Università degli Studi di Firenze
Dipartimento di Ingegneria dell'Informazione (DINFO)
Corso di Dottorato in Ingegneria dell'Informazione

Curriculum: Telecomunicazioni

Image and video source identification in the era of mobile devices and social media

Candidate
Omar Alshaya

Supervisor
Prof. Alessandro Piva

PhD Coordinator
Prof. Luigi Chisci

XXXI, 2015-2018

Università degli Studi di Firenze, Dipartimento di Ingegneria
dell'Informazione (DINFO).

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Information Engineering.

Ancora imparo

Abstract

Modern world brought many technological advancements. Smartphone devices stand out among them, due to their popularity caused by having a large span of options available in a small, portable device. Their usage is increased in recent years, due to social media platforms, which became an inevitable part of everyday life, providing the ability to share information instantly with the selected audience. As popularity commonly causes vulnerability, neither smartphones, nor social media are spared of it. Multimedia content acquired by these devices and shared with other users is often altered for entertainment or malicious purposes, thus raising questions about its originality and authenticity.

Source identification is one of the burning issues that multimedia forensics copes with. Recent studies have shown that identification procedure can be successfully conducted relying on the characteristics of camera sensor noise, thus making it an interesting research approach. However, in order to obtain reliable results, multimedia tools need to be tested using an appropriate number and variety of multimedia information. Having in mind the constant development of today's portable devices, currently available databases became outdated, making the whole procedure difficult.

This Thesis introduces three novel image and video datasets, taking into account different types of multimedia and its alterations caused by the exchange through popular social media platforms. The first one is MOSES mobile application, proposed as an elegant option for providing an expandable, up-to-date video database. While the initial MOSES dataset contains SDR (*Standard Dynamic Range*) videos, the second dataset, named VISION, combines both SDR and HDR (*High Dynamic Range*) images and videos, thus providing the ability of comparison of different types of multimedia acquired by the same device. Due to their rising popularity, special attention is paid to HDR images. The third proposed dataset is one of the largest currently available HDR-based datsets and it enables SDR and HDR images comparison.

All the created datasets are used for source identification purposes, employing well-known PRNU (*Photo-Response Non-Uniformity*)-based methods. Exchanging multimedia content through social media platforms, using more complex multimedia types, such as HDR, as well as different camera movements, is shown to affect PRNU-based source identification procedure. Its reliability is shown to be dependable on the previously mentioned factors, thus opening space for further research in the field of multimedia forensics.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Technology has rapidly developed in the past few decades, simplifying many processes and providing users easier and faster options to produce desired result. While mobile phones could offer only text messaging and voice calls two decades ago, today's devices have overcome providing only telecommunication services. Digital camera, Internet access, mobile applications and all the available Internet services are included in addition to the previously existing options when smartphone devices were introduced. Having a wide range of services available in a small, portable device, caused higher technology usage. Users started capturing more photographs, recording videos, editing them in one of the large span of image or video processing applications and posting them to social media platforms. This enabled digital media to become available to the world in only a few clicks.

Having in mind that first mobile phone devices did not have a digital camera and that later versions started introducing ones with a very low pixel resolution, the only way for capturing quality photographs was using standard digital camera devices. As most of them were big, robust and heavy, they were not practical for every day usage, especially for non-professional photographers. Smartphones provided the advantage of having a high-resolution in-built camera, available anytime when carrying the smartphone itself. Furthermore, image and video processing software programs were available only on computer devices and laptops until the recent times. Therefore, multimedia content editing was time-consuming, editing software programs were not wide-spread among the users and more processor and memory power were required. However, editing is now approachable to the average smartphone user, with very low hardware and software requirements.

Introduction of Wi-Fi network provided the ability for uploading and downloading digital content on a specific location covered by Wi-Fi signal. The problem of being tied to some location in order to access the Internet has been overcome by the appearance of mobile networks, which are progressing in terms of speed, availability and battery consumption over the years. As a result, a big number of the world's population has the ability to be reachable through the Internet, regardless of their current location.

Although technology development brought many advantages, it can be used for malicious purposes, representing more curse than blessing of today's world. Digital media is now considered to be the main source of all the information globally. It has the coverage of almost every information related to every case, scenario and field. Therefore, it is often used in court, as an evidence of criminal activity or as an alibi. This fact puts the high importance on content originality, which is often harder to examine in comparison to non-digital evidences. In contrast to the printed media, manipulation of digital content is much easier due to its vast exposure and dependability. Information acquired from the digital media devices such as smartphones, camcorders, cameras etc. can easily be transferred to other devices and edited to change its perspective altogether. With the development of various post-processing techniques and software programs, information distortion became very common. Although the programs for digital content manipulation were developed for simplification of jobs related to camera-work, they are commonly used in forgery and fraud purposes. This has led to various difficulties related to the authentication of the information shared through the world in the form of multimedia content.

While images can be altered in terms of adding or removing an object or a group of them, videos are easily modified by cutting out a number of frames from the original content. This can be performed only for fun, to mislead the public or to cause harm to an individual or a group. Digital content altered only for the purpose of entertainment usually can be identified as unreliable even with the bare eyes and ears, because it commonly contains awkwardly replaced parts of an image or changed audio parts where voice is non-synchronized with lip movements. On the other hand, misleading the public with altered digital content is sometimes performed by journalists, especially yellow press, to produce sensationalist news. One form of delusion spread by media are also retouched photographs of models and celebrities, where their look is brought to perfection. While the last mentioned example of alteration causes no harm to the individuals shown in the images, hiding some objects from the image or introducing the nonexistent ones can be very harmful in case the image represents an evidence on the court or is used for the purposes of defamation.

Various incidents have witnessed in the recent past that a personal judgment can be based only upon the seen multimedia files. Internet consumers are often warned to keep themselves protected and to pay attention to the information they share through the network. Despite following these security measures ensures the safety of user's account and allows only him to share the information he wants, images and videos can still be endangered. Since the shared content is commonly downloadable, it becomes accessible for malicious users who can edit it and re-post it afterwards. Having in mind the crucial importance of digital information security and credibility, strong measures have to be taken in order to prevent the information manipulation and to provide authentication of distributed multimedia for the efficient communication and information storage.

Forensics is the field which investigates cases of tampering and crime. Multimedia forensics is one of its branches, which has the important role in investigation of information security and which acts as a key technology of digital evidence authentication [7]. Its domain ranges from the investigation to the recovery of damage caused intentionally or unintentionally to the parent information [8,9]. It is one of the cornerstones to accumulate and fetch data regarding criminal activities [10], content manipulation and security breaches, as well as the sharing of tampered data. Moreover, it is important to note that multimedia forensics sometimes faces the problem of information manipulation in such high rates that it is very difficult to distinguish the original content from the tampered or fake one. Having the information as dynamic variable, it becomes very hard to investigate it in the minimum time possible.

Forensics does not operate and gain the results on their own. The investigation procedures follow certain codes and links to identify the initial information and separate it from the mixed one. Statistical analysis is often conducted, due to the valuable results it can give in the process of detection of data alternation. Authentication of the content involves tracing of specific links, logos, ambience lighting, or any sort of clue which was present in the original content. This process can also include several types of information or data preservation. Most common ones are digital watermarks, data information console, copyright registration and trademark registration. Previously listed techniques are used in the field of multimedia forensics as the basis for further analysis. Information stored around the base points is then accumulated, in order to diagnose manipulation of the images, videos or other forms of multimedia content.

This thesis focuses on source identification procedure used in multimedia forensics, taking into account the specifics of today's portable devices and popular social media platforms.

The thesis is organized as follows: Chapter 2 describes the prerequisites for source identification process. Multimedia forensics and its tasks are presented, as well as characteristics of HDR (*High Dynamic Range*) multimedia, which is of a special interest for the conducted research. In order to help understanding the processes behind the multimedia forensics algorithms performed on images and videos, the process of their formation is explained, as well as the impacts of camera movements on the obtained multimedia files. Chapter 3 provides literature review, focusing on the existing algorithms for forgery detection and source identification, as well as on the existing image and video datasets, which are of a special importance for any kind of multimedia forensics analysis. PRNU-based approach in source identification is described in Chapter 4, while Chapter 5 presents MOSES mobile application and its initial video dataset, as well as the PRNU-based source identification experiments conducted on the mentioned collection of videos. Chapter 6 presents a novel VISION dataset of images and videos, which are used for source pattern noise fingerprints comparison. Similarly, Chapter 7 presents PRNU-based source identification over a novel dataset of HDR images and analyzes the obtained results. Finally, Chapter 8 gives the conclusion and guidance for further research on the topics engaged in this thesis.

1.1 The objective

This thesis aims to investigate the results of multimedia source identification in the challenging conditions caused by rapid technology development and popularity of social media platforms. While technological progress brought a wide range of options for capturing images and recording videos, thus introducing difficulties in content originality and authenticity detection, social media platforms enabled rapid sharing of those multimedia files. Both of the processes leave their marks on the original content, making it possible to investigate if the forgeries or malversations occurred.

As the current state-of-the-art literature does not provide a large number of multimedia files acquired by modern portable devices, the first problem this thesis copes with is up-to-date, large enough dataset formation. Considering that modern devices introduced differences in the capturing and recording processes, as well as novel possibilities for producing visually more realistic and appealing multimedia, suitable dataset creation was necessary for further multimedia forensics investigations.

Not only complexity of multimedia files introduces difficulties for multimedia forensics, but also the files exchange, due to different compression levels, algorithms and number of compression times performed. This became a burning problem, since social media platforms reached a popularity they have today. Due to the very easy multimedia sharing, it became of a huge importance to check multimedia content's originality. As source identification is one of the possibilities to perform that, this thesis focuses on investigating the impacts of possible obstacles introduced by modern multimedia on the well-known source identification algorithms.

1.2 Contributions

During the research work for this thesis, three novel datasets were introduced for the purposes of carrying out multimedia forensics algorithms for source identification and forgery detection. Various experiments were conducted using the introduced datasets, thus providing valuable results of the well-known source identification algorithms executed on multimedia files acquired by modern smartphone devices. Differences between standard SDR (*Standard Dynamic Range*) multimedia and its more complex HDR (*High Dynamic Range*) counterpart were specially considered during the analysis, as well as the problems occurring on the multimedia files transferred through social media platforms.

The first contribution is MOSES mobile application. It was developed for the purposes of video recording and storing, hence producing up-to-date video dataset, including a large variety of contents acquired by a wide range of smartphone devices. Application offers choosing the capturing motion and scenario type before recording, and stores the information about the record and source device afterwards. The initial dataset was formed using the MOSES application, in order to test its usage, as well as to proceed the multimedia forensics algorithms and investigate their results. Dataset is then expanded by exchanging a number of original videos through social media platform (i.e. by uploading to and downloading from YouTube), which resulted in a total of 1,209 SDR videos. By including both original and exchanged files in the dataset, testing the influence of an introduced compression and other possible modifications was enabled. PRNU-based source identification was conducted on this dataset and obtained results have shown significant differences between original and exchanged videos. Therefore, the research conducted in this thesis can serve as a starting point for further investigation of impact of video exchange through social media platforms on the original file. Moreover, MOSES shows a potential of becoming one of the largest video datasets, due to its world-wide availability and the idea of easy expandability. Since it enables anyone with the installed mobile application to upload their video, not only the database can be expanded, but the information about devices can also be obtained. This can help in coping with the problem of unknown devices.

VISION is the second created dataset, which includes both SDR and HDR images and videos. The number of images contained is 34,427, while the number of videos is 1,914. Unlike the first introduced dataset, which contains only videos, VISION provides combination of both types of multimedia files in SDR and HDR formats, thus providing the ability to investigate source identification based on different types of multimedia. Furthermore, researches outside of the field of multimedia forensics can be conducted using VISION dataset. For example, differences between image and video creation using the same camera can be investigated for a large set of modern smartphone devices. The focus of this thesis was on investigation in terms of multimedia forensics, specifically PRNU-based source identification, differences in PRNU estimates between different types of multimedia files and impact of social media exchange on images and videos contained in VISION. Storing the same information about the multimedia files and their acquiring devices, VISION is MOSES-compatible dataset and can therefore be extended with videos and frames (images) obtained by MOSES in the future.

Finally, the third introduced dataset consists of total of 5,415 HDR and SDR images and thus represents one of the largest currently available datasets focusing on HDR images. This dataset was formed using variety models of modern smartphone devices. Using the introduced dataset, HDR analysis and multimedia forensics researches can be conducted taking into account device specifications such as resolution, operating system, camera movements, etc. All the previously mentioned parameters can affect the final results and there is a need for investigation of their influences. This thesis provides an analysis of PRNU-based source identification, and confirms that camera movements and device properties have a significant impact on identifying the acquiring device.

Chapter 2

Image and video source identification prerequisites

The aim of this chapter is to describe the processes that lay underneath the problem of image and video source identification. Tasks of digital forensics and its branches are described in the first part of the chapter. Characteristics of High Dynamic Range images are presented afterwards. Finally, principles of multimedia content creation using camera devices and impact of camera movements on the obtained multimedia are explained in the last two sections of the chapter.

2.1 Introduction

Forensic sciences can be divided by their domain of evidence, which is used in further analysis. Since we are living in an analog world, classical *analog forensics* explores physical evidences, while *digital forensics* traces digital ones [11]. Digital evidences appear to be abstract to the individuals outside the branches related to computer sciences, in contrast to physical evidences, which are usually intuitive. Underneath the visible and audible content, digital evidence is written, using binary system, in the form of bit sequences, which can contain a lot more information than it can be seen or heard. This sets a difficult task for digital forensic sciences, which analyze all the aspects of complex digital information.

In the past, due to less exposure and interference, storage and security of the information were comfortable and less hectic. Before digital revolution, analog evidences were the only ones that could be endangered. Numerous of them are still thought to be in their original shape and structure, while minority is considered as being tampered. Authentic and pure information maintenance is the reason humans have diagnosed the true essence of life, universe, religions, social ethics, living creatures, etc. Therefore, analog forensics has the important task to investigate the physical evidences in order to provide the trustworthy assessment of information authenticity. There are two main principles used in this field of forensics: *divisibility of matter* and *exchange principle* [12]. Divisibility of matter implies that all parts of the same object remain having the same characteristics as the object as a whole. Exchange principle refers to the fact that when an object is transferred between individuals, each of them can leave some mark on it, such as fingerprint, clothing fiber, etc. However, digital forensics is much more complex and it has been divided to a number of branches in order to cope with the burning issue of digital evidence investigation.

