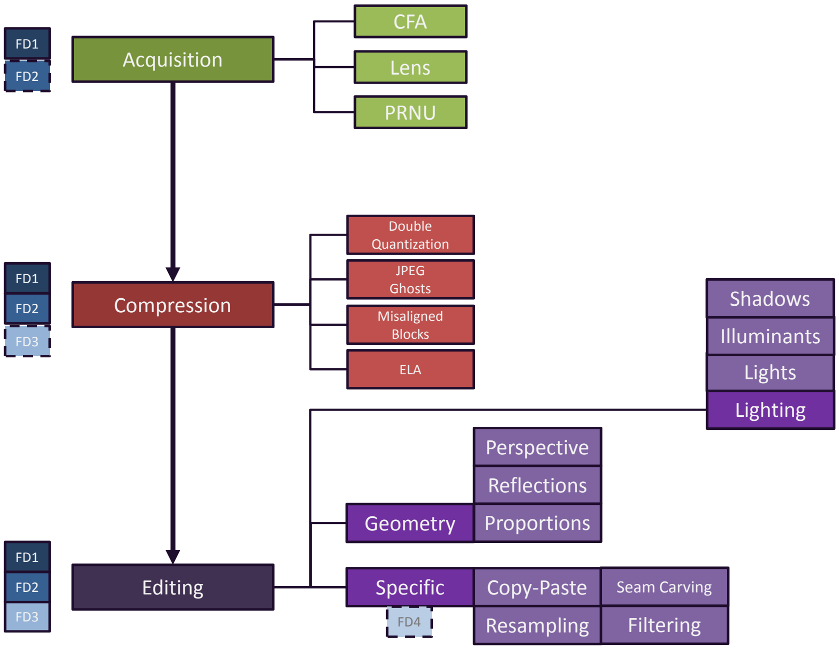


Figure 2.1: Forensic techniques classification. Each type of trace is organized under its correspondent phase on the forgery process. The techniques themselves were omitted for the sake of clarity, but would appear as leaf nodes under their analyzed traces. On the left, the relation to the FD scale is displayed. Only by analyzing specific editing traces it would be possible to achieve FD4.

illuminant, or the light source. The key observation is that if an object is spliced and the original image had different light conditions, such as indoor or outdoor lighting, or even incandescent vs. fluorescent lights, this trace can be used to identify it. The **features** used are the estimated illuminant colors and the light intensity on the edges, for the different analyzed regions of the image. The **decision** process uses a Support Vector Machine (SVM) to classify the image as either spliced (FD3) or inconclusive (FD0) based on the features.

Forensic tools classification is based on the traces they analyze. Piva [19] distinguishes between traces left by three different steps of the image formation process: acquisition, coding and editing. Another intuitive classification has been proposed by Farid [3] where the forensic techniques are grouped into five main categories: pixel-based, format-based, camera-based, physically-based and geometric-based. We propose a classification based on Piva's approach (Fig. 2.1), but with greater specificity to Farid's.

The most relevant traces and correspondent tools developed by the forensic community will be discussed in the following subsection. The FD scale will be used to describe which level of assessment can be expected when examining an image using a specific tool.

Acquisition Traces (AT)

Native images come to life with distinctive marks (artifacts, noise, inconsistencies) due to the acquisition process. Both hardware (e.g., lens, sensor) and software components (e.g., demosaicing algorithm, gamma correction) contribute to the image formation, introducing specific traces into the output (native) image. When a native image is processed some of these traces can be deteriorated or destroyed, exposing evidence of tampering. Forgery detection using acquisition traces generally falls in one of two categories:

1. *Global*: The analyzed trace is a global camera signature. Non-native images can be exposed when this signature does match with the supposed source device. For instance, in [21] non-native JPEG images are exposed by analyzing quantization tables, thumbnails and information embedded in EXIF metadata. In [22] the reference pattern noise of the source device is taken as a unique identification fingerprint. The absence of the supposed pattern is used as evidence that the image is non-native.

2. *Local*: The analyzed trace has a local structure in the image. Its inconsistencies in some portion of the image can be exploited to localize the tampering. For instance, Ferrara [7] uses demosaicking artifacts that form due to color interpolation. They can be analyzed at a local level to derive the tampering probability of each 2×2 image block. Fridrich [8] reveals forgeries by detecting the absence of the PRNU on specific regions of the investigates image.

Let us note that some traces can be considered both at a global or local level (e.g., demosaicing artefacts and PRNU), allowing to identify non-native images (FD1) or to localize forgeries (FD2). The analysis of acquisition traces is usually limited for the matching a known pattern, and they can be easily disrupted. For this reason, FD2 is the highest we can expect to achieve on the FD scale using acquisition traces. The analysis of acquisition traces generally requires some additional information about the source device. In some cases this information depends on the source device model or manufacturer (e.g., color filter array pattern, quantization tables), and can be easily obtained to assess image nativity [23]. In other cases these traces are unique camera fingerprints (e.g. PRNU) and can be obtained by having the source device available, or be estimated by using different images captured by the same device.

Coding Traces

Lossy compression often happen during digital images life-cycle:

- Native images of non professional cameras and smartphones usually come to life in JPEG format;
- When uploading a photo on a social network lossy compression is possibly applied to the image;
- When a JPEG image is altered and saved again in JPEG, double lossy compression occurs.

For this reason, most of the literature has focused on studying the traces left by single and multiple JPEG-compressions. This is a very prolific field of study in forensics, with a wide variety of techniques. Fan [24] and Luo [25] provide efficient methods to determine whether an image has been previously JPEG compressed, and, if so, are able to estimate some of the compression parameters. Further advances have been also provided by Li et al. [9] to identify high-quality compressed images basing on the analysis of noises in multiple-cycle JPEG compression. On Bianchi's technique [10], original and forged regions are discriminated in double compressed images, either aligned (A-DJPG) or nonaligned (NA-DJPG) even when no suspect region is detected. Yang et al. [26] propose an error-based statistical feature extraction scheme to face the challenging case where both compressions are based on the same quantization matrix.

In most cases the analyst can exploit coding traces to disclose non-native images or to localize the tampering, reaching FD1 and FD2 in the forensic scale; FD3 has not been deeply investigated but, as shown in literature, coding traces can reveal something more than mere localization of the tamper. Farid [27] shows that, when combining two images with different JPEG compression quality, it may be possible to recover information of the original compression quality of the tampered region. A stronger compression in later stages usually deteriorates the traces of previous compressions, compromising the effectiveness of these techniques. This technique has been proved effective only if the tamper was initially compressed at a lower quality than the rest of the image; on the contrary, when the compression is stronger in the latter stage, the traces of the first compression are probably damaged and the detection fail.

A general rule of signal based techniques is that stronger compression in later stages can possibly ruin the traces of previous traces and processing, even another compression, thus making the tools ineffective.

Editing Traces

Image editing modifies the visual information of the image and the scene depicted, introducing traces in several domains of the image such as pixel, geometric, and physical. Editing traces are the most numerous, and can be split into subcategories (Fig. 2.1) according to these domains.

Image illumination inconsistencies (light source direction, cast and attached shadows) are powerful traces considering that it is hard to achieve a perfect illumination match when composing two images. The are two main approaches for illumination techniques: geometric and illuminant. The first are based on the geometric constraints of light, trying to use scene elements as cues to determine if the arrangement of lights [28][29][30][13][31] or shadows [11][32] are plausible. Illuminant techniques exploit the color, intensity and temperature aspects of the illumination, and are able to detect if a region or object in the image was differently lighted [33][34].

Similarly, geometric relations within an image (e.g., object proportions, reflections) are based on the perspective model defining the projection of the 3D real scene onto the image plane. This process is commonly modelled through the pin hole camera model [35]. Any deviation from this model can be exploited as evidence of tampering. Perspective constrained method, proposed by Yao [14], is used to compare the height ratio between two objects in an image. Without the knowledge of any prior camera parameter, it is possible to estimate the relative height of objects and eventually identify whether one of those have been inserted on the scene without properly respect the perspective rule. An extension has been proposed by Iuliani et al. [15] to apply the technique on images captured under general perspective conditions. Conotter [36] describes a technique for detecting if a text on a sign or billboard has been digitally inserted on the image. The method looks if the text shape satisfies the expected geometric distortion due to the perspective projection of a planar surface. The authors show that, when the text is manipulated, it is unlikely to precisely satisfy this geometric mapping.

When an editing trace exposes evidence of forgery, we can expect to infer something about its nature (FD3): if an object has as shadow inconsistent with the scene, he was probably inserted; if the illuminant color is inconsistent, the object could have been either spliced or retouched.

Obtaining other specific information about the techniques involved in the tampering process (FD4) is a very challenging task. There are two main reasons for this. The development of a technique for detecting the use of a specific tampering process/tool may require a strong effort compared to its applicability in a narrow range. Secondly, proprietary algorithms have undisclosed details about their implementation, making hard to develop analytical models for their traces. A first step toward this kind of assessment has been proposed by Zheng et al. [37] to identify the feather operation used to smooth the boundary of pasted objects.

Here above we defined the forensic arsenal and what kind of evidence it can possibly provide. To have a complete picture we should also consider the capability of the available techniques to create forges. A deeper investigation of this topic is tackled in Chapter 4 where the image composition arsenal is surveyed and some forensic techniques are tested over photorealistic forgeries.

In the next chapter we focus on how forensic technologies should be applied by the forensic expert when an image is involved in legal action and a possible methodology for image investigation.

CHAPTER 3

Methodologies and Standards for Image Forensics

3.1 Images in the court

In digital investigations images are more and more frequently analysed through Image Forensic tools and presented as potential digital evidences to the court. In this scenario it is necessary to take into account salient aspects, such as the chain of custody, data authentication, application of scientific methods, documentation and reporting. Due to the rapid growth of multimedia technologies and the ever changing situations in the digital field, presently there are no uniform procedures to face with all such issues, although guidelines and best practices are beginning to be proposed.

As described in the previous chapter, academic research has developed many techniques for image analysis, but from the point of view of applying these techniques in the courtroom, an important gap must be still bridged. This gap includes the poor communication between the legal and the scientific actors, as well as the not fully maturity of technologies, that are often tested in laboratory conditions and not in real-world scenarios. Then, when a tool "provides" some kind of evidence, scientific results have to be transmitted to the legal part. This poses a problem in proving and reporting the pertinence of the applied method and the reliability of the provided results.

The previous chapter rounded up the current Image Forensic methods for image analysis and authentication. In this chapter we will focus on the available standards and best practices emerged so far relating to digital investigation where images are involved.

3.2 Standards and Guidelines

The ISO/IEC JTC1 Working Group 4 has put its effort in developing some standards giving guidance on several aspects of the digital investigations. The main International Standards which affect the investigative process are:

- ISO/IEC 27035 (published in 2011, to be revised in 2013): Information Security Incident Management
- ISO/IEC 27037 (published in 2012): Guidelines for the Identification, Collection, Acquisition and Preservation of Digital Evidence
- ISO/IEC 27041 (published in 2015): Guidance on Assuring the Suitability and Adequacy of Incident Investigative Methods
- ISO/IEC 27042 (published in 2015): Guidelines for the Analysis and Interpretation of Digital Evidence
- ISO/IEC 27043 (published in 2015): Incident Investigation Principles and Processes

The fundamental goal of these Standards is to promote good procedures and methods for investigating digital evidences and to encourage the adoption of similar digital forensics approaches internationally, thus making easier comparison and combination of results coming from different people and organizations, also across different jurisdictions. ISO/IEC 27035 (in its three parts) defines the steps that should be taken prior and during an incident, in order to ensure that investigations can be conducted readily; it discusses the means by which those involved in the early stages of the investigation can ensure that sufficient potential digital evidences are captured, allowing the investigation to proceed appropriately. ISO/IEC 27037 addresses the problem of maintaining the integrity of potential digital evidences during all their life-cycle by adopting a correct chain of custody; it discusses the steps which should be taken immediately following an incident. ISO/IEC 27041 deals with methods by which the processes adopted at all stages of the investigation can be shown to be appropriate; it offers guidance on assuring the suitability and adequacy for all the stages of the investigation process. ISO/IEC 27042 provides indications on the important phases of analysis and interpretation of digital evidences; it discusses fundamental principles which are intended to ensure that tools and techniques adopted for the analysis and interpretation are selected appropriately. It provides a guidance on the analysis and interpretation ensuring continuity, validity, reproducibility, repeatability. ISO/IEC 27043 defines the basic principles and processes underlying the investigation of incidents, providing a general overview of the incident investigation process. The indications coming from all these Standards highlight that a harmonised investigation process model is needed (both in a criminal prosecution and in other frameworks such as information security incident) even if it can be customized in different investigation scenarios.

From the United States some important indications on digital investigation come from the Scientific Working Group on Digital Evidence (SWGDE) [38] and the Scientific Working Group Imaging Technology (SWGIT) [39]. SWGDE brings together organisations actively engaged in the field of digital and multimedia evidence, in order to promote communication and cooperation among them and to ensure quality and consistency within the forensic community.

On the other hand, SWGIT focuses on imaging technology and aims to facilitate the integration of imaging systems within the criminal justice system by providing best practices and guidelines for the capture, storage, processing, analysis, transmission, output of image and archiving. SWGIT makes available several documents, i.e. 24 Sections, presenting guidelines, best practices and recommendations. The most interesting among them for our purpose of investigating (i.e. analysis and interpretation) visual data are sections addressing the problem of forensic video analysis (Section 7), forensic image analysis (Section 12) and image authentication (Section 14).

Regarding forensic image (and video) analysis, the process is seen as composed by three main tasks: technical preparation, examination, interpretation. Technical preparation concerns all those steps that are necessary to prepare videos/images for the other tasks (examination and interpretation) as well as the preparation of the outputs obtained from the forensic analysis process. Examination represents the core activity of the analysis: it regards the application of techniques for extracting the information conveyed by the video/image itself, such as hidden messages, intrinsic device noise, manipulations, as well as anthropometric measures that can be evidenced by some video/image enhancement processing. Interpretation regards the visual analysis of digital content by specific subject matter experts, providing conclusions about the subjects/objects depicted in the observed video/image. Besides the description of the general tasks, these documents also suggest a set of best practices for the implementation of such activities, including indications on the chain of custody procedures, the appropriate documentation for any analysis step, the demonstration of the analyst competency, the Standard Operating Procedures (SOPs) describing the work flow of all the actions performed during the forensic analysis. Regarding image authentication, Section 14 provides other guidelines to perform image trustworthiness verification through appropriate practices. Authenticity indicates that videos/images are an accurate representation of the original event. The tasks performed during the authentication process include detection of manipulations, analysis of metadata included in the image file, identification of provenance, etc., and all the conclusions coming from this process (in terms of both numerical probabilities and more frequently subjective criteria) should be detailed in a final report.

In the next Section, starting from the just mentioned standards and guidelines, we focus on the specific task of investigating an image exploiting Image Forensics technologies.

3.3 Proposed Methodology

The purpose of the proposed methodology is to accomplish the appropriate application of Image Forensics technologies to acquire information from the inspected image. Our effort focuses on digital images but the proposed methodology can be applied to any multimedia content [40] (digital audio and video).
From now on we refer to the Forensic Analyst (FA) as the expert able to apply such technologies following the proposed methodology, and to make a synthesis of the multiple results. We will focus on the analysis of the image content itself leaving aside the well known problem of the correct chain of custody for digital data supposing that the FA is well trained and “fully experienced” in the international standards ISO/IEC 27037 to deal with digital contents in forensic contexts.

In Fig. 3.1 we sum up the proposed methodology for the investigation of digital images.
The image file can be seen as a package composed by two main parts: i) the header, containing a set of data (known as metadata) including some information about the file content; ii) the content itself, that is the stream forming the audio-visual signal (`avs`). The analysis process can be resumed as follow:

1. metadata extraction and analysis
2. audio-visual inspection of the signal followed by
   - source identification
   - authentication assessment
   - content enhancement and/or analysis
3. result analysis
4. reporting

The steps followed in the proposed methodology are now described in more detail:

Metadata Extraction and Analysis Images come to life with their own metadata containing information about the image itself. Depending on the image format (JPEG, PSD, Raw, . . .), different types of metadata can be embedded into the file (Exif, XMP, PLUS . . .). Embedded information including

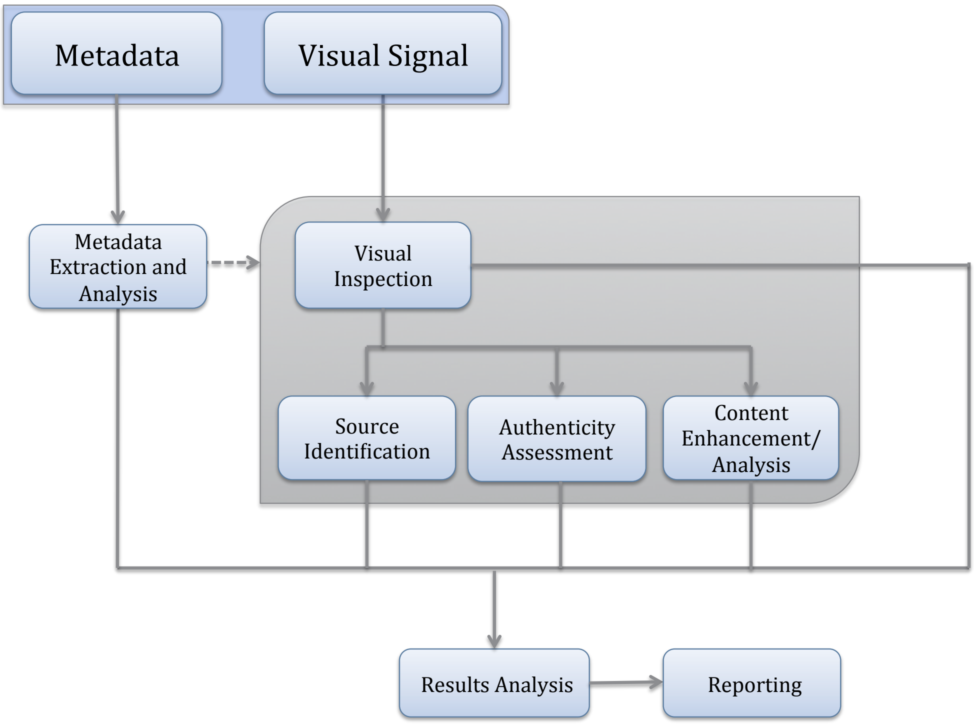


Figure 3.1: Methodology scheme

data source device, colour space, resolution and compression parameters, time and GPS coordinates, ... Table 3.1 shows some information included in the Exif metadata of a native image taken as an example.

When an image is opened or modified in someway, some traces may be left in the embedded metadata, thus providing traces of the image history. Anyway there are two main drawbacks in their use for forensic application: i) metadata can be easily modified even by non expert users by free available softwares to provide misleading traces; ii) some common processing (as social media platforms upload) usually delete most interesting metadata from images.

Then, a first task in metadata analysis consists in the assessment of the compatibility, completeness and coherence of the extracted information in the considered scenario. For instance, the absence of most metadata has nothing strange in a Facebook image while is really suspect in an image that is supposed to be native.

Visual Inspection The process of examining the picture through the visual inspection essentially consists of: i) the interpretation of its content and ii)

Table 3.1: Some metadata information included in the Exchangeable Image File Format (Exif) header of an image taken as an example.

Tag	**Value**
File Name	3948_122.jpg
File Size	381 kB
File Modification Date/Time	2014-07-16 16:06:14+02:00
File Access Date/Time	2014-07-17 13:16:36+02:00
File Inode Change Date/Time	2014-07-16 16:06:18+02:00
File Type	JPEG
Make	Canon
Camera Model Name	Canon EOS REBEL T1i
Software	Adobe Photoshop Lightroom
Modify Date	02/12/10 05:51 PM
Exposure Time	132
ISO	100
Date/Time Original	02/08/10 04:15 AM
Create Date	02/08/10 04:15 AM
Shutter Speed Value	132
Focal Length	18.0 mm
Focal Plane X Resolution	5315.436242
Focal Plane Y Resolution	5342.32715
Compression	JPEG (old-style)
Thumbnail Offset	728
Thumbnail Length	10155
Color Space	sRGB
Aperture	4
Flash	Off, Did not fire
Image Size	1000x761
Shutter Speed	132
Focal Length	18.0 mm
Hyperfocal Distance	4.28 m
Light Value	-3

the identification of relevant details within depicted event. The interpretation regards the contextual analysis of the image to understand what is happening, the subjects/objects involved in the event, the depicted environment and all the semantic information derivable from a human inspection; the identification of relevant details is referred to the visual anomalies both on the signal level (e.g., block or color artefacts) and scene level (e.g. light direction, shadows or perspective). This phase may help the FA in developing the best strategy for further analysis.

Source Identification The source identification process aims to recover information about the source device (e.g. the camera or the recorder) of the inspected image. A first classification consists in determining whether the image comes from a camera, a scanner, a mobile phone or it has been generated using computer graphics [41]. Furthermore, techniques also exist to discriminate between different camera brands or even models [42]. Anyway the most attractive is the so called *image ballistic*: given a group of cameras, even of the same brand and model, the goal is to determine which one was used to capture the image.

Authenticity Assessment The authentication problem addresses the task of establishing if the image is an accurate rendition of the original event. The criteria to define what is an accurate rendition is linked to the specific analysis context. For instance, the contrast enhancement of a face could be considered a forgery in a photographic contest; on the contrary, the same processing can support the correct interpretation of the original event in a surveillance video. In this phase signal and scene traces are investigated through the forensic tools to determine the whether the image has been tampered and in case, where and how. In fact an image can be globally a forgery, i.e., is computer generated, or locally spliced with another image. Recent researches also showed that benefits can be also obtained by using several tools together in a synergic way. Existing forensic tools are far from ideal and often give uncertain or even wrong answers, so, whenever possible, it is wise to employ more than one tool searching for the same trace. Furthermore, it may also be the case that the presence of one trace inherently implies the absence of another, because the traces are mutually exclusive by definition (e.g., aligned and not-aligned double compression). In this case a decision fusion strategy [43] can be exploited to merge the output of several tools into a final decision.

Content Enhancement/Analysis The FA exploits the image forensics tools to extract content information and enhance the intelligibility of the image. In visual content the process regards evaluations about people, objects and the environment of the depicted scene. A non-complete list of interventions is: signal enhancement to reveal details, extraction of dimensional relations or parameters (such as the height of a subject), photographic comparison to link known objects with something depicted in the scene.

Results Analysis The results achieved in each step of the analysis may be too weak or sometimes misleading due to the unreliability of tools under certain environments. Furthermore, by means of the current counter-forensics technologies, expert users may deceive forensics tools and mislead the FA. Results analysis aims at putting into relation the different outputs coming from each analysis step to avoid errors and produce more complete, accurate and robust conclusions. In fact, while it may be easy for a skilled, possibly "forensic aware" image retoucher to conceal some traces of his work, it would be far more difficult for him to fool an heterogeneous set of analysis tools that account for many different traces. Then, chances for the analyst to reveal the manipulation increase significantly when many different clues are put together, making the "perfect crime" much harder to accomplish. As an example, the results achieved in the source identification phase can be related to some Exif, e.g., *Make*, *Camera Model Name*, *Software*, *Focal Length* to determine their coherence. Generally speaking the reliability of the results is stronger if it is sustained by different kind of analyses

Reporting The process of reporting the results involves the communication of scientific considerations to a legal part. A clear and correct reporting of the applied methodology must satisfy lots of requirements otherwise the whole analysis could be invalidated. Such requirements essentially regard the validation of the digital data and of the applied Multimedia Forensics technologies in the legal procedure. In order to make the image accepted as digital data the FA must guarantee that the whole chain of custody for digital data has been respected during the whole investigation process. International Standard ISO/IEC 27037 can be used as reference for such a purpose. On the other hand the validation of the Multimedia Forensics technologies is not clear yet. Few general guidances come from ISO/IEC 27041 in the assurance for digital evidence investigation methods; some other principles (*Daubert* principles) regarding the admissibility of expert witnesses testimony come from United States. They provide a set of general observations that are considered relevant for establishing the validity of scientific testimony:

- Empirical testing: whether the theory or technique is falsifiable, refutable, and/or testable
- Whether it has been subjected to peer review and publication
- The known or potential error rate
- The existence and maintenance of standards and controls concerning its operation.

- The degree to which the theory and technique is generally accepted by a relevant scientific community.

Anyway there's no clear standard for the validation of the applied Multimedia Forensics technologies (and technological tools in general).
In the next section we apply the proposed methodology in a case study.

3.4 A Case Study

In this Section we present a practical case study, where we walk through the steps of the proposed methodology. We received a set of digital images that were seized by the police and had to be used within a trial; we were asked to determine whether images underwent manipulations such as insertion or removal of objects, possibly locating manipulated regions. In the following we focus on one JPEG image (see Fig. 3.2) for sake of brevity, and show the methodology and the technological instruments we used to answer the questions about it.

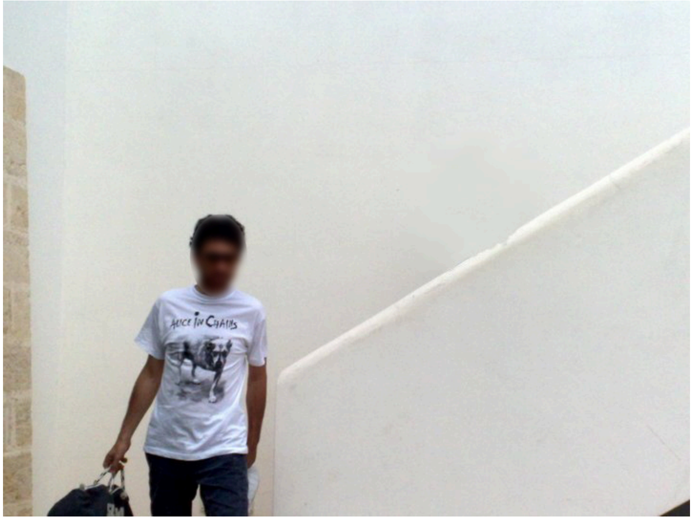


Figure 3.2: Case study image (face was blurred on purpose).

Metadata Extraction and Analysis We began our analysis from the first step, that is metadata extraction and analysis. By using the `exiftool` software [44], we extracted the information stored in the Exif header of the image. Unfortunately, no relevant metadata were present regarding the source device nor the acquisition/modification date nor the gps coordinate of picture.

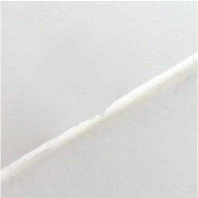


Figure 3.3: Wall detail.

Table 3.2: Output of forgery detection tools and their fused score according to [49].

Tool Name	**ROI$_1$**	**ROI$_2$**
Aligned JPEG [46]	65.5%	61.6%
Not-Aligned JPEG [48]	0%	8%
JPEG Ghost [47]	76.7%	68.2%
Fused score [49]	**72.8%**	**52.7%**

Visual Inspection By visual inspection some anomalies became evident: the edge of the wall is “crunched” near the center of the picture, and the grain of the picture above that region is much smoother compared to the rest.

Authenticity Assessment Following visual inspection, we resort to image forensics tools for forgery localization and detection. First, we use the tool presented in [45] for forgery detection and localization. The tool allows two kinds of analysis: forgery *localization*, where a map associating each 8×8 block of pixels to its probability of being tampered is generated using the forensics algorithm proposed in [46], and forgery *detection*, where the analyst manually selects suspect regions and the tool runs three different forensics algorithms [47, 46, 48], combining their decisions through a decision fusion engine tailored to image forensics [49]. Localization results are reported in Fig. 3.4: we see that the region that raised suspects actually contains inconsistent traces of double quantization compared to the rest of the background; moreover, the region corresponding to the person’s shirt drawing also shows anomalous traces; we will denote these two regions, respectively, with ROI$_1$ and ROI$_2$ in the following (Fig. 3.4). To further investigate the authenticity of these regions, we employ the forgery detection tool and the decision fusion system; results are reported in Table 3.2, where in the first three rows the scores of each used algorithm for both ROIs are shown, and in the final row the corresponding fused scores.