Digital revolution brought many advantages, but also provided possibility to easily perform harmful actions. Information and media manipulation has become the greatest threat for security and storage of the original information. Possible reason of this chaos is the ability of each individual to store, analyze and republish information very quickly and easily. Cheap and affordable devices for information recording and storage, which are widespread and easily accessible nowadays, have largely contributed to this occurrence. This has posed threats to information security and media credibility like never before. As a consequence, the process of proving the originality of any information has become very painstaking and the number of verification points have become minimized. Digital forensics deals with this situation and analyzes available digital evidences in order to prove their authenticity or alteration.

The main branches of digital forensic sciences are [12]:

- computer forensics,
- mobile device forensics,
- database forensics,
- network forensics,
- multimedia forensics.

Computer forensics is often employed for piracy detection, as well as in investigation of child pornography and in the process of tracing the source computer that contains controversial files which can lead to the person who committed the crime. This process usually includes isolating the suspect's computer or laptop, searching through the files, hidden content and web history, and making a copy of its hard disk to perform more complex actions which can reveal the contentious content and participation in criminal activities.

Similarly, mobile device forensics is a branch of digital forensic science, which deals with the problem of fraud detection by investigating the information obtained using mobile devices. Besides from the phone call logs, SMS (*Short Message Service*) messages, instant messaging logs, photographs, textual, audio and video files, mobile devices can offer GPS (*Global Positioning System*) tracks, which can be of a high importance in case of kidnapping or a mobile device theft.

Database forensics has a different approach to frauds detection in comparison to the previously two described digital forensics branches. It analyzes database properties, i.e. the data that gives more information about the database, or so-called *metadata*. Using this information, it can be detected when did some change, which is a possible fraud, occur.

Network attacks became very frequent since the Internet emerged. The task of network forensics is to analyze and detect frauds occurred in both local and external (Internet) networks. For those purposes, traffic capturing is performed and the information captured in a form of small units called *packets* is investigated afterwards. Network security is of a high importance for every individual, because its disruption can lead to files hijacking and identity theft. In case of companies and especially banks, endangered network security leads to huge financial losses.

Finally, multimedia forensics analyzes multimedia files, such as images, audios and videos, in order to check their authenticity. The Thesis focuses on this branch of digital forensics, which is thus explained in more detail in the following chapter.

2.2 Multimedia forensics

Multimedia forensics is a branch of digital forensic sciences which is employed when authenticity of multimedia file is questioned [13]. It gathers various data points upon images, audio and video files to correlate their existence and behavior. This approach benefits the probability of assessing tampering performed to the investigated information.

All the approaches used in multimedia forensics can be divided into two groups, based on the information they obtain, having a digital evidence. Those are *active* and *passive* approaches [14]. Active ones cope with the information added to a multimedia file, such as digital watermark, or digital signature. Watermarks are inserted by some camera devices on all the photographs and videos recorded by that device. Digital signature is, on the other hand, used in digital forms of textual documents. In contrast to active approaches, passive ones do not possess an active information, and they are based on the assumption that there is some kind of pattern included in all the multimedia files obtained by the same device [15]. These approaches are presented in more detail in the remainder of this section.

2.2.1 Active approaches

Inserting the additional information into original multimedia content is a helpful technique in source identification and content authentication. The added information is considered as *active* and multimedia forensics approach for the analysis conduction based on it is therefore called *active approach*. The most common form of active information in multimedia files are *digital watermarks* [16] and *digital signatures* [17, 18].

Digital watermark refers to a digital code induced in the file before its delivering. For example, digital cameras, whose manufacturers included watermarking procedure in the photograph or video creation, add a specific digital code to multimedia content before it gets to the final user [16]. On the other hand, digital signature can be added to a textual document which was previously created and available to the user [19]. By adding digital signature to a multimedia file, the file is secured for sharing. In the first given example, a special hardware is needed, while the second one requires post-processing, which explains the term *active approach*.

Digital signature is an external digital code, which is generated from the original content and usually encrypted to produce hash values [20]. During the process of its generation, user's private key is required for association of the original content with the signature. Once a digitally signed multimedia file is received by other user, he can verify if the content is changed by using sender's public key. This key in combination to the received content enables creating another hash code. If the two created hash codes are identical, multimedia content was not altered. On the other side, difference in only one bit of the two generated hash codes signalizes data alteration.

Digital watermarks and signatures are commonly added for copyright protection, but they can also serve as an element of fraud detection, content authentication or source identification. Watermarks are sometimes not visible to the eye, but they can be extracted by image or video post-processing [16]. Once the watermark is extracted, it can be compared to the original one that was added in the process of multimedia file creation, similarly as it was described in case of digital signatures. Approaches based on watermarking and digital signatures are of a high importance for contents shared through the Internet without owner's permission. Movies are often copied and shared among the Internet users, who download them without paying any money [21]. Copyrights of the owner are thus violated and he has right to sue the user for illegally handling his file. Digital watermarks and signatures can serve as an evidence on the court. This applies not only to the video files, but also to scientific papers, books [22], images [23], or any other protected file [24]. Considering previous statements, it becomes clear that watermarks resistance to any kind of manipulations is of a high importance. Manipulations do not only refer to the frauds and malicious actions, but also to compression algorithms used on social media platforms, as well as in other programs used on Internet, that include uploading and/or downloading options.

Although active multimedia forensic approach represent an elegant way of proving who is the content owner and did any malicious manipulation occur, the major drawback are high requirements. As it is mentioned earlier, either more complex hardware, or post-processing is needed in order to embed a watermark in the multimedia content.

It is already described in short in this section how digital signatures and watermarks can be extracted. However, in order to understand complexity of active multimedia forensics tasks, it is important to get to know the principles of content hiding (steganography [25]) and its revelation (steganalysis). Therefore, the following subsection describes how digital watermarks, signatures, and other hidden data can be extracted from the analyzed content.

Steganalysis - retrieving hidden files/data

Steganography is a technique of hiding information in a visual content and, as such, is the subject of analysis in the field of multimedia forensics. Steganographic content may become visible in different ways and the aim of steganalysis is to discover that hidden, imprinted content. A file may require a key, stegokey, or a password to retrieve the secret information, and it is available only to the intended recipients.

Techniques such as watermarking and digital signatures, as well as cover channels and secret communication channels, are used to retain files secrecy. Steganalysis enables forensic technicians to detect those kind of hidden data embedded in multimedia files [26–30]. For example, techniques such as EnCase and Ilook Investigator [31] can help in identification of hidden content in storage devices which contain suspicious empty space. On the other hand, hidden messages in high-resolution digital images can be detected using higher-order magnitude and phase image statistics [32]. As they are commonly employed in the field of multimedia forensics in general, *Support Vector Machines* (SVM) and Markov chains can also be used in steganalysis. While the empirical transition matrices of Markov chain can serve as image features, SVM can be utilized as a classifier in steganalysis procedure performed on thresholded prediction-error image [33]. This method has shown to be able to detect more than 85% of the hidden content. Prediction-error images are also used in combination with neural networks and wavelet decomposition [34], in order to achieve the same result. Similarly, steganalysis can be performed on digital video sequences using the same method [35], as well as performing inter-frame collusion technique, that exploits the temporal statistical visibility of a hidden message [36, 37].

However, more complex steganographic techniques can even prevent recognition of the existence of hidden files [38,39], putting a difficult task ahead of steganalysis and multimedia forensics itself. Recent studies have developed powerful steganographic algorithms resistant to the well-known staganalytic attacks, as well as the ones used on HDR images [40–42], which are of a special interest for this research. Having that in mind, it is very important to keep the steganalysis methods up-to-date in order to cope with the problem of altered data.

2.2.2 Passive approaches

In contrast to the active approaches in multimedia forensics, passive ones do not require specific hardware, nor post-processing in order to add digital signature to a multimedia file. Passive approaches are based on the assumption that original content contains an inherent pattern introduced in the very process of multimedia file formation. According to this assumption, all the originals acquired by the same device should contain the same pattern. Deviation from the pattern leads to the conclusion that the content has been changed.

Two main tasks of passive multimedia forensics are *source identification* and *tampering detection*. Besides from the mentioned, passive approaches are commonly used for discriminating between computer generated and real-world generated multimedia content.

Process of source identification is conducted relying on the assumption that all files obtained by the same device include a pattern specific for that device. That pattern consists of a noise introduced in the multimedia file formation process and is referred to as fingerprint in literature, because it uniquely identifies device, just like fingerprint uniquely identifies human beings. Source identification is of a special interest for this thesis, and its concepts, including fingerprint estimation, are described in detail in the following sections.

Forgery and tampering detection are very hard processes, considering that alterations are often invisible to the eye and can sometimes be hard to detect even by employing post-processing algorithms. In order to understand their complexity, wider description is provided in the following subsection. It is worth noting that techniques used in forgery detection and in source identification cannot be distinctively separated. Some of the algorithms developed for the purposes of forgery detection can successfully identify the source device, and vice versa.

Forgery detection

Forgery detection enables confirmation of multimedia content authenticity [43]. It largely uses techniques that can detect inconsistencies in acquisition and coding fingerprints, or a total absence of acquisition and coding fingerprints. The latter is a sure way of confirming that the content of interest had undergone tampering.

Several techniques can be employed during the forgery detection. *Meta tag* data can reveal a plethora of information like source device, editing software, time of capturing or recording, time of editing (if any), and *geo tags* can identify the exact location where image or video was recorded. However, both *meta tag* and *geo tag* data can be tampered, and hence may provide false leads to the investigators.

Image and video processing can contain a large number of actions which result in changing an image or a frame, or obtaining and analyzing the information it contains. As videos are composed of frames, which are nothing but images themselves, all the processes that can be conducted on an image can be performed on a video, as well. Therefore, when any sort of image processing is mentioned in the remainder of this paper, it is important to notice that it also applies to the video frames.

Currently operating tools for image forgery detection can be classified into five major classes [44]:

- pixel-based techniques,
- format-based techniques,
- camera-based techniques,
- physically-based techniques,
- geometric-based techniques.

All the previously mentioned techniques are applied in the specific circumstances and they have a main contribution in the forensics analysis of the information.

Digital images and video frames are represented as a set of points, called pixels, with corresponding values that describe the color of that pixel. Therefore, pixels are considered as elementary units for these multimedia files. Pixel-based techniques investigate statistical behavior developed at that elementary level of images or videos.

Malicious image pixel-level editing is often performed by using *cloning* tools which enable extraction of one part of an image and cloning it to some other location, in order to hide the original content. By using statistical analysis and finding correlation between different picture elements, multimedia forensics can cope with these kind of frauds, but it is not always easy to detect them. An example of *cloning* forgery is shown in Fig. 2.1.

The other common pixel-level editing method is inserting fragments to an image from the same or some other source, or combining two or more images, which is usually called *splicing*. Splicing often requires resizing, rotation, or stretching a part of an image, in order to produce realistic composite image. This process implies that the originals have to be *resampled*, introducing specific periodic correlation that is unlikely to occur naturally [44], which helps in detection of these kind of frauds. Techniques such as higher-order Fourier statistics and artificial intelligence can be employed for coping with this problem. While detection of disruption of higher-order Fourier statistics implies that splicing has occurred, techniques employing artificial intelligence

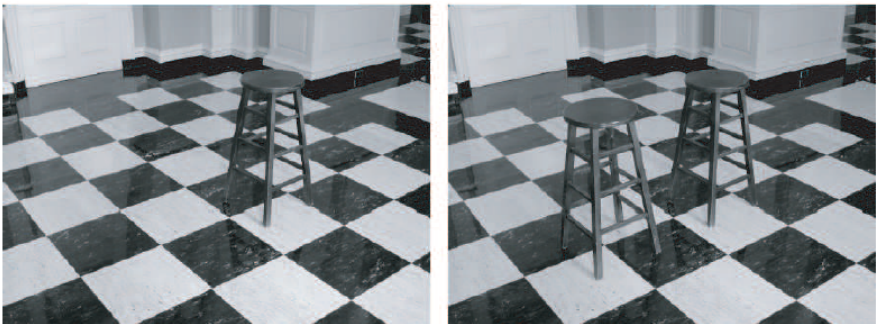


Figure 2.1: Example of cloning. Forged image (left) is created by hiding parts of the original content (right) [1].

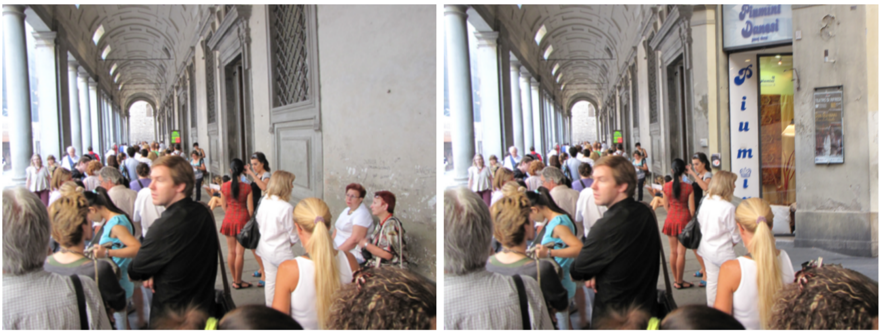


Figure 2.2: Example of splicing. Original image is shown on the left and spliced image on the right [2].

enable machines to learn how visual data may appear or change in the near future, and hence have a predictive element inbuilt, which is used for fraud detection. An example of splicing is given in Fig. 2.2.

Format-based techniques for forgery detection are relying on the format of multimedia file. In case of images, the most commonly used format is JPEG, while MPEG format is used for storing video files by most of the camera devices. Considering that both JPEG and MPEG formats use lossy compression, manufacturers typically configure their devices differently in order to balance compression and quality of the resulting files [44]. This fact can help not only in forgery detection, but can also serve for source identification purposes [45]. Furthermore, considering JPEG and MPEG popularity, there is a high probability that both original and forged multimedia files will be saved in the same format. Therefore, forged files will be double compressed, which in case of JPEG and MPEG formats means they will irretrievably lose on their quality twice, which is the fact multimedia forensics uses in the forgery investigation.

Camera-based tools allow highlighting camera module's characteristics, artificial artifacts in the parent information, as well as contribution of specific camera lenses and sensors. In order to understand these techniques, it is important to know how cameras work. Detailed explanation is therefore given in Section 2.3. All hardware parts included in the process, such as *color filter arrays* and *sensors*, can leave their mark on a produced multimedia file at some stage of image or video processing, before the result gets to the user capturing an image or recording a video. Thanks to these marks, source identification can be performed.