Results Analysis We obviously skip metadata analysis, since they were not present for the image at hand. If they were available, particular attention

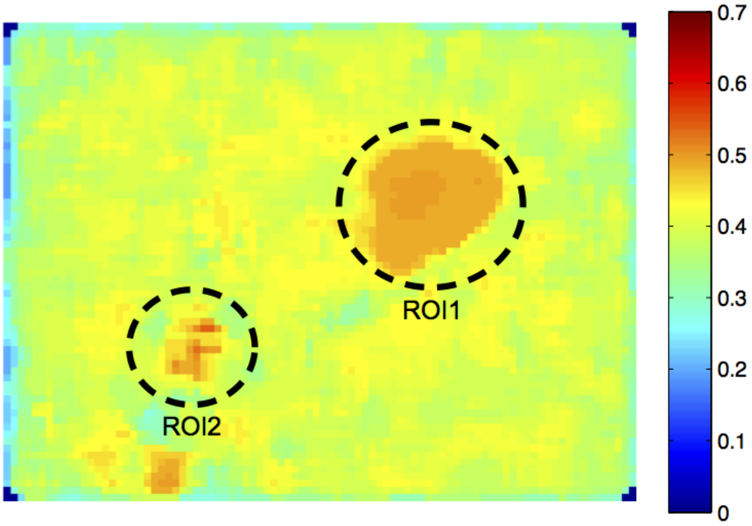


Figure 3.4: Forgery localization map produced by the tool in [46] for the case study image. Suspect regions are highlighted by a black circle.

would have been devoted to comparing the `DateTime` and `DateTimeOriginal` fields, since they are typically different when the image has been opened and re-saved. Another telltale tag is `Software`, which usually contains the name of the last software that wrote metadata. Turning to the visual inspection, the anomalies visible on the wall are compatible with an incautious use of the smudge tool, whose low-pass nature alters the natural image grain. Suspects about ROI_1 are also confirmed by the instrumental analysis: traces of double compression were not detected in that region while they were detected in the rest of the image. This is again consistent with the hypothesis that some processing tool was used on those pixels, removing traces of the previous JPEG compression. As to the other region (ROI_2), the output of other tools and the fused score is much lower (Table 3.2), and no anomalies are visible. By comparing the localization map and the visual content of the image, we can state that probably the region was signaled as suspect because it contains many bright-to-dark transitions, a situation that is known to make the analysis less reliable [46]. Therefore, there is not clear evidence of manipulation for that particular region.

Finally, based on our analysis, we can state that the complete analysis supports the hypothesis that the image was manipulated, probably by erasing someone/something that was present in ROI_1.

In the next Chapter we introduce our contributions to the forensic technologies for image analysis and investigation.

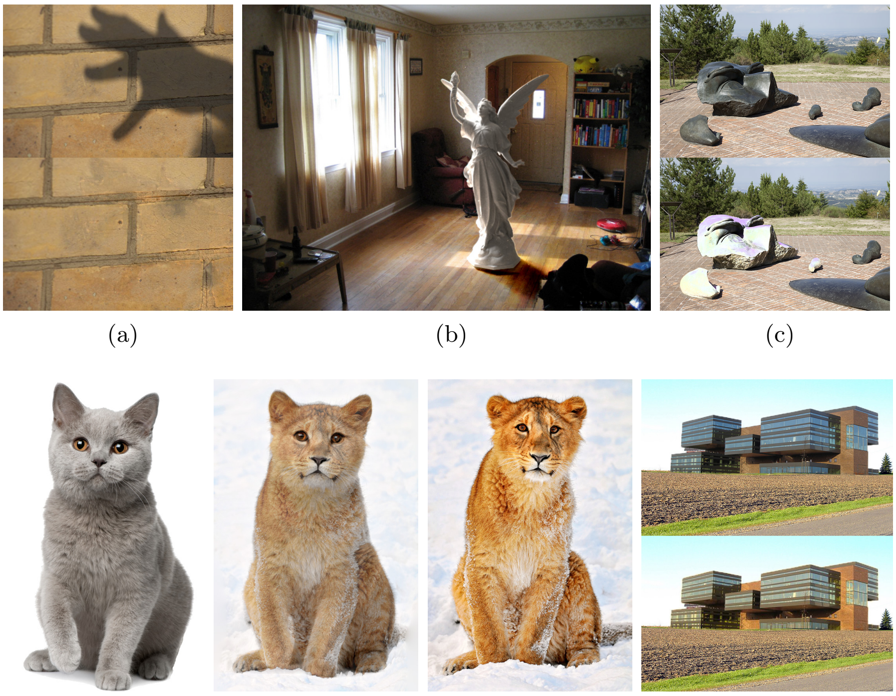
CHAPTER 4

Image Tampering Detection faces Image Composition

4.1 Introduction

Current research suggests that people are not very keen on discerning between real and edited pictures [50]. This poses a critical problem, as softwares such as Adobe Photoshop [51] and GIMP [52] allow anyone to easily create high-quality composites. In such a scenario, an arms race between forgers and forensics analysts is in progress [3]. While new and more sophisticated forges are being conceived, forensic techniques keep evolving to catch them. Most image manipulations, however, are neither malicious nor dangerous. There are plenty of legitimate reasons to edit images, such as for marketing and design. Unfortunately, sophisticated tools developed for these tasks can be used by forgers and the analysts have to struggle to catch up. In this section we analyze the current state of this arms race between the field of image forensics and image composition techniques. Here, *image composition* is used as an umbrella term for all techniques from areas such as computer graphics, computational photography, image processing and computer vision, that could be used to modify an image. More specifically, we discuss works that have the potential to either be used to perform or hide forges in digital images.

While many modern image-composition techniques could be used to make sophisticated forgeries, almost none of them have been scrutinized by forensic works. There is, however, a large body of forensic tools that could be used for this task. Next section (4.2) surveys and classify the image composition techniques, basing on the type of forgery they can perform, to identify the best strategies to analyze these novel forgeries. Then both qualitative and quantitative test are performed to assess forgery detection effectiveness against image composition.



(a) (b) (c)

(d) (e) (f) (g)

Figure 4.1: Different ways in which composition techniques can be used to alter images. a) Removing soft shadows [53]. The hand shadow from the top image has been removed on the bottom image. b) Inserting synthetic objects [54]. The marble angel in the picture is not real, it was rendered into the scene along with its complex light interactions. c) Performing edge-aware filtering [55]. The bottom image was filtered to perform a localized color editing on some of the stone statues. d-f) Morphing two different objects together to create a blend [56]. The cat in Figure 4.1e is a composite of the cat in Figure 4.1d and the lion in Figure 4.1f. g) Transferring an object from one image to another, adjusting its illumination according to the target scene [57]: the building was spliced on the field in the top image, and in the bottom it had its lighting adjusted to match the composition.

4.2 Image Composition Arsenal

The term “Image Composition” is used to encompass different fields such as Computational Photography, Image Processing, Image Synthesis, Computer Graphics and even Computer Vision. Recent works on all of these fields were surveyed to determine which ones could be used to aid in forgery. For this purpose, techniques that a forger could use to perform any form of operation were considered, from splicing to highly creative operations.

The techniques were classified in five general classes based on the type of forgery they could perform:

- **Object Transfering**: transfering an object or region from one image to another image, or even to the same image. This is the most common type of forgery, and encompasses both splicing and copy-and-paste operations. It is mainly divided into *Alpha Matting*, *Cut-Out*, *Gradient Domain*, *Structurally Changing* and*Inpainting*;

- **Object Insertion and Manipulation**: inserting synthetic objects into an image or manipulating an existing object to change its properties. It is divided into *Object Insertion*, *Object Manipulation* and *Hair*;

- **Lighting**: altering image aspects related to lights and lighting. It is divided into *Global Reillumination*, *Object Reillumination*, *Intrinsic Images*, *Reflections*, *Shadows* and *Lens Flare*;

- **Erasing**: removing an object or region from the image and concealing it. It is divided into *Image Retargeting* and *Inpainting*;

- **Image Enhancement and Tweaking**: this is the most general class of forgery, and is related to what is considered retouching on the forensics literature. It is divided into *Filtering*, *Image Morphing*, *Style Transfer*, *Recoloring*, *Perspective Manipulation* and Restoration/Retouching.

It must be noted that some of the surveyed techniques could be used to perform more than one type of forgery in the classification. Erasing, for instance, is often performed by copy-pasting regions of the image to conceal an object. In this sense, a technique under the *Object Transfering* classification can be also considered on the *Erasing* class.

In the following, we discuss each of the different forgery classes and their relation to the forensic traces and techniques.

4.2.1 Object Transferring

This class contains techniques that can be used with an end goal of transferring objects between images or in the same image. A fundamental task of transferring an object or region is defining its boundaries, and techniques that can help on making good contours are classified as *Cut-out* [58][59]. These techniques do not change the content from the source or target images, they only aid in selecting a pixel area.

Most techniques to detect splicing or copy-and-paste are well-suited against *Cut-out* forgeries, because the pixel content is unaltered. From a forensics point of view, well-defined boundaries on the transferred region reduce the amount of information being carried from the original image. This might alter some traces and affect the performance of techniques based on those traces [60]. A bad cut can also be easy to note visually, without the use of additional tools.

One of the main limitation of transferring objects by cut-and-paste is that transparency is ignored. Hair, thin fabrics, glass, and edges may contain a mix of colors from the foreground and background of the source image. This can cause visual artifacts on the resulting composition, and the presence of foreign colors that can be used for traces. *Alpha Matting* techniques can estimate the transparency of a region in the image, which can be used to better extract it from the source image, and then composite on the target image (Fig. 3e-h). The visual aspect is the most critical on the use of alpha matting for object transferring, as it blends colors on borders and transparent regions, making convincing forgeries. On most cases greater transparency is present only on a small part of the composition, such as borders. The majority of the composited area remains unaffected as a regular splicing. The most sophisticated object transferring techniques are *Gradient Domain* ones. These techniques aim to combine the gradient of the transferred object with the target image, making a complex blend. The simplest technique is Poisson Image Editing [61], which matches the gradients by solving a Poisson equation from the boundaries of the transferred region. The resulting object has different colors and gradient, blending with the scene. Poisson Image Editing, also commonly referred to as *Seamless Cloning*, spawned several works that improved its basic idea of solving differential equations for gradient matching on transferred regions [62][63].

Works such as Sunkavalli's [64] focus on the Laplacian Pyramid as the main component for sophisticated blends between images, being able to maintain the noise and texture of the target image (Figures 4.2e through 4.2h) to some degree. This kind of approach was generalized [65] and improved [66] by other authors.

Gradient Domain techniques can blend the whole transferred area and merge the images on a profound level. There are big variations on the inner workings of each technique, and the results are very dependent on the images to be combined. Furthermore, most of these techniques can be finely tuned. This makes them hard to be analysed from a forensics point of view. The safest way to detect forgeries of this kind would be focusing on high-level traces such as shadows and geometry. Light-based traces could help in cases where a full object is being transferred, because the resulting colors after the blending may create irregular lighting. When transferring parts of objects, such as changing faces on an existing head (Figure 4.2d), it is possible that the result can have plausible lighting and illuminant traces.

Object Transferring techniques are arguably the most relevant to the forensics community, because they can be used to perform both splicing and copy-pasting. Figure 4.2 shows an *Alpha Matting* (top row), and a *Gradient Domain*

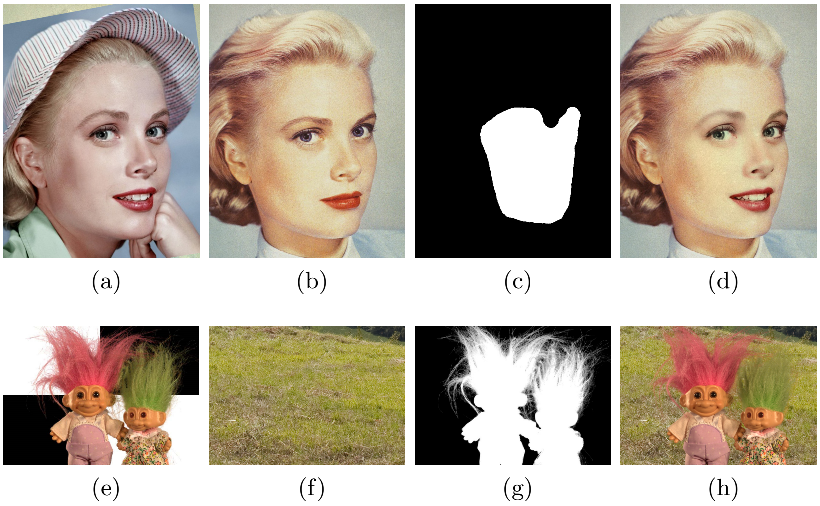


Figure 4.2: Example of splicing using object transferring techniques. The top row represents *Alpha Matting*, and uses the Shared Matting technique [67]. The bottom row corresponds to *Gradient Domain*, and uses Multi Scale Harmonization [64]. The source images are on the first column, the target images on the second colum, the transference masks are on the third column, and the final result is displayed on the fourth column for each technique.

(bottom row) splicing. Both forgeries are visually unnoticeable. Notice how the *alpha matte* (Figure 4.2g) contains very precise information about the transparency of each hair, and the mixture of colors on the final composition (Figure 4.2h). The *Gradient Domain* composition exemplified does not use transparency information (Figure 4.2c), but is able to transfer some of the color and texture of the target image (Figure 4.2b) into the transferred region of the source region (Figure 4.2a). The final result (Figure 4.2d) is a very convincing composition.

4.2.2 Object Insertion and Manipulation

Images are 2D projections of a 3D scene, with complex interactions of light and geometry. To insert a new object into the image, or to manipulate existing objects, the properties of the 3D scene must be known. This is a very challenging task. Techniques under this category focus on estimating characteristics of the 3D scene or its objects, providing means to alter them on a visually convincing way.

Rendering a synthetic object into an image is a simple task if the scene lighting and camera parameters are known. Additional knowledge about scene geometry also helps to increase realism. The challenge is to obtain this information from a single image. The most advanced techniques for object insertion, developed by Karsch, are able to estimate perspective, scene geometry, light sources and even occlusion between objects. In [68] heavy user input was needed to aid the parameter estimation, whereas in a second work (Figure 4.1b) [54] most input tasks were replaced with computer vision techniques to infer scene parameters.

The manipulation of objects in an image suffers from similar problems than insertion. Scene lighting, camera parameters and geometry are required for a visually convincing composition, and the geometry of the object being modified must be also known. A slight advantage in relation to rendering synthetic objects is that the photographic texture of the modified object can be used, providing a more photo-realistic touch. It is possible to perform resizing operations on objects without directly dealing with its 3D geometry [69], but most techniques will focus on modeling it.

The easiest way to work with the geometry of objects in an image is to limit the scope to simple primitives. Zheng [70] focus on cube-like objects, modeling them through "cuboid proxies", which allow for transformations such as scale, rotation, and translation in real time. Chen's work [71] uses user input to model an objects geometry through swipe operations. This technique works specially well on objects with some kind of symmetry, such as a candelabrum or a vase, and allows changes in the geometry itself. Another solution for dealing with object geometry is to use a database of 3D models, and find one that fits with the object depicted in the image [72].

Manipulating human body parts in images is a specially hard task, because human bodies vary greatly in shape, and clothes affect the geometry. This type of manipulation, however, is of special interest due to its applications in marketing photography and modeling. Zhou [73] uses a parametric model of the human body, and fits a photography to a warped 3D model, achieving a correspondence between body parts in the image and 3D geometry. This allows the reshaping of body parts, making a person in a picture look thinner, stronger, taller, etc. Hair manipulation is also a hot topic in image composition, with a special focus on changing hair styles after the picture has been taken [74][75].

Even though state-of-the-art techniques in image insertion and manipulation can create visually convincing results, they should not pose a problem for modern forensic techniques. Distinguishing between real and synthetic images is a very debated topic [76][77], and there are forensic techniques that focus on identifying them [78][79].

The weak point for this category of image composition is in the acquisition traces. The process of rendering a synthetic object is different from capturing it from a camera, so the acquisition traces should point to the manipulation, providing FD1 or FD2 results (see section 2.2.1). Similarly, when performing transformations on an object (scaling, rotating, deforming, etc.), its pixels have to be resampled, changing the acquisition traces. Resampling detection also could be used to obtain an FD3 result in these cases, while compression-based techniques could identify this type of manipulation if the original image was compressed. Kee has demonstrated that object insertion might be able to fool geometry-based lighting techniques [32], which could also extend to object manipulation. The reason for this is that the same lighting parameters estimated to verify the integrity of the scene were used to generate the composition.

4.2.3 Erasing

An manipulation is usually called erasing when an element of the image is intentionally removed or hidden, and not a consequence of other editing. This category is comprised mostly of Inpanting and Image Retargeting techniques.

Inpainting techniques are used to complete a region in an image, filling it with appropriate content [80]. By selecting a region that one wants erased as the region to be completed, inpainting can make objects disappear. Several works on inpainting are focused on stitching different parts of images together [81][82], or filling large gaps [83]. There are implementations of inpainting techniques already available on commercial editing sofware, such as Photoshop's Spot Healing Brush and Content Aware Fill tools. The main limitation of inpainting is filling regions with high amount of details, or using image features which are not local in the filling. Huang's [84] work is capable of identifying global planar structures in the image, and uses "mid-level structural cues" to help the composition process.

Image retargeting is a form of content-aware image resizing. It allows to rescale some elements in an image and not others, by carving seams in the image, i.e. removing non-aligned lines or columns of pixels [85]. The seams usually follow an energy minimization, removing regions of "low-energy" from the image. The objects and regions that have seams removed will shrink, while the rest of the image will be preserved. This can be used to remove regions of the image by forcing the seams to pass through certain places instead of strictly following the energy minimization. Most research on image retargeting focus on better identifying regions in the image to be preserved, and choosing the optimal seam paths [86][87].

Erasing manipulations should behave in a similar fashion to object insertion and manipulation, as the modified region will not come from a photograph, but from an estimation. This affects acquisition and compression traces, provided the original images were compressed. Image retargeting has already been analyzed from the point of view of image anonymization [88], and there is even a specific technique for its detection [89]. Detecting that a seam carving has been done in an image would constitute and FD4 in our scale.

4.2.4 Lighting

Lighting techniques are capable of changing the lighting of scenes [90] and objects [91][92], inserting light effects such as reflections [93][94], lens flare [95][96], and even manipulating shadows [97][98]. From a forensics point of view, lighting techniques are dangerous due to their potential of concealing other forgeries. After splicing an object in an image, for instance, a forger could add convincing shadows and change its lighting, making it harder for both human analysts and forensic techniques to detect it. Indeed, it is a concern in image composition when the source and target lighting conditions are different, and there are works focused on correcting this issue [99][100].

Due to the variety of lighting techniques, it is hard to make a general statement about them from a forensics point of view. As always, it seems plausible that at least an FD2 result can be achieved if compression is involved in the forgery. Techniques that add shadows or change the lighting in a visually convincing way, but do not account for all lighting parameters of the scene, could fail to deceive geometry and light-based forensics analysis. Specifically identifying light inconsistencies is an FD3 in our scale.

4.2.5 Image Enhancement/Tweaking

This is a broad classification for techniques that perform image modifications and are too specific to have their own category. Image morphing techniques [56][101] can fuse objects together, creating a composite that is a combination of them. Style transfer techniques are able to transform an image to match the style of another image [102], a high-level description of a style [103], or an image collection [104][105]. In the same vein, recoloring techniques can add or change the color of image elements [106], and even simulate a different photographic process [107].

Filtering techniques can be very flexible, allowing for a wide variety of effects. They can be used to remove noise or detail from images (Figure 4.1c) [108], [109], or even to add detail [55] while preserving edges. Different filters may be designed to obtain different effects. From a forensics point of view, filtering techniques can be used to remove low-level traces. A simple median or gaussian filter is able to remove compression and CFA traces, but it is easily detectable, as it softens edges. Edge-aware filtering, however, can be used to destroy such traces preserving edges. If used in a careful way, it can remove the aforementioned traces in a visually imperceptible way.

Perspective manipulation techniques allow an user to change the geometry [110], and perspective [111] of a scene, or to recapture an image from a different view point [112]. Its uses are mostly artistic and aesthetic, but these techniques could be used to forge photographic evidence. The final type of manipulation that will be discussed is Retouching. Retouching techniques aim to perform adjusts on image properties such as white balance [113][114], focus [115], or several at the same time [116]. They can also aid in performing adjustments in several images at the same time [117].

4.3 Experiments

In addition to the general analysis in Section 4.2, to understand how image composition affects forensics traces, we performed a set of experiments. This task was challenging since there are no available implementations for most composition techniques. Firstly, we performed test a more qualitative analysis considering a broad variety of techniques, and a more quantitative experiment focusing on a few state-of-the art techniques. The following two subsections discuss both phases in detail.

4.3.1 Qualitative Analysis

Firstly, to test on a broader scope how different image composition techniques affect forensics traces, we gathered images from 12 different works on image composition, either from the publication website or directly from the authors. Approximately 80 images generated with 9 different types of forgery described on Section 4.2 were studied. Our main goal was to analyze the images directly before and after the composition has been applied.

In particular, we applied forensics techniques that analyze traces of CFA [7], PRNU [8], Double JPEG compression [10], ELA, and high-frequency noise [1]; all these techniques generate as output a detection that can be used to visually identify if the composition had any outstanding impact on the corresponding traces. No objective conclusion can be achieved from the provided images considering their limited number, the different amounts of pre-processing and compression applied to them, the lack of knowledge about the tuning details of the algorithms. However, the most interesting results and their descriptions are shown in Figure 4.3. Then we focused the analysis on few available composition techniques and on a well studied tampering detection approach to provide quantitative results.

4.3.2 Quantitative Analysis based on JPEG Artifacts

For our quantitative experiment we focused on one of the most generally effective forensic approaches: image forgery localization via block grained analysis of JPEG artifacts, as proposed in [119]. This approach, by assuming that tampered images present a double JPEG compression, either aligned (ADJPG) or nonaligned (NADJPG), can be used to detect a suspect region. We replicated the experiments by considering the scenario where half of the image has undergone manipulation; but while in the original paper only splicing was

[1] https://29a.ch/photo-forensics

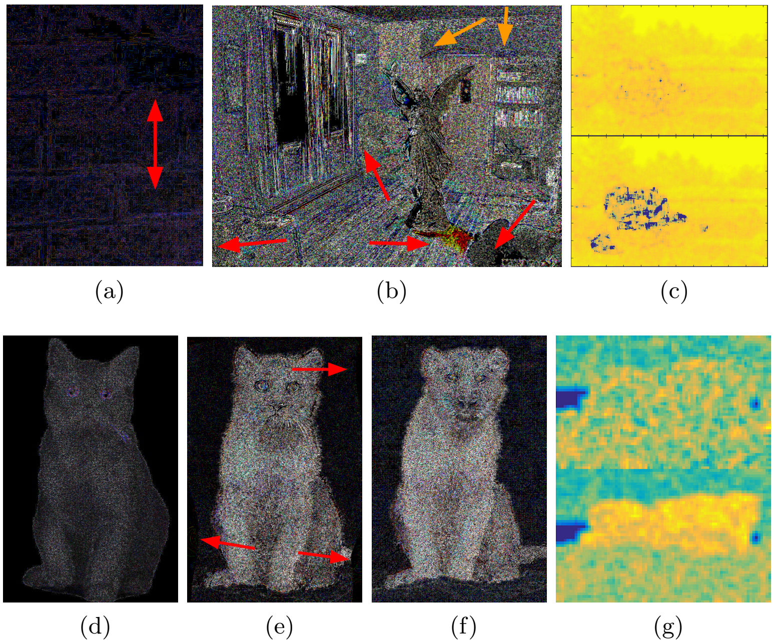


Figure 4.3: Results of analyzing different traces for the images on Figure 4.1. (a) ELA [118] of Soft Shadow Removal. In this case, it is not possible to identify any irregularity in the composited image. (b) Noise analysis of object insertion. The first identifiable irregularity is that the noise pattern for the shadow cast by the synthetic object greatly differs from other shadowed regions in the image (red arrows). The indirect illumination estimated after the scene's light interactions with the object appear as salient planes in the noise map (orange arrows). (c) PRNU analysis of localized recoloring. The more yellow, higher is the correlation between the region and the cameras sensor pattern noise. On the first image, there are some false positives thorough the image caused by high frequency areas. On the recolored image, the probability map shifts completely to the altered region. (d)-(f) Noise analysis of image morphing. The morphing process creates distinct warping artifacts on the noise pattern. (g) Double JPEG compression analysis of reillumination. The more yellow, higher the probability that the region has undergone double JPEG compression. While the top image shows a very noisy pattern, in the bottom image the uniform interpretation of a salient portion suggest that different compression traces (single and double) are present in the image.

considered, here we compared its performance considering three object transferring approaches: Splicing (SP), Alpha Composition (AC) [67] and Seamless Cloning (SC) [64].

Similarly to [119] we considered uncompressed TIFF images belonging to three different cameras (Nikon D90, Canon 5D, Lumix G2): 100 images were used for SP and AC while only 10 images for SC, due to its heavy computational cost[2]. They were acquired with the highest possible resolution and their cen-

[2] It must be noted that these ten base images actually produced 1100 sample test images.

tral portion 1024×1024 was cropped. Then the following steps were performed for each image to produce A-DJPEG artefacts: i) JPEG compression with QF_1 was applied, ii) the left half of the image was replaced with the original TIFF applying each different object transferring technique, iii) JPEG compression with QF_2 was applied. The NA-DJPG artifacts are produced by removing a random number of rows and columns between one and seven before step (ii). The QF_1 and QF_2 are taken from the sets $[50, 55, \ldots, 95]$ and $[50, \ldots, 100]$ respectively. We performed our analysis using 6 DCT coefficients. The results were evaluated using the area under the ROC curve (AUC) by varying QF_2 (exactly as defined in [119]): AUC usually assumes values between 0.5 (random classification) and 1 (exact classification). In the following we discuss the achieved results, that are summarized in Figures 4.4 and 4.5 for the aligned and not-aligned cases respectively.

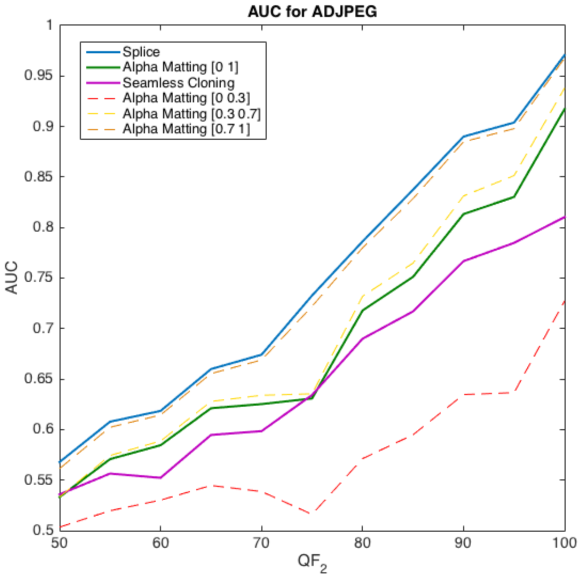


Figure 4.4: (Best viewed in colours): Performance comparison of the A-DJPEG analysis for different object transferring approaches. Dotted lines show the results of different alpha value ranges for alpha composition.

Alpha Composition: Since the result of the composition is strongly influenced by the value of α defining the transparency of the tampering pixel by pixel (see Section 4.2), to test all the possible outcomes we applied a linear transparency gradient mask from the bottom left to the upper right corner of the tampering (see example in Fig. 4.6a), with four different α ranges: i) $[0, 1]$ - average response; ii) $[0, 0.3]$ - high transparency; iii) $(0.3, 0.7]$ - mid transparency; iv) $(0.7, 1]$ - low transparency. The results confirm that both A-DJPEG and NA-DJPEG performance are strongly influenced by the α value:

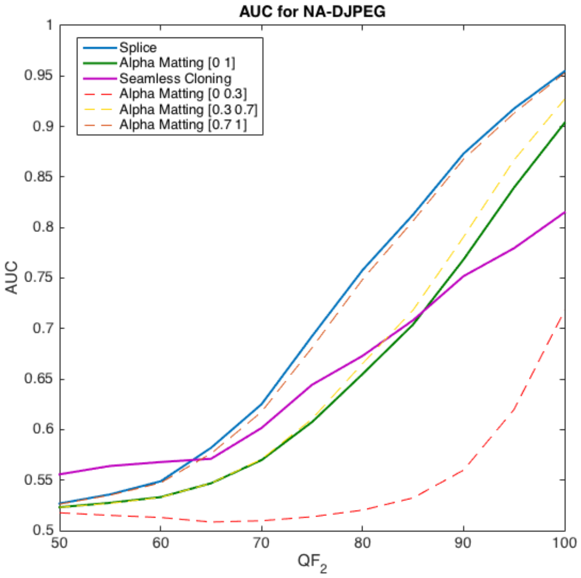


Figure 4.5: (Best viewed in colours): Performance comparison of the NA-DJPEG against splicing, alpha composition and seamless cloning tampering. Dotted lines show the impact of tampering transparency on the performances.

transparent objects can be hardly detected unless the last compression is really slight. Conversely, in case of low transparency objects, there is no real difference between SP and AC. Considering that, in most real cases, high transparency is applied only on a small percentage of the composition (like borders or hair), we expect that the use of this technique would not degrade the performance of the detection.

Seamless Cloning: The multi-scale technique allows to transfer the appearance of one image to another. It aims to to harmonize the visual appearance of images before blending them. Furthermore seamless boundary conditions are imposed to produce a highly realistic result. To exploit the peculiarity of this technique, the tampering region was slightly reduced according to the mask frame shown in Fig. 4.6b. The achieved results show that, similarly to the SP case, the detector produces an almost random output when the second compression is too strong. Anyway, when QF_2 is high, the detector is still able to detect the tampering, although with lower accuracy with respect to the SP case.

4.4 Conclusions

In this section we surveyed the fields of Image Composition and we assessed the applicability of an effective forensic technique for splicing detection

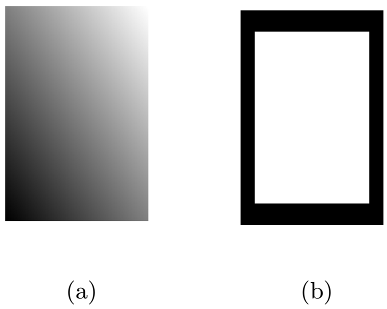


Figure 4.6: (a) One of the considered gradient transparency mask (α from 0 to 1) applied for AC composition; (b) the mask frame adopted to reduce the tampering region for SC composition.

against different kind of object transferring techniques quantifying how the performances are affected both on the kind of artifact (aligned or not-aligned double compression) and the parameters introduced by the composition technique (e.g., transparency factor in alpha composition).

A natural extension for this work would be to increase the number of techniques surveyed and tested, considering other traces and forensic approaches. Since both fields are in an "eternal arms race" the list of available works to be compared will keep increasing each year.

CHAPTER 5

Geometric Based Tampering Detection

It is well known that, when an image has undergone several processing as filtering and compression (e.g., it has been uploaded to a social network or modified with in-camera apps), the performances of most tools based on signal-level traces may be strongly affected, in an unpredictable way. Furthermore, when the image has been strongly compressed in the end of its life cycle, most of these traces are ruined by the quantisation operations. In these cases, techniques based on scene-level traces, e.g., geometry, have proved to be much more robust to common processing. In this chapter we investigate two techniques based on geometrical properties of the image. Specifically, in Section 5.1 we introduce the geometrical model for the mapping of the 3D scene on the image plane; in Section 5.2 and 5.3 we discuss two techniques for Splicing and Cropping Detection respectively. The first is based on a general perspective constraints, the second on image principal point shift introduced by image cropping. For both techniques we also investigate their applicability on images exchanged through a social network.

5.1 Introduction to Pinhole Model

A digital image is the outcome of a 3D world scene mapped onto a 2D plane. This process can be modelled by means of the *pinhole model* consisting in a central projection of space points onto a plane.
Let us consider a Euclidean coordinate system and the plane $Z = f$, called *image plane* or *focal plane*. As shown in Fig. 5.1, a world point $\mathbf{X}$ is mapped to the point on the image plane where a line joining the point $\mathbf{X}$ to the centre of projection meets the image plane. The center of the projection is usually

called *camera centre* or *optic centre*. The line from the camera centre perpendicular to the image plane is called the *principal axis* of the camera, and the intersection between the principal axis and the image plane is called the *principal point*.

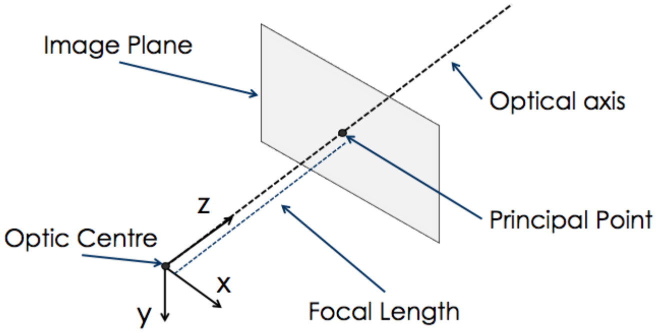


Figure 5.1: Pinhole Model

If we represent the world and image points by homogeneous vectors, the central projection of the 3D scene onto the image plane can be simply expressed as a linear mapping [35]:

$$\mathbf{x} = K[I|0]\mathbf{X} \tag{5.1}$$

where $\mathbf{X} = (X, Y, Z, 1)$ and $\mathbf{x} = (x, y, 1)$ are the homogeneous coordinates of world and image points respectively, whereas K is the *camera calibration matrix* containing the *internal* camera parameters. The general form of K is

$$K = \begin{bmatrix} f_x & s & p_x \\ 0 & \rho f_y & p_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{5.2}$$

where f is the *focal length*, while the aspect ratio ρ and skew s take into account the actual shape of a pixel. Lastly, (p_x, p_y) are the coordinates of the principal point.

Modern cameras have reached a high level of quality, with unity aspect ratio and zero skew. So, without significant loss of accuracy, the K matrix can be often modelled with $\rho = 1$ and $s = 0$, passing from 5 to 3 degrees of freedom. In the proposed model, the camera is assumed to be located at coordinate system origin with the principal axis of the camera pointing straight down the Z-axis with the points expressed in this coordinate system, called *the camera coordinate frame*. In general, points in space can be expressed in terms of a different Euclidean coordinate frame, known as the *world coordinate frame*.

The two coordinate frames are related via a rotation R and a translation t. In this case the general projection rule takes the form:

$$\mathbf{x} = K[R|t]\mathbf{X} \tag{5.3}$$

with $[R|t]$ defining the *extrinsic matrix*. In general we refer to the camera matrix $P = K[R|t]$, which is usually called *projection matrix*.
In the following section we briefly review the theory behind the vanishing points that are required for the proposed applications.

5.1.1 Vanishing points estimation

It is well known that the perspective image projection of parallel lines in the 3D world intersect at a vanishing point (VP). It can be easily shown that the vanishing point $\mathbf{v}_d$ for a 3D direction $\mathbf{d} = (d_x, d_y, d_z)^\top$ —expressed in a coordinate frame with its origin in the camera center and its Z-axis coincident with the optical axis—is

$$\mathbf{v}_d = K\mathbf{d} \tag{5.4}$$

In a practical scenario, if more than two concurrent lines are available, their intersection will not be unique (see Fig. 5.2)—since noise can perturb the image line detection—and the VP have to be estimated with an optimization algorithm. In our experiments we employ the solution reported in [35], Chapt 8, where, after initializing the VP by solving a linear least square problem, a non-linear optimization is carried out.

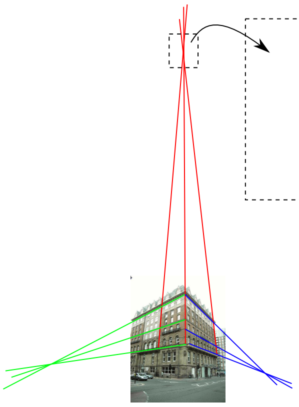


Figure 5.2: (Best viewed in color) In red, green and blue three sets of image lines corresponding to orthogonal 3D directions. Since noise can perturb the line orientations the intersection can be not unique, as shown in the magnified area.

5.2 Splicing Detection based on General Perspective Constraints

In this section we present a method for forgery detection based on perspective constraints; similar techniques have been proposed in the past but they are effective only when the image is captured with no tilt and no roll thus been unusable in most natural scenes. Here, this solution is extended to include these cases, and we show its applicability even when the image is exchanged through a social network (specifically Facebook and Twitter) where the image is subjected to heavy compression and resizing. This section is organized as follow: in Section 5.2.1 recent works based on geometric traces are reported; in Section 5.2.2 we summarize the method proposed in [120] to detect spliced subjects in low perspective image; in Section 5.2.4 we introduce the theoretical model to obtain height ratio in general perspective images; in Section 5.2.5 the proposed method is explained in details Finally, in Section 5.2.6 experimental results are shown on native images and on the same images exchanged through social media platforms, namely Facebook and Twitter. Finally, in Section 5.2.8 some conclusions are drawn.

5.2.1 Related Works

In last years different kinds of geometric constraints have been exploited to expose spliced images: in [121] the authors demonstrate that in presence of translation of a person or of an object, the principal point (the projection of the camera center onto the image plane) is shifted proportionally. Differences in the estimated principal point across the image can then be used as evidence of manipulation. The manipulation may concern not only the splicing of people and objects but also the tampering of other details in the picture. For instance a text on a sign or billboard is relatively easy to do in a perceptually convincing way.

In [122] Conotter et al. show how to determine whether the depicted text precisely satisfies the geometric mapping of a plan under perspective projection. Any deviations from the model are exploited to expose the tampering.

In [120] a perspective-constraint method is proposed to detect image splicing. The method is based on the computation of the height ratio of two subjects in an image starting from the vanishing line of the plane on which both subjects of interest are placed, and the vertical vanishing point, whereas the knowledge of camera parameters is not required. The authors observe that while pasting a subject into an image, it is difficult to properly size it in such a way to respect the principles of visual perspective. Then, if the estimated ratio exceeds a tolerable interval, then it is revealed that one of the two subjects was spliced into the scene. Unfortunately this detection method can be applied only if the picture is taken with no tilt and no roll, resulting almost useless on many natural images.

We propose then a generalization of this last method in general perspective condition. Furthermore we test its effectiveness on images that have been downloaded from most relevant social network, namely Facebook and Twitter, thus proving its applicability in social scenario and its robustness to compression.

5.2.2 Height Ratio Estimation in low perspective images

Yao et al. [120] describe a method to determine whether two subjects in an image have proper relationship in size satisfying the perspective rules. This is done by estimating the ratio of their height in uncalibrated scene and by checking its consistency with respect to the supposed known ratio.
In the following we describe the procedure pushing back the theoretical model that will be deeply analyzed in the next section.
The authors consider a simplified scenario in which the two subjects, namely A and B are placed on the same plane (called reference plane) and the scene is taken with no roll or tilt of the camera so that the optical axis is parallel to the reference plane. In this case the ratio of their respective heights, say $\frac{Z_A}{Z_B}$, can be easily estimated from the vanishing line of the reference plane and is independent of the camera's intrinsic parameters. The procedure is composed by four steps: objects selection, vanishing line detection, height ratio computation and consistency measure.

Objects selection The user manually selects in the image coordinates the top and bottom of the two subjects to be checked, obtaining on the vertical axis v the points having coordinates (t_A, b_A) and (t_B, b_B) respectively (as shown in Fig. 5.3);

Vanishing line detection With no tilt or roll the vanishing line VL of the reference plan is horizontal and thus can be determined from a single vanishing point having ordinate v_0. Since usual scenes contain several lines that in the real world are parallel (like road sides, buildings, furniture), the vanishing point can be obtained from the intersection of these lines in the image plain.

5.2.3 Height ratio estimation

Given the above coordinates, the height ratio β of the two subjects can be easily estimated (see Eq. (7) in [120]) as:

$$\beta = \frac{Z_A}{Z_B} = \frac{(t_A - b_A)(v_0 - b_B)}{(v_0 - b_A)(t_B - b_B)} \tag{5.5}$$

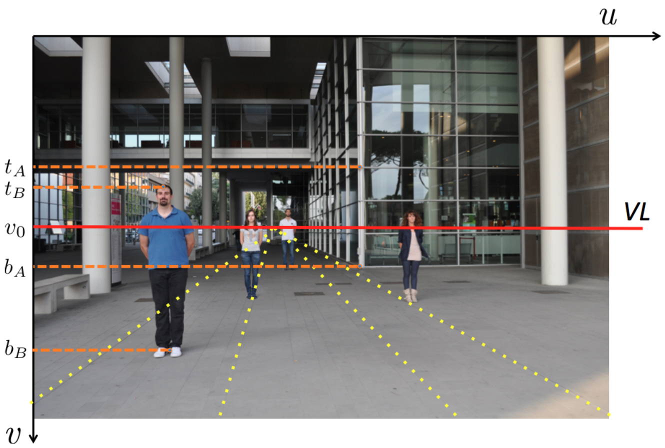


Figure 5.3: Two subjects, A and B, are placed on the reference plane. t_A, b_A, t_B, b_B represent their top and bottom ordinates in the image coordinates; v_0 is the ordinate of the vanishing line (VL) of the reference plane.

Consistency measure The estimated ratio β is then compared with the correct ratio α obtained by some a priori knowledge about the real size of the subjects under analysis or indirectly derived from some reference object with known height having the same depth of the two targets. In absence of tampering, the estimated ratio obeys the Gaussian distribution with a mean equal to the correct ratio α, such that $(\beta - \alpha) \sim N(0, \sigma^2)$. Then it is possible to define a consistency measure C as

$$C = 2F(-|\beta - \alpha|; 0, \sigma^2) \tag{5.6}$$

where $F(x)$ is the cumulative distribution function of $(x - \alpha)$. As it is defined, the measure $C \in [0, 1]$, reaches its maximum when $\beta = \alpha$ (when the estimated and expected value are exactly the same) while decreases when the difference between β and α increases. Thus, if $C < T$, where T is a properly defined threshold, then it is derived that one of the subjects under analysis is doctored.
According to the authors of [120], this method is effective in tampering detection even with low tilt or roll of the image but cannot handle images captured under general perspective. In the next section, we show how the estimation can be extended to more general conditions.

5.2.4 Height Ratio Estimation in uncalibrated scenes

Let us consider a reference plane with the X and Y-axes spanning the plane, and a direction Z not parallel to the plane (see Fig. 5.4). Then the first three columns of the projection matrix P represent the vanishing points of the directions X,Y and Z respectively (see Lemma 1), so that $\mathbf{v}_X = \mathbf{p_1}$,

$\mathbf{v}_Y = \mathbf{p_2}$ and $\mathbf{v}_Z = \mathbf{p_3}$. It can be easily shown that we can set $\mathbf{p_4} = \frac{\mathbf{l}}{||\mathbf{l}||} = \bar{\mathbf{l}}$ with $\mathbf{l} = \mathbf{v}_X \times \mathbf{v}_Y$ (because in natural photos the mapping of the reference plan $Z = 0$ onto the image is not degenerate so that the homography defined by $[\mathbf{p_1}, \mathbf{p_2}, \mathbf{p_4}]$ is full rank).
Therefore the projection matrix can be rewritten as:

$$P = [\mathbf{v_x}\ \mathbf{v_y}\ \alpha\mathbf{v_z}\ \bar{\mathbf{l}}] \tag{5.7}$$

with α the factor defining the scale of the projection.

Lemma 1. *Given the world reference system in the space XYZ and $\mathbf{v}_X$, $\mathbf{v}_Y$, $\mathbf{v}_Z$ the vanishing points corresponding to the directions of X, Y and Z axes respectively. Let $P = [\mathbf{p_1}\ \mathbf{p_2}\ \mathbf{p_3}\ \mathbf{p_4}]$ the projection matrix. Then $\mathbf{v}_X = \mathbf{p_1}$, $\mathbf{v}_Y = \mathbf{p_2}$ and $\mathbf{v}_Z = \mathbf{p_3}$.*

Proof. A world line can be parameterized as $X(\lambda) = A + \lambda D$, with A a point on the line and $D = (d^T, 0)$ a direction in the space. X spans all the points on the line, including the point to the infinity D when $\lambda \to \infty$. The world point $X(\lambda)$ is mapped on $x(\lambda)$ by the projection rule

$$x(\lambda) = P(A + \lambda D) = a + \lambda PD \tag{5.8}$$

and the vanishing point for direction D as

$$\lim_{\lambda \to \infty} x(\lambda) = PD. \tag{5.9}$$

Then, the vanishing point $\mathbf{v}_X$ of the direction $D_X = (1, 0, 0, 0)^T$ is

$$\mathbf{v}_X = [\mathbf{p_1}\ \mathbf{p_2}\ \mathbf{p_3}\ \mathbf{p_4}] \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \mathbf{p_1} \tag{5.10}$$

and similarly that $\mathbf{v}_Y = \mathbf{p_2}$ and $\mathbf{v}_Z = \mathbf{p_3}$. □

Let us consider a target A placed on the reference plane, such that its base and top affine coordinates in the 3D space are $\mathbf{X}_A = (X_A, Y_A, 0)$ and $\mathbf{X}'_A = (X_A, Y_A, Z_A)$ respectively. If P is the projection matrix, these points are mapped onto the image points through $\mathbf{x}_A = P\mathbf{X}_A$, $\mathbf{x}'_A = \mathbf{X}'_A$.
As shown in [123] the height Z_A of the target A can be determined up to the scale factor α as

$$\alpha Z_A = \frac{||\mathbf{x}_A \times \mathbf{x}'_A||}{(\bar{\mathbf{l}} \cdot \mathbf{x}_A)||\mathbf{v}_Z \times \mathbf{x}'_A||} \tag{5.11}$$

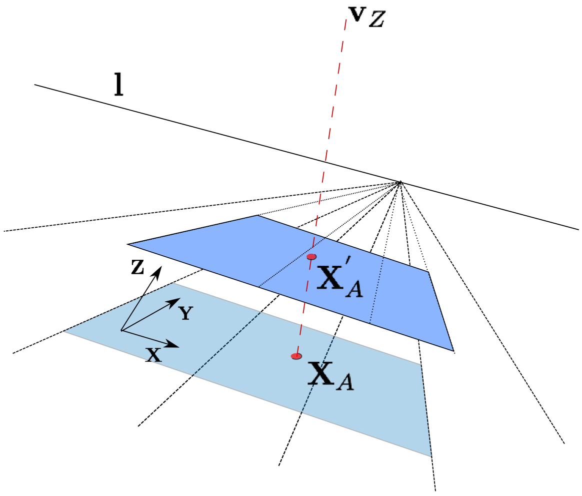


Figure 5.4: World reference system XYZ with vertical direction Z not parallel to the reference plane XY, the vertical vanishing point $\mathbf{v}_Z$ and the vanishing line of the reference plane $\mathbf{l}$.

where $(\cdot)$ and $(\times)$ are scalar and cross products respectively.
Then, given a second target B on the reference plan, height ratio $\frac{Z_A}{Z_B}$ can be easily determined using eq. (5.11) as

$$\frac{Z_A}{Z_B} = \frac{||\mathbf{x}_A \times \mathbf{x}'_A||}{||\mathbf{x}_B \times \mathbf{x}'_B||} \frac{(\bar{\mathbf{l}} \cdot \mathbf{x}_B)||\mathbf{v}_Z \times \mathbf{x}'_B||}{(\bar{\mathbf{l}} \cdot \mathbf{x}_A)||\mathbf{v}_Z \times \mathbf{x}'_A||} \tag{5.12}$$

Equation (5.12) shows that the height ratio between targets A and B, both placed on the reference plane, can be computed from their image coordinates, if we assume the knowledge of the vanishing line of the reference plane and the vanishing point of the vertical direction.

5.2.5 Proposed Method

The proposed method can be decomposed (as shown in Fig. 5.5) in two main steps: i) an user interaction step, consisting in the selection of points to determine the vanishing line of the reference plane, the vanishing point of the vertical direction and the targets borders, ii) the automatic procedure in which the height ratio is estimated and it's consistency is evaluated.

Vanishing line/points detection First step is the estimation of the vanishing line $\bar{\mathbf{l}}$ of the reference plane and the vertical vanishing point $\mathbf{v}_Z$. The

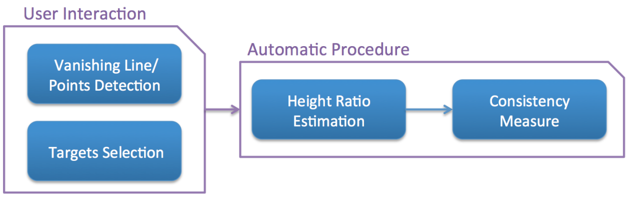


Figure 5.5: Scheme of the proposed method: in the user interaction step the needed image point coordinates are acquired, while in the automatic part the height ratio is computed and the consistency measure derived.

vanishing line can be commonly identified by the cross product of the vanishing points of two non parallel directions of the reference plane (e.g., $\mathbf{v}_X$ $\mathbf{v}_Y$); geometric elements or drawings present on the reference plane can fit for this purpose. Similarly, $\mathbf{v}_Z$ can be determined from the intersection of vertical lines of buildings and furniture.
At least two lines are needed to estimate a vanishing point but if more are available, then their intersection can be robustly estimated as shown in [35]. In our experiments we choose the vanishing point as the one that minimizes the sum of its euclidean distances from the given lines.

Targets Selection Top and bottom of the two targets A and B are manually selected by the user, thus obtaining $\mathbf{x}^{'}_A$, $\mathbf{x}_A$, $\mathbf{x}^{'}_B$ and $\mathbf{x}_B$. Each couple of top and bottom should be aligned with $\mathbf{v}_Z$ (being the target aligned to the vertical direction), then the points selected by the user have to be corrected to satisfy the geometric constraint. In our experiment, given a couple $\mathbf{x}'$, $\mathbf{x}$, we detect the correct points $\mathbf{y}'$, $\mathbf{y}$ by solving

$$\min_{\mathbf{y}',\mathbf{y}}[||\mathbf{x}' - \mathbf{y}'||_2^2 + ||\mathbf{x} - \mathbf{y}||_2^2] \tag{5.13}$$

subjected to the alignment constraint

$$\mathbf{v}_Z \cdot (\mathbf{y}' \times \mathbf{y}) = 0 \tag{5.14}$$

Height Ratio Estimation Height ratio β can be estimated through Eq. (5.12) using the detected vanishing line $\bar{\mathbf{l}}$, the vertical vanishing point $\mathbf{v}_Z$ and the two corrected points $\mathbf{y}'$, $\mathbf{y}$ of both considered subjects.

Consistency Measure The computed ratio β is compared with the supposed ratio α and the consistency factor C is obtained through Eq. (5.6). Similarly to [120] we set $\sigma = 0.1\alpha$. Finally, given a threshold T, the image is classified as tampered if $C < T$.

5.2.6 Experiments

The proposed method has been tested on native images and then on the same images exchanged through social networks, namely Facebook and Twitter. All the pictures have been captured in a general perspective with tilt and roll of the camera with respect to the reference plane[1]. The subjects depicted in the scene consist both in objects and people thus taking into consideration the possible errors introduced by border selection on different types of targets. The dataset is composed by 7 high resolution images (6-8 Mpixel) containing both authentic and tampered elements, specifically:

- n. 5 images containing 4 authentic and 2 tampered elements;
- n. 1 image containing 6 authentic and 3 tampered elements (reported as an example in Fig. 5.6);
- n.1 image containing 6 authentic elements.

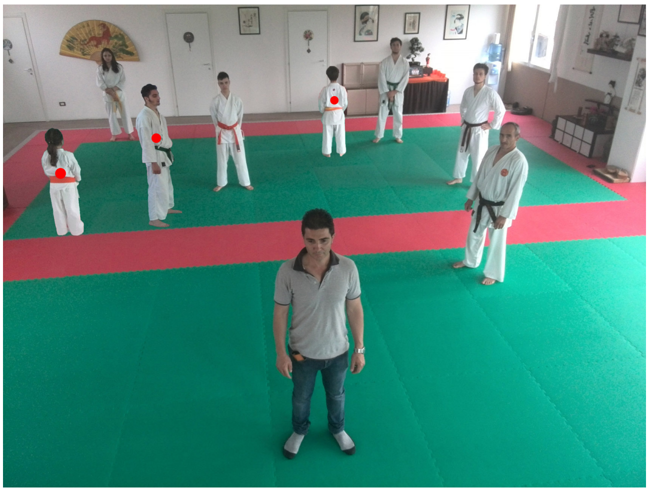


Figure 5.6: Example of image from the dataset used in the tests. Tampered elements are highlighted by red dots.

As shown in Table 5.1, with the considered dataset, 118 couples of subjects have been evaluated. Specifically 60 authentic couples (in which both elements are really depicted in the scene) and 58 tampered couples (in which only one of the two elements is really depicted into the scene while the other is tampered) have been tested. Considering that the performance of the method are influenced by user interaction in the vanishing points and border selection, the test has been carried out by three different users to average their introduced error.

[1] the dataset is available at `https://iapp.dinfo.unifi.it/index.php/english/materials_en/datasets_en`.

The height of each analyzed subject is known, so that the real ratios α_i, with $i = 1, ..., 118$, between any two elements can be exactly determined; then the same ratios have been estimated with the proposed method obtaining β_i. Finally the consistency factors C_i have been computed through the Eq. (5.6) for each couple of elements in the same image.

Authentic Targets	Tampered Targets	Number of Pictures	Autentic Couples	Tampered Couples
4	2	5	30 = 6x5	40 = 8x5
6	3	1	15	18
6	0	1	15	0
		TOT	**60**	**58**

Table 5.1: Details of the subjects considered in each picture and corresponding number of the estimated consistency factors evaluated during the tests.

In Fig. 5.7 we report the Receiver Operating Characteristic (ROC) curve showing the detection rate against the false alarm rate obtained by varying the threshold. The results show that a with a false alarm rate of 1.7%, the algorithm yields a detection rate of 98%, thus validating the effectiveness of the proposed idea.

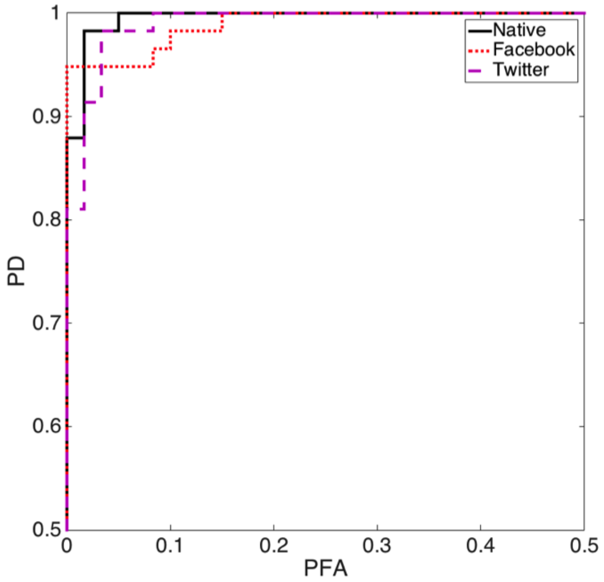
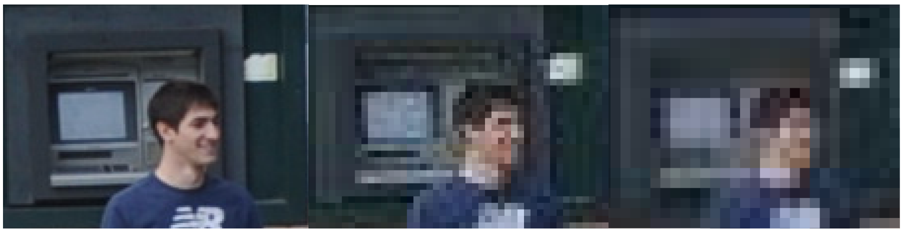


Figure 5.7: The ROC curve describing the performance of the method in terms of correct detection vs. false alarm probability.