Physically-based procedures allow uplifting of physical characteristic of an image or video by interlinking physical parameters such as light, lenses and camera unit. As it is hard to balance the light from multiple different images, these procedures are especially focused on investigation of lightning characteristics in potentially forged image or video.

Finally, geometric-based methods calculate geometric perspective of the parent information related to the positions and locations relative to the information recording device. Projection of the camera center onto the image plane is called *principal point* [45] and it is the most interesting subject of analysis in geometric-based forgery detection. It is shown that translation of an object in the image causes a proportional principal point movement [46]. Comparing the estimated position of principal point to the calculated one, tampering can be detected.

Previously mentioned editing processes can be very complicated and performed at a high level, using less known and unexplored techniques, which require higher level of investigation in order to determine the content's originality. Latest developments in technology have brought not only visual inputs, but also thermal and other sensory data into the gamut of what computer vision can analyze [47]. However, multimedia forensic science continues to develop and copes with the newly introduced problems.

2.3 Digital image formation

In order to understand how it is possible to perform source identification based on some features extracted from an image, it is important to understand the way digital capturing devices work. Although devices themselves can have different purposes and different implementations, some processes can be more or less generalized in case of image capturing, regardless of the manufacturer and device type (digital camera, mobile phone, tablet, or any other device with a capturing option).

Block diagram of a typical digital camera is given in Fig. 2.3. All the included components can be divided into three groups: *optical and mechanical subsystem*, an *image sensor* and an *electronic subsystem* [3]. The process starts when the light passes through the camera lenses. It travels further through shutter and diaphragm, anti-aliasing filters and color filter arrays, before reaching the most important component for digital image creation - *imaging sensor* or *image sensor*. Shutter and diaphragm are in charge for making the exposure by briefly uncovering the camera aperture. While anti-aliasing filters are optical low-pass filters used in order to prevent frequency components overlapping, color filter arrays filter out some spectrum ranges to provide that each pixel detects only one color. That way, the photons are being prepared for the imaging sensor, which is sensitive only to monochromatic light. Imaging sensor then collects filtered photons and converts them into voltages. The sensor's output is analog signal, which needs to be processed by analog pre-processor, which contains sample-and-hold circuits for sampling and quantization, and performs operations such as color separation, *Automatic Gain Control* (AGC), tone adjustment, etc. [3]. Processed signal is finally converted to its digital counterpart using *Analog-to-Digital* (A/D) converter. For the purpose of getting the image in color, signal is demosaiced or interpolated by *digital signal processors* (DSP) or microprocessors. These components can also scale the signal to achieve proper white balance [48].

Most of the capturing and recording devices include display, as well as memory card socket and connectors. These components can be connected to DSP through a data bus. Apart from the mentioned camera elements, block diagram shown in Fig. 2.3 contains a system controller, which is in charge of controlling the camera operations, such as auto-focus and automatic exposure.

Mathematical formation of the previously described procedure before de-

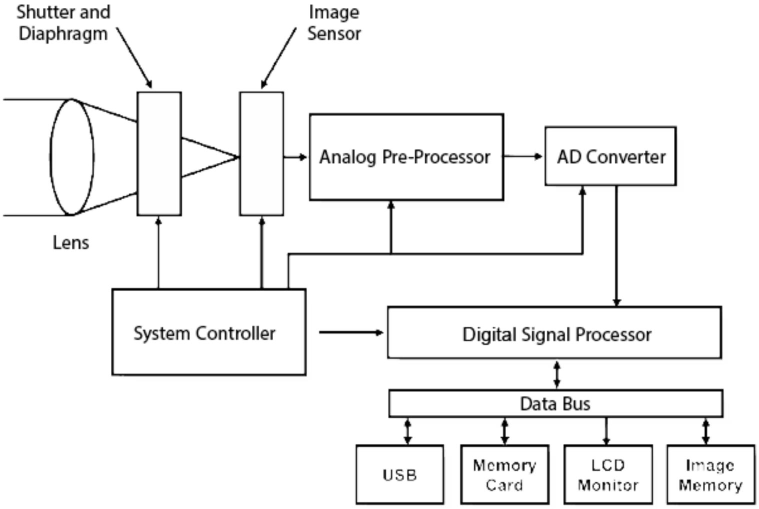


Figure 2.3: Block diagram of a typical digital camera [3].

mosaicking process can be described by the relation (2.1), which applies to each pixel of an image. The equation represents a simplified output model of the image sensor. Symbol I in the equation denotes the quantized luminance value at analyzed pixel, K is the PRNU *(Photo Response Non Uniformity)* factor, Y represents the incident light intensity, g is the channel color gain factor, γ stands for gamma-correction factor, Θ_q is quantization noise, while Λ includes combination of other noise sources. PRNU factor K is the most interesting element for the analysis conducted in this research and will be further explained in one of the following sections. At this place, it is only important to note that it is a noise-like signal responsible for the fingerprint [49], which enables source identification.

$$I = g^{\gamma} \times [(1 + K)Y + \Lambda]^{\gamma} + \Theta_q \tag{2.1}$$

By performing some basic mathematical operations, the sensor output model described by relation (2.1), can be simplified in order to calculate the factor K. This factor can be used in further analysis for source identification. Procedure of its calculation is described in Chapter 4.

2.4 HDR multimedia characteristics

Images are digitally represented as a collection of tiny dots, colored and arranged in a pattern of pixels, that a computer has to understand and envisage as a concrete and recognizable object within the backdrop of space. This is what a basic image recognition software does [50]. However, owing to the complexity of digital image and video reproduction, it is likely that an image (which can also be a video frame) itself captures only a smaller percentage of the actual data that exists from the object.

Images are likely to be degraded, as far as the presentation of color, detailing or texture is concerned. With advances in technology, it is possible to get HDR (*High Dynamic Range*) images, which give a closer representation of the natural object. Nowadays, a large number of images and videos are captured/recorded and processed using HDR technology. This leads to the problem of forensic detection of such images and tracing history of digital image.

Majority of today's multimedia devices enable HDR option when user is capturing a photograph or recording a video. The abbreviation HDR, which stands for *High Dynamic Range*, shortly describes its difference in comparison to standard profile for capturing and recording. HDR introduces wider range of luminance in multimedia content, providing more realistic captures. While standard capturing profile, better known as SDR (*Standard Dynamic Range*), does not allow big luminance adjustments and is therefore sensitive in cases of bad lighting conditions and facing the source of light, HDR profile copes with these problems and simulates the way human's visual system adjusts to these kind of lighting changes.

One of the examples of adjustment that HDR image introduces in case when camera device is facing the source of light is given in Fig. 2.4. In case of SDR images captured in the same conditions, results cannot reach the quality of HDR ones, even with brightness and contrast adjustment.

In digital world, images and videos are represented using three color channels: red, green and blue. Each of the channels normally employs eight bits for color representation, having $2^8 = 256$ possibilities for channel value. Combination of values of all the channels results in total of 1.6 million different colors that can be represented using SDR profile, which seems like an enormously big number. However, our visual system can perceive much larger number of colors, and HDR profile provides that in multimedia files by using floating point representation of values, instead of integers used in

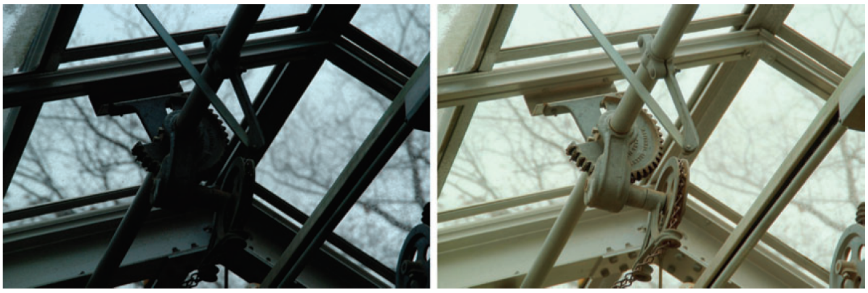


Figure 2.4: Example of SDR (left) and HDR (right) captures taken while camera was facing light source and capturing an object against it [4].

8-bit SDR channels. Each pixel in an image or a video frame is represented as 16-bit or 32-bit floating number in HDR representation.

In order to see the original HDR image, special devices are needed. Therefore, it is worth noting that only a small number of devices have the ability of showing original HDR content and that the printed HDR image always possesses reduced dynamic range [4]. This reduction is performed by specific algorithms and is often referred to as *tone-mapping* in literature.

Creation of HDR images can be performed by employing one of the following methods:

- rendering algorithms and other digital graphics techniques,
- employing conventional SDR cameras by capturing a static scene multiple times, with varying exposure time, and combining the captures afterwards [4].

Most of the capturing devices, especially mobile phones and tablets, create HDR images using the latter method. It is worth noting that one HDR image represents an HDR frame in a video, and therefore the previously described process applies to the videos, as well. The only difference is in higher requirements for the execution time of one HDR image creation when HDR videos are recorded, for the purposes of real-time recording and processing.

Due to the need of combining the captures in order to obtain one HDR image, it is important to avoid any camera movements between different shots. If a camera device is not still, the final result will contain visible parts of the images that were combined and displaced in relation to the other.

2.5 Impact of camera movements on the obtained multimedia files

While capturing images of a landscape at night, most people faced a problem of getting blurry images as a result. This problem is caused by shaking camera device in the moment of capturing and is known as *camera shake.* It occurs even in the daylight images, but is less noticeable to a human eye in case of good lighting conditions. Images captured at night or in case of bad indoor lighting are vulnerable to motion blur because of the necessity of longer exposure times [51]. Taking into account that HDR images are mostly produced as a combination of SDR images captured with variable exposure times, it is natural to assume that they are more vulnerable to the artifacts than their standard SDR counterparts. The occurred errors accumulate when combining SDR images, which makes the post-processing procedures for blur suppression more complex. Therefore, the information about camera movements is valuable in the process of source identification using HDR images.

Camera shake can be modeled as a blur kernel, describing the camera motion during exposure, convolved with the image intensities [52]. A large number of post-processing algorithms for blur reduction have been created, but in most cases, it is important for the user to capture the image without a need for post-processing. Camera shake can be prevented by using a tripod when employing conventional digital cameras, but as the light-weighted mobile devices with high camera resolutions are available at relatively small price nowadays, tripod is not a common equipment in case of images captured on daily basis. Moreover, tripod is not a guarantee for an artifact-free image. Even pressing the capturing button or exposure time change causes camera movements which can produce visible blurring effect [53].

Blur is a result of pixel offsets occurred in the process of image formation in camera device, destroying details in the capture. It is worth noting that offset can be produced not only in case of camera movements, but also in case of moving the object that is being captured. The latter often produces so-called *ghost effect*, because of the shades which form a ghost-alike object.

Previously mentioned side-effects of image capturing and video recording can seriously endanger the processes carried out in multimedia forensics. Therefore, it is important to examine their influence when the analysis in terms of source identification or forgery detection is conducted.

Chapter 3

Literature review

This chapter aims to discuss state-of-the-art on algorithms employed in multimedia forensics, specific tools and technologies used in source identification and forgery detection, as well as their applications and future trends that can be expected in both source identification and forgery detection processes. The last section of the chapter engages in the analysis of available datasets of images and videos that could be used in multimedia forensics.

3.1 Forgery detection algorithms

Multimedia forensic analysts recently started to study statistical properties of pixels, in order to improve currently available methods and algorithms used for forgery detection. One of the results of such researches is design of contrast enhancement detectors using pixel-graylevel histogram's peak-gap artifacts introduced in the process of forgery detection. Unfortunately, this approach did not yield accurate results. However, a recent research [54] has introduced new variants of the contrast enhancement operators that enable better detection.

In contrast to the previously mentioned pixel-based method, format-based techniques can use quantization tables, employed in JPEG compression, for detection of image tampering [55]. One of them is forgery detection software that uses nine Benford features extracted from quantized *Discrete Cosine Transform* (DCT) coefficients of original and morphed images, both JPEG compressed. Features are afterwards fitted to a logarithmic curve [56].

This enables tracing the changes that were imposed by morphing a fake image, and even a single parameter of the logarithmic curve is sufficient to find a difference between the original and morphed image. This software is expected to find extensive usage in security agencies, where facial recognition is based on photo IDs, and may be compromised if the used IDs include tampered images.

A number of scientific papers have focused on physically-based multimedia forensics approach, using light detection methods to identify images that have been used in tampering process. The example of this approach are methods that detect light sources within images and can predict how an image may appear when observed in different viewing environments. This technology enables isolating pixel and identifying a subset of pixels associated with the same light source and then configuring a pre-determined parameter to generate the color that a reproduced image should posses. This, in turn, enables identification of portions of image that have been faked or morphed [57].

Malicious alteration of images is mostly performed only in some regions (added or removed objects), which leaves a digital mark on an image, even if it cannot be noticed with a bare eye. Forgery localization is shown to be feasible using DCT coefficients [58–61], DWT (*Discrete Wavelet Transformation*) coefficients [62] and SVD (*Singular Value Decomposition*) [63], as well as image matching techniques, such as SIFT (*Scale-Invariant Feature Transform*) [64] and SURF (*Speeded Up Robust Features*) [65] descriptors. Numerous algorithms were developed for these purposes, and they keep up being improved by the researchers. However, all of them were created mainly for standard SDR images.

By generating a 3D model using some digital image and juxtaposing it on Google map, it is possible to verify if the image in question is the original and authentic one, or it is fake. Using the backdrop of landscape, and considering the time of the day, as well as weather conditions, it is possible to determine if an image was taken at the time claimed and by the source claimed. 3D modelling technology enables an accurate assessment of the genesis of an image and also helps in differentiating between the original and tampered version [66].

In another attempt to identify recaptured images and differentiate them from the original ones, researchers Yin & Fang [67] found that recaptured images posses changed statistics, which can be characterized using Markov process-based features. These features were extracted using DCT coefficient arrays. SVM (*Support Vector Machines*) training was then employed in order to identify differences between a dataset containing 3,994 recaptured images and to compare them against a similar number of originals.

A large number of algorithms employed in image forensics uses a single image in the analysis of possible frauds. However, a group of images can be analyzed [68] to explore their mutual dependencies which can provide a valuable information about the image history. This approach introduced more similarity to image and video forensics, considering the analysis is usually performed on a set of frames in case of video forensics. While image forgeries usually occur on a specific image region, video tampering is commonly performed on a frame level, either by removing the existing or introducing new frames.

Detection of tampering in video frames is even more complicated, as duplication is cumbersome and too time-consuming to justify its usage. Researchers Wang and Farid [69] invented an algorithm that made it time-efficient to detect duplicated frames, as well as duplicated regions within video frames. Detection of duplicated regions that is suggested by the mentioned authors was based on the work of Popescu and Farid [70]. Previous researches of the same group of authors resulted in development of techniques that depended on assessing MPEG (*Moving Picture Experts Group*) compression [71] and using interlaced and de-interlaced videos [69].

Besides from the above mentioned algorithms, there are numerous other known approaches for detection of frame insertion or deletion. Some of them are employing machine learning techniques for feature-based detection [72], computing the total motion residual of video frame [73], using the fact that inter-frame forgery will disturb the optical flow consistency [74] and detecting MCEA (*Motion-Compensated Edge Artifact*) [75].

Despite the large number of algorithms developed for image and video forensics purposes, they keep up being improved by the research community and new approaches are frequently presented in the literature. However, majority of them are created mainly for standard SDR images and videos, which leaves the space for further investigation, especially considering the powerful options and properties of today's smartphone devices and other portable devices.

3.2 Source identification algorithms

As it was stated in one of the previous sections, source identification and forgery detection principles can overlap, thus causing that both processes can be performed using a single algorithm, or at least relying on the same principle. Authors in [76] have shown that format-based approach used in forgery detection can also serve as source identifier. By relying on the fact that manufacturers usually develop their own algorithms for JPEG compression, it is demonstrated in [76] that choice of JPEG quantization table acts as an effective discriminator between model series, with a high level of differentiation. Understanding this procedure requires understanding forgery possibilities and principles followed in order to detect them. They are therefore described in Sections 2.2.2 and 3.1.

In the recent times, camera model identification based on captured image or recorded video became a standard procedure in multimedia forensics. However, techniques that can actually identify the exact camera device that made the capture/recording are still being explored. Most of the today's source identification techniques aim to identify the acquisition traces from multimedia files. State-of-the-art tools available for multimedia forensics analysis therefore focus on extracting the acquisition fingerprinting data and comparing it with some pre-developed dataset of fingerprints that have already traced genealogy to specific camera model or brand [77].

Researchers came to the conclusion that the source device, from which HDR multimedia file originated, can be accurately determined by isolating the fingerprint of the HDR-induced effects and running them through SVM classifier [78]. As it is described in Section 2.3, each camera device and model introduces its unique fingerprint through the lens, sensor and color filter array. More specifically, each lens is unique and has certain characteristics or aberrations, like the lateral chromatic aberration, which results in different wavelengths of light to focus on different sections of the image plane [79]. This information can be used for the purposes of tracing to a specific camera device.

Similarly, sensor related aberration, or noise, is unique to each camera device as it is a result of some imperfections in the image sensor. These imperfections create some differences between the scene and image captured by the camera [80], leaving a unique mark. Moreover, each camera sensor has a distinct radiometric response which is likely to be similar across the same brand [8].

Just like the camera sensor, color filter array contributes with its unique mark as it enables interpolation of the color scheme in an image [81]. As it was mentioned earlier, marks, better known as fingerprints, can be traceable to the exact camera device. Therefore, multimedia forensics algorithms based on noise are of a special interest for this research. The ones based on *Photo-Response Non-Uniformity* (PRNU) noise have shown the great success in source identification and are widely used in practice. While authors in [82–84] employed PRNU noise for source identification using images, video source identification was performed following the same principle in [85, 86]. Characteristics of PRNU, as well as the process of PRNU-based source identification are described in more detail in Chapter 4.

Source identification can be automatized using deep learning methodology, which is a very popular approach nowadays. In order to get reliable results, deep learning algorithms require a large set of information about the available devices. This fact implies that the list of known devices and their features has to be frequently updated, so that the device can be correctly identified. Otherwise, if source device is unknown, deep learning algorithm can only detect a wrong device, whose features have the highest correlation value with features of all the known devices. This problem is addressed in [87], where the authors described a process of its overcoming by identifying unknown camera models.

Learning features of source devices is conducted using convolutional neural networks, which is a complex computational model partially based on human neural system and its functioning [88]. Features that are used in learning process are the ones that are specific for a source device, mostly artifacts produced during the image or video acquisition. This means that deep learning methodology can be combined with PRNU-based source identification procedure, making it automatized. Authors in [88] have proved that dividing an image into several patches can be useful for PRNU detection and that deep learning methodology can result in highly reliable source identification in this case. However, a drawback of this method is its computational complexity and time needed for the algorithm execution.

Recent studies have shown to be able to detect and identify not only the source capturing device, but also from which embedded camera the image was captured [89]. Since both of the procedures can be performed with a high accuracy, deep learning methodology employed for multimedia forensics purposes is expected to be used even more in the future, improving currently existing source identification algorithms.

3.3 Data origin classification

Convolutional neural networks can be used not only for source identification purposes, but also for data origin classification. In today's world, where social media has a great impact on society, it is of a huge importance to be aware of the data origin and to distinguish if an image or a video was downloaded or acquired by some user device. This information can be of a crucial importance in court, when investigating digital evidences. Tracing images back to their social network of origin is analyzed in [90, 91], where the authors proposed methods based on convolutional neural networks to determine whether an image originates from a social network, a messaging application or directly from a photocamera. Features were extracted in the image frequency domain and then used in the training phase of the process, in order to identify the origin of the image among different social networks. It was shown that this method is able to identify the social platform of provenance.

Since PRNU-based techniques for source identification proved to be very robust and accurate, researches came to an idea to use PRNU fingerprint for origin social network detection [92], as well. It was demonstrated that PRNU is diversely modulated by different social networks and that it can therefore be adopted as a feature for training convolutional neural network and later detection of the social network of origin.

3.4 Image and video datasets overview

Several projects were undertaken, and several are in progress, to develop databases of fully annotated images [93], which can act as an evaluation point for forensic analysts. Most of these databases are available in the public domain and they find extensive usage in forensic analysis. Researchers Gloe and Böhme [94] have documented an image database consisting of over 14,000 images that were acquired using controlled situations which made them traceable to 73 different types of digital camera devices. Database was supplemented with additional information regarding specific noise pattern of each camera device and model-specific JPEG (*Joint Photographic Experts Group*) compression. This database, known as *Dresden Image Database*, can be used by researchers and analysts as a benchmark in identifying source camera devices.

HMDB (*The Human Motion DataBase*) [95] contains total of 6,766 video clips extracted from a wide range of sources. This database was introduced in 2011 for the purposes of action recognition and its robustness under various conditions, such as camera motion, viewpoint, video quality and occlusion. Taking into account that a large number of source devices were used, this database can provide some valuable information about their characteristics and serve for the purposes of source identification.

Unlike HMDB, SULFA (*Surrey University Library for Forensic Analysis*) is a video dataset created for the purposes of multimedia forensics investigations, specifically localization of cloned regions. SULFA contains 150 videos in low resolution (320×240) pixels, with the 10 seconds duration. The original videos included in the dataset are acquired using three different camcorders, while forged videos were created using Adobe software. However, technology has rapidly developed in the recent years and there is a need of updating the video and image databases, in order to include more complex, high-resolutioned multimedia content, provided by today's devices.

While many SDR image and video datasets are accessible on-line and for free, there is a small number of image datasets which contain HDR images and videos, due to the complexity of their formation.

One of the most commonly used HDR image dataset dates from 2007 and was created by Fairchild, under *HDR Photographic Survey* project [96]. The other known datasets mostly include several different types of images and/or videos, not focusing only on the HDR profile. The Fairchild's dataset consists of a total of 106 HDR images, but its shortage is lack of information about camera calibration, as well as the fact that it is not up-to-date anymore, since devices have changed rapidly in the past decade. Image properties have become more complex, starting from the resolution, over camera zooming and filtering options, to the number of bits used for color representation and the procedure of image creation.

DEIMOS (DatabasE of Images: Open Source) database [93] was formed more recently, in 2011, and it contains a large number of different types of images and videos. At the very beginning, this database contained about 70 HDR images, but it allowed the expansion of this set. In 2015, Korshunov et al. created a database of 20 HDR images for the purposes of testing different types of compression methods and performing subjective quality assessment of compressed HDR images [97]. Funt et al. created a novel HDR dataset containing images of 105 scenes [98], providing a larger number of available images. However, they were all captured using one single device model - Nikon D700 professional camera. As the images are mostly created by smartphone devices nowadays, this research did not focus on professional cameras identification, rather on more commonly used devices.

In 2015, database of 8,156 RAW images named RAISE (*RAw ImageS datasEt*) [99] was developed to aid in multimedia forensics detection of fake images. RAISE includes complete information on image sources and metadata, and allows a basic benchmark for analysts to match the images under observation and arrive at their source of origination. It is also found useful by researchers who aim to develop detection algorithms as it provides the basic dataset of images that they can be useful for comparisons.

DML-HDR video database [100] was introduced in 2014, due to the lack of representative HDR video dataset. DML-HDR consists of five HDR videos, all captured by professional camera, capable of capturing HDR videos [101]. Stuttgart HDR Video Database [102] contains a slightly bigger number of recordings, providing a total of 16 HDR videos showing different scenes. This dataset was formed for the purposes of evaluation of temporal tone mapping operators and HDR-displays.

Considering the complexity of HDR videos, it is understandable that currently available datasets include only a small number of them. However, for the purposes of carrying out the compatible research on video source identification, a larger number of available HDR videos is needed.

Despite the existence of a number of HDR datasets, none of the previously mentioned ones was designed for the purposes of testing the possibility of source identification, which requires a large number of images and videos, employment of bigger number of capturing devices and some specific capturing conditions, such as the good lighting, off-flash mode and existence of flat surfaces. This fact has been a motivation for creating the novel datasets described in the following chapters of this thesis.

Chapter 4

Multimedia forensics based on sensor noise

Studies have shown that one of the most successful approaches in source identification procedure is based on camera reference noise extraction. This chapter aims to describe how can the camera be identified using its own generated noise and to explain the further procedure for source identification based on noise extraction.

4.1 Photo-Response Non-Uniformity noise

Each digital camera device or any other capturing device produces so-called *pattern noise*. As the name suggests, it is a characteristic noise of each image capturing sensor, which remains approximately the same on all the photographs of the same scene captured using that sensor. Two types of the pattern noise can be differed: *Photo-Response Non-Uniformity* (PRNU) noise and *Fixed Pattern Noise* (FPN). The latter is also called *dark current* noise, because it appears when sensor is not exposed to the light. In contrast, PRNU is caused by sensor's reaction to the light and it is a dominant part of the sensor pattern noise. The major PRNU component is *Pixel Non-Uniformity* (PNU) noise, which appears due to different sensitivity of pixels to the light and it has much better resistance to image processing in comparison to fixed pattern noise collected from the sensor [103]. The other component contains all the low-frequency defects, caused by the usage of zooming option, light refraction, etc.

In order to investigate the characteristics of PNU, as the main part of PRNU noise, authors in [103] have conducted an experiment on a set of images of uniformly illuminated surface, captured by the same camera device. Low frequency components were first filtered and images were averaged afterwards, which has shown to reduce random noise and accumulate the sensor pattern noise. Furthermore, the experiment has proved that PNU noise is suppressed in very dark image areas, leaving FPN noise as a dominant part of the pattern noise. While PNU noise is not preeminent in case of dark areas, it cannot exist at all in saturated areas.

4.2 PRNU-based source identification

The first step in source identification using PRNU noise is estimation of PRNU factor K. In order to make a good estimation, a large number of images captured by the same digital capturing device is needed. The reason for this requirement is better random noise suppression, which can increase reliability of source identification conducted using PRNU method. In case of video analysis, it is easier to obtain the required number of images, as each video frame represents an image itself. However, if video recordings are not available, a large image database of N images is needed, where N should satisfy condition $N > 50$ [103]. Although improved PRNU estimators [49] require a smaller number of images, empirical results available in literature show that reliability is higher if the larger number N is employed. As it is shown that the image averaging results in accumulated sensor pattern noise, the idea is to use a large number of images and to compute their average in order to get PRNU.

The best results can be obtained if the images are smooth and do not contain many details. Flat surfaces, such as clear sky or uniformly illuminated flat objects, are the most flattering image contents when it comes to PRNU factor estimation.

As it is stated in Chapter 2, the adopted model for image camera acquisition is represented by the equation (4.1).

$$I = g^{\gamma} \times [(1+K)Y + \Lambda]^{\gamma} + \Theta_q \tag{4.1}$$

The procedure starts with improving *Signal-to-Noise Ratio* (SNR) in each employed image from the set of N images, by employing host signal rejection. This way, the difference between noisy and noiseless parts is enhanced.

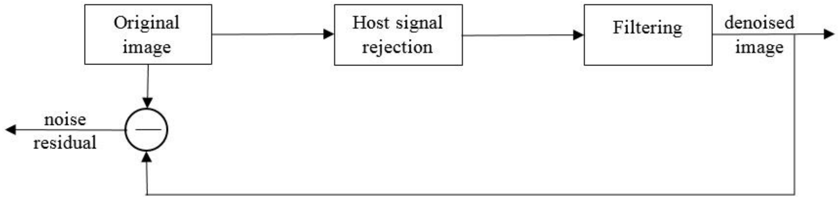


Figure 4.1: Scheme of the noise residual extraction.

After that, filtering process can be performed using wavelet-based denoising filter [104, 105]. This process results in a denoised image, which is further used to extract a noise component from the original image. The extraction can be performed by subtracting the previously computed denoised image from the original one, as it is shown in Fig. 4.1. The signal left after the previously described procedure is *noise residual* W. As it contains enhanced information about the sensor pattern noise, W is averaged instead of the raw images.

Partial derivation of the log-likelihood $L(K)$ of ratio $\frac{W}{I}$ solved for K is computed in order to obtain maximum likelihood estimate $\hat{K}$, as it is described by the relation (4.2). Factor σ^2 in the relation denotes variance of *White Gaussian Noise* (WGN). Although the real systems contain much more complex forms of noise, WGN can be accepted as a simplified model of the noise term, without a significant impact on the results.

$$\frac{\delta L(K)}{\delta K} = \sum_{k=1}^{N} \frac{W_k/I_k - K}{\sigma^2/(I_k)^2} = 0 \implies \hat{K} = \frac{\sum_{k=1}^{N} W_k I_k}{\sum_{k=1}^{N} (I_k)^2} \tag{4.2}$$

Estimate $\hat{K}$ contains a valuable information about the PRNU, but it also includes some artifacts that are common to multiple cameras, due to the implementation of image formation process and the sensor design itself. Having the same characteristics included in the maximum likelihood estimate of more than one device results in high possibility of false source identification. Therefore, it is advisable to reduce the unwanted similarities as much as possible. Suppression of artifacts effect on $\hat{K}$ can be performed by manipulation of pixel values with aim of producing PRNU factor with zero mean in each row and column of pixels [49]. This method is shown to be able to reduce color interpolation artifacts, as well as the artifacts produced by row-wise and column-wise operations of sensors and processing circuits. The result is significant reduction of correlation between PRNU factors of different devices.