(a) Original photo (b) Facebook version (c) Twitter version

Figure 5.8: Detail of one picture in its native version (a), Facebook version (b) and Twitter version (c).

5.2.7 Application to social network images

We tested the applicability of the proposed method in a social web scenario, specifically when the image under analysis is exchanged through Facebook and Twitter. The study has been carried out on images of the same dataset that have been uploaded on Facebook (choosing the worst quality) and Twitter, and then downloaded from them. In Table 5.2 we report an example of the signal degradation introduced by Facebook and Twitter processing on one image of the dataset. The quality factor (QF) of the compression has been estimated through Jpeg Snoop [124] and reported in the last column. An example of visual degradation introduced by the processing associated to the upload on the social platform is shown in Fig. 5.8.

Table 5.2: Signal degradation introduced into one image of the dataset by exchange through social networks

Type	Resolution	Compression	Estimated
	(MPixel)	(in %)	QF
Native	6.9	0	93
Facebook	0.4	94,2	77
Twitter	0.2	97,1	74

Table 5.3: Area Under Curve for Native, Facebook and Twitter Images.

Type	AUC
Native Dataset	0.9807
Facebook Dataset	0.9776
Twitter Dataset	0.9778

The test has been carried out again by three different users, and the results in terms of ROC have been reported in Fig. 5.7. The curves are almost unchanged with respect to the one obtained on the native dataset, thus showing that the method is robust to the degradation introduced by resize and compression. To have a quantitative evaluation, in Table 5.3 the Area Under the Curve (AUC) is reported for all the three cases.

5.2.8 Discussion

We proposed a method for forgery detection based on general perspective constraints and we showed its applicability even when the image is downloaded from a social network (specifically Facebook and Twitter) where the image is subjected to heavy compression and resize. This method improves a similar technique proposed in the past that is effective only when the image is captured with no tilt and no roll thus been unusable in most natural scenes. The method is based on the comparison of two targets placed on the same reference plane that is not always possible. Future works will allow the possibility of compare targets placed on parallel planes or to compare more than two targets to obtain more accurate results. Another issue that has to be investigated is the modelling of the uncertainty due to the user selection of vanishing lines and borders of the targets. This study is needed to assess the reliability of the achieved results.

In the next section we'll consider another physical trace, namely the image principal point and we'll focus on assessing the reliability of its estimation under different environments.

5.3 Cropping Detection based on Principal Point Estimation

In this section we consider another scene-level trace, namely the image principal point (PP). Forensic community developed techniques, based on PP, to expose both image splicing and cropping [121, 125, 126]. Here, focusing on the latter application we evaluate the reliability of its estimation under different conditions, showing how state of the art results may be affected. In particular, we studied the estimation of the PP, by exploiting vanishing points related to three mutually orthogonal directions [35]. Several tests were performed, on synthetic as well as real images, by varying both the point of view—so as to obtain different perspective conditions—and the number and position of the extracted features. A critical study of the obtained results led us to define a novel feature, referred to as *Minimum Vanishing Angle* (MVA), allowing us to measure the reliability of the PP identified into the image under analysis. Using the MVA concept, it's possible to establish a criterion to select lines on the image to achieve better performance. Specifically, one should just care about to choose lines providing the widest possible MVA, since the accuracy of PP estimation relies on MVA amplitude rather than on the number of image lines used.

The section is organized as follows: in Section 5.3.1 the related works are reported and in Sect. 5.3.2 we briefly review the theory behind the adopted PP estimation method. In Section 5.3.3 we introduce the MVA and its relation with the image perspective conditions. Then in Sect. 5.3.4 an in deep analysis of the reliability of the method is provided. Section 5.3.5 presents the forensic applications of the PP for cropping detection. Section 5.3.6 concludes summarizes the contributions in light of the achieved results.

5.3.1 Related Works

The estimation of the image PP from a single image is a known issue in computer vision and photogrammetry, usually embedded into the camera calibration problem, where the intrinsic parameters of the camera taking images of the scene (including focal length, principal point, pixel skew and aspect ratio) need to be estimated [127, Chapter 2]. In order to calibrate the camera, accurate off-line techniques usually require a known pattern into the scene [128, 129]. Other methods use video sequences or multiple images to self-calibrate the camera while solving the Structure from Motion problem [130]. In addition, solutions that exploit specific characteristics of the scene or particular objects in it have been proposed. In [131], coaxial circles are extracted from objects in the scene and used to estimate the calibration. In [132, 133, 134], the structural layout of a Manhattan World scene [135] is used instead to recover the intrinsic parameters by detecting the vanishing points corresponding to mutually orthogonal directions.

All the reported methods assume to use genuine images only, without any malicious modification. This hypothesis allows the authors to impose constraints on the parameters to ease and improve the estimation (for example, the PP is often initialized in the image center). Obviously, in the forensic application scenario, this assumption doesn't hold: no a priori information on the parameters can be supposed in the estimation process. Moreover, we have to typically deal with single images already acquired, so it's often impossible to employ solutions that require multiple images or a calibration pattern in the scene. The only viable approach is to find useful characteristics already present in the images. Given the abundance of line elements in pictures of real scenes, we focus on techniques based on vanishing points computation.

Given these difficulties, forensic literature presented only few methods that try to exploit the camera PP as a cue for tampering detection. In [121], the authors presented a method based on the estimation of the homography mapping a person's eyes to the image plane. Then, the PP is recovered by homography decomposition (supposing focal length is known) and exploited for splicing detection. A similar approach, that exploits circles in the image to obtain the PP position, is presented in [125]. In [126], the authors notice that asymmetric cropping of an image introduces a correspondent shift of the principal point. Hence, the distance between the estimated PP and the image center can be exploited as evidence of cropping. Slightly different, but still related to this topic, is the approach described in [136] where, instead of estimating the PP, tampering detection is based on the direct observation of the vanishing points of different 3D structures (e.g. buildings).

5.3.2 Principal Point Estimation

The image PP can be estimated by determining three vanishing points corresponding to three orthonormal directions. Let $\mathbf{d}_1$, $\mathbf{d}_2$ two orthogonal directions in the 3D space and $\mathbf{v}_1$, $\mathbf{v}_2$, their correspondent vanishing points respectively. Using Eq.(5.4), it holds

$$0 = \mathbf{d}_1^\top \mathbf{d}_2 = (K^{-1}\mathbf{v}_1)^\top (K^{-1}\mathbf{v}_2) = \mathbf{v}_1^\top \omega \mathbf{v}_2 \quad , \tag{5.15}$$

where $\omega = (KK^T)^{-1}$ is the *image of the absolute conic*, depending on the three camera parameters f and (p_x, p_y). Given three vanishing points corresponding to three orthogonal directions, we can thus define three independent constraints and finally estimate ω by solving a linear homogeneous system. Eventually K can be obtained using the Cholesky factorization of ω, from which both focal length and principal point can be estimated [35].

Summarizing, the estimation of the PP on a single image requires three main steps: (1) selection of three groups of concurrent image lines, corresponding to mutually orthogonal direction in the scene; (2) estimation of vanishing points; (3) computation of ω and recovery of f and (p_x, p_y). The first step can be done in a manual or semi-automatic way. To the best of our knowledge, no fully automatic method is actually available to detect mutually orthogonal lines, if no a priori information can be used, like in a forensic scenario. Indeed, in the computer vision field, many works have appeared dealing with the problem of line selection and grouping. Typically, these methods firstly retrieve the image line segments (e.g., using the Canny method [137]), then execute a clustering step to group lines converging to the same vanishing point. This can be achieved through iterative refinement with an Expectation-Maximization approach [138], by using a schema based on the Hough transform [139], or using a robust estimator based on the Random Sample Consensus (RANSAC) method [140], usually modified to be able to deal with multiple models, such as the J-Linkage algorithm [141], employed in [142].

The main issue here is the selection of mutually orthogonal line clusters: if the camera calibration is known, this problem can be solved with less effort, even simultaneously with the line clustering by including orthogonality constraints [143], and operating directly in the image space [144], or in particular accumulation spaces, such as the Gaussian sphere [145]. On the other hand, if neither the focal length, nor the principal point can be fixed a priori, it isn't possible to check the vanishing point orthogonality without an user intervention, or anyway under the hypothesis that the three most populated line clusters are related to orthogonal directions. Given these criticisms, in this work we preferred to use a manual line selection scheme. Moreover, notice that also in [136] parallel lines are validated by the user; while in [126] no specific indication is given about the method to automatically detect orthogonal vanishing points.

5.3.3 Perspective Analysis

In this section, we evaluate the performance of the PP estimation algorithm under different perspective conditions, so as to determine if and how its accuracy changes when passing from *weak* to *strong* perspective images. The following two subsections report the results of synthetic and real world tests respectively.

Synthetic tests

In order to carry out extensive tests, a synthetic dataset featuring 248 representative camera poses was built as follows. A 3D cube with unit length sides was placed in the center of the world coordinate frame with its X, Y, Z axes aligned with the cube. Then, 248 camera center positions were sampled over a sphere of radius r, by varying their azimuth by an angle $\alpha \in (0, \pi/4]$ and their altitude by an angle $\beta \in (0, \pi/2)$ with steps of $\frac{\pi}{32}$ and $\frac{\pi}{64}$ respectively; all other perspective conditions can be deduced by symmetry. Since the VPs are invariant to translation, the camera distance with respect to the world coordinate frame (i.e. the radius r) was kept fixed. In the camera coordinate frame, the z-axis is the line passing through the camera center and the world coordinate origin. The x-axis is perpendicular to the z-axis and parallel to the world plane defined by X and Y and, finally, the y-axis is obtained from the cross product between the unit vectors of the z and x axes (see Fig. 5.9).

We excluded extrema positions — i.e. when $\alpha = 0$, $\beta = 0$, $\beta = \pi/2$ — that produce orthographic images of the cube, thus leading to known degeneracies in VP estimation. Likewise, camera roll was not taken into account considering that, as any pure rotation, no parallax effects are induced, thus leaving the perspective appearance of the image unaltered. From each camera pose $P(\alpha, \beta)$, an image of the cube was acquired by using a virtual camera with known PP and focal length. With noise-free measurements (i.e., line points are selected with no error), the PPs were estimated with an Euclidean error with respect to the ground truth lower than 10^{-9} pixels in all the positions. The behaviour in the presence of noise was then evaluated by carrying out a Monte Carlo simulation: for each pose we collected 1000 principal points $PP(\alpha, \beta) = \{PP_1(\alpha, \beta), \ldots, PP_{1000}(\alpha, \beta)\}$ by perturbing the line points with a noise from a zero mean Gaussian distribution with standard

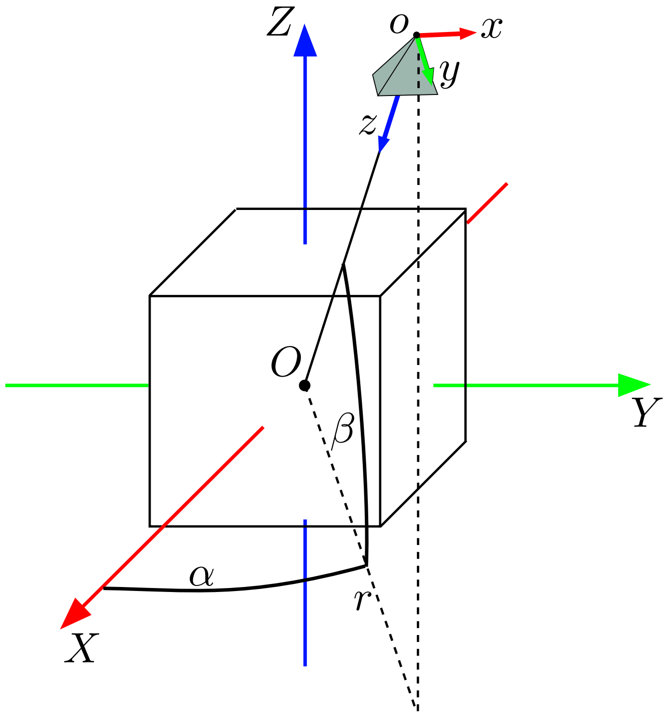
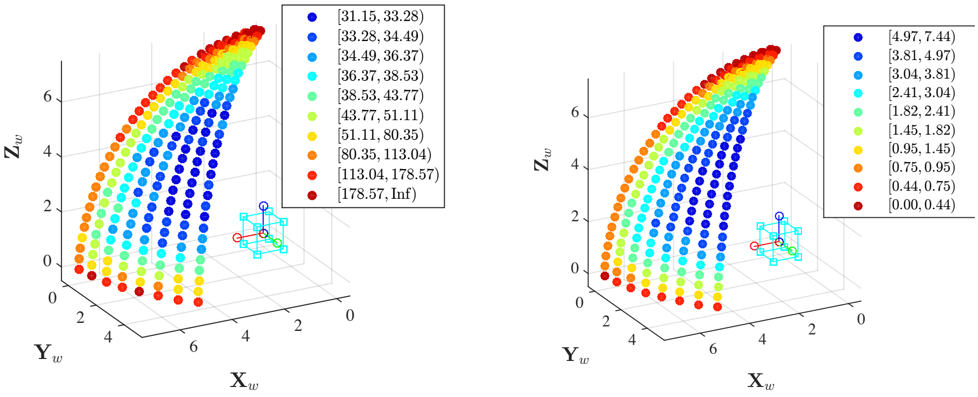


Figure 5.9: (Best viewed in color) Synthetic data setup. A cube is placed at the center of the world coordinate system O, with its sides aligned with the axis X,Y,Z. The image is taken from the camera — represented here as a pyramid — with center $o(\alpha, \beta)$ with a relative coordinate system x, y, z.

deviation $\sigma = 0.5$ pixel — representing an uncertainty of at most 1.5 pixel radius in points selection. For each test we determined a robust index for the dispersion of the collected $PP(\alpha, \beta)$ as follows: we trimmed the 5% of the points with highest distance from the ground truth PP, then we calculated the standard deviations (STD_x, STD_y) of the remaining points along the x and y axes and we chose their maximum as a dispersion index of the estimated PP for that position.

Results are graphically reported in Fig. 5.10a, where the synthetic cube is placed in the origin of the coordinate frame aligned with the orthogonal axes, while each point represents a camera position, colored according to the correspondent estimated dispersion. Notice that the scattering of the estimated PPs is strictly related to the image perspective: Most of the poses have comparable uncertainty, except when marginal α or β occurs. In those cases, the computation accuracy of the VPs strongly drops, and the PP estimates become unreliable and virtually useless for forensic purposes.

These results suggest the possibility to define a novel image feature to be used by the forensic analyst to evaluate the expected accuracy. Firstly, given



(a) (b)

Figure 5.10: (Best viewed in color) 3D plots representing results obtained with the synthetic data setup: in both figures, the virtual cube is placed in the origin of the coordinate system, aligned with the orthogonal axis. Colored points represent the tested camera positions. In (a) we report the maximum STD (between x and y-axis) of the estimated PP: the PP dispersion is bigger for reddish and, lower for blueish points. In (b) the same camera poses are reported but with color related to the MVA: poses with wider MVA are reported in blue, while poses with narrower MVA are in red. Note that poses with lower STD are characterized by wider MVA, and vice-versa. In both plots, the thresholds used to assign colors are obtained from the deciles (i.e. ten quantile with step of 10%) of the respective distribution (STD and MVA).

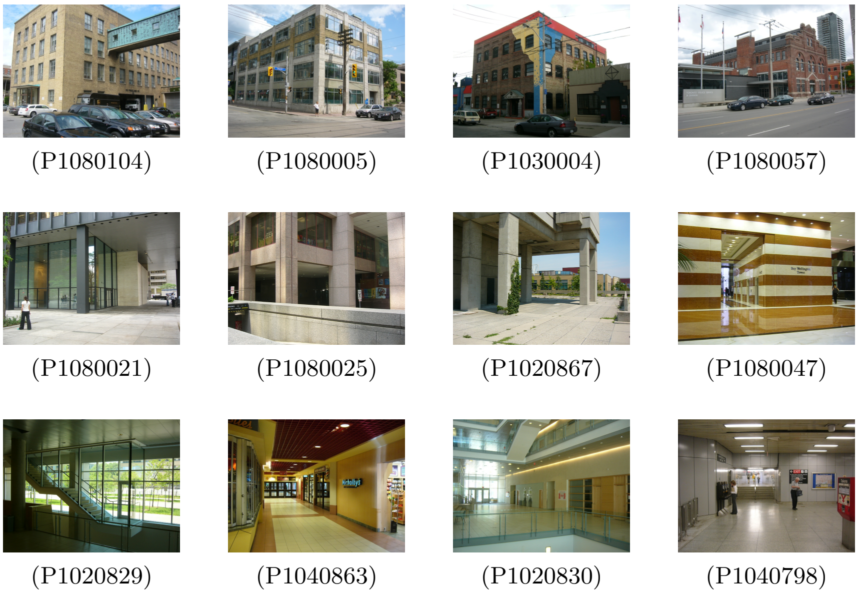
a vanishing point $\mathbf{v}_i$, let θ_i be the widest angle among those obtained from the pairwise intersection of lines concurrent to $\mathbf{v}_i$ (see Fig. 5.12). Then, given θ_1, θ_2, and θ_3, related to three mutually orthogonal VPs, we can define the *Minimum Vanishing Angle* (MVA) as

$$\text{MVA} = \min(\theta_1, \theta_2, \theta_3) \tag{5.16}$$

A visual representation of the MVA values for different camera poses is reported in Fig. 5.10b. Its comparison with the results in Fig. 5.10a confirms our intuition that the proposed feature is a sensible indicator of PP dispersion. Indeed, small MVAs are associated to marginal poses characterized by a weaker perspective.

Tests on real images

To compare the synthetic data with real experiments we clustered the 248 synthetic poses in three groups according to their correspondent MVAs: Weak Perspective ($MVA < 1.5°$), Mid Perspective ($1.5° \leq MVA < 4°$), and Strong Perspective ($MVA \geq 4°$). Then we considered 12 images from the York Urban Database [146] spanning several MVAs between 0° and 7.52°. For each image 25 different PPs were computed, as described in Section 5.3.2, by letting 25 different users to select three lines for each direction. In Figure 5.11



(P1080104) (P1080005) (P1030004) (P1080057)

(P1080021) (P1080025) (P1020867) (P1080047)

(P1020829) (P1040863) (P1020830) (P1040798)

Figure 5.11: Twelve images, and their names, from the York Urban Database [146], used in the real test to corroborate results achieved with the synthetic cube dataset. Top row shows images with strong perspective, with MVAs spanning from 7.52° to 5.53°. Second row includes mid perspective images with MVAs from 3.96° to 2.11°. Finally, the last row shows images with low perspective and MVAs from 1.09° to ~ 0.00 °. MVA here reported are the mean value of the MVAs computed on each image during the tests, since any user can select different lines and obtain slightly dissimilar MVA.

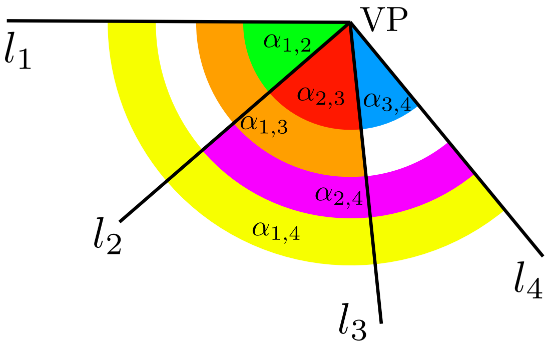


Figure 5.12: (Best viewed in color) Graphical visualization of angles obtained from the pairwise intersection of lines concurrent to the same VP. In this case θ_i correspond to $\alpha_{1,4}$ since it is widest angle available.

we reported the name of the selected images, their MVAs estimated by users selection and the perspective group they belong to (Weak, Mid or Strong).

The achieved results are compared in Figure 5.13. Crosses represent the estimated PPs on real images: in red, green and blue for the images belong-

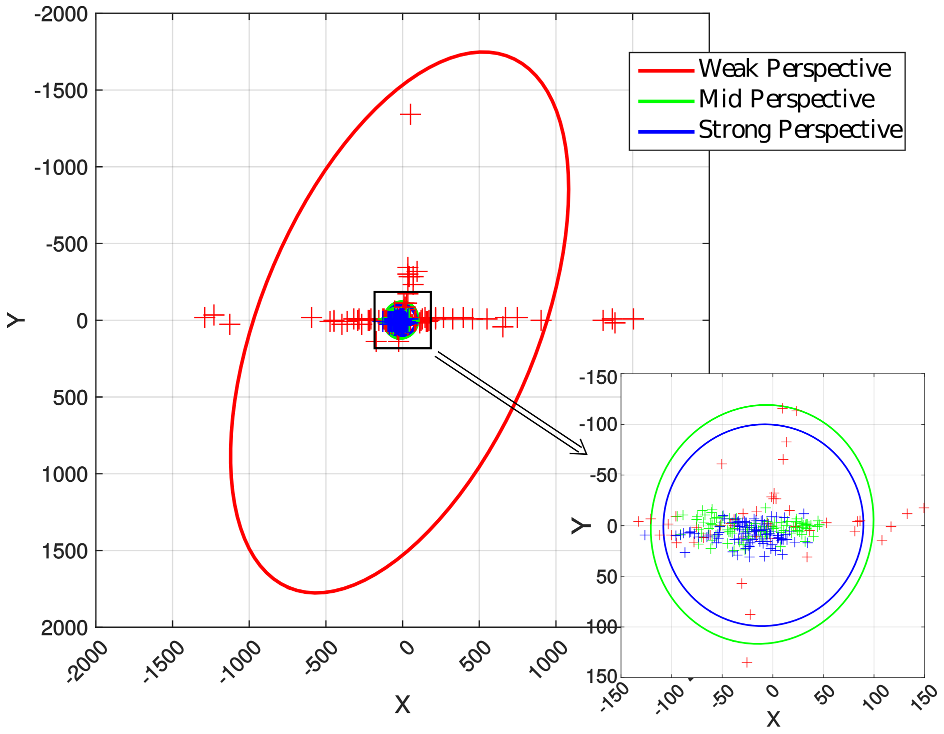


Figure 5.13: (Best viewed in color) Comparison of results achieved from synthetic and real images. Crosses represent the estimated PPs (red for *subway*, green for *hall*, blue for *building*). Ellipses enclose the PPs distribution obtained in synthetic tests.

ing to Weak, Mid and Strong perspective groups respectively. The plotted ellipses represent the 95% confidence ellipses estimated on the corresponding synthetic clusters. Synthetic results show that the estimation is expected to be extremely noisy on the Weak perspective cluster while more accuracy and stability is expected on the Mid and Strong cluster where the MVA is wide enough. Real data confirm the synthetic prediction (STD$_x$ is 435.69, 38.52 and 29.69 pixels on Weak, Mid and Strong perspective clusters respectively). Looking at the picture, a horizontal dispersion of the real data sticks out. This is due to the fact that the images of the considered dataset are characterized by small altitudes, while the synthetic data is built considering all possible viewing angles.

5.3.4 Image Characteristic Analysis

In the previous section we defined the MVA feature, after observing a strong relationship between the amplitude of the vanishing angles and the PP estimation accuracy. In practical cases, the scene may allow the forensic analyst to extract more lines for each direction and possibly forming even wider MVAs. In this section we investigate more deeply the estimation accuracy

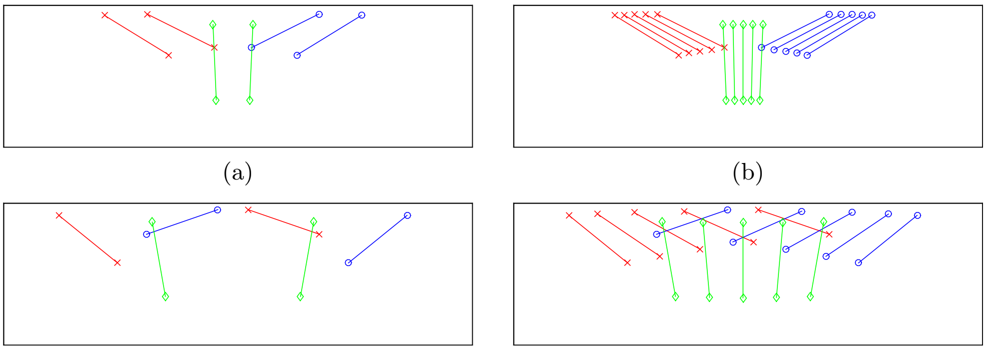


Figure 5.14: (Best viewed in color) Example images produced to test the PP estimation algorithm with reference to the extracted features. On the left, images with two lines for each VP, with different minimum vanishing angle (i.e. MVA={5,20}); on the right, similar images but with five lines. Lines with the same color converge to the same vanishing point.

with reference to the MVA amplitude. For this purpose, we take into account only MVAs with sufficient amplitude able to provide reliable results, and we evaluate how its increase affects the estimation accuracy.

We also study how the performance is sensitive to an increase in the number of lines intersecting in the same VP: Since VPs are obtained by minimization, we expect an accuracy improvement when more data are available. As for the tests of Section 5.3.3, a synthetic image dataset is used first, then tests on real images are carried out to corroborate the synthetic results.

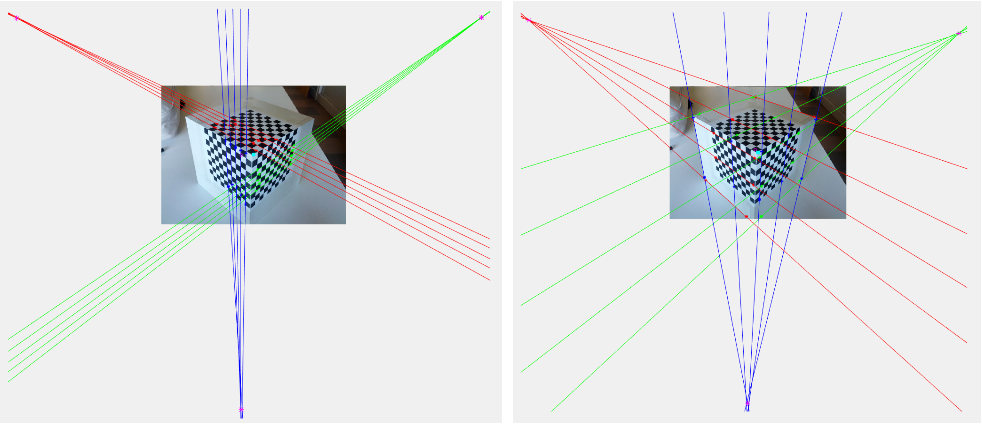
Synthetic tests

We generated different MVAs with different numbers of lines: starting with two lines for each VP, with an angle of incidence of 5°, we progressively added new lines into the image and increased the angle. More specifically, we used $n = \{2, 3, 4, 5\}$ lines, with a length of 200px, and angles $\theta = \{5°, 10°, 15°, 20°\}$ (see Fig. 5.14 for some synthetic image examples). Gaussian noise with zero mean and standard deviation $\sigma = 0.5$ pixel was added to the point coordinates, and the evaluation was repeated 1000 times for each image.

Table 5.4 shows the maximum STDs (as defined in Section 5.3.3) for the estimated PPs, along the x and y image directions. As clearly visible, the accuracy is almost stable when adding new lines, while it significantly grows using well spaced lines (i.e., wider MVAs).

Table 5.4: Max STD of estimated PPs between x and y direction

		MVAs			
		5°	**10°**	**15°**	**20°**
#Lines	**2**	18.55	10.15	7.09	5.98
	3	19.54	9.85	6.59	5.56
	4	18.67	9.53	6.17	5.12
	5	17.03	8.74	6.12	4.79



(a)

(b)

Figure 5.15: Examples of lines selected by the user on the real image searching for (a) narrow and (b) wide MVAs.