In case that the zero mean PRNU factor contains visually identifiable patterns, it is suggested to translate the processed signal into Fourier domain and perform Wiener filtering to filter out all the components except from the noise [49].

Once the PRNU factor is estimated and processed in order to suppress all the unnecessary information, source identification procedure can begin. Computed PRNU is a unique stochastic fingerprint of imaging sensor and it serves as a basis for further procedure. Similar procedure that has been conducted on N images for calculating PRNU has to be conducted on an image that needs to be classified as the result of capturing by specific capturing device. The image is first processed to extract the noise which is going to be correlated with the computed fingerprint.

The problem of image source identification is formulated as a hypothesis testing with the aim of PRNU detection in the noise residual. As shown in Fig. 4.2, the zeroth hypothesis H_0 is that the noise residual contains only random noise without any other components, while the first hypothesis H_1 is that there are more components related to the estimate of the same capturing device, except from the random noise. In other words, if hypothesis H_0 is true, the image that is analyzed over a fingerprint of some capturing device was not obtained by that device. On the other hand, if hypothesis H_1 is true, analyzed image has the same or similar characteristics as the images which produced PRNU fingerprint of capturing device, and device is therefore identified to be the source of the analyzed image.

Mathematical representations of the hypotheses can differ, depending on the noise model that is taken into consideration. As white Gaussian noise is the simplest noise type, which does not appear in the real systems, it is better to operate with more complex ones, such as colored Gaussian noise η. Furthermore, it is important to keep in mind that PRNU factor estimation $\hat{K}$ may be attenuated due to the previously described PRNU processing procedure. Therefore, if the sensor output model defined by equation (4.1) is modified in accordance to the previous statements and defined by the relation (4.3), hypotheses can be formulated as it is presented in relation 4.4. While T represents pixel-wise multiplicative attenuation factor, $X = I\hat{K}$ is

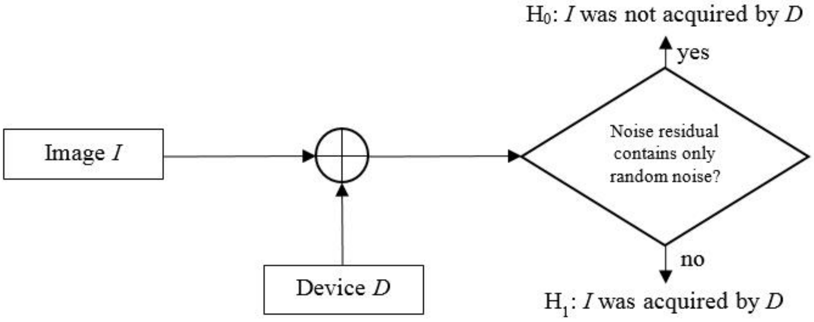

H0: I was not acquired by D
yes
Image I
Noise residual contains only random noise?
Device D
no
H1: I was acquired by D

Figure 4.2: Scheme of the image source identification problem formulation: hypothesis testing with the aim of PRNU detection in the noise residual.

the non-attenuated PRNU factor value [49].

$$W = TX + \eta \tag{4.3}$$

$$H_0 : W = \eta, H_1 : W = TX + \eta \tag{4.4}$$

At this stage, attenuation factor T and unequal variances σ_c^2 of Gaussian variables that form colored Gaussian noise are the unknown variables. Their estimation is not easy, as each pixel has its own values of the previously mentioned factors. As it would be computationally and time exhausting to perform the estimation procedure at each pixel of an image, it is advisable to divide image into a number of blocks and perform the computations for each of them. This simplification implies that all the pixels from the same block have the same values of T and σ_c^2.

Normalized *Generalized Matched Filter* (GMF) is the optimal detector for the problem set up by previously formed hypotheses [49]. It is defined by relation (4.5), where M is the total number of image blocks. Normalized correlation between non-attenuated PRNU factor and noise residual can be derived from this equation. Simplified form is given by the relation (4.6), where ρ_b denotes the normalized correlation, which is defined in (4.7). The other component β_b is defined by relation (4.8).

$$\rho = \frac{\sum_{b=1}^{M} \frac{\hat{T}_b}{\hat{\sigma}_b^2}(X_b \times W_b)}{\sqrt{\sum_{b=1}^{M} \frac{\hat{T}_b^2}{\hat{\sigma}_b^2}||X_b||^2}\sqrt{\sum_{b=1}^{M} \frac{1}{\hat{\sigma}_b^2}||W_b||^2}} \tag{4.5}$$

$$\rho = \sum_{b=1}^{M} \beta_b \rho_b \tag{4.6}$$

$$\rho_b = \frac{X_b \times W_b}{||X_b||||W_b||} = corr(X_b, W_b) \tag{4.7}$$

$$\beta_b = \frac{\frac{\hat{T}_b}{\hat{\sigma}_b^2}||X_b||||W_b||}{\sqrt{\sum_{i=1}^{M} \frac{\hat{T}_i^2}{\hat{\sigma}_i^2}||X_i||^2}\sqrt{\sum_{i=1}^{M} \frac{1}{\hat{\sigma}_i^2}||W_i||^2}} \tag{4.8}$$

The problem of estimating values for attenuation factor and variances of Gaussian variables from the optimal detector requires a known value of normalized correlation ρ_b. As it is available only under the hypothesis H_1, predictor of values ρ_b can be constructed based on the known PRNU factor estimate and features from the image block of interest, under this hypothesis. Finally, Neyman-Pearson approach can be employed for deciding if the analyzed image was captured by the device whose fingerprint is used in the described process, or not.

Due to the dependence of correlation factor on the image size, it is not suitable parameter for further analysis of the results. Peak to Correlation Energy ratio (PCE) is a better comparison factor [106] and it can be defined by the relation (4.9), where s_{peak} denotes coordinates of the peak, m and n are the image dimensions, and M is a small neighborhood around the peak [106].

$$PCE = \frac{\rho(s_{peak}; X, Y)^2}{\frac{1}{mn-|M|}\sum_{s \notin M} \rho(s; X, Y)^2} \tag{4.9}$$

PCE considers a possible special shift s between the fingerprint and the noise extracted from the image due to a possible cropping or use of the image. Then a correlation is conducted for each shift, and if a correlation prove is found, corresponding shift is considered to give the correct output.

The above described procedure of PRNU-based source identification refers to images, but having in mind that video represents a sequence of images, it is easy to conclude that the same procedure applies to videos, as well. N images used in process of PRNU factor estimation are video frames in this case, and they can be extracted from the same video, without need to record multiple of them, as in case of images.

4.3 Advantages and vulnerabilities of PRNU-based source identification

Due to the great performances it showed, PRNU-based source identification became a popular approach in multimedia forensics. However, as all the other methodologies, approaches and algorithms, source identification based on PRNU extraction has some vulnerabilities, apart from all the advantages it provides. This section aims to provide an analysis of both positive and negative sides of using PRNU fingerprint for multimedia forensic purposes.

Advantages can be summarized in five major categories: stability, generality, universality, dimensionality and robustness [107]. Regardless of the physical conditions and time lapse, PRNU fingerprint remains stable, which represents its first advantage. Since it is contained in every image and every video file, no matter which source device is used for multimedia acquisition, PRNU-based source identification follows the generality principle. Having in mind that all types of sensors exhibit PRNU, it is also universal. Furthermore, dimensionality, or uniqueness, is achieved, due to the large number of information contained in each fingerprint. Since many features characterize device's fingerprint, it is unlikely that they will be similar for two different sensors. Finally, this approach is robust, because the fingerprint can survive a wide range of multimedia manipulations, such as filtering and lossy compression [107].

Disadvantages of PRNU-based methods are computation load and sensitiveness to modifications, or so-called de-synchronization attacks [108]. Having in mind that using PRNU fingerprint can lead to the exact source device identification, it is clear that one of the requirements for such result has to be familiarity of the identified source device. Since there is a huge amount of different camera devices, which increases as time lapses, database of information about the devices and their fingerprints gets larger and larger, requiring tremendous physical storage [108]. De-synchronization is the other vulnerability of this method, caused by geometric distortion attacks such as scaling and cropping, which spatially de-synchronize target PRNU with reference PRNU [109]. This problem is investigated and researches came to the conclusion that using scale- and rotation-invariant transforms [109] or using the *Generalized Likelihood Ratio Test* directly in the spatial domain and finding the maximum of the test statistics using brute force [107] can help in overcoming this issue.

Taking into account the benefits it offers, vulnerabilities of PRNU-based approach in source identification are actively explored and research community still develops novel, improved algorithms, which suppress the recognized issues.

Chapter 5

MOSES mobile application for video dataset collection

For the purposes of video dataset expandability, MOSES mobile application is presented in this chapter. After description of the initial dataset, impact of the social media exchange on the original video files, as well as the ability of PRNU-based source identification on the presented dataset are investigated.

5.1 Introduction

As it is addressed in Chapter 3, one of the major problems for the research community is the lack of convenient and up-to-date datasets which can be used for multimedia forensic purposes. Having recognized this problem, three novel datsets of images and videos were created as part of this thesis. First of them is mobile application named MOSES [5]. The aim of this application is to provide a video dataset that will contain videos from a large number of smartphone devices, recorded using different camera specifics and showing a large span of scenes, with the advantage of being up-to-date.

Initial dataset was made using the implemented application and it contains 1,209 videos captured with 35 different devices. This dataset is expandable, because MOSES provides users to capture and upload their own videos to the dataset stored on the Florence University server. Dataset is currently not publicly accessible, but it can gain a public access for the researching purposes.

Previously described, up-to-date and expandable, video dataset would establish an excellent test environment for multimedia forensics techniques, especially in the field of source identification. Unlike the other datasets, which provide a certain number of files taken under controlled conditions, MOSES can exceed a large number of video files captured by a various smartphone devices. Different capturing scenarios can also benefit the investigation processes, introducing a big variety of contents.

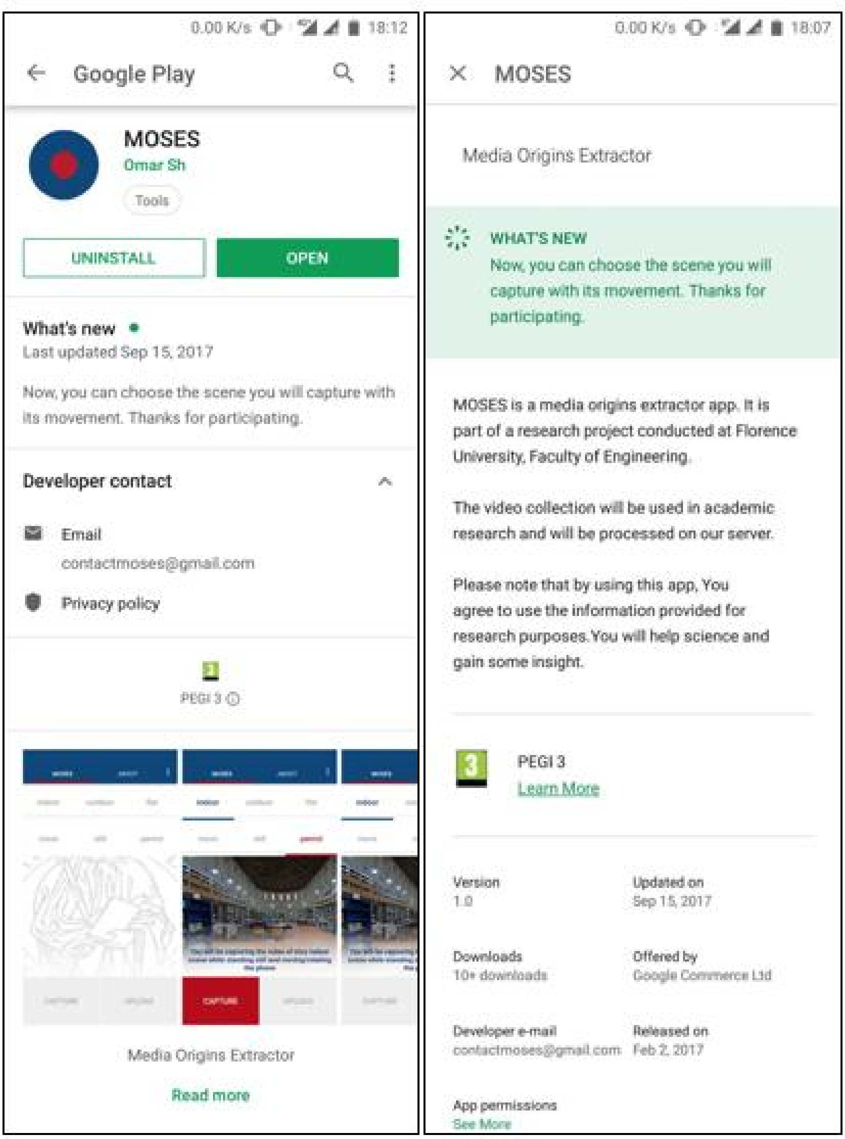
As video frames are images themselves, this application automatically provides a large database of images, which is of a special importance for PRNU estimation used for source identification purposes. Although MOSES does not initially contain HDR recordings, allowance of dataset expansion enables users to upload HDR videos, which can easily be transferred in the sequences of HDR images for the analysis purposes.

5.2 Guide for using the MOSES mobile application

Currently, iOS and Android versions of the MOSES application are available. Android application can be downloaded using the following URL:
https://play.google.com/store/apps/details?id=com.vmoses.metadata (Fig. 5.1), while iOS version can easily be found under iMoses name in the Italian Apple store. Readers are invited to download the application and follow standard installation procedure to contribute in the currently available dataset expansion. It is worth noting that the Android version is available worldwide, while its iOS counterpart can now be used only by the Italian users.

The Android graphical user interface for MOSES application is represented in Fig. 5.2, while Fig. 5.3 represents its iOS GUI.

After installation of MOSES mobile application and starting it, selection of one of the three scenario types: *indoor*, *outdoor* or *flat* is needed. After that, user selects one of the three camera motion types. The available choices are: *move*, *still* and *panrot*. *Move* refers to the case when a person is walking while recording a video. In the *still* movement scenario, video is recorded by a steady hand, in a still position, while *panrot* scenario refers to the case when a video is acquired while standing still, but combining pan-movements and rotation of the device. Acquisition starts by pressing the RECORD button, which calls the native camera application, that performs recording.