Tests on real images

As before, the results achieved with the synthetic data were validated on real tests with the help of 25 different users, having to select up to five lines per VP, with quasi regular spacing. For this purpose, the image of a cube-like checkerboard pattern was used. The considered image allows the user to select either narrow or wide MVAs of approximatively 5° and 20° respectively. 25 PPs were collected in both cases — i.e. the narrow (Fig. 5.15a) and wide (Fig. 5.15b) selection schemes — and the results were evaluated with respect to MVA amplitude and number of lines.

The PPs estimated on real images are represented as colored dots in Fig. 5.16a — in red for angles of 5°, in blue for wider angles (20°). The 95% confidence ellipses of PPs obtained during the synthetic tests (see Section 5.3.4) are also shown, with the same color coding. In Fig. 5.16b, a similar plot considering instead the line number is presented. Almost all PPs obtained on the real images fall inside the associated ellipse, confirming that synthetic results are in close agreement with the real ones. Furthermore, these tests corroborate the observation that increasing the MVA clearly improves the estimation stability (Fig. 5.16a), while adding more lines does not significantly affect the performance (Fig. 5.16b).

In conclusion, results obtained in Sections 5.3.3 and 5.3.4 can be summarized in two main outcomes: (i) Images characterized by a narrow MVA should not be used for forensic analysis based on PP; (ii) To improve accuracy, the selection of few well spaced lines is preferable over many, closely spaced lines.

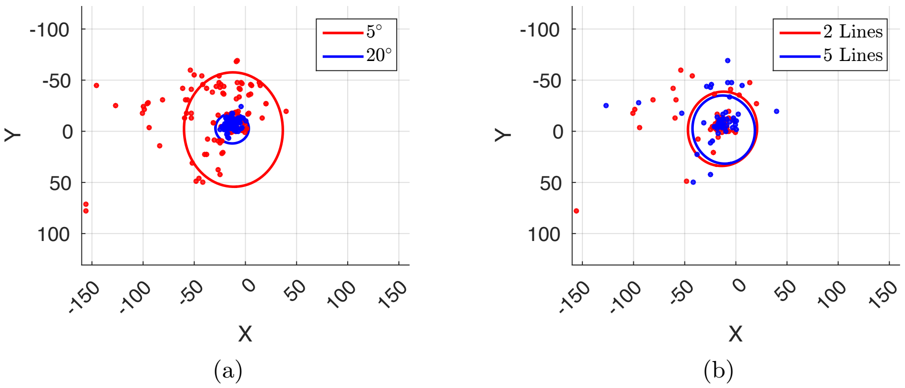


Figure 5.16: (Best viewed in color) Results on real images obtained by varying MVA ans and line number. In (a) dots represent estimated PPs, clustered with respect to the MVA, while in (b) PPs are grouped by the line number. Reported ellipses represent the PP dispersion on the synthetic data. The coordinate system is centered in the ground truth PP.

5.3.5 Forensic Case Studies

In [126] the distance between the PP and the image center is exploited to identify asymmetrically cropped images (see Fig. 5.17). Once computed, the image and the PP are normalized in the interval $[-1, 1]$. A cropping threshold (CT) — i.e. the radius of a circle centered in the estimated PP — is defined, and the image is labeled as cropped if the distance of the PP from the image center exceeds CT. In the following tests we show how the achieved results can support the analyst in assessing the cropping detection performance:

- *Perspective-based Test*: we verify that the MVA amplitude can suggest whether the cropping detection is applicable on a query image. The test is performed on the synthetic and real data defined in Section 5.3.3 and confirms that the technique cannot be applied on images with a narrow MVA;
- *Characteristic-based Test*: we assess the performance variations when more lines and wider MVAs are available on the image. The test is performed on the synthetic and real data defined in Section 5.3.4;

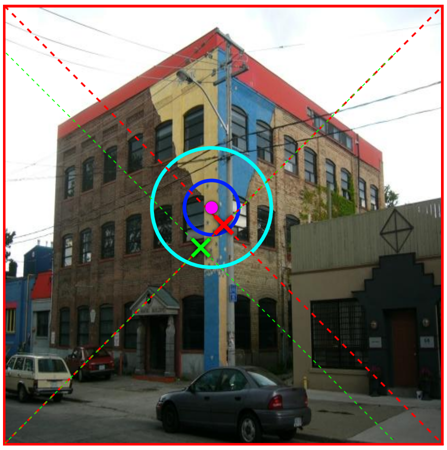


Figure 5.17: (Best viewed in color) In a pristine image (surrounded by a red border) the image center (red cross) falls near the PP (purple dot). On the other hand, if an upper-right cut (green area) is performed, the image center (green cross) shifts falling away from the PP, that remains fixed. The green area is related to the cropping percentage (CP). Blue and cyan circles, centered on the PP, represent instead two cropping thresholds (CT): note that in this example, using the smaller CT (blue circle) the cropping will be successfully detected, since the center of the cropped image center (green cross) fall outside the circle. On the other hand, using the bigger threshold (cyan circle), the image will be erroneously labeled as pristine. Note that in this figure we changed the aspect ratio of the original image (Fig. 5.11(P1030004)) so to visualize the normalization process in [-1,1].

- *Robustness Test*: we verify the robustness of the cropping detection to image compression and resizing. We consider a practical case where the image has been exchanged through Facebook at low quality, thus having been resized and compressed.

In our experiments we consider both cropping percentage (CP) — i.e. the size of the cut — and CT from 0% to 50% of the image size, with steps of 5%. Results are reported for an upper-left cropping only, where both dimensions of the image have been cut with the same percentage, thus leaving unchanged the image aspect ratio. However, tests were performed on all the other eleven cases of asymmetric cropping too (upper, left, right, bottom, upper-left, upper-right, bottom-left, bottom-right, left-upper-right, upper-right-bottom, right-bottom-left, bottom-left-upper). These results are summarized in the Appendix A where is shown that performances significantly increase between Weak and Mid perspective in all the cropping cases, confirming that the proposed feature allows the analyst to decide whether the cropping detection can possibly be applied to a query image.

When useful, the performance was evaluated using the Receiver Operating Characteristic (ROC) curve, where each point corresponds to True Positive (TP) and False Alarm (FA) rates for a given CT. The Area Under Curve

Table 5.5: AUC for Perspective based Test on synthetic and real data

Synthetic Data				Real Data			
CP	Weak	Mid	Strong	CP	Weak	Mid	Strong
<25%	0.60	0.70	0.72	<25%	0.56	0.77	0.81
25%-50%	0.82	0.97	0.99	25%-50%	0.73	1.00	1.00

(AUC) is used to compare the overall performance under different conditions: the more the AUC is close to one, the better is the detector accuracy. In some cases the mean accuracy, computed as the average of TP and TN rates on all considered cropping percentages, was also reported. For the sake of presentation, results have been grouped into two clusters, corresponding to slightly cropped (lower than 25% of the image) or strongly cropped (between 25% and 50%) images.

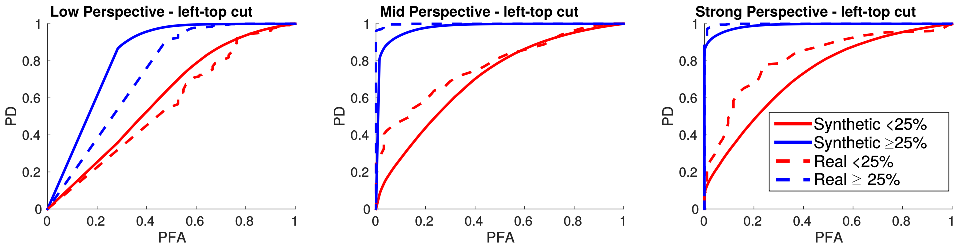
Perspective-based Test

In this test we assess the performance of the cropping detection with reference to perspective conditions. We considered both synthetic and real PPs acquired in section 5.3.3. The cropping detection performance was evaluated separately on the three clusters (Weak, Mid and Strong Perspective) for both synthetic and real PPs. In Figure 5.18 we reported the ROC curves considering slightly and strongly cropped images, while in Table 5.5 we reported the AUC values. In Table 5.6 we summarize the cropping detection performance on the three clusters for different CTs, namely: FA rate, TP rate for both slight and strong cropping, and the mean accuracy. Note that we only report results considering the CTs in $[0.05, 0.25]$, since we noticed a progressive performance drop for higher CTs.

These results suggest that, given a threshold, the false alarm rate may strongly depend on the MVA. For instance, a false alarm of 0.03 on the Mid perspective cluster (real data) corresponds to a threshold of 0.25 of the image. However, the same threshold on the Weak perspective cluster corresponds to a false alarm of 0.73. Both synthetic and real results confirm that the cropping detection can hardly be applied on Weak perspective images and a threshold on the MVA can be chosen to discern unusable images (AUC passes from 0.73 to 1 from Weak to Mid perspective on real images). Furthermore we notice that, on images characterized by decent perspective ($MVA > 1.5°$), the technique is extremely effective when the applied cropping is greater than 25% of the image.

Characteristic-based Test

In this test we assess the performance of the cropping detection with reference to the number of lines and their MVAs. We tested the cropping detection on the synthetic PPs acquired in Sections 5.3.4 (for angles of 5° or 20°, and



(a) (b) (c)

Figure 5.18: (Best viewed in color) ROC curves of the cropping detection for synthetic and real data. The results are reported for (a) Weak, (b) Mid and (c) Strong cluster separately.

Table 5.6: Cropping detection on both synthetic and real data, considering Weak (a,b), Mid (c,d), and Strong perspective (e,f)

(a)

Synthetic Weak Perspective				
CT	FA	TP (<25%)	TP (25%-50%)	Mean Accuracy
0.05	0.96	0.99	1.00	0.52
0.10	0.86	0.96	1.00	0.56
0.15	0.73	0.90	1.00	0.61
0.20	0.62	0.81	1.00	0.65
0.25	0.53	0.71	0.99	0.67

(b)

Real Weak Perspective				
CT	FA	TP (<25%)	TP (25%-50%)	Mean Accuracy
0.05	0.97	0.99	1.00	0.52
0.10	0.90	0.96	1.00	0.56
0.15	0.80	0.90	1.00	0.61
0.20	0.75	0.81	1.00	0.65
0.25	0.73	0.71	0.99	0.67

(c)

Synthetic Mid Perspective				
CT	FA	TP (<25%)	TP (25%-50%)	Mean Accuracy
0.05	0.92	0.98	1.00	0.54
0.10	0.72	0.93	1.00	0.62
0.15	0.53	0.81	1.00	0.70
0.20	0.37	0.67	1.00	0.75
0.25	0.25	0.51	0.99	0.77

(d)

Real Mid Perspective				
CT	FA	TP (<25%)	TP (25%-50%)	Mean Accuracy
0.05	0.82	0.98	1.00	0.54
0.10	0.53	0.93	1.00	0.62
0.15	0.30	0.81	1.00	0.70
0.20	0.15	0.67	1.00	0.77
0.25	0.03	0.51	0.99	0.77

(e)

Synthetic Strong Perspective				
CT	FA	TP (<25%)	TP (25%-50%)	Mean Accuracy
0.05	0.90	0.98	1.00	0.55
0.10	0.67	0.91	1.00	0.65
0.15	0.45	0.78	1.00	0.73
0.20	0.30	0.62	1.00	0.77
0.25	0.10	0.45	0.99	0.80

(f)

Real Strong Perspective				
CT	FA	TP (<25%)	TP (25%-50%)	Mean Accuracy
0.05	0.83	0.99	1.00	0.58
0.10	0.45	0.88	1.00	0.75
0.15	0.22	0.70	1.00	0.83
0.20	0.12	0.50	1.00	0.84
0.25	0.07	0.31	0.97	0.82

with 2 or 5 lines) and on the real data acquired in Section 5.3.4. Firstly, we compared the results obtained when the VPs are estimated from $5°$ and $20°$ MVAs; the performances are shown through the ROC curves in Fig. 5.19a and 5.19b. Secondly, we compared the results achieved using 2 or 5 lines to

Table 5.7: AUC for Characteristic based Test on synthetic and real data

(a)

Synthetic Data				
CP	2 lines	5 lines	∼ 5° MVA	∼ 20° MVA
<25%	0.87	0.96	0.87	0.99
25%-50%	1.00	1.00	1.00	1.00

(b)

Real Data				
CP	2 lines	5 lines	∼ 5° MVA	∼ 20° MVA
<25%	0.86	0.89	0.81	1.00
25%-50%	0.99	1.00	0.99	1.00

Table 5.8: Mean Accuracy for Characteristic based Test on synthetic and real data

(a)

Synthetic Data				
CT	2 lines	5 lines	5° MVA	20° MVA
0.05	0.77	0.80	0.63	0.91
0.10	0.90	0.91	0.79	0.97
0.15	0.91	0.91	0.86	0.93
0.20	0.89	0.89	0.88	0.89
0.25	0.86	0.86	0.86	0.86

(b)

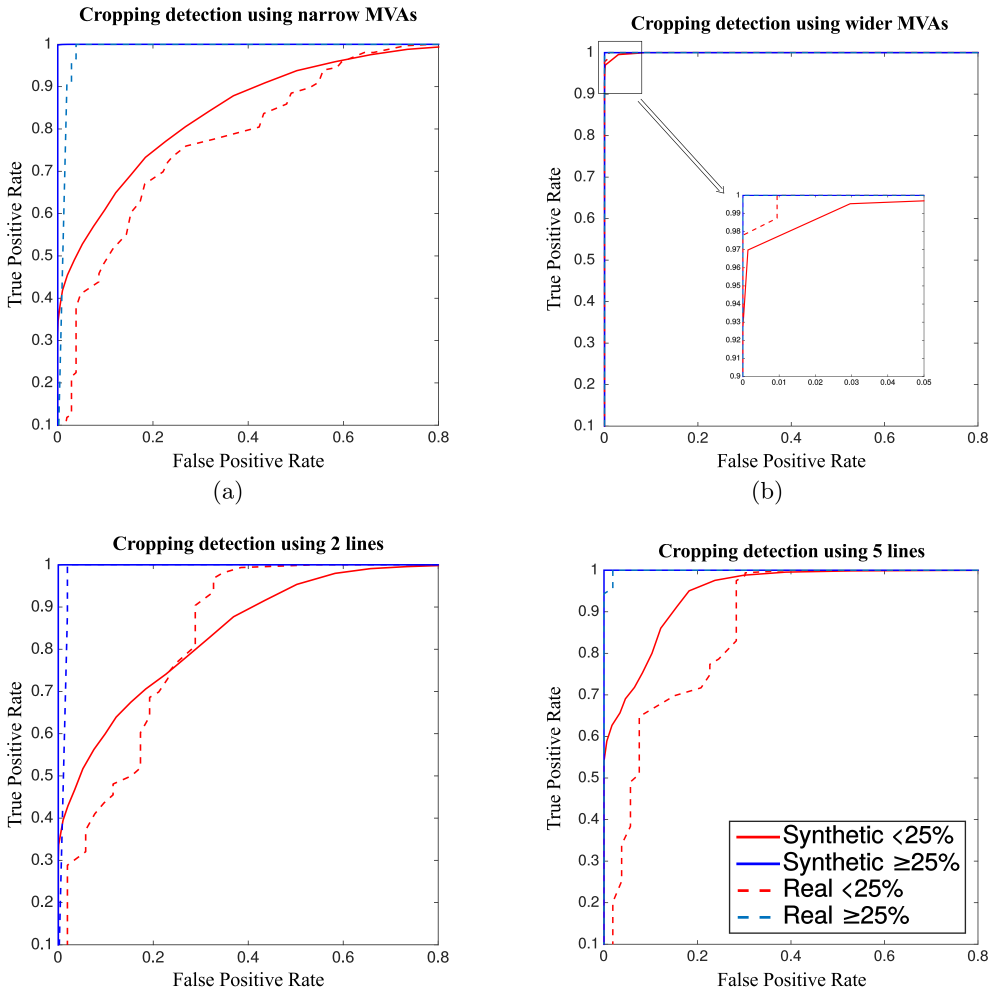
Real Data				
CT	2 lines	5 lines	∼ 5° MVA	∼ 20° MVA
0.05	0.59	0.60	0.58	0.63
0.10	0.82	0.85	0.67	0.99
0.15	0.82	0.83	0.72	0.94
0.20	0.83	0.86	0.79	0.92
0.25	0.82	0.87	0.84	0.87

detect each vanishing point; the corresponding ROC curves are reported in Fig. 5.19c and 5.19d. In Table 5.7 the AUCs for the two experiments have been reported to compare the overall performances. To be consistent with the previous test we briefly report in Table 5.8 the mean accuracy at varying CT for each of the cases. The achieved results show that wider MVAs produce a significant improvement in the detection rate. For instance, with a CT of 0.10, the mean accuracy passes from 0.79 to 0.97 on the synthetic data. This behaviour is confirmed by real data: with the same CT the mean accuracy passes from 0.67 to 0.99. As expected, performances are slightly affected by increasing line numbers. Indeed mean accuracy improvements are always at most 5% for all the synthetic and real cases.

In [126] the authors state that a CT of 0.1 and 0.15 can fit different demands. Anyway this threshold is set regardless of image content. The achieved results suggest instead that a more fitting threshold could be selected according to the available MVA. Synthetic results show that the best performances are obtained with a CT of 0.20 when a 5° MVA is available on the image. Conversely, with a 20° MVA, a CT of 0.10 should be preferred to achieve the best accuracy. Real data confirmed that two different thresholds should be considered according to MVA amplitude: 0.25 for a 5° MVA and 0.10 for a 20° MVA.

Robustness test

In this test we assess whether the technique is usable when the image has been resized and/or compressed. We consider a practical case where the image (considered in the characteristic-based test) was uploaded on Facebook at low quality version and then downloaded: the resolution changes from 2592×1944



(a)

(b)

(c)

(d)

Figure 5.19: ROC curve on synthetic and real data with different cropping percentage using (a) narrow vanishing angles and (b) wider vanishing angles, and then using (c) 2 lines and (d) 5 lines to detect each vanishing point.

to 972 × 729, and its size from 1.4 MB to 80 KB. 25 PPs were collected on the downloaded image (similarly to Section 5.3.4) and the cropping detection was applied as in the characteristic-based test. In Tables 5.9 and 5.10 we report the AUC and the mean accuracy at varying CT: by comparison with the results achieved in the characteristic-based test, we notice that performances are almost unchanged, with the only exception of slightly cropped images, when only narrow MVAs are available, in which case performance drops, with the AUC passes from 0.81 to 0.66. This result once more confirms that the MVA amplitude is crucial to determine the usability of this technique.

Table 5.9: AUC for on Facebook Data

Facebook Data				
CP	2 lines	5 lines	∼ 5° MVA	∼ 20° MVA
<25%	0.82	0.82	0.66	1.00
25%-50%	0.99	1.00	0.99	1.00

Table 5.10: Mean Accuracy on Facebook Data

Facebook Data				
CT	2 lines	5 lines	∼ 5° MVA	∼ 20° MVA
0.05	0.61	0.59	0.51	0.71
0.10	0.76	0.80	0.55	1.00
0.15	0.82	0.81	0.69	0.94
0.20	0.84	0.81	0.73	0.92
0.25	0.82	0.82	0.77	0.87

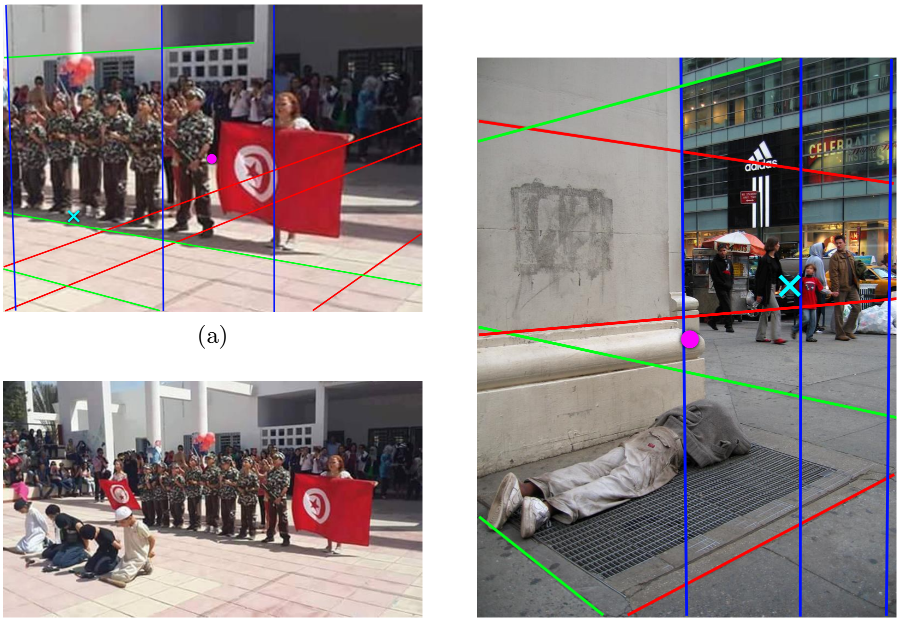
A practical example of cropping detection

We now show how MVA analysis can practically support the forensic analyst to assess whether an image has been cropped. Let us consider the images in Fig. 5.20a and 5.20c, downloaded from the web. The analyst estimates the PP on both images selecting lines that intersect with the widest possible angles. As a result he/she obtains that in both cases the normalized distance of the estimated PP from image center is anomalous (0.3875 and 0.2585 respectively). At first glance this fact leads to the conclusion that both images have been cropped. On the other hand, the analyst notices that the MVAs are 4.83 and 1.21 respectively. This means that he can be much more confident with the first result while the PP estimation on Fig. 5.20c is subjected to strong noise. More specifically, with such a small MVA the estimated PP is unreliable for the purpose. Then the analyst concludes that Fig. 5.20a is probably cropped while no evidence can be provided on Fig. 5.20c by this single test.

In figure 5.20b we report the original version of 5.20a that can be found on the web, confirming the achieved results.

5.3.6 Discussion

In this chapter we presented an in deep assessment of the reliability of a physical-based feature for forensic image authentication. In particular we focused on the estimation accuracy of the principal point of an image and its application to the forensic scenario. By observing the principal point estimation accuracy under different perspective conditions, we were able to define a novel feature, the minimum vanishing angle (MVA), strictly related to principal point uncertainty. Then we further investigated the MVA influence on the estimation accuracy by comparing it with respect to the number of detected lines, exploited for the estimation of the PP. Results underlined that the use of wider vanishing angles leads to higher accuracy, while by employing more


(a)

(b)

(c)

Figure 5.20: (Best viewed in color) Two examples of cropping detection (a,c), with lines of mutually orthogonal directions in red, green and blue. The purple dot indicates the image center, while the cyan cross shows the estimated position of the PP. In both images the MVA is the angle related to the vertical direction (blue lines): in (a) MVA=4.83, in (c) MVA=1.21. In (b) the original version of (a) is presented

lines only slight uncertainty reductions are achieved. As shown in the case studies presented in the previous Sections, the application of our criteria to cropping detection allows the analyst to easily exclude an image that is not suitable for the application of this technique. Moreover we verified that on resized and compressed images — as for example pictures downloaded in low quality from Facebook — the performance only slightly decreases, provided that wide MVAs are available.

In future work the proposed MVA will be exploited to analytically compute a likelihood score to provide more than a binary decision on the integrity of the examined image. Moreover, we are planning to deeply investigate the relation between the MVA and the best cropping threshold to be used, to control the false alarm rate. For this purpose, automatic techniques for principal point localization — so as to remove the human-in-the-loop — will be investigated to perform tests on huge amount of real data.

CHAPTER 6

Social Media Profile Linking by means of Hybrid Source Identification

Digital videos (DVs) are steadily becoming the preferred means for people to share information in an immediate and convincing way. Recent statistics showed a 75% increase in the number of DVs posted on Facebook in the last year [147] and posts containing DVs yields more engagement than their text-only counterpart [148]. Interestingly, the vast majority of such contents are captured using smartphones, whose impact on digital photography is dramatic: in 2014, compact camera sales dropped by 40% worldwide, mostly because they are being replaced by smartphone cameras, which are always at your fingertips and makes sharing much easier [149].

In such a scenario, it is not surprising that digital videos gained importance also from the security, forensic and intelligence point of view: videos have been recently used to spread terror over the web, and many critical events like Boston bombing[1] have been filmed and shared by thousands of users. In such cases, investigating the digital history of DVs is of paramount importance in order to recover relevant information, such as acquisition time and place, authenticity, or information about the source device.

In particular, the source identification problem - that is, univocally linking the digital content to the device that captured it - received great attention in the last years. Currently, the most promising technology to achieve this task exploits the detection of the sensor pattern noise (SPN) left by the acquisition device [150]. This footprint is universal (every sensor introduces one) and unique (two SPNs are uncorrelated even in case of sensors coming from two

[1] https://en.wikipedia.org/wiki/Boston_Marathon_bombing

cameras of same brand and model). As long as still images are concerned, SPN has been proven to be robust to common processing operations like JPEG compression [150], or even uploading to social media platforms (SMP) [151].

Research on source device identification for DVs is not as advanced. This is probably due to: i) the computational and storage effort required for video analysis, ii) the differences in video coding standards with respect to images, and iii) the absence of sizeable datasets available to the community for testing video source device identification algorithms. Since their origin, DV source identification methods borrowed both the mathematical background and the methodology from the still image case [152]: like for images, thus, assessing the origin of a DV requires the analyst to have either the source device or some training DVs captured by that device, from which to extract the reference SPN. In the case of a device that captures both images and videos, no other method has been proposed in the state of the art than to compute two reference SPNs, one for still images and one for videos, which contrasts with the fact that the device has just one sensor. This approach is anachronistic if we consider that today 85% of shared media are captured using smartphones, whose camera captures *both* still images and videos, although at different resolution. If we manage to use the same reference SPN both for images and videos, several advantages arise: first, only one SPN has to be computed and stored for each device, which makes more sense and is more computationally convenient; moreover, new investigative possibilities are enabled: for example, one could link two different media sharing accounts (e.g., YouTube and Flickr) by checking whether the SPN in YouTube videos matches with the SPN in Flickr images.

This chapter targets exactly this problem: we propose an hybrid approach allowing to perform video source identification using a reference SPN obtained from still images. This also allows the use of an image fingerprint for both image and video source identification, such that the available datasets of image fingerprints are ready to use for video source identification. There's no need to build a video fingerprint dataset that also requires computational effort and the availability of the devices or reference videos. We show that such method yields comparable or even better performance than the current strategy of using a reference SPN calculated from DVs. As a second step, we investigate the robustness of the proposed approach when the reference SPN is estimated from images downloaded from Facebook and tested against videos downloaded from YouTube; this is an extremely interesting scenario from an investigative point of view.

The chapter is organized as follows: Section 6.1 introduces SPN based source device identification, and reviews the state of the art for DV source identification; Section 6.2 formalizes the considered problem and describes the proposed hybrid approach; Section 6.3 presents the proposed dataset and discusses some YouTube/Facebook technical details related to SPN; Section

6.4 is dedicated to the experimental validation of the proposed technique, including comparison with existing approaches; finally, Section 6.5 draws some final remarks and outline future works.