MOSES
Omar Sh
Tools

UNINSTALL
OPEN

What's new ●
Last updated Sep 15, 2017

Now, you can choose the scene you will capture with its movement. Thanks for participating.

Developer contact

Email
contactmoses@gmail.com

Privacy policy

PEGI 3

Media Origins Extractor

Read more

MOSES

Media Origins Extractor

WHAT'S NEW
Now, you can choose the scene you will capture with its movement. Thanks for participating.

MOSES is a media origins extractor app. It is part of a research project conducted at Florence University, Faculty of Engineering.

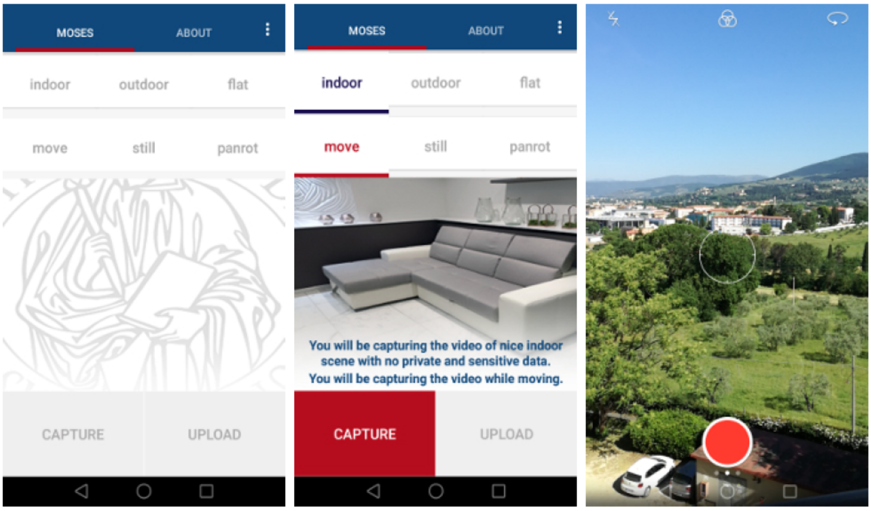
The video collection will be used in academic research and will be processed on our server.

Please note that by using this app, You agree to use the information provided for research purposes.You will help science and gain some insight.

PEGI 3
Learn More

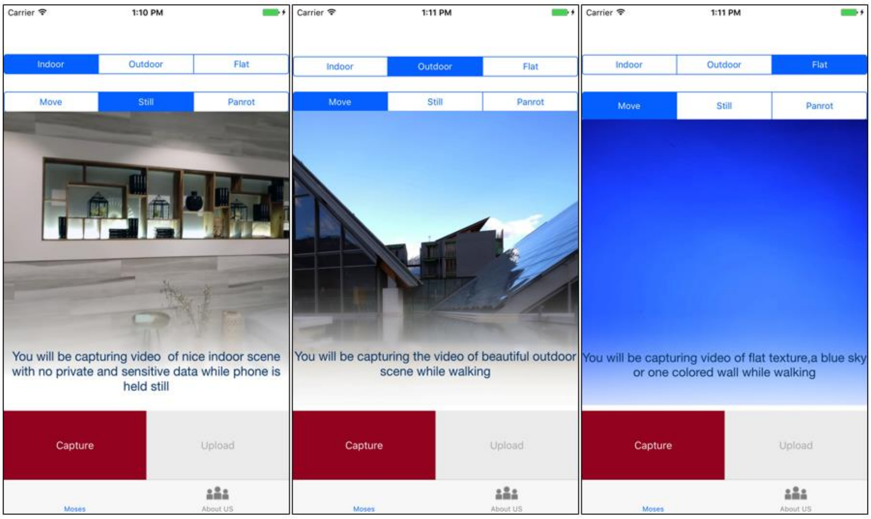
Version 1.0	Updated on Sep 15, 2017
Downloads 10+ downloads	Offered by Google Commerce Ltd
Developer e-mail contactmoses@gmail.com	Released on Feb 2, 2017
App permissions See More	

Figure 5.1: MOSES application in Google Play Store.



You will be capturing the video of nice indoor scene with no private and sensitive data. You will be capturing the video while moving.

Figure 5.2: The Android interface for MOSES application [5].



You will be capturing video of nice indoor scene with no private and sensitive data while phone is held still

You will be capturing the video of beautiful outdoor scene while walking

You will be capturing video of flat texture,a blue sky or one colored wall while walking

Figure 5.3: The iOS interface for MOSES application.

Considering the limited storage and bandwidth capacities, duration of videos that users can record using MOSES is set to 30 seconds. After that time, user can choose whether to upload a file to the existing dataset or cancel the procedure. Uploading is performed by pressing the UPLOAD button, which sends the content via *File Transfer Protocol* (FTP) to the servers, without any further processing.

5.3 Implementation details

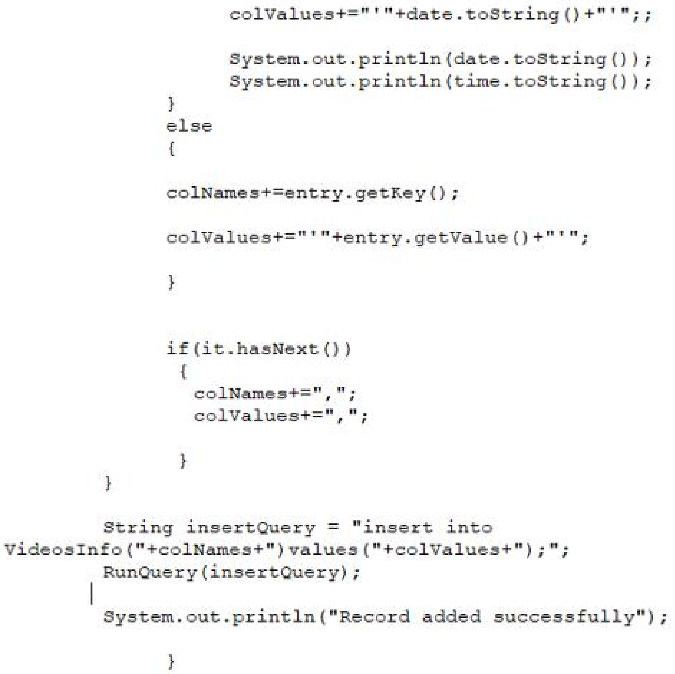
Besides from offering the ability of recording and storing videos in the dataset, MOSES collects the available information about camera device from which the video was recorded. This information is being stored in an XML file, which is obtained by analyzing the video metadata, and can later be used in process of source identification. Obtained information are as follows:

- manufacturer,
- operating system and its version,
- model of the device,
- frame rate in *fps*,
- resolution (video width and height),
- rotation of the display during the acquisition,
- acquisition timestamp (start of recording),
- creation timestamp (time of storing),
- information about video stabilization.

An example of XML metadata is given in Listing 5.1. Although it contains a large variety of information, it is not useful in a form where a quick analysis of the information cannot be performed. In order to enhance its usefulness, an SQLite database was created through Java script that stores the contents of XML files in the form of table, creating database of information that can be extracted using SQL queries. In order to perform this, XML file needs to be parsed to convert the information from the shown structure to a relational database, which provides the ability to analyze the dataset more efficiently. SQLite database consists of a single table that contains information of each XML in a single row.

Listing 5.1. An example of XML metadata from a video acquired with MOSES in the Android version [5].

```xml
<track>
...
<bitRate>5255207</bitRate>
<captureTime>20170525214034</captureTime>
...
<dateCreated>20170525T194047.000Z</dateCreated>
<device>
<brand>Lenovo</brand>
<device>P70-A</device>
<deviceID>865897020799766</deviceID>
<display>P70-A S138 151020 16G L ROW</display>
<manufacturer>LENOVO</manufacturer>
<model>Lenovo P70-A</model>
<os>
<name>LOLLIPOP MR1</name>
<release>5.1</release>
<sdk>22</sdk>
</os>
<product>P70-A</product>
<user>unknown</user>
</device>
<duration>5</duration>
<frameRate>0.0</frameRate>
...
<hasAudio>yes</hasAudio>
<hasVideo>yes</hasVideo>
<height>720</height>
<location>43.7937419,11.2304078</location>
<mimeType>video/mp4</mimeType>
...
<resolution>921600</resolution>
<rotation>90</rotation>
<scene>indoor</scene>
<movement>flat</movement>
<timeSubdivision>00:00:05</timeSubdivision>
<title>86589702079976620170525214034 indoor</title>
<videoStabilizationMode>0,1</videoStabilizationMode>
<width>1280</width>
<year>2017</year>
</track>
```



```
                        colValues+="'"+date.toString()+"'";;

                        System.out.println(date.toString());
                        System.out.println(time.toString());
                   }
                   else
                   {

                   colNames+=entry.getKey();

                   colValues+="'"+entry.getValue()+"'";

                   }


                   if(it.hasNext())
                    {
                      colNames+=",";
                      colValues+=",";

                    }
              }

              String insertQuery = "insert into
       VideosInfo("+colNames+")values("+colValues+");";
              RunQuery(insertQuery);
             |
              System.out.println("Record added successfully");

                   }
```

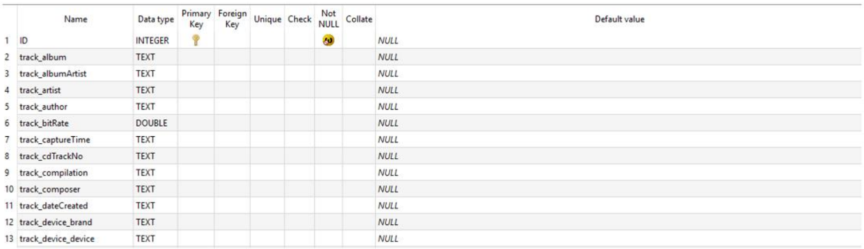
Figure 5.4: Snapshot of Java script for conversion of XML files to SQLite database.

Following steps are followed in order to extract data from XML files to SQLite database table:

- reading XML files,
- parsing XML files,
- inserting data in table.

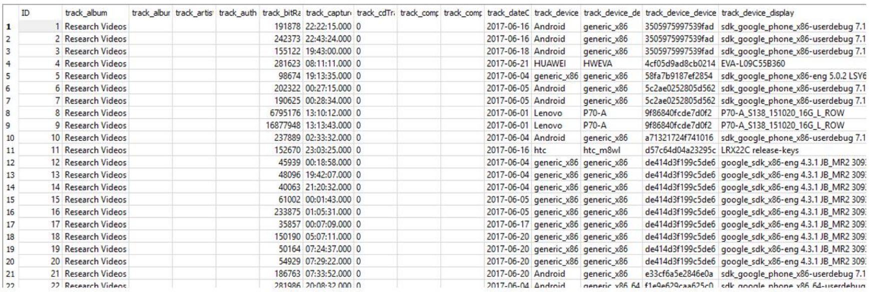
Java script has been written in order to perform the previously listed steps. The only input for the script is a path to a dataset of XML files. The script first reads all the files available in a given path recursively. Once the reading is done, the script parses a file according to the nodes of XML file. When the information from XML file is retrieved, script creates an SQL insert query based on the nodes data and creates a record, i.e. a row in the database table. A snapshot of Java script is given in Fig. 5.4, while the fields of SQLite table are shown in Fig. 5.5.

Once the script completes its execution, it produces a result in the form



	Name	Data type	Primary Key	Foreign Key	Unique	Check	Not NULL	Collate	Default value
1	ID	INTEGER	🔑				✓		NULL
2	track_album	TEXT							NULL
3	track_albumArtist	TEXT							NULL
4	track_artist	TEXT							NULL
5	track_author	TEXT							NULL
6	track_bitRate	DOUBLE							NULL
7	track_captureTime	TEXT							NULL
8	track_cdTrackNo	TEXT							NULL
9	track_compilation	TEXT							NULL
10	track_composer	TEXT							NULL
11	track_dateCreated	TEXT							NULL
12	track_device_brand	TEXT							NULL
13	track_device_device	TEXT							NULL

Figure 5.5: Partial representation of fields in created SQLite database table.



	ID	track_album	track_albur	track_artis	track_auth	track_bitRa	track_captur	track_cdTr	track_comp	track_comp	track_dateC	track_device	track_device_de	track_device_device	track_device_display
1	1	Research Videos				191878	22:22:15.000	0			2017-06-16	Android	generic_x86	3505975997539fad	sdk_google_phone_x86-userdebug 7.1
2	2	Research Videos				242373	22:43:24.000	0			2017-06-16	Android	generic_x86	3505975997539fad	sdk_google_phone_x86-userdebug 7.1
3	3	Research Videos				155122	19:43:00.000	0			2017-06-18	Android	generic_x86	3505975997539fad	sdk_google_phone_x86-userdebug 7.1
4	4	Research Videos				281623	08:11:11.000	0			2017-06-21	HUAWEI	HWEVA	4cf05d9ad8cb0214	EVA-L09C55B360
5	5	Research Videos				98674	19:13:35.000	0			2017-06-04	generic_x86	generic_x86	58fa7b9187ef2854	sdk_google_phone_x86-eng 5.0.2 LSY6
6	6	Research Videos				202322	00:27:15.000	0			2017-06-05	Android	generic_x86	5c2ae0252805d562	sdk_google_phone_x86-userdebug 7.1
7	7	Research Videos				190625	00:28:34.000	0			2017-06-05	Android	generic_x86	5c2ae0252805d562	sdk_google_phone_x86-userdebug 7.1
8	8	Research Videos				6795176	13:10:12.000	0			2017-06-01	Lenovo	P70-A	9f86840fcde7d0f2	P70-A_S138_151020_16G_L_ROW
9	9	Research Videos				16877948	13:13:43.000	0			2017-06-01	Lenovo	P70-A	9f86840fcde7d0f2	P70-A_S138_151020_16G_L_ROW
10	10	Research Videos				237889	02:33:32.000	0			2017-06-04	Android	generic_x86	a71321724f741016	sdk_google_phone_x86-userdebug 7.1
11	11	Research Videos				152670	23:03:25.000	0			2017-06-16	htc	htc_m8wl	d57c64d04a23295c	LRX22C release-keys
12	12	Research Videos				45939	00:18:58.000	0			2017-06-04	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
13	13	Research Videos				48096	19:42:07.000	0			2017-06-04	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
14	14	Research Videos				40063	21:20:32.000	0			2017-06-04	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
15	15	Research Videos				61002	00:01:43.000	0			2017-06-05	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
16	16	Research Videos				233875	01:05:31.000	0			2017-06-05	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
17	17	Research Videos				35857	00:07:09.000	0			2017-06-17	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
18	18	Research Videos				150190	05:07:11.000	0			2017-06-20	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
19	19	Research Videos				50164	07:24:37.000	0			2017-06-20	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
20	20	Research Videos				54929	07:29:22.000	0			2017-06-20	generic_x86	generic_x86	de414d3f199c5de6	google_sdk_x86-eng 4.3.1 JB_MR2 309
21	21	Research Videos				186763	07:33:52.000	0			2017-06-20	Android	generic_x86	e33cf6a5e2846e0a	sdk_google_phone_x86-userdebug 7.1
22	22	Research Videos				281986	20:08:32.000	0			2017-06-04	Android	generic_x86_64	f1e9e629caa625c0	sdk_google_phone_x86_64-userdebug

Figure 5.6: Example of XML files parsed to SQLite database table.

of the one given in Fig. 5.6. This form of a relational database allows user to easily run queries and get results. Furthermore, aggregated results are easily producible in a single query in this case, in contrast to a very time-consuming process in case of raw XML files.

5.4 Initial dataset formation

The initial dataset was created using MOSES application, which was first downloaded to devices from Google Play and Apple Store applications. All the recorded videos were then uploaded to the server through the installed smartphone applications.

Created dataset consists of 622 native videos and 587 videos exchanged through YouTube social platform, resulting in total of 1,209 videos. Video exchange had been conducted in order to provide ability for forgery detection testing using format-based techniques, as well as to provide conducting other multimedia forensics tests which rely on MPEG compression. Furthermore, exchanged videos can be used for analysis of fingerprints inserted by different social media platforms, which can be useful in provenance analysis of the shared data. It is worth noting that uploading video and downloading it at the maximum resolution available is meant by the term exchange.

Obtained videos include both indoor and outdoor scenes, as well as flat scenes. The latter scenery is provided in order to enable PRNU-based source identification, due to the needs of this method, described in Chapter 4.

Dataset was created using 35 devices from 11 different manufacturers. The number of devices per each manufacturer was as follows:

- 13 Apple devices,
- 8 Samsung devices,
- 5 Huawei devices,
- 2 One Plus devices,
- 1 Asus device,
- 1 Lenovo device,
- 1 LG electronics device,
- 1 Microsoft device,
- 1 Sony device,
- 1 Wiko device,
- 1 Xiaomi device.

Previously listed devices use Android and iOS operating systems. Employed versions of Android operating system span from 5.x to 7.x, while versions span from 7.x to 10.x for the used iOS-operating devices. Depending on the device model, camera resolutions are different, and the produced videos therefore have full HD, HD, or 480p resolution. All of them were obtained using rear-camera at the maximum resolution possible, with the exception of Asus device, for which the highest provided resolution was not employed. The dataset resulted in containing videos from 24 devices that provided full HD resolution, 9 that produced videos in HD resolution and 2 which provided 480p resolution. Summary of device and video characteristics are given in the Table 5.4, which also provides distinguishing different device models for those devices who share the same manufacturer. It is worth noting that the majority of recorded videos last longer than 60 seconds, with the exception of small number of videos obtained by devices D5 and D27, which are 25 seconds long. Videos with duration longer than 30 seconds were allowed in the initial dataset formation, but as it is stated earlier, limitation

Figure 5.7: Video frame samples from the initial dataset obtained using MOSES mobile application [5].

of video duration is set afterwards, due to the limited server's storage.

For the purposes of providing the ability of testing the impact of camera movements on the recorded videos, each scenario was captured in three different camera motions. As videos are mostly obtained without using tripod, this case was not included as a test scenario. The first set of recordings was made in the still camera motion, where only small movements due to still hand acquisition were present. Recording of the same scene was repeated in the walking motion, where the person was walking at the time of recording a video. The third set of videos is recorded while the person recording a video was standing still and simultaneously combining pan-movement and rotation of the device. The examples of frames extracted from the recorded videos are shown in Fig. 5.7.

Having in mind the power of social platforms in today's world, it is of a special interest for multimedia forensics to explore their influences on the original content. Therefore, in addition to the native contents described above, initial video dataset includes a subset of videos exchanged through YouTube platform. After creating a YouTube account, native videos were uploaded into playlists (one playlist has been created for each device) using the *Public flag*. Downloading process was carried out by executing *youtube-dl 6 command-line* free software. Related playlist was downloaded for each of the employed devices by selecting the best possible resolution. Using the above mentioned software, this can be performed by specifying the following parameter: *-f 137+140/bestvideo+bestaudio*. Exchanged videos were stored in the dataset afterwards.

Table 5.1: Main features of the devices employed in initial MOSES dataset and video files obtained by them [5].

Brand	Model	ID	Video resolution	#Videos
Apple	iPad 2	D13	1280 × 720	16
Apple	iPad mini	D20	1920 × 1080	16
Apple	iPhone 4	D09	1280 × 720	19
Apple	iPhone 4S	D02	1920 × 1080	13
Apple	iPhone 4S	D10	1920 × 1080	15
Apple	iPhone 5	D29	1920 × 1080	19
Apple	iPhone 5	D34	1920 × 1080	18
Apple	iPhone 5c	D05	1920 × 1080	19
Apple	iPhone 5c	D14	1920 × 1080	19
Apple	iPhone 5c	D18	1920 × 1080	13
Apple	iPhone 6	D06	1920 × 1080	15
Apple	iPhone 6	D15	1920 × 1080	18
Apple	iPhone 6 Plus	D19	1920 × 1080	19
Asus	Zenfone 2 Laser	D23*	640 × 480	19
Huawei	Ascend G6-U10	D33	1280 × 720	18
Huawei	Honor 5C NEM-L51	D30	1920 × 1080	19
Huawei	P8 GRA-L09	D28	1920 × 1080	19
Huawei	P9 EVA-L09	D03	1920 × 1080	19
Huawei	P9 Lite VNS-L31	D16	1920 × 1080	19
Lenovo	Lenovo P70-A	D07	1280 × 720	19
LG electronics	D290	D04	800 × 480	19
Microsoft	Lumia 640 LTE	D17	1920 × 1080	10
OnePlus	A3000	D25	1920 × 1080	19
OnePlus	A3003	D32	1920 × 1080	19
Samsung	Galaxy S III Mini GT-I8190	D26	1280 × 720	16
Samsung	Galaxy S III Mini GT-I8190N	D01	1280 × 720	16
Samsung	Galaxy S3 GT-I9300	D11	1920 × 1080	19
Samsung	Galaxy S4 Mini GT-I9195	D31	1920 × 1080	19
Samsung	Galaxy S5 SM-G900F	D27	1920 × 1080	19
Samsung	Galaxy Tab 3 GT-P5210	D08	1280 × 720	34
Samsung	Galaxy Tab A SM-T555	D35	1280 × 720	16
Samsung	Galaxy Trend Plus GT-S7580	D22	1280 × 720	16
Sony	Xperia Z1 Compact D5503	D12	1920 × 1080	19
Wiko	Ridge 4G	D21	1920 × 1080	19
Xiaomi	Redmi Note 3	D24	1920 × 1080	19

5.5 Experiments

The experiment was conducted using the obtained video files to perform source identification based on PRNU noise. For those purposes, the same procedure described in detail in Chapter 4 was conducted.

Camera fingerprint was first estimated from the first 100 frames of a referent *flat* video in *panrot* camera motion. Performing the same algorithm, fingerprint was afterwards estimated for the query video, as well, using its available frames. PCE factor was then calculated for the query video and compared to the threshold value of originating to the examined source device or not. All the available matching cases (videos from the same device) and the same number of mismatching cases (videos randomly chosen from other devices) were considered.

In order to investigate if the compression introduced by exchanging videos through YouTube platform influences the result of source identification, experiments were run on both original and exchanged video files.

5.6 Results

The achieved results are shown in the form of ROC curve in Fig. 5.8. This figure represents true positive (TP) and false alarm (FA) rates compared at varying thresholds. This kind of results representation allows quantification of the performance drop when YouTube compression was involved in the process.

It can be noticed that the results are not as good as expected, considering usually very good performances of PRNU-based methods in source identification, especially in case of original multimedia files. The possible reason for results degradation is the fact that several employed devices contain in-camera digital stabilization, which has a negative impact on fingerprints alignment during the process of its estimation. Therefore, the experiment was repeated using only devices without this feature. The obtained results for this case are shown in Fig. 5.9.

Experiment which excluded devices with in-camera digital stabilization has produced better results for both native and YouTube exchanged videos. This confirms the assumption that the previously mentioned feature introduces some difficulties in the PRNU-based source identification. Therefore, it is important to take this characteristic into account when performing tech-

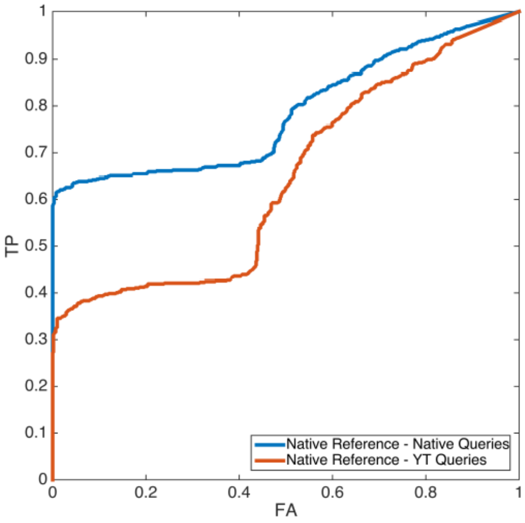


Figure 5.8: ROC curve of video source identification performances on native and YouTube exchanged videos [5].

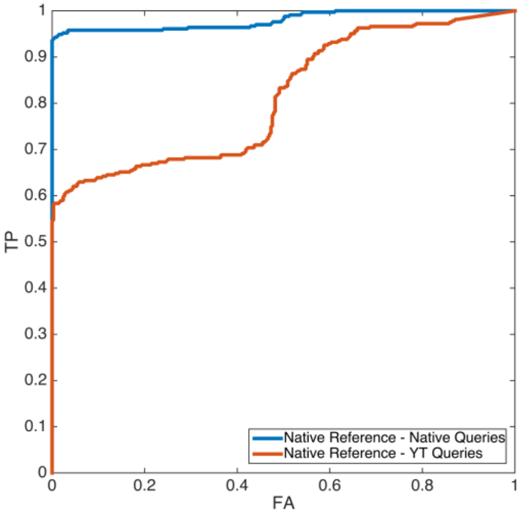


Figure 5.9: ROC curve of video source identification performances on native and YouTube exchanged videos with limitation of the analysis to non-stabilized videos [5].

niques for identification of video source device.

However, YouTube exchanged videos showed relatively low TP rate even in case when only devices without in-built digital stabilization were employed. While native videos were shown to have TP rate in the range from 0,94 to 1, TP rate of the exchanged videos dropped even to 0,58 in some cases. This result leads to the conclusion that exchanged videos are harder to trace to their original source device. Moreover, it confirms the assumption that social media platforms induce their own fingerprint to the exchanged content, which makes it different from the original. This fact opens up an interesting topic of types and characteristics of the marks created by different social media platform and their possible recognizability, which should be further investigated. MOSES mobile application provides research community a large span of videos which can be used for these purposes.

Chapter 6

VISION dataset

A novel dataset of images and videos is presented in this chapter. Having different types of multimedia files obtained by the same devices, VISION dataset provides investigation of differences between PRNU estimates obtained using image and video files. Besides from the estimation analysis, the chapter describes results of source identification based on PRNU estimates and impact of the social media exchange on the original files.

6.1 Introduction

Motivated by the results and derived conclusions after conducting the source identification analysis on initial MOSES video dataset, described in Chapter 5, a novel database including both images and videos exchanged through the social network platforms was created. The dataset is named VISION and its characteristics are described in more detail in this chapter.

Creation of a novel dataset of images and videos was performed due to the lack of an adequate dataset of this type, which can be used for source identification purposes. Although many image and video databases are available in state-of-the-art, as described in Section 3.4, to the best of our knowledge, none of them is up-to-date database with large variety of scenarios and multimedia types acquired by diverse modern smartphone devices.

One of the largest and newest datasets which includes both images and videos is IARPA Janus Benchmark-B Face Dataset, presented in [110]. However, as the dataset's name suggests, it is designed for the face analysis and therefore includes relatively similar scenery in all the included multimedia files. Since there is an assumption that most of the source identification algorithms, including PRNU-based method, depend on an image/video content at some extent, this kind of database is not adequate for source identification purposes, even though it includes a large number of multimedia files, having 21,798 still images and 7,011 videos.

DEIMOS (DatabasE of Images: Open Source) database [93] is expandable set of images and videos, briefly described in Section 3.4. Although this dataset includes variety of scenes, not many of the multimedia files included are acquired by modern smartphone devices, especially in HDR mode. Having in mind that HDR capturing and recording options gets more and more popular with the abilities provided by the newest capturing and recording devices, it is of a big importance for multimedia forensics to have an access to a large variety of such multimedia.

Furthermore, none of the known image and video datasets include both regular, spontaneously obtained multimedia files, taken in different conditions and using different scenarios and a large number of flattish, untextured surfaces, which could enable better PRNU estimation. Considering all the obstacles encountered analyzing state-of-the-art databases, a novel database named VISION was created. This database provides both image and video files in standard and HDR mode, as well as a large variety of captured and recorded scenes, including flattish, monotonous surfaces, convenient for PRNU-based methods.

Using VISION, multimedia forensics tests based on PRNU factor estimation for source identification were run in order to investigate the impact of social media platform exchange on the obtained images. The analysis also includes comparison of PRNU factors estimated from video frames and from images acquired by the same devices. PRNU factors estimated from different multimedia types are usually hard to match, which can represent a problem in source identification field. Therefore, this analysis can show dependability of PRNU factor on the multimedia files used for its estimation. Considering that both original (generic, native) images and videos, as well as their exchanged counterparts are included in the analysis, influence of compression procedures can also be investigated in this case.

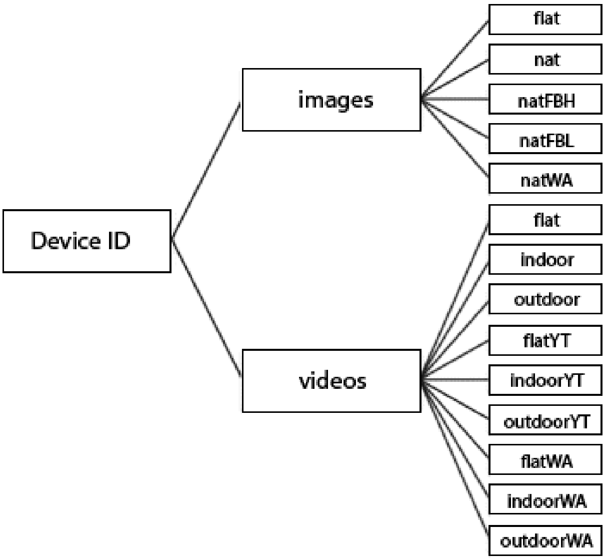


Figure 6.1: Structural organization of VISION dataset [6].