From now on, vectors and matrices are denoted in bold as $\mathbf{X}$ and their components as $\mathbf{X}(i)$ and $\mathbf{X}(i,j)$ respectively. All operations are element-wise, unless mentioned otherwise. Given two vectors $\mathbf{X}$ and $\mathbf{Y}$ we denoted as $||\mathbf{X}||$ the euclidean norm of $\mathbf{X}$, as $\mathbf{X} \cdot \mathbf{Y}$ the dot product between $\mathbf{X}$ and $\mathbf{Y}$, as $\bar{\mathbf{X}}$ the mean values of $\mathbf{X}$, as $\rho(s_1, s_2; \mathbf{X}, \mathbf{Y})$ the normalized cross-correlation between $\mathbf{X}$ and $\mathbf{Y}$ calculated as

$$\rho(s_1, s_2; \mathbf{X}, \mathbf{Y}) = \frac{\sum_i \sum_j (\mathbf{X}(i,j) - \bar{\mathbf{X}})(\mathbf{Y}(i+s_1, j+s_2) - \bar{\mathbf{Y}})}{||\mathbf{X} - \bar{\mathbf{X}}||||\mathbf{Y} - \bar{\mathbf{Y}}||} \tag{6.1}$$

If $\mathbf{X}$ and $\mathbf{Y}$ dimensions mismatch a zero down-right padding is applied. Furthermore its maximum, namely the $\max\limits_{\mathbf{s_1},\mathbf{s_2}} \rho(s_1, s_2; \mathbf{X}, \mathbf{Y})$, is denoted as $\rho_{peak}(\mathbf{X}, \mathbf{Y}) = \rho(\mathbf{s}_{peak}; \mathbf{X}, \mathbf{Y})$. The notations are simplified in $\rho(s_1, s_2)$ and in ρ_{peak} when the two vectors cannot be misinterpreted.

6.1 Introduction to Video Source Identification Based on Sensor Pattern Noise

The task of blind source device identification has gathered great attention in the multimedia forensics community. Several approaches were proposed to characterize the capturing device by analyzing traces like sensor dust [153], defective pixels [154], color filter array interpolation [42]. A significant breakthrough was achieved when Lukas et al. first introduced the idea of using Photo-Response Non-Uniformity (PRNU) noise to univocally characterize camera sensor [150]. Being a multiplicative noise, PRNU cannot be effectively removed even by high-end devices; moreover, it remains in the image even after JPEG compression at average quality. The suitability of PRNU-based camera forensics for images retrieved from common SMPs has been investigated in [151], showing that modifications applied either by the user or by the SMP can make the source identification based on PRNU ineffective. The problem of scalability of SPN-based camera identification has been investigated in several works [155, 156]. Noticeably, in [155] authors showed that the Peak-to-Correlation Energy (PCE) provides a significantly more robust feature compared to normalized correlation. The vast interest in this research field fostered the creation of reference image databases specifically tailored for the evaluation of source identification [157], allowing a thorough comparison of different methods [158]. Recently, authors of [159] addressed the problem of reducing the computational complexity of fingerprint matching, both in terms of time and memory: they propose to use random projections to compress the fingerprints, thus allowing the storage of a large database of fingerprints at the price of a small reduction in matching accuracy.

All the methods mentioned so far have been thought for (and tested on) still images. Although research on video source identification began almost at the same time (the first attempt dates back to 2007, [152]), the state of the art is much poorer. In their pioneering work, Chen et al. proposed to extract the SPN from each frame separately and then merge the information through a Maximum Likelihood Estimator; as to the fingerprint matching phase, the PCE was recommended [152]. The experimental results showed that resolution and compression have an impact on performance, but identification is still possible if the number of considered frames can be increased (10 minutes are required for low resolution, strongly compressed videos). Two years later, Van Houten et al. investigated the feasibility of camcorder identification with videos downloaded from YouTube [160], yielding encouraging results: even after YouTube recompression (which will be discussed later), source identification was possible. However, results in [160] are outdated, since both the quality of acquisition devices and the complexity of video coding algorithms have evolved significantly since then. This study was extended by Scheelen et al. [161], considering more recent cameras (with resolution up to full-HD) and coding algorithms (such as H.264 and MPEG-4). Results confirmed that source identification is possible, however authors clarify that the reference pattern was extracted from reference and natural videos before re-encoding. Concerning reference pattern estimation, Chuang et al. [162] firstly proposed to treat differently the SPN extracted from video frames based on the type of their encoding; the suggested strategy is to weigh differently intra- and inter-coded frames, based on the observation that intra-coded frames are more reliable for PRNU fingerprint estimation, due to less aggressive compression. Finally, a recent contribution from Chen et al. [163] considered the problem of source identification for video surveillance systems where the video is transmitted over an unreliable wireless channel. This scenario can lead to videos affected by blocking artifacts, which hinder pattern estimation; authors propose a way to automatically detect such noisy regions in frames and exclude them from the analysis, thus re-establishing the reliability of source identification.

As the reader may have noticed, all the mentioned works discuss source identification *either* for still images or videos, and in the vast majority of works the reference pattern is estimated from native contents, meaning images or frames as they exit from the device, without any alteration due to re-encoding or (even worse) upload/download from SMPs. This approach seriously limits the applicability of source device identification, since it assumes that either the device or some original content is available to the analyst. In the following sections we show how to exploit the available mathematical frameworks to determine the source of a DV basing on a reference derived by still images and the new opportunities introduced by the proposed strategy.

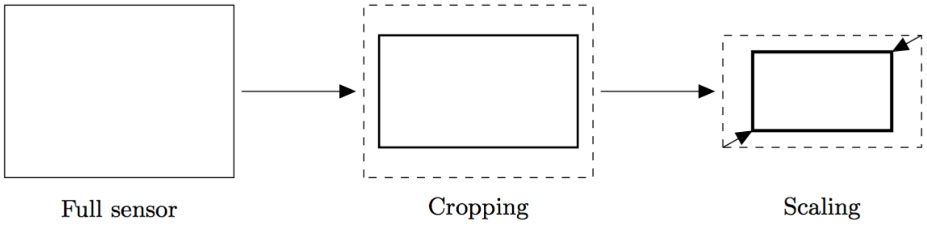


Figure 6.1: Geometric transformations from the full frame to the output format.

6.2 Hybrid Sensor Pattern Noise Analysis

Digital videos are commonly captured at a much lower resolution than images: top-level portable devices reach 4K video resolution at most (which means, 8 Megapixels per frame), while the same devices easily capture 20 Megapixels images. During video recording, a central crop is carried so to adapt the sensor size to the desired aspect ratio (commonly 16:9 for videos), then the resulting pixels are scaled so to match exactly the desired resolution (Figure 6.1). As a direct consequence, the sensor pattern noise extracted from images and videos cannot be directly compared and most of the times, because of cropping, it is not sufficient to just scale them to the same resolution.

The hybrid source identification (HSI) process consists in identifying the source of a DV basing on a reference derived from still images. The strategy involves two main steps: i) The PRNU fingerprint is derived from still images acquired by the source device (reference); ii) the fingerprint is estimated from the investigated video (query) and then compared with the reference to determine the video origin.

The camera fingerprint $\mathbf{K}$ can be derived from N images $\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}$ captured by the source device. A denoising filter ([150], [164]) is applied to each frame and the noise residuals $\mathbf{W}^{(1)}, \ldots, \mathbf{W}^{(N)}$ are obtained as the difference between each frame and its denoised version. Then the fingerprint estimation $\widetilde{\mathbf{K}}$ is derived by the maximum likelihood estimator [165]

$$\widetilde{\mathbf{K}} = \frac{\sum_{i=1}^{N} \mathbf{W}^{(i)} \mathbf{I}^{(i)}}{\sum_{i=1}^{N} (\mathbf{I}^{(i)})^2}. \tag{6.2}$$

The fingerprint of the video query is estimated in the same way by the available video frames.

Denoting by $\widetilde{\mathbf{K}}_R$ and $\widetilde{\mathbf{K}}_Q$ the reference and query fingerprints, the source identification is formulated as a two-channel hypothesis testing problem [166]

$$\begin{aligned} H_0 &: \mathbf{K}_R \neq \mathbf{K}_Q \\ H_1 &: \mathbf{K}_R = \mathbf{K}_Q. \end{aligned}$$

In the considered case, $\widetilde{\mathbf{K}}_R$ and $\widetilde{\mathbf{K}}_Q$, are derived from still images and video frames respectively, thus differing in resolution and aspect ratio due to the cropping and resize operations occurring during the acquisition process (see Fig. 6.1). Then, the test statistic is built as follows: the two-dimensional normalized cross-correlation $\rho(s_1, s_2)$ is calculated for each of the possible spatial shifts (s_1, s_2) determined by the feasible cropping parameters [167]. Then, given the peak ρ_{peak}, its sharpness is measured by the Peak to Correlation Energy (PCE) ratio [155] as

$$PCE = \frac{\rho(\mathbf{s}_{peak})}{\frac{1}{mn-|\mathcal{N}|}\sum\limits_{\mathbf{s}\notin\mathcal{N}} \rho(\mathbf{s})} \tag{6.3}$$

where $\mathcal{N}$ is a small set of peak neighbours.
In order to consider the different scaling factor of the two fingerprints - videos are usually resized - Goljan et. al. [167] showed that a brute force search can be conducted considering the PCE as a function of the plausible scaling factors $r_0, \ldots, r_m$. Then its maximum

$$P = \max_{r_i} PCE(r_i) \tag{6.4}$$

is used to determine whether the two fingerprints belong to the same device. Practically, if this maximum overcomes a threshold τ, H_1 is decided and the corresponding values $\mathbf{s}_{peak}$ and r_{peak} are exploited to determine the cropping and the scaling factors. The authors showed that a theoretical upper bound for False Alarm Rate can be obtained as

$$FAR = 1 - (1 - Q(\sqrt{\tau}))^k \tag{6.5}$$

where Q is the cumulative distribution function of a normal variable N(0,1) and k is the number of tested scaling and cropping parameters.
This method is expected to be computationally expensive, namely for large dimension images. Anyway, if the source device is available, or its model is known, the resize and cropping factors are likely to be determined by the camera software specifics or by experimental testing. Then, most of the computational effort can be avoided. In Section 6.3 the cropping and scaling factors of 14 smartphones have been reported.
In the next section we show how the proposed method can be also applied to link images and videos retrieved on different SMPs, even when the source device is not available.

6.2.1 Extension to contents shared on social media platforms

Let us consider a user publishing, with an anonymous profile, videos with criminal content through a SMP. At the same time this user, say Bob, is leading his virtual social life on another social network where he publicly shares

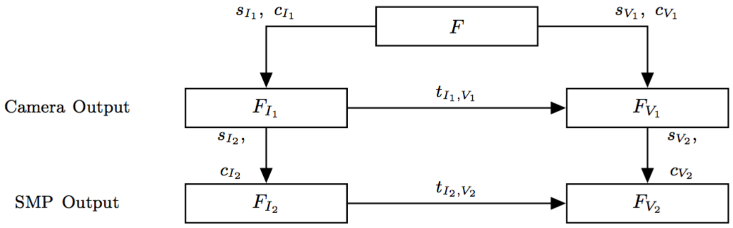


Figure 6.2: Geometric transformations applied to the sensor pattern from the full frame to the image/video outputs on SMPs.

his everyday's pictures. Unaware of the traces left by the sensor pattern noise (and their robustness to several filtering and compression operations), he captures with the same device the contents he shares on both profiles. Then, the fingerprints derived from the images and videos on the two social platforms can be compared with the proposed method to link Bob to the criminal videos. Noticeably, analyzing multimedia content shared on SMPs is not a trivial task. Indeed, besides stripping all metadata, SMPs usually re-encode images and videos during upload. For example, Facebook policy is to down-scale and re-compress images so to obtain a target bit-per-pixel value [168]; Youtube also scales and re-encodes digital videos [169]. Needless to say, forensic traces left in the signal are severely hindered by such processing, which acts as an unintented counter-forensic step. Sensor pattern noise, however, is one of the most robust signal-level features, and it can survive down-scaling followed by compression. Nevertheless, when it comes to link the SPN extracted from, say, a Youtube video and a Facebook image, a new problem arises: since both content have been scaled/cropped by an unknown amount, such transformation must be estimated in order to align the patterns. Interestingly, the hybrid approach can be applied considering that the image and video fingerprints derived from the SMPs contents can be still matched through the right cropping and rescaling parameters. In Fig. 6.2 the geometric transformations occurring on the sensor frame are summarized: the full frame F is scaled and cropped - with factors s_{I_1} and c_{I_1} respectively - by the image acquisition process to produce F_{I_1}. The uploading process over the SMP applies similar transformations - with factors s_{I_2} and c_{I_2} respectively - thus producing F_{I_2}. In a similar way, the video F_{V_1} is generated from the camera and F_{V_2} is uploaded onto another SMP - applying cropping and scaling factors of s_{V_1}, c_{V_1} and s_{V_2}, c_{V_2} respectively. It can be easily deduced that, for both native and uploaded contents. image and video fingerprints are linked by a geometric transformation consisting in a cropping and rescaling operation. Then, the hybrid approach to determine the transformation t_{I_1,V_1} to align the fingerprints of two native contents can be also applied to determine t_{I_2,V_2}, thus linking F_{I_2} to F_{V_2}.
Two main drawbacks are expected for this second application. Firstly the

compared contents have been probably compressed twice and the SPN traces are likely deteriorated. Furthermore it may be hard to guess the right scaling and cropping parameters just from F_{I_2} and F_{V_2}. In these cases an exhaustive search of all plausible scaling and cropping factors is required. In section 6.3 the proposed application is tested to link the images of a Facebook profile to the videos of a YouTube profile.

This is just one of the possible scenario where the HSI can be applied: the trace of the same sensor could be seeked in multimedia contents belonging to different SMPs to link two profiles. Thus, the proposed approach is very versatile and one could easily imagine how F_{I_1} could be linked to F_{V_2} (as F_{I_2} to F_{V_1}).

6.3 PoDIS: a new dataset for video source identification

We considered a brand new dataset (Portable Device Imaging Sensor, PoDIS) consisting of 4019 flat field and natural images and 265 videos captured by 14 devices from different brands (Apple, Samsung, Huawei). For each device a flat field video was also acquired disabling, when possible, the digital stabilization. The Facebook and YouTube versions of all contents were also included. In the following we detail the dataset structure and how we obtained its Facebook and YouTube version.

Native contents We considered 14 different modern devices, both smartphones and tablets. Pictures and videos have been acquired with the default phone settings that, for some models, include the automatic digital video stabilization. In Table 6.1 we report the considered models, their standard image and video resolution and if they're equipped with digital stabilization. From now on we'll refer to these phones with the names $C1, \ldots, C14$ as defined in the table. For each device we collected at least:

- 100 (reference) images depicting the sky
- 150 (query) images of indoor and outdoor scenes
- 1 (reference) video of the sky captured with slow camera movement for more than 160 seconds
- 18 (query) videos of flat textures, indoor and outdoor scenes, captured with both still and moving camera for more than 60 second each.

For each of the video categories (flat textures, indoor and outdoor) at least 6 different videos have been captured considering various acquiring scenario: i) still camera, ii) walking operator and iii) panning and rotating camera. We'll refer to this types as *still*, *move* and *panrot* videos respectively.

ID	model	image resolution	video resolution	digital stabilization
C1	Samsung Galaxy S3	3264 × 2448	1920 × 1080	off
C2	Samsung Galaxy S3 Mini	2560 × 1920	1280 × 720	off
C3	Samsung Galaxy S3 Mini	2560 × 1920	1280 × 720	off
C4	Samsung Galaxy S4 Mini	3264 × 1836	1920 × 1080	off
C5	Samsung Galaxy Tab 3 10.1	2048 × 1536	1280 × 720	off
C6	Samsung Galaxy Tab A 10.1	2592 × 1944	1280 × 720	off
C7	Samsung Galaxy Trend Plus	2560 × 1920	1280 × 720	off
C8	Huawei Ascend G6	3264 × 2448	1280 × 720	off
C9	Ipad 2	960 × 720	1280 × 720	off
C10	Ipad Mini	2592 × 1936	1920 × 1080	on
C11	Iphone 4s	3264 × 2448	1920 × 1080	on
C12	Iphone 5c	3264 × 2448	1920 × 1080	on
C13	Iphone 5	3264 × 2448	1920 × 1080	on
C14	Iphone 6	3264 × 2448	1920 × 1080	on

Table 6.1: Considered devices with their default resolution settings for image and video acquisition respectively.

Facebook and YouTube sharing platforms Images have been uploaded on Facebook in both low and high quality (LQ and HQ respectively). The upload process eventually downscales the images depending on their resolutions and the selected quality [168]. Videos have been uploaded to YouTube through its web application and then downloaded through KeepVid [170] selecting the best available resolution (corresponding to the native one). When possible, videos were downloaded in multiple resolutions. The metadata orientation has been removed from all of the images and videos to avoid unwanted rotation during the contents upload.

6.4 Experimental validation

The effectiveness of the HSI for $C1 - C9$ has been tested by performing the video source identification based on a fingerprint derived from still images. $C10 - C14$ have been excluded, consisting of videos acquired with active digital stabilization. Source identification on these kind of video isn't solved yet. More details are given in Section 6.6, The results have been compared with the standard video source identification based on a fingerprint derived by the frames of a video reference. The same challenge has been also faced when the contents have been exchanged through YouTube and Facebook respectively.

6.4.1 Fingerprints matching parameters

We firstly derived the cropping and the scaling parameters to compare image and video fingerprints of native contents. For each device we estimated the reference fingerprint $\widetilde{\mathbf{K}}_I$ by means of 60 images randomly chosen from the flat-field pictures available for that device. Similarly, the reference fingerprint $\widetilde{\mathbf{K}}_V$ was derived by means of 60 frames randomly chosen from the reference video available for that device. The P statistic (Eq. 6.4) between the image and the video fingerprint has been calculated to determine the cropping and scaling parameters for each device. In Table 6.2 we reported the obtained scaling factor and the consecutive central crop applied to align the image to the video fingerprint. In the table we reported only the up-left corner of the cropping along x and y axes. For instance, $C1$ image fingerprint should be downscaled with a factor 1.7 and then cropped of 180 pixels on both sides along y axis to match the video fingerprint. $C9$ is a pretty unique case in which the video is produced by upscaling and then cropping by -160 along the x axis with respect to the native image size. The proposed procedure can be applied to any new device to determine the image and video fingerprint matching parameters.

Given a video query and a reference fingerprint, the matching is performed as follow: the query fingerprint is derived by 900 randomly selected frames ($\sim$ 30 seconds) of the video query and the test statistic (Eq. 6.4) is evaluated considering the cropping and scaling parameters of the candidate device. We refer to the test statistic as P_I, meaning that the reference fingerprint is derived from still images. For comparison purposes we also consider the statistic P_V,

ID	scaling	central crop along x and y axes
C1	1.7	[0 180]
C2	2	[0 114]
C3	2	[0 114]
C4	1.7	[0 0]
C5	1.6	[408 354]
C6	2.025	[0 122]
C7	2	[0 120]
C8	2.55	[0 120]
C9	0.75	[-161 0]
C10	1.35	[0 179]
C11	1.7	[0 180]
C12	1.7	[0 180]
C13	1.7	[0 180]
C14	1.7	[0 180]

Table 6.2: Rescaling and cropping parameters for the considered devices from image to video output.

meaning that the reference fingerprint is derived from a video reference. The latter is the standard procedure for video source identification and does not require cropping and rescaling parameters (both fingerprints are obtained from videos).

6.4.2 HSI Performance

For each device we obtained at least 18 values of the test statistic P_I by comparing matching pairs (the reference and the query have the same source device) and at least 247 values by comparing mismatching pairs (the reference and the query belong to different source devices). We refer to these statistics as mP_I and mmP_I respectively. Similarly, the statistics mP_V and mmP_V are obtained basing on a reference fingerprint derived by video frames. In Fig. 6.3 we report for each device: i) the statistics mP_I and mP_V of matching pairs; ii) $\overline{mmP_I}$ and $\overline{mmP_V}$, the maximum of the statistics for the mismatching case.

The two statistics are completely separated for all the devices and several thresholds τ produce identical performance. We consider a rational choice to set $\tau = 62.9$, the mass center between the maximum of mismatching and the minimum of matching statistics respectively. In this case the cropping and scaling factors are known, so that the corresponding upper bound of false alarm rate is approximately 10^{-15} (obtained by Eq. 6.5 with $k = 1$). The HSI method correctly identifies the source in all the considered videos through a reference derived by still images; furthermore the HSI yields comparable and sometimes better performance than the current video source identification strategy. Furthermore, the two compared approaches have comparable com-

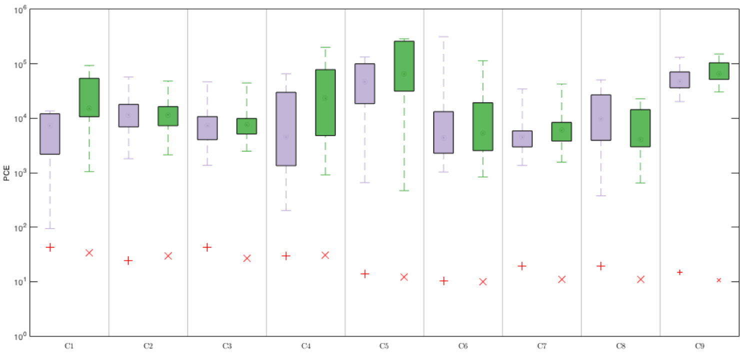


Figure 6.3: (Best viewed in colors) Matching statistics mP_V and mP_I are represented by the green and purple boxplot respectively. The tails correspond to the minimum and maximum of the statistics and the box to the first and third quartile of data. Red crosses are the maximum of the correspondent statistics in the mismatching case.

putational cost, excluding the effort to determine the matching parameters. Anyway this step have to be computed once for each device model as shown in Section 6.4.1.

6.4.3 Results on contents from SMPs

Images and videos acquired by $C1, \ldots, C9$ have been exchanged through Facebook and YouTube respectively, as described in section 6.3. Then the correspondent image and video fingerprints have been estimated as described in the above section. The Facebook image fingerprint has been estimated from images uploaded in both HQ and LQ. In most cases the resolution of the uploaded contents has been modified by the SMPs so that the cropping and scaling factors were estimated by exhaustive search. The statistics mP_I and $\overline{mmP_I}$ were obtained for both LQ and HQ cases and reported in Fig. 6.4. By comparison with Figure 6.3 we notice a performance decay that can be attributed to two main facts: i) images and videos are resized and recompressed by the SMPs thus lowering the quality of both the estimated fingerprint and the query; ii) the exhaustive research of the scaling an cropping factors increases the PCE values of the mismatching pairs. In Table 6.3 we report, for each device, the best achievable accuracy and its corresponding PCE threshold. We notice that the technique accuracy varies according to the devices. For instance, at low quality, the best achievable accuracy is 1 for $C9$ while it's 0.82 for $C6$. Furthermore, as shown in Table 6.3 a different threshold should be considered according to the device. Unfortunately, in most cases, we cannot take advantage of any a priori information considering that contents

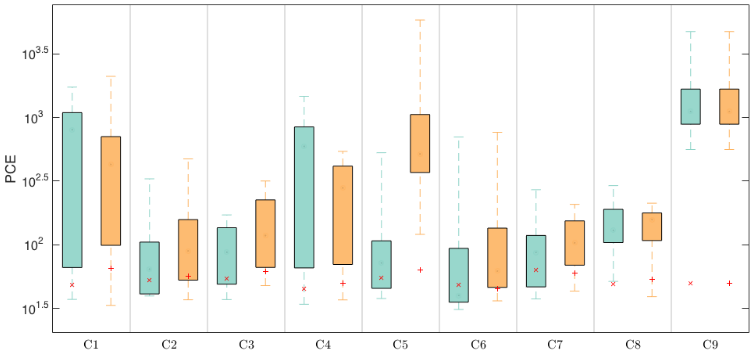


Figure 6.4: (Best viewed in colors) Matching statistics mP_I obtained using image reference from LQ and HQ uploaded contents (green and orange boxplot respectively). The maximum of the correspondent statistics in the mismatching case are shown by the red crosses.

Table 6.3: Best achievable accuracy and correspondent PCE threshold.

	Low Quality		High Quality	
Device	Threshold	Best Accuracy HQ	Threshold	Best Accuracy LQ
C1	71.82	0.98	49.63	0.98
C2	78.51	0.97	49.87	0.96
C3	54.54	0.96	47.02	0.92
C4	69.95	0.97	47.75	0.97
C5	76.68	1.00	55.04	0.92
C6	46.19	0.94	52.18	0.82
C7	60.53	0.96	44.92	0.88
C8	106.43	0.99	49.49	1.00
C9	109.34	1.00	102.71	1.00
Overall	60.22	0.97	98.89	0.95

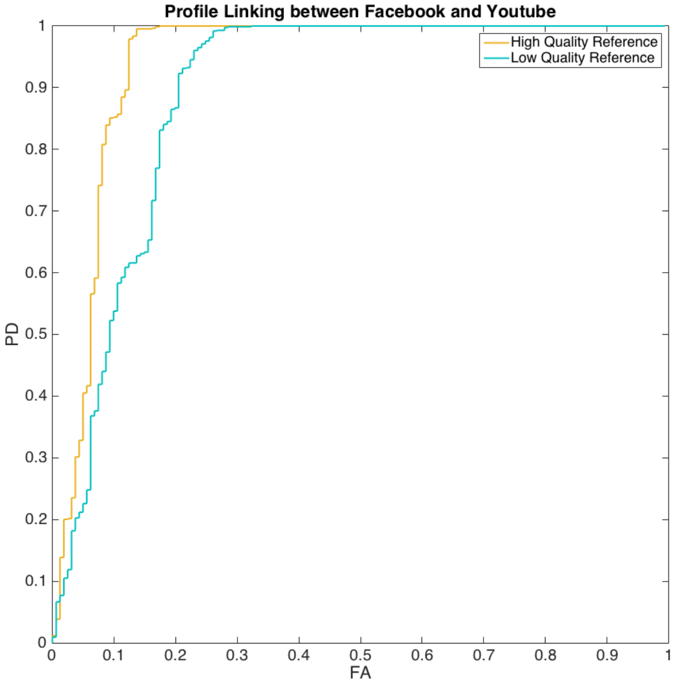


Figure 6.5: Performance of profile linking between Facebook and Youtube considering both LQ and HQ Facebook image reference.

are usually resized by Facebook/YouTube and most of metadata are deleted. Therefore, we report in the last column of Table 6.3 the best achievable accuracy and its correspondent PCE threshold determined by considering all the devices together. The method yields similar performance on both HQ and LQ images. Such a similarity is explained by the fact that Facebook uploading system tries to maintain a constant bit-per-pixel; thus, when an image is uploaded using HQ mode (more pixels), more aggressive quantization is employed to keep BPP at the desired value. In Figure 6.5 we also reported the overall Receiver Operating Characteristic (ROC) curve where true positive and true negative are compared on varying threshold. These results confirm that the performances are similar both in case of high and low quality image reference: the Area Under the Curve (AUC) slightly improves from 0.88 (LQ) to 0.93 (HQ).

6.5 Discussion

In this chapter we proposed an hybrid approach to video source identification using a reference fingerprint derived from still images. Such a method yields comparable or even better performance than the current strategy of using a video reference. Applying the proposed method, the available datasets of image fingerprints are ready to use for video source identification. Furthermore a single fingerprint is needed for both image and video source identification, saving the computational effort to build a video fingerprint dataset. We also showed that the trace of the same sensor could be found in image and video contents belonging to different social media platforms, even when the source device is not available. This application allows the linking of two profiles and opens new investigation opportunities on the web. This last application requires a higher computational effort, since rescaling and cropping parameters have to be determined to match the query and the reference fingerprints; however its effectiveness has been proved to link Facebook images to YouTube videos.

6.6 Future Works: Investigating Stabilised Videos

The problem of source identification based on SPN has not been solved on devices with digital stabilization. Most recent camera softwares include this technology to reduce the impact of shaky hands on captured videos. By estimating the impact of the user movement the software adjusts which pixels on the camcorder's image sensor are being used. Image stabilization can be usually turned on and off by the user on devices based on Android OS while in iOS devices this option cannot be modified by the camera software. The source identification of videos captured with active digital stabilization cannot be accomplished by direct comparison of the PRNU fingerprint: in fact the process disturbs the fingerprints alignment that is a *sine qua non* condition for the identification process. HSI solves the problem on the reference side (it's estimated from still images) but the issue still exists on the query side. Preliminary results proved that digital stabilization can be possibly approximated with with a time-varying cropping and resize of the sensor thus opening the chance to extend the HSI to identify the source of digital stabilised videos.

CHAPTER 7

Discussion and Open Issues

This thesis, starting from the available technologies, focused on the applicability of image forensic techniques in the wild. Firstly, after summarising the most relevant available techniques, a new general forensic scale was defined to classify tools capability in terms of the kind of evidence they are able to provide. Furthermore, basing on the best practices and standards available, a new methodology for digital image investigation was defined. This is aimed to support forensic experts to properly apply several tools to an image and to provide results to be presented in court. We also started to consider how forensic tools can be affected by emerging composition technologies, able to produce realistic forgeries. After surveying the most relevant composition tools, qualitative and quantitative tests were performed. Specifically, quantitative tests provided indicators on how some tools performance are affected by the application of advanced techniques.

Moreover, new forensic applications have been proposed. Focusing on geometric-based features, two contributions were provided: i) a generalisation of a perspective-based technique for tampering detection and ii) the reliability assessment of a cropping detection technique based on principal point estimation. The former proved to be strongly effective to identify subjects spliced without respecting the perspective rules. Its effectiveness was also proved on images exchanged through Facebook and Twitter. In the latter, we started from a known cropping detection technique, subjected to strong false alarm rate. We defined a new feature, namely the minimum vanishing angle, and we proved that its amplitude can be exploited to predict the tool reliability.

A new application was also provided basing on the promising trace of the sensor pattern noise: we introduced a method to link social media profiles where images and/or videos are captured with the same device. Basing on a dataset on recent smartphone we proved how the proposed technique may strongly help in linking Facebook and YouTube profiles which contents were acquired

with the same sensor.
However, given the achieved results, there are some open issues and unsolved problems in the considered research areas:

- new composition techniques should be considered to further evaluate the effectiveness of available forensic tools. This could be challenging considering that these techniques can be hard to find, they may require high computational cost and usually have several tuning parameters, making hard to accurately evaluate the performance.

- the cropping detection tool is not automatic yet, it still requires an user identifying and selecting parallel lines in the scene. Automatic principal point estimation from automatic line detection is a challenging task considering that no a priori information can be supposed on the image.

- the source identification is still unsolved on digital stabilised videos. This is a critical issue considering that a big portion of sold smartphones automatically applies it.

Starting from the achieved results, all these topics will be addressed in the future to further improve the credibility and reliability of forensic techniques in the wild.

Publications

Iuliani, M., Fanfani, M&., Colombo, C., & Piva, A. (2016). Reliability Assessment of Principal Point Estimates for Forensic Applications. *Journal of Visual Communication and Image Representation.*

Iuliani, M., Fabbri, G., & Piva, A. (2015, November). Image splicing detection based on general perspective constraints. In *Information Forensics and Security (WIFS)*, 2015 IEEE International Workshop on (pp. 1-6). IEEE.

De Rosa, A., Piva, A., Fontani, M., & Iuliani, M. (2014, October). Investigating multimedia contents. In 2014 *International Carnahan Conference on Security Technology (ICCST)* (pp. 1-6). IEEE.

Iuliani, M., Rossetto, S., Bianchi, T., De Rosa, A., Piva, A., & Barni, M. (2014, February). Image counter-forensics based on feature injection. In *IS&T/SPIE Electronic Imaging (pp. 902810-902810). International Society for Optics and Photonics.*

Acknowledgements

The research activity developed during the PhD course was partially supported by GNSAGA of INdAM.

Results achieved in Chapter 3 are based on research partially supported by FORLAB, Multimedia Forensics Laboratory, PIN S.c.r.l - Prato (IT).

Results achieved in Chapter 5 are based on research partially sponsored by the Air Force Research Laboratory and the Defense Advanced Research Projects Agency under agreement number FA8750-16-2-0188.

APPENDIX A

Complete Results for Asymmetric Cropping Detection

In section 5.3.5 we investigated the reliability of image (asymmetric) cropping detection basing on a new proposed feature, namely the Minimum Vanishing Angle (MVA). Hereafter, for sake of completeness, we reported the achieved results on all twelve possible asymmetric cropping. We considered both synthetic and real data collected in Section 5.3.3, clustered in three groups according to their correspondent MVAs: Weak Perspective ($MVA < 1.5°$), Mid Perspective ($1.5° \leq MVA < 4°$), and Strong Perspective ($MVA \geq 4°$). We report the ROC curves and correspondent AUC for slightly ($5\% - 20\%$ of images size) and strongly ($25\% - 50\%$ of images size) cropped images.

The AUCs increase between Weak and Mid perspective in all the cases confirms that the proposed feature allow the analyst to decide whether the cropping detection can be possibly applied to a query image. More specifically, if we limit the comparison to crop size wider than 25%, the AUC on synthetic data always increase of at least 0.13 when passing from Weak to Mid perspective. Real data always confirm this prediction with even greater improvements (AUC increases of even 0.30 in some cases).

Notice that results with real data have in same cases slight discrepancies with respect to the their synthetic counterparts. This is due, on the one hand, to the obvious limitations of generating synthetic data that perfectly model the user behaviour. On the other hand, this is due to the impossibility of performing real tests on huge amount of data, as done with synthetic data. Anyway we expect to overcome this limit in our future works by investigating automatic way to blindly localize an image PP.

Table A.1: AUC values for Synthetic and Real Tests combining three different perspective conditions, slightly and strongly cropped images along different single cropping directions

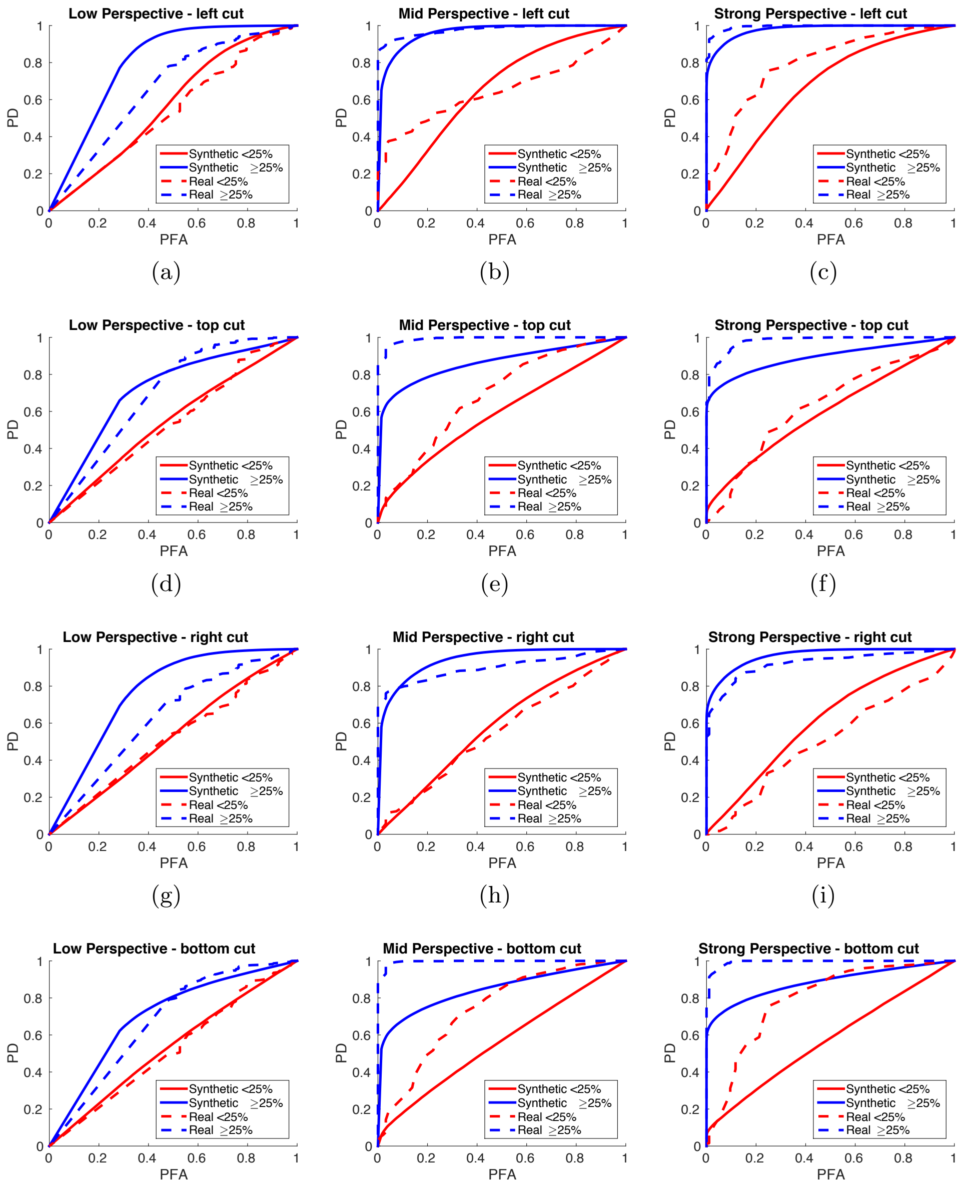
	LEFT		TOP		RIGHT		BOTTOM	
	5%-20%	25%-50%	5%-20%	25%-50%	5%-20%	25%-50%	5%-20%	25%-50%
SYNTH								
Weak	0.57	0.80	0.54	0.71	0.53	0.77	0.53	0.70
Mid	0.66	0.96	0.59	0.86	0.58	0.94	0.56	0.84
Strong	0.67	0.98	0.60	0.89	0.61	0.97	0.57	0.88
REAL								
Weak	0.54	0.66	0.54	0.69	0.52	0.63	0.52	0.68
Mid	0.65	0.98	0.69	0.99	0.55	0.90	0.74	0.99
Strong	0.80	0.99	0.63	0.98	0.50	0.92	0.79	0.99

Table A.2: AUC values for Synthetic and Real Tests combining three different perspective conditions, slightly and strongly cropped images along two cropping directions

	TOP-LEFT		TOP-RIGHT		BOTTOM-LEFT		BOTTOM-RIGHT	
	5%-20%	25%-50%	5%-20%	25%-50%	5%-20%	25%-50%	5%-20%	25%-50%
SYNTH								
Weak	0.60	0.82	0.57	0.81	0.59	0.82	0.56	0.80
Mid	0.70	0.97	0.65	0.97	0.68	0.97	0.63	0.96
Strong	0.72	0.99	0.68	0.98	0.71	0.99	0.66	0.98
REAL								
Weak	0.56	0.73	0.55	0.71	0.55	0.72	0.54	0.71
Mid	0.77	1.00	0.70	0.99	0.80	1.00	0.76	1.00
Strong	0.81	1.00	0.63	0.99	0.88	1.00	0.77	1.00

Table A.3: AUC values for Synthetic and Real Tests combining three different perspective conditions, sslightly and strongly cropped images along three cropping directions

	LEFT-TOP-RIGHT		TOP-RIGHT-BOTTOM		RIGHT-BOTTOM-LEFT		BOTTOM-LEFT-TOP	
	5%-20%	25%-50%	5%-20%	25%-50%	5%-20%	25%-50%	5%-20%	25%-50%
SYNTH								
Weak	0.56	0.75	0.57	0.80	0.55	0.75	0.61	0.82
Mid	0.60	0.90	0.64	0.96	0.58	0.88	0.70	0.97
Strong	0.61	0.92	0.66	0.98	0.59	0.91	0.72	0.99
REAL								
Weak	0.56	0.72	0.53	0.66	0.55	0.70	0.54	0.69
Mid	0.73	1.00	0.57	0.92	0.78	1.00	0.76	0.98
Strong	0.67	0.99	0.54	0.94	0.81	1.00	0.77	1.00



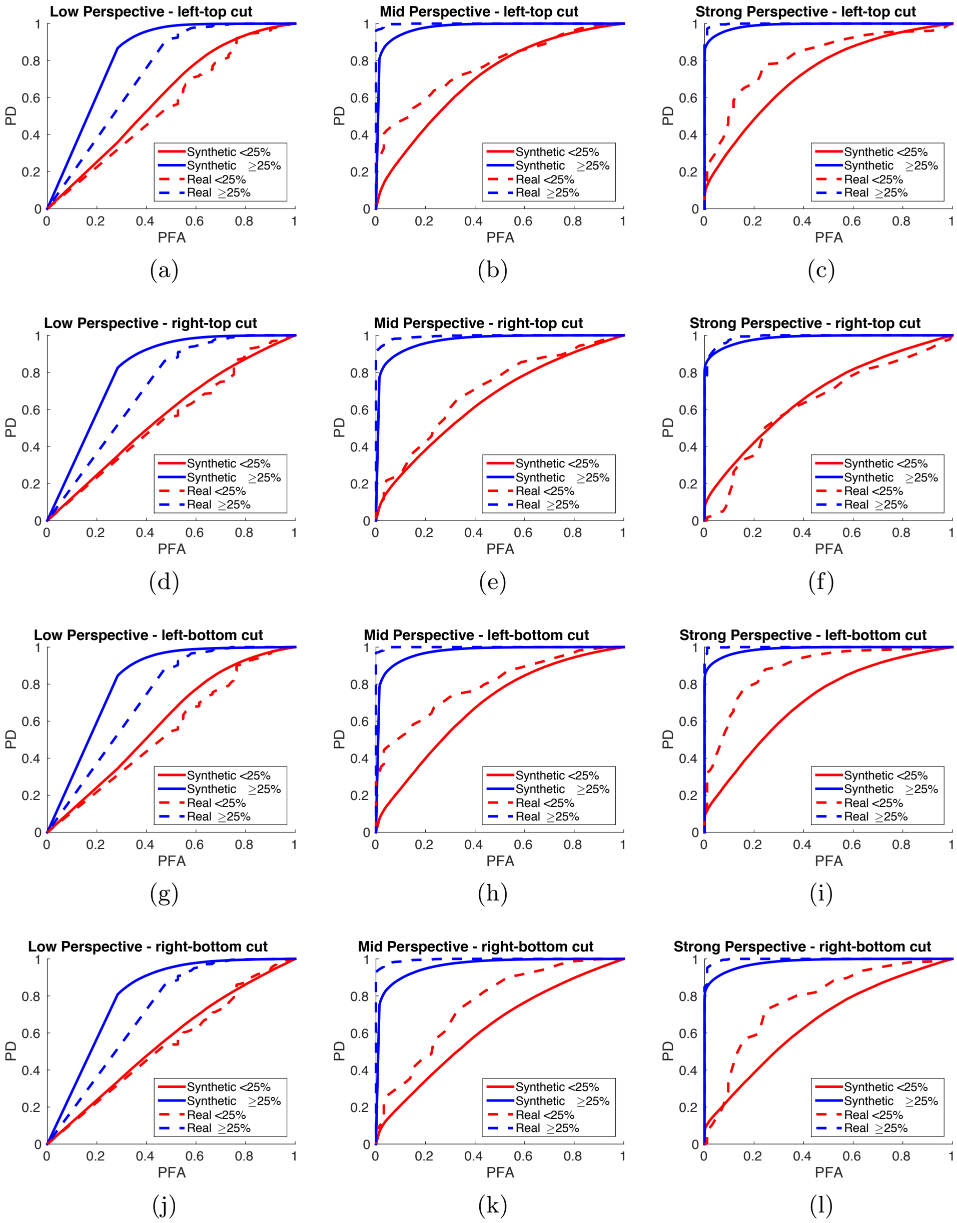
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.1: (Best viewed in color) ROC curves for Synthetic and Real Tests combining three different perspective conditions, slightly and strongly cropped images along different single cropping directions. In columns Low, Mid and Strong perspective respectively; in rows left, top, right and bottom cuts. Red and blue are used for slightly and strongly cropped images respectively. Continuous and dashed lines are used for synthetic and real data respectively.



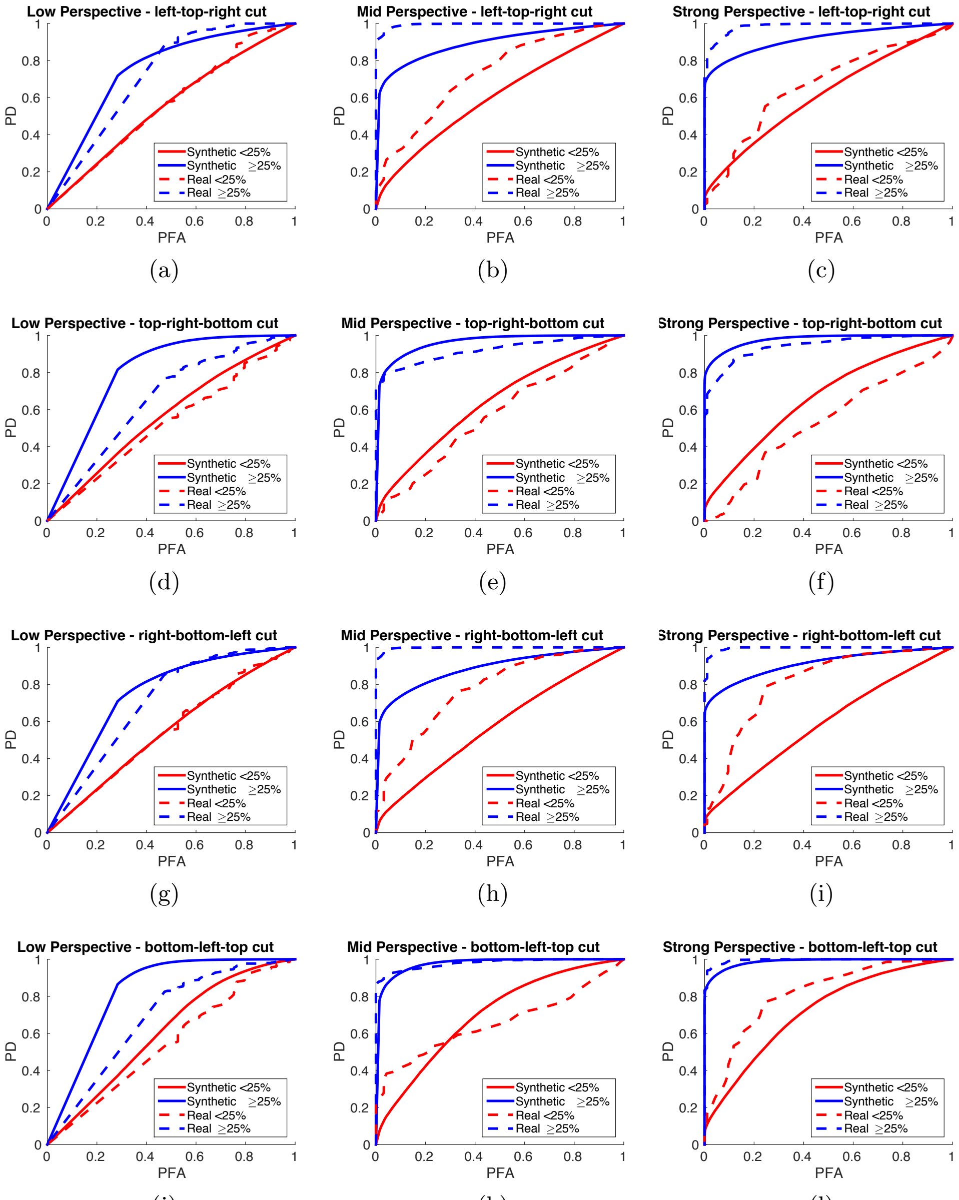
(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.2: (Best viewed in color) ROC curves for Synthetic and Real Tests combining three different perspective conditions, slightly and strongly cropped images along two cropping directions. In columns Low, Mid and Strong perspective respectively; in rows left-top, right-top, left-bottom and right-bottom cuts. Red and blue are used for slightly and strongly cropped images respectively. Continuous and dashed lines are used for synthetic and real data respectively.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Figure A.3: (Best viewed in color) ROC curves for Synthetic and Real Tests combining three different perspective conditions, slightly and strongly cropped images along three cropping directions. In columns Low, Mid and Strong perspective respectively; in rows left-top-right, top-right-bottom, right-bottom-left, bottom-left-top cuts. Red and blue are used for slightly and strongly cropped images respectively. Continuous and dashed lines are used for synthetic and real data respectively.

Bibliography

[1] S. Inc., "Statista." [Online]. Available: http://www.statista.com/statistics/263437/global-smartphone-sales-to-end-users-since-2007/

[2] F. led initiative, "internet.org." [Online]. Available: http://internet.org/about

[3] H. Farid, "A survey of image forgery detection," *IEEE Signal Processing Magazine*, vol. 26(2), pp. 16–25, 2009.

[4] M. Stamm, M. Wu, and K. Liu, "Information forensics: An overview of the first decade," *IEEE Access*, vol. 1, pp. 167–200, 2013.

[5] G. K. Birajdar and V. H. Mankar, "Digital image forgery detection using passive techniques: A survey," *Digital Investigation*, vol. 10, no. 3, pp. 226 – 245, 2013.

[6] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, pp. Article ID 496 701, 22 pages, 2013. [Online]. Available: http://www.hindawi.com/isrn/sp/2013/496701/

[7] P. Ferrara, T. Bianchi, A. De Rosa, and A. Piva, "Image forgery localization via fine-grained analysis of cfa artifacts," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 5, pp. 1566–1577, Oct 2012.

[8] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 1, pp. 74–90, March 2008.

[9] B. Li, T.-T. Ng, X. Li, S. Tan, and J. Huang, "Revealing the trace of high-quality jpeg compression through quantization noise analysis," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 3, pp. 558–573, March 2015.

[10] T. Bianchi and A. Piva, "Image forgery localization via block-grained analysis of jpeg artifacts," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 3, pp. 1003–1017, June 2012.

[11] 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. [Online]. Available: http://doi.acm.org/10.1145/2487228.2487236

[12] T. Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. de Rezende Rocha, "Exposing digital image forgeries by illumination color classification," *IEEE Transactions on Information Forensics and Security*, pp. 1182–1194, 2013.

[13] M. Johnson and H. Farid, "Exposing digital forgeries in complex lighting environments," *Information Forensics and Security, IEEE Transactions on*, vol. 2, no. 3, pp. 450 –461, sept. 2007.

[14] H. Yao, S. Wang, Y. Zhao, and X. Zhang, "Detecting image forgery using perspective constraints," *Signal Processing Letters, IEEE*, vol. 19, no. 3, pp. 123–126, March 2012.

[15] M. Iuliani, G. Fabbri, and A. Piva, "Image splicing detection based on general perspective constraints," in *Proceedings of the Information Forensics and Security (WIFS), 2015 IEEE International Workshop*, 2015.

[16] M. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris, "Detecting image splicing in the wild (web)," in *Proc. IEEE Int Multimedia & Expo Workshops (ICMEW) Conf*, 2015, pp. 1–6.

[17] T. Carvalho, H. Farid, and E. Kee, "Exposing photo manipulation from user-guided 3d lighting analysis," in *Proc. SPIE*, vol. 9409, 2015, pp. 940 902–940 902–10.

[18] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Comput. Surv.*, vol. 43, no. 4, pp. 26:1–26:42, Oct. 2011. [Online]. Available: http://doi.acm.org/10.1145/1978802.1978805

[19] A. Piva, "An overview on image forensics," *ISRN Signal Processing*, vol. 2013, pp. 1–22, 2013.

[20] Z. Dias, A. Rocha, and S. Goldenstein, "Image phylogeny by minimal spanning trees," *Information Forensics and Security, IEEE Transactions on*, vol. 7, no. 2, pp. 774–788, 2012.

[21] E. Kee, M. Johnson, and H. Farid, "Digital image authentication from jpeg headers," *Information Forensics and Security, IEEE Transactions on*, vol. 6, no. 3, pp. 1066–1075, Sept 2011.

[22] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *Information Forensics and Security, IEEE Transactions on*, vol. 1, no. 2, pp. 205–214, June 2006.

[23] C. Hass, "Jpegsnoop." [Online]. Available: http://www.impulseadventure.com/photo/jpeg-snoop.html

[24] Z. Fan and R. de Queiroz, "Identification of bitmap compression history: Jpeg detection and quantizer estimation," *Image Processing, IEEE Transactions on*, vol. 12, no. 2, pp. 230–235, Feb 2003.

[25] W. Luo, J. Huang, and G. Qiu, "Jpeg error analysis and its applications to digital image forensics," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 3, pp. 480–491, Sept 2010.

[26] J. Yang, J. Xie, G. Zhu, S. Kwong, and Y.-Q. Shi, "An effective method for detecting double jpeg compression with the same quantization matrix," *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 11, pp. 1933–1942, Nov 2014.

[27] H. Farid, "Exposing digital forgeries from jpeg ghosts," *Information Forensics and Security, IEEE Transactions on*, vol. 4, no. 1, pp. 154–160, March 2009.

[28] E. Kee and H. Farid, "Exposing digital forgeries from 3-d lighting environments," *IEEE International Workshop on Information Forensics and Security*, 2010.

[29] P. Saboia, T. Carvalho, and A. Rocha, "Eye specular highlights telltales for digital forensics: A machine learning approach," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, sept. 2011, pp. 1937 –1940.

[30] E. Kee and H. Farid, "Detecting photographic composites of famous people," Dartmouth College, Tech. Rep., 2009.

[31] T. Carvalho, H. Farid, and E. Kee, "Exposing photo manipulation from user-guided 3d lighting analysis," in *IS&T/SPIE Electronic Imaging*, vol. 9409, 2015, pp. 940 902–940 902–10.

[32] E. Kee, J. F. O'brien, and H. Farid, "Exposing photo manipulation from shading and shadows," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 165:1–165:21, Sep. 2014. [Online]. Available: http://doi.acm.org/10.1145/2629646

[33] C. Riess and E. Angelopoulou, "Scene illumination as an indicator of image manipulation," in *Proceedings of the 12th international conference on Information hiding*, 2010.

[34] T. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. de Rezende Rocha, "Exposing digital image forgeries by illumination color classification," *Information Forensics and Security, IEEE Transactions on*, vol. 8, no. 7, pp. 1182–1194, July 2013.

[35] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[36] V. Conotter, G. Boato, and H. Farid, "Detecting photo manipulation on signs and billboards," in *Image Processing (ICIP), 2010 17th IEEE International Conference on*, Sept 2010, pp. 1741–1744.

[37] J. Zheng, T. Zhu, Z. Li, W. Xing, and J. Ren, "Exposing image forgery by detecting traces of feather operation," *Journal of Visual Languages & Computing*, vol. 27, pp. 9–18, 2015.

[38] SWGDE, "Scientific Working Group on Digital Evidence," accessed: 2014-07-02. [Online]. Available: https://www.swgde.org/documents/CurrentDocuments

[39] SWGIT, "Scientific Working Group on Imaging Technology," accessed: 2014-07-02. [Online]. Available: https://www.swgit.org/documents/CurrentDocuments

[40] A. De Rosa, A. Piva, M. Fontani, and M. Iuliani, "Investigating multimedia contents," in *Security Technology (ICCST), 2014 International Carnahan Conference on*, 2014, pp. 1–6.

[41] N. Khanna, A. K. Mikkilineni, G. T. Chiu, J. P. Allebach, and E. J. Delp, "Forensic classification of imaging sensor types," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 65 050U–65 050U.

[42] S. Bayram, H. T. Sencar, N. Memon, and I. Avcibas, "Source camera identification based on cfa interpolation," in *IEEE International Conference on Image Processing (ICIP)*, vol. 3. IEEE, 2005, pp. III–69.

[43] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A framework for decision fusion in image forensics based on dempster–shafer theory of evidence," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 593–607, 2013.

[44] "Exiftool," http://www.sno.phy.queensu.ca/~phil/exiftool/, accessed: 2014-07-02.

[45] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A forensic tool for investigating image forgeries," *International Journal of Digital Crime and Forensics (IJDCF)*, vol. 5, no. 4, pp. 15–33, 2013.

[46] T. Bianchi, A. De Rosa, and A. Piva, "Improved DCT coefficient analysis for forgery localization in JPEG images," in *ICASSP 2011, IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, CZ, 2011.

[47] H. Farid, "Exposing digital forgeries from JPEG ghosts," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 1, pp. 154–160, 2009.

[48] T. Bianchi and A. Piva, "Detection of non-aligned double JPEG compression with estimation of primary compression parameters," in *ICIP 2011, IEEE International Conference on Image Processing*, Brussels, BE, September 2011, pp. 1929 –1932.

[49] M. Fontani, T. Bianchi, A. De Rosa, A. Piva, and M. Barni, "A Framework for Decision Fusion in Image Forensics Based on Dempster-Shafer Theory of Evidence," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 4, pp. 593–607, 2013.

[50] V. Schetinger, M. M. Oliveira, R. da Silva, and T. J. Carvalho, "Humans are easily fooled by digital images," *CoRR*, vol. abs/1509.05301, 2015. [Online]. Available: http://arxiv.org/abs/1509.05301

[51] A. Systems, "Adobe photoshop." [Online]. Available: http://adobe.com/photoshop

[52] P. M. Spencer Kimball, "Gnu image manipulation program." [Online]. Available: www.gimp.org

[53] M. Gryka, M. Terry, and G. J. Brostow, "Learning to remove soft shadows," *ACM Transactions on Graphics*, 2015.

[54] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, H. Jin, R. Fonte, M. Sittig, and D. Forsyth, "Automatic scene inference for 3d object compositing," *ACM Trans. Graph.*, vol. 33, no. 3, June 2014.

[55] E. S. L. Gastal and M. M. Oliveira, "High-order recursive filtering of non-uniformly sampled signals for image and video processing," *Computer Graphics Forum*, vol. 34, no. 2, pp. 81–93, May 2015, proceedings of Eurographics 2015.

[56] J. Liao, R. S. Lima, D. Nehab, H. Hoppe, P. V. Sander, and J. Yu, "Automating image morphing using structural similarity on a halfway domain," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 168:1–168:12, Sep. 2014. [Online]. Available: http://doi.acm.org/10.1145/2629494

[57] S. Xue, A. Agarwala, J. Dorsey, and H. Rushmeier, “Understanding and improving the realism of image composites,” *ACM Transactions on Graphics*, vol. 31, no. 84, pp. 84:1 – 84:10, 07/2010 2012. [Online]. Available: http://doi.acm.org/10.1145/2185520.2185580

[58] E. N. Mortensen and W. A. Barrett, “Intelligent scissors for image composition,” in *Proceedings of the 22Nd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH ’95. New York, NY, USA: ACM, 1995, pp. 191–198. [Online]. Available: http://doi.acm.org/10.1145/218380.218442

[59] H. Huang, L. Zhang, and H.-C. Zhang, “Repsnapping: Efficient image cutout for repeated scene elements,” *Computer Graphics Forum*, vol. 30, no. 7, pp. 2059–2066, 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2011.02044.x

[60] P. Sutthiwan, Y. Q. Shi, W. Su, and T.-T. Ng, “Rake transform and edge statistics for image forgery detection,” in *IEEE International Conference on Multimedia and EXPO*, 2010.

[61] P. Pérez, M. Gangnet, and A. Blake, “Poisson image editing,” *ACM Trans. Graph.*, vol. 22, no. 3, pp. 313–318, Jul. 2003. [Online]. Available: http://doi.acm.org/10.1145/882262.882269

[62] M. W. Tao, M. K. Johnson, and S. Paris, “Error-tolerant image compositing,” in *European Conference on Computer Vision (ECCV)*, 2010. [Online]. Available: http://graphics.cs.berkeley.edu/papers/Tao-ERR-2010-09/

[63] M. Ding and R.-F. Tong, “Content-aware copying and pasting in images,” *Vis. Comput.*, vol. 26, no. 6-8, pp. 721–729, Jun. 2010. [Online]. Available: http://dx.doi.org/10.1007/s00371-010-0448-8

[64] K. Sunkavalli, M. K. Johnson, W. Matusik, and H. Pfister, “Multi-scale image harmonization,” *ACM Transactions on Graphics*, vol. 29, no. 4, pp. 125:1–125:10, 2010.

[65] Z. Farbman, R. Fattal, and D. Lischinski, “Convolution pyramids,” *ACM Trans. Graph.*, vol. 30, no. 6, pp. 175:1–175:8, Dec. 2011. [Online]. Available: http://doi.acm.org/10.1145/2070781.2024209

[66] S. Darabi, E. Shechtman, C. Barnes, D. B. Goldman, and P. Sen, “Image Melding: Combining inconsistent images using patch-based synthesis,” *ACM Transactions on Graphics (TOG) (Proceedings of SIGGRAPH 2012)*, vol. 31, no. 4, pp. 82:1–82:10, 2012.

[67] E. S. L. Gastal and M. M. Oliveira, "Shared sampling for real-time alpha matting," *Computer Graphics Forum*, vol. 29, no. 2, pp. 575–584, May 2010, proceedings of Eurographics.

[68] K. Karsch, V. Hedau, D. Forsyth, and D. Hoiem, "Rendering synthetic objects into legacy photographs," in *Proceedings of the 2011 SIGGRAPH Asia Conference*, ser. SA '11. New York, NY, USA: ACM, 2011, pp. 157:1–157:12. [Online]. Available: http://doi.acm.org/10.1145/2024156.2024191

[69] H. Wu, Y.-S. Wang, K.-C. Feng, T.-T. Wong, T.-Y. Lee, and P.-A. Heng, "Resizing by symmetry-summarization," *ACM Transactions on Graphics (SIGGRAPH Asia 2010 issue)*, vol. 29, no. 6, pp. 159:1–159:9, December 2010.

[70] Y. Zheng, X. Chen, M.-M. Cheng, K. Zhou, S.-M. Hu, and N. J. Mitra, "Interactive images: Cuboid proxies for smart image manipulation," *ACM Transactions on Graphics*, vol. 31, no. 4, pp. 99:1–99:11, 2012.

[71] T. Chen, Z. Zhu, A. Shamir, S.-M. Hu, and D. Cohen-Or, "3sweepp: Extracting editable objects from a single photo," *ACM Trans. Graph.*, vol. 32, no. 6, pp. 195:1–195:10, Nov. 2013. [Online]. Available: http://doi.acm.org/10.1145/2508363.2508378

[72] N. Kholgade, T. Simon, A. Efros, and Y. Sheikh, "3d object manipulation in a single photograph using stock 3d models," *ACM Transactions on Computer Graphics*, vol. 33, no. 4, 2014.

[73] S. Zhou, H. Fu, L. Liu, D. Cohen-Or, and X. Han, "Parametric reshaping of human bodies in images," *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH)*, vol. 29, pp. Article No. 126, 1–10, 2010.

[74] M. Chai, L. Wang, Y. Weng, X. Jin, and K. Zhou, "Dynamic hair manipulation in images and videos," *ACM Trans. Graph.*, vol. 32, no. 4, pp. 75:1–75:8, Jul. 2013. [Online]. Available: http://doi.acm.org/10.1145/2461912.2461990

[75] Y. Weng, L. Wang, X. Li, M. Chai, and K. Zhou, "Hair interpolation for portrait morphing," *Computer Grap*, vol. 32, 2013.

[76] H. Farid and M. Bravo, "Perceptual discrimination of computer generated and photographic faces," *Digital Investigation*, 2012.

[77] D.-T. Dang-Nguyen, "Discrimination of computer generated versus natural human faces," Ph.D. dissertation, University of Trento, UNITN, 2014.

[78] F. Peng and D. lan Zhou, "Discriminating natural images and computer generated graphics based on the impact of cfa interpolation on the correlation of prnu" *Digital Investigation*, no. 0, pp. –, 2014.

[79] F. Peng, J. Li, and M. Long, "Discriminating natural images and computer generated graphics based on compound fractal features," *Journal of Computational Information Systems*, vol. 9, no. 13, pp. 101–5108, 2013.

[80] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH '00. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 417–424. [Online]. Available: http://dx.doi.org/10.1145/344779.344972

[81] H. Huang, K.Yin, M. Gong, D. Lischinski, D. Cohen-Or, U. Ascher, and B. Chen, "Mind the gap: Tele-registration for structure-driven image completion," *ACM Transactions on Graphics (Proceedings of SIGGRAPH ASIA 2013)*, vol. 32, pp. 174:1–174:10, 2013.

[82] J. Kopf, W. Kienzle, S. Drucker, and S. B. Kang, "Quality prediction for image completion," *ACM Trans. Graph.*, vol. 31, no. 6, pp. 131:1–131:8, Nov. 2012. [Online]. Available: http://doi.acm.org/10.1145/2366145.2366150

[83] M. Daisy, D. Tschumperlé, and O. Lézoray, "A fast spatial patch blending algorithm for artefact reduction in pattern-based image inpainting," in *SIGGRAPH Asia 2013 Technical Briefs*, ser. SA '13. New York, NY, USA: ACM, 2013, pp. 8:1–8:4. [Online]. Available: http://doi.acm.org/10.1145/2542355.2542365

[84] J.-B. Huang, S. B. Kang, N. Ahuja, and J. Kopf, "Image completion using planar structure guidance," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2014)*, vol. 33, no. 4, p. to appear, 2014.

[85] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, Jul. 2007. [Online]. Available: http://doi.acm.org/10.1145/1276377.1276390

[86] D. Panozzo, O. Weber, and O. Sorkine, "Robust image retargeting via axis-aligned deformation," *Computer Graphics Forum (proceedings of EUROGRAPHICS)*, vol. 31, no. 2, pp. 229–236, 2012.

[87] L. Liu, Y. Jin, and Q. Wu, "Realtime Aesthetic Image Retargeting," in *Computational Aesthetics in Graphics, Visualization, and Imaging*, P. Jepp and O. Deussen, Eds. The Eurographics Association, 2010.

[88] A. Dirik, H. Sencar, and N. Memon, “Analysis of seam-carving-based anonymization of images against prnu noise pattern-based source attribution,” *Information Forensics and Security, IEEE Transactions on*, vol. 9, no. 12, pp. 2277–2290, 2014.

[89] T. Yin, G. Yang, L. Li, D. Zhang, and X. Sun, “Detecting seam carving based image resizing using local binary patterns,” *Computers & Security*, vol. 55, pp. 130 – 141, 2015.

[90] R. Wanat and R. K. Mantiuk, “Simulating and compensating changes in appearance between day and night vision,” *ACM Trans. Graph.*, vol. 33, no. 4, pp. 147:1–147:12, Jul. 2014. [Online]. Available: http://doi.acm.org/10.1145/2601097.2601150

[91] P.-Y. Laffont, A. Bousseau, S. Paris, F. Durand, and G. Drettakis, “Coherent intrinsic images from photo collections,” *ACM Transactions on Graphics (SIGGRAPH Asia Conference Proceedings)*, vol. 31, 2012. [Online]. Available: http://www-sop.inria.fr/reves/Basilic/2012/LBPDD12

[92] S. Bell, K. Bala, and N. Snavely, “Intrinsic images in the wild,” *ACM Trans. Graph.*, vol. 33, no. 4, pp. 159:1–159:12, Jul. 2014. [Online]. Available: http://doi.acm.org/10.1145/2601097.2601206

[93] Y. Endo, Y. Kanamori, Y. Fukui, and J. Mitani, “Matting and compositing for fresnel reflection on wavy surfaces,” *Computer Graphics Forum (Proc. of Eurographics Symposium on Rendering 2012)*, vol. 31, no. 4, pp. 1435–1443, 2012.

[94] S. Sinha, J. Kopf, M. Goesele, D. Scharstein, and R. Szeliski, “Image-based rendering for scenes with reflections,” *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2012)*, vol. 31, no. 4, p. to appear, 2012.

[95] M. Hullin, E. Eisemann, H.-P. Seidel, and S. Lee, “Physically-based real-time lens flare rendering,” *ACM Trans. Graph.*, vol. 30, no. 4, pp. 108:1–108:10, Jul. 2011. [Online]. Available: http://doi.acm.org/10.1145/2010324.1965003

[96] S. Lee and E. Eisemann, “Practical real-time lens-flare rendering,” in *Proceedings of the Eurographics Symposium on Rendering*, ser. EGSR ’13. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2013, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1111/cgf.12145

[97] R. Guo, Q. Dai, and D. Hoiem, “Single-image shadow detection and removal using paired regions,” in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition*, ser. CVPR

’11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 2033–2040. [Online]. Available: http://dx.doi.org/10.1109/CVPR.2011.5995725

[98] G. D. Finlayson, S. D. Hordley, C. Lu, and M. S. Drew, “Removing shadows from images,” in *In ECCV 2002: European Conference on Computer Vision*, 2002, pp. 823–836.

[99] S. Xue, A. Agarwala, J. Dorsey, and H. E. Rushmeier, “Understanding and improving the realism of image composites,” *ACM Trans. Graph.*, vol. 31, no. 4, pp. 84:1–84:10, 2012. [Online]. Available: http://doi.acm.org/10.1145/2185520.2185580

[100] J. Lopez-Moreno, S. Hadap, E. Reinhard, and D. Gutierrez, “Compositing images through light source detection,” *Computers & Graphics*, vol. 34, no. 6, pp. 698 – 707, 2010, graphics for Serious GamesComputer Graphics in Spain: a Selection of Papers from {CEIG} 2009Selected Papers from the {SIGGRAPH} Asia Education Program. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0097849310001299

[101] P. Kaufmann, O. Wang, A. Sorkine-Hornung, O. Sorkine-Hornung, A. Smolic, and M. H. Gross, “Finite element image warping.” *Comput. Graph. Forum*, vol. 32, no. 2, pp. 31–39, 2013. [Online]. Available: http://dblp.uni-trier.de/db/journals/cgf/cgf32.html

[102] Y. Shih, S. Paris, C. Barnes, W. T. Freeman, and F. Durand, “Style transfer for headshot portraits,” *ACM Trans. Graph.*, vol. 33, no. 4, pp. 148:1–148:14, Jul. 2014. [Online]. Available: http://doi.acm.org/10.1145/2601097.2601137

[103] P.-Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays, “Transient attributes for high-level understanding and editing of outdoor scenes,” *ACM Transactions on Graphics (proceedings of SIGGRAPH)*, vol. 33, no. 4, 2014.

[104] Y. HaCohen, E. Shechtman, D. B. Goldman, and D. Lischinski, “Optimizing color consistency in photo collections,” *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2013)*, vol. 32, no. 4, pp. 85:1 – 85:9, 2013.

[105] Y. Liu, M. Cohen, M. Uyttendaele, and S. Rusinkiewicz, “AutoStyle: Automatic style transfer from image collections to users’ images,” *Computer Graphics Forum (Proc. Eurographics Symposium on Rendering)*, vol. 33, no. 4, Jun. 2014.

[106] R. Carroll, R. Ramamoorthi, and M. Agrawala, “Illumination decomposition for material recoloring with consistent interreflections,” *ACM Trans. Graph.*, vol. 30, no. 4, pp. 43:1–43:10, Jul. 2011. [Online]. Available: http://doi.acm.org/10.1145/2010324.1964938

[107] J. I. Echevarria, G. Wilensky, A. Krishnaswamy, B. Kim, and D. Gutierrez, “Computational simulation of alternative photographic processes,” *Computer Graphics Forum (Proc. EGSR 2013)*, vol. 32, no. 4, 2013.

[108] E. S. L. Gastal and M. M. Oliveira, “Adaptive manifolds for real-time high-dimensional filtering,” *ACM TOG*, vol. 31, no. 4, pp. 33:1–33:13, 2012, proceedings of SIGGRAPH 2012.

[109] H. Cho, H. Lee, H. Kang, and S. Lee, “Bilateral texture filtering,” *ACM Trans. Graph.*, vol. 33, no. 4, pp. 128:1–128:8, Jul. 2014. [Online]. Available: http://doi.acm.org/10.1145/2601097.2601188

[110] H. Lieng, J. Tompkin, and J. Kautz, “Interactive multi-perspective imagery from photos and videos,” in *Computer Graphics Forum (Proceedings of Eurographics 2012)*, vol. 31, no. 2pt1, May 2012, pp. 285–293. [Online]. Available: http://dx.doi.org/10.1111/j.1467-8659.2012.03007.x

[111] R. Carroll, A. Agarwala, and M. Agrawala, “Image warps for artistic perspective manipulation,” *ACM Trans. Graph.*, vol. 29, no. 4, pp. 127:1–127:9, Jul. 2010. [Online]. Available: http://doi.acm.org/10.1145/1778765.1778864

[112] K.-T. Lee, S.-J. Luo, and B.-Y. Chen, “Rephotography using image collections,” *Computer Graphics Forum*, vol. 30, no. 7, pp. 1895–1901, 2011, (Pacific Graphics 2011 Conference Proceedings).

[113] I. Boyadzhiev, K. Bala, S. Paris, and F. Durand, “User-guided white balance for mixed lighting conditions,” *ACM Trans. Graph.*, vol. 31, no. 6, pp. 200:1–200:10, Nov. 2012. [Online]. Available: http://doi.acm.org/10.1145/2366145.2366219

[114] E. Hsu, T. Mertens, S. Paris, S. Avidan, and F. Durand, “Light mixture estimation for spatially varying white balance,” *ACM Trans. Graph.*, vol. 27, no. 3, pp. 70:1–70:7, Aug. 2008. [Online]. Available: http://doi.acm.org/10.1145/1360612.1360669

[115] M. W. Tao, J. Malik, and R. Ramamoorthi, “Sharpening out of focus images using high-frequency transfer,” *Computer Graphics Forum (Eurographics 2013)*, 2013. [Online]. Available: http://graphics.berkeley.edu/papers/Tao-SOO-2013-05/

[116] N. Joshi, W. Matusik, E. H. Adelson, and D. J. Kriegman, “Personal photo enhancement using example images,” *ACM Trans. Graph.*, vol. 29, no. 2, pp. 12:1–12:15, Apr. 2010. [Online]. Available: http://doi.acm.org/10.1145/1731047.1731050

[117] K. Yücer, A. Jacobson, A. Hornung, and O. Sorkine, “Transfusive image manipulation,” *ACM Transactions on Graphics (proceedings of ACM SIGGRAPH ASIA)*, vol. 31, no. 6, pp. 176:1–176:9, 2012.

[118] “Krawets, neil. ”a picture’s worth: Digital image analysis and forensics,” http://www.hackerfactor.com/papers/bh-usa-07-krawetz-wp.pdf, accessed: 2016-10-17.

[119] T. Bianchi and A. Piva, “Image forgery localization via block-grained analysis of jpeg artifacts,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 3, pp. 1003 – 1017, 2012.

[120] H. Yao, S. Wang, Y. Zhao, and X. Zhang, “Detecting image forgery using perspective constraints,” *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 123–126, 2012.

[121] M. K. Johnson and H. Farid, “Detecting photographic composites of people.” in *IWDW*, ser. Lecture Notes in Computer Science, Y. Q. Shi, H.-J. Kim, and S. K. 0001, Eds., vol. 5041. Springer, 2007, pp. 19–33.

[122] V. Conotter, G. Boato, and H. Farid, “Detecting photo manipulation on signs and billboards,” in *International Conference on Image Processing*, 2010. [Online]. Available: www.cs.dartmouth.edu/farid/publications/icip10.html

[123] A. Criminisi, I. Reid, and A. Zisserman, “Single view metrology,” in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 1, 1999, pp. 434–441 vol.1.

[124] C. Hass, “Jpegsnoop.” [Online]. Available: http://www.impulseadventure.com/photo/jpeg-snoop.html

[125] J. Hu, Y. Li, S. Niu, and X. Meng, “Exposing digital image forgeries by detecting inconsistencies in principal point,” in *Computer Science and Service System (CSSS), 2011 International Conference on*, June 2011, pp. 404–407.

[126] X. Meng, S. Niu, R. Yan, and Y. Li, “Detecting photographic cropping based on vanishing points,” *Chinese Journal of Electronics*, vol. 22, pp. Article ID 496 701, 22 pages, 2013.

[127] G. Medioni and S. B. Kang, *Emerging Topics in Computer Vision.* Upper Saddle River, NJ, USA: Prentice Hall PTR, 2004.

[128] B. Caprile and V. Torre, "Using vanishing points for camera calibration," *Int. J. Comput. Vision*, vol. 4, no. 2, pp. 127–140, May 1990.

[129] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000.

[130] R. Toldo, R. Gherardi, M. Farenzena, and A. Fusiello, "Hierarchical structure-and-motion recovery from uncalibrated images," *Comput. Vis. Image Underst.*, vol. 140, no. C, pp. 127–143, Nov. 2015.

[131] C. Colombo, D. Comanducci, and A. Del Bimbo, "Camera Calibration with Two Arbitrary Coaxial Circles," in *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. Proceedings, Part I.* Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–276.

[132] E. Guillou, D. Meneveaux, E. Maisel, and K. Bouatouch, "Using vanishing points for camera calibration and coarse 3d reconstruction from a single image," *The Visual Computer*, vol. 16, no. 7, pp. 396–410, 2000.

[133] J. Deutscher, M. Isard, and J. Maccormick, "Automatic camera calibration from a single manhattan image," in *Eur. Conf. on Computer Vision (ECCV*, 2002, pp. 175–205.

[134] R. Pflugfelder and H. Bischof, "Online auto-calibration in man-made worlds," in *Digital Image Computing: Techniques and Applications (DICTA'05)*, Dec 2005, pp. 75–75.

[135] J. M. Coughlan and A. L. Yuille, "Manhattan world: compass direction from a single image by bayesian inference," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, 1999, pp. 941–947 vol.2.

[136] Y. LI, Y. jian ZHOU, K. guo YUAN, Y. cui GUO, and X. xin NIU, "Exposing photo manipulation with inconsistent perspective geometry," *The Journal of China Universities of Posts and Telecommunications*, vol. 21, no. 4, pp. 83 – 104, 2014.

[137] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, Nov 1986.

[138] J. Košecka and W. Zhang, "Efficient computation of vanishing points," in *Robotics and Automation, 2002. Proceedings. ICRA '02. IEEE International Conference on*, vol. 1, 2002, pp. 223–228 vol.1.

[139] T. Tuytelaars, L. V. Gool, M. Proesmans, and T. Moons, “The cascaded hough transform as an aid in aerial image interpretation,” in *Computer Vision, 1998. Sixth International Conference on*, Jan 1998, pp. 67–72.

[140] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[141] R. Toldo and A. Fusiello, “Robust multiple structures estimation with j-linkage,” in *Proceedings of the 10th European Conference on Computer Vision: Part I*, ser. ECCV ’08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 537–547.

[142] J. P. Tardif, “Non-iterative approach for fast and accurate vanishing point detection,” in *2009 IEEE 12th International Conference on Computer Vision*, Sept 2009, pp. 1250–1257.

[143] J. C. Bazin, Y. Seo, C. Demonceaux, P. Vasseur, K. Ikeuchi, I. Kweon, and M. Pollefeys, “Globally optimal line clustering and vanishing point estimation in manhattan world,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, June 2012, pp. 638–645.

[144] C. Rother, “A new approach for vanishing point detection in architectural environments,” in *In Proc. 11th British Machine Vision Conference*, 2000, pp. 382–391.

[145] J. C. Bazin and M. Pollefeys, “3-line ransac for orthogonal vanishing point detection,” in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2012, pp. 4282–4287.

[146] P. Denis, J. H. Elder, and F. J. Estrada, “Efficient edge-based methods for estimating manhattan frames in urban imagery,” in *Computer Vision – ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part II*, D. Forsyth, P. Torr, and A. Zisserman, Eds. Springer Berlin Heidelberg, 2008, pp. 197–210. [Online]. Available: http://www.elderlab.yorku.ca/YorkUrbanDB/

[147] T. Peterson, “Facebook Users Are Posting 75% More Videos Than Last Year,” http://adage.com/article/digital/facebook-users-posting-75-videos-year/296482/, 2015, [Online; accessed 20-October-2015].

[148] M. Beck, “Reversal Of Facebook: Photo Posts Now Drive Lowest Organic Reach,” http://marketingland.com/want-maximum-reach-facebook-dont-post-photos-118536, 2015, [Online; accessed 20-October-2015].

[149] R. Maxwell, “Camera vs. Smartphone: Infographic shares the impact our smartphones have had on regular cameras,” http://www.phonearena.com/news/.

[150] J. Lukas, J. Fridrich, and M. Goljan, “Digital camera identification from sensor pattern noise,” *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[151] A. Castiglione, G. Cattaneo, M. Cembalo, and U. F. Petrillo, “Experimentations with source camera identification and online social networks,” *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 2, pp. 265–274, 2013.

[152] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, “Source digital camcorder identification using sensor photo response non-uniformity,” in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 65 051G–65 051G.

[153] A. E. Dirik, H. T. Sencar, and N. Memon, “Digital single lens reflex camera identification from traces of sensor dust,” *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 3, pp. 539–552, 2008.

[154] Z. J. Geradts, J. Bijhold, M. Kieft, K. Kurosawa, K. Kuroki, and N. Saitoh, “Methods for identification of images acquired with digital cameras,” in *Enabling Technologies for Law Enforcement*. International Society for Optics and Photonics, 2001, pp. 505–512.

[155] M. Goljan, J. Fridrich, and T. Filler, “Large scale test of sensor fingerprint camera identification,” in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2009, pp. 72 540I–72 540I.

[156] G. Cattaneo, G. Roscigno, and U. F. Petrillo, “A scalable approach to source camera identification over hadoop,” in *IEEE International Conference on Advanced Information Networking and Applications (AINA)*. IEEE, 2014, pp. 366–373.

[157] T. Gloe and R. Böhme, “The dresden image database for benchmarking digital image forensics,” *Journal of Digital Forensic Practice*, vol. 3, no. 2-4, pp. 150–159, 2010.

[158] B.-b. Liu, X. Wei, and J. Yan, “Enhancing sensor pattern noise for source camera identification: An empirical evaluation,” in *Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security*, ser. IH&#38;MMSec ’15. New York, NY, USA: ACM, 2015, pp. 85–90. [Online]. Available: http://doi.acm.org/10.1145/2756601.2756614

[159] D. Valsesia, G. Coluccia, T. Bianchi, and E. Magli, "Compressed fingerprint matching and camera identification via random projections," *Information Forensics and Security, IEEE Transactions on*, vol. 10, no. 7, pp. 1472–1485, July 2015.

[160] W. Van Houten and Z. Geradts, "Source video camera identification for multiply compressed videos originating from youtube," *Digital Investigation*, vol. 6, no. 1, pp. 48–60, 2009.

[161] J. v. d. L. Scheelen, Yannick, Z. Geradts, and M. Worring, "Camera identification on youtube," *Chinese Journal of Forensic Science*, vol. 5, no. 64, pp. 19–30, 2012.

[162] W.-H. Chuang, H. Su, and M. Wu, "Exploring compression effects for improved source camera identification using strongly compressed video," in *IEEE International Conference on Image Processing (ICIP)*. IEEE, 2011, pp. 1953–1956.

[163] S. Chen, A. Pande, K. Zeng, and P. Mohapatra, "Live video forensics: Source identification in lossy wireless networks," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 1, pp. 28–39, 2015.

[164] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, vol. 6. IEEE, 1999, pp. 3253–3256.

[165] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, "Determining image origin and integrity using sensor noise," *Information Forensics and Security, IEEE Transactions on*, vol. 3, no. 1, pp. 74–90, 2008.

[166] C. R. Holt, "Two-channel likelihood detectors for arbitrary linear channel distortion," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 35, no. 3, pp. 267–273, 1987.

[167] M. Goljan and J. Fridrich, "Camera identification from scaled and cropped images," *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819, p. 68190E, 2008.

[168] M. Moltisanti, A. Paratore, S. Battiato, and L. Saravo, "Image manipulation on facebook for forensics evidence," in *Image Analysis and Processing - ICIAP 2015 - 18th International Conference, Genoa, Italy, September 7-11, 2015, Proceedings, Part II*, 2015, pp. 506–517. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-23234-8_47

[169] Z. P. Giammarrusco, "Source identification of high definition videos: A forensic analysis of downloaders and youtube video compression using a group of action cameras." Ph.D. dissertation, University of Colorado, 2014.

[170] "Keepvid," www.keepvid.com.