6.2 Dataset formation

VISION dataset employs all 35 devices used for creation of initial videos included in MOSES application. For the purposes of image capturing and video recording, the best-quality camera available was employed. While the structure of video part of the dataset remained the same as for the MOSES, including three types of recorded scenarios (*indoor*, *outdoor* and *flat*), image part of the dataset included two scenarios: *flat* and *nat*. *Flat* scenario implies captures of flattish, uniform surfaces, such as walls or skies, just as it was case for the same scenario in the video part of the dataset. On the other hand, *nat* scenario involves both *indoor* and *outdoor* scenes. The abbreviation *nat* refers to native, original images. Separation of the images based on the environment they were captured in (indoor or outdoor) was not performed in this case. Furthermore, three camera motions: *still*, *moving* and *panrot* available in MOSES application are available in videos part of the VISION dataset, making it compatible with MOSES-obtained videos.

Total of 11,732 native images were collected for the purposes of VISION dataset creation. 7,565 of them were shared through Facebook, in both high and low quality, as well as through WhatsApp. This resulted in a total of 34,427 images. It is worth noting that HDR images were obtained from devices which have had the ability of HDR capturing, while the remaining images were obtained in standard SDR mode provided by other employed devices. For the purposes of videos collection, 648 originals were recorded. 622 of them were also shared through YouTube at the maximum available resolution, while 644 originals were shared through WhatsApp, resulting in a total of 1,914 videos. It should be noted that images and videos shared through social media platforms were rescaled by them, leading to lower multimedia files resolutions than the ones shown in Table 6.1 for original multimedia.

The structure of VISION dataset is shown in Fig. 6.1. The obtained images and videos were first sorted by the device model, then after the obtained multimedia type (image or video), and finally, by categories implying the scenarios in which the files were obtained. The explanation of each category name is given below:

- *flat* in images category: images of flat scenes,
- *nat*: native images including both indoor and outdoor scenes,
- *natFBH*: native images exchanged through Facebook platform in high quality,
- *natFBL*: native images exchanged through Facebook platform in low quality,
- *natWA*: native images exchanged through WhatsApp platform,
- *flat* in videos category: videos of flat scenes,
- *indoor*: videos recorded in the indoor environment,
- *outdoor*: videos recorded in the outdoor environment,
- *flatYT*: videos of flat scenes exchanged through YouTube in high quality,
- *indoorYT*: videos recorded in the indoor environment exchanged through YouTube platform in high quality,
- *outdoorYT*: videos recorded in the outdoor environment exchanged through YouTube platform in high quality,
- *flatWA*: videos of flat scenes exchanged through WhatsApp,
- *indoorWA*: videos recorded in the indoor environment exchanged through WhatsApp,
- *outdoorWA*: videos recorded in the outdoor environment exchanged through WhatsApp.

Summarized features of the complete dataset are provided in Table 6.1.

As it can be seen from the table, duration of videos obtained in this dataset is much shorter in comparison to the duration of videos from initial MOSES dataset. Moreover, it should be noticed that this research takes into account some of the inherit camera device characteristics, such as the ability of automatic digital stabilization and HDR capturing. Special attention should be paid on differences between image and video resolutions, although both types of files were captured and recorded by the very same devices, using the same camera and its configuration.

The example of images included in the VISION dataset is shown in Fig. 6.2. Differences in image and video resolutions were the main cause of PRNU factor estimate nonconformity, which was investigated afterwards and is described in the following sections.

6.2.1 Multimedia files exchange through social media platforms

For the purposes of exchanging images through Facebook, two photo-albums were opened on this social platform. Both of them were used only for *nat* images, but one of the albums contained images uploaded in low quality (*natFBL*), while the other one contained their counterparts uploaded in high quality (*natFBH*). Uploading processes significantly differ for these two quality levels, due to different compression methods employed, which is explained in detail in [111].

Downloading images was performed in two different manners: single image and a whole album. This test has been conducted in order to explore if the internally set downloading processes differ for these two cases. It was shown that there is no difference between the downloaded contents.

The exchanged images were saved in the same format as the originals and they follow the analogous naming convention, thus providing that the native image and its exchanged counterpart can be immediately recognized. For example, the first native image captured by D01 device was named *D01_I_nat_0001.jpg*, while the same image exchanged through the Facebook platform with high quality upload was named *D01_I_natFBH_0001.jpg.* It is easily noticeable that the format of storing was *deviceID_multimediaType_sceneType_ordinalNumber.fileFormat.* Variable *deviceID* takes a value from the ID tab of the Table 6.1, while *multimediaType* can be set to "I" or "V", which stands for image and video, respectively. Argument *sceneType* can take any value from the last branch of the dataset structure hiererarchy rep-

Table 6.1: Main features of the devices employed in VISION dataset and multimedia files obtained by them. DS tab shows the presence or absence of digital stabilization on the acquired content, HDR indicates whether the device supports HDR recording/capturing or not, VR refers to the video resolution, while IR stands for image resolution [6].

Brand	Model	ID	DStab	HDR	VR	#Videos	IR	#Images	#Flat	#Nat
Apple	iPad 2	D13	Off	F	1280 × 720	16	960 × 720	330	159	171
Apple	iPad mini	D20	On	F	1920 × 1080	16	2592 × 1936	278	119	159
Apple	iPhone 4	D09	Off	T	1280 × 720	19	2592 × 1936	326	109	217
Apple	iPhone 4S	D02	On	T	1920 × 1080	13	3264 × 2448	307	103	204
Apple	iPhone 4S	D10	On	T	1920 × 1080	15	3264 × 2448	311	133	178
Apple	iPhone 5	D29	On	T	1920 × 1080	19	3264 × 2448	324	100	224
Apple	iPhone 5	D34	On	T	1920 × 1080	32	3264 × 2448	310	106	204
Apple	iPhone 5c	D05	On	T	1920 × 1080	19	3264 × 2448	463	113	350
Apple	iPhone 5c	D14	On	T	1920 × 1080	19	3264 × 2448	339	130	209
Apple	iPhone 5c	D18	On	T	1920 × 1080	13	3264 × 2448	305	101	204
Apple	iPhone 6	D06	On	T	1920 × 1080	17	3264 × 2448	281	149	132
Apple	iPhone 6	D15	On	T	1920 × 1080	18	3264 × 2448	337	110	227
Apple	iPhone 6 Plus	D19	On	T	1920 × 1080	19	3264 × 2448	428	169	259
Asus	Zenfone 2 Laser	D23*	On	F	640 × 480	19	3264 × 1836	327	117	210
Huawei	Ascend G6-U10	D33	Off	T	1280 × 720	19	2448 × 3264	239	84	155
Huawei	Honor 5C NEM-L51	D30	Off	T	1920 × 1080	19	4160 × 3120	351	80	271
Huawei	P8 GRA-L09	D28	Off	T	1920 × 1080	19	4160 × 2336	392	126	266
Huawei	P9 EVA-L09	D03	Off	F	1920 × 1080	19	3968 × 2976	355	118	237
Huawei	P9 Lite VNS-L31	D16	Off	T	1920 × 1080	19	4160 × 3120	350	115	235
Lenovo	Lenovo P70-A	D07	Off	F	1280 × 720	19	4784 × 2704	375	158	217
LG electronics	D290	D04	On	F	800 × 480	19	3264 × 2448	368	141	227
Microsoft	Lumia 640 LTE	D17	Off	T	1920 × 1080	10	3264 × 1840	285	97	188
OnePlus	A3000	D25	On	T	1920 × 1080	19	4640 × 3480	463	176	287
OnePlus	A3003	D32	On	T	1920 × 1080	19	4640 × 3480	386	150	236
Samsung	Galaxy S III Mini GT-I8190	D26	Off	F	1280 × 720	16	2560 × 1920	210	60	150
Samsung	Galaxy S III Mini GT-I8190N	D01	Off	F	1280 × 720	22	2560 × 1920	283	78	205
Samsung	Galaxy S3 GT-I9300	D11	Off	T	1920 × 1080	19	3264 × 2448	309	102	207
Samsung	Galaxy S4 Mini GT-I9195	D31	Off	T	1920 × 1080	19	3264 × 1836	328	112	216
Samsung	Galaxy S5 SM-G900F	D27	Off	T	1920 × 1080	19	5312 × 2988	354	100	254
Samsung	Galaxy Tab 3 GT-P5210	D08	Off	F	1280 × 720	37	2048 × 1536	229	61	168
Samsung	Galaxy Tab A SM-T555	D35	Off	F	1280 × 720	16	2592 × 1944	280	126	154
Samsung	Galaxy Trend Plus GT-S7580	D22	Off	F	1280 × 720	16	2560 × 1920	314	151	163
Sony	Xperia Z1 Compact D5503	D12	On	T	1920 × 1080	19	5248 × 3936	316	100	216
Wiko	Ridge 4G	D21	Off	T	1920 × 1080	11	3264 × 2448	393	140	253
Xiaomi	Redmi Note 3	D24	Off	T	1920 × 1080	19	4608 × 2592	486	174	312

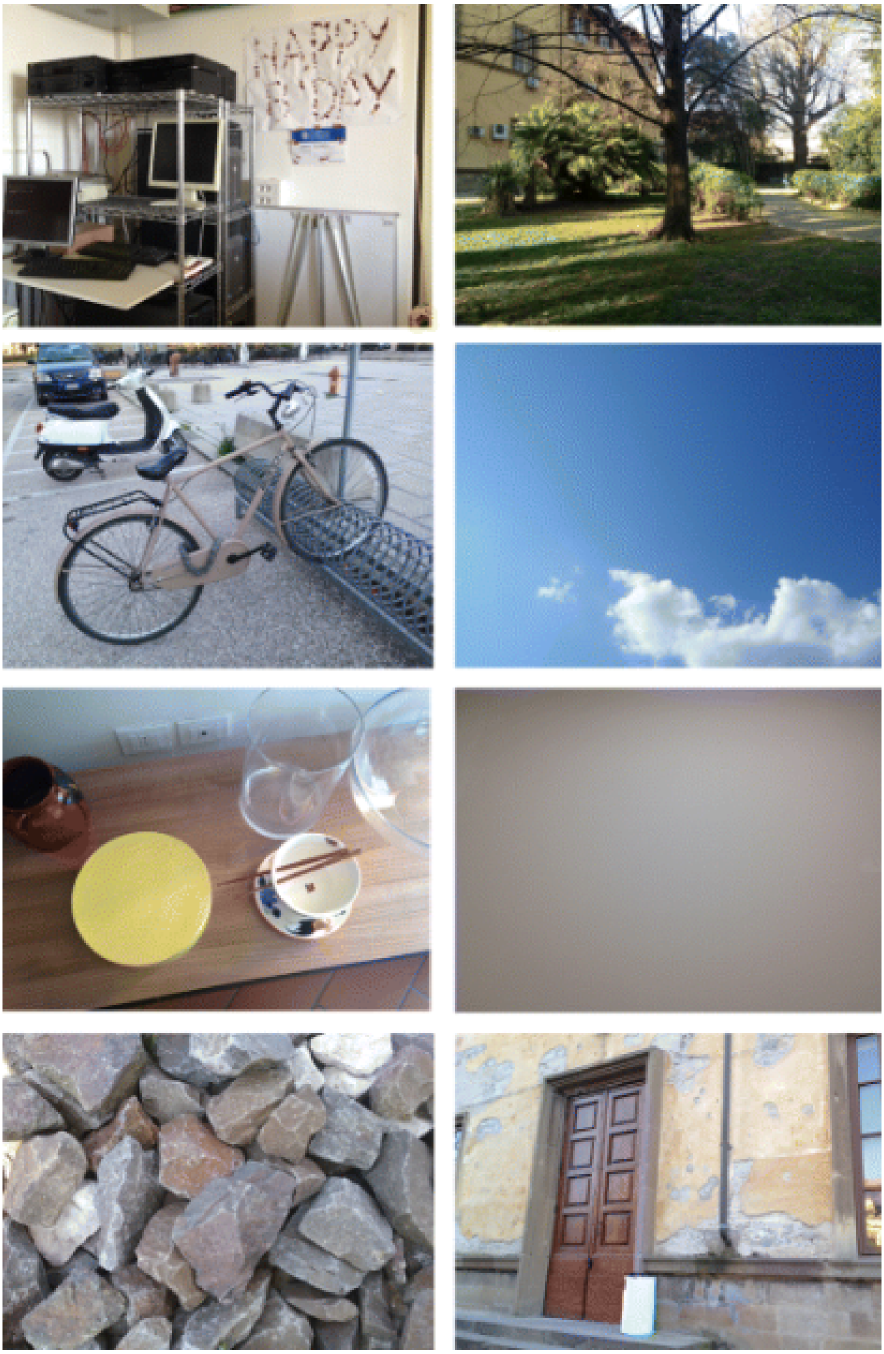


Figure 6.2: Samples of images from the VISION dataset [6].

resented in Fig. 6.1. Finally, *ordinalNumber* represents the ordinal number of multimedia file captured or recorded by the device with specified *deviceID*, while *fileFormat* represents a format in which the images or videos were stored by the concrete device. The most common value of *fileFormat* for images is JPEG, while MPEG is most frequently used format for videos.

It is important to note that the previously described naming convention includes one more parameter in case of videos, because they were obtained using three different camera motions and can be sorted thereafter. For example, if the previously mentioned device D01 had recorded a video of *indoor* scenario, using *panrot* camera motion, stored as a fifth file, its name would have been *D01_I_indoor_panrot_0005.mpeg.*

YouTube web platform was used for uploading and downloading original videos obtained by the employed camera devices. This process is referred to as exchanging, just as it was case for images. Exchanging was performed in the same manner as it was conducted for the purposes of creating initial video dataset for MOSES mobile application. *Public privacy* flag was used for uploading in the high resolution mode.

Besides from *youtube-dl* tool used in case of downloading the exchanged videos obtained by MOSES application, *ClipGrab*[1] tool was employed in order to investigate are there any differences between the resulting contents in case of downloading using two different tools. Previously mentioned tools produced the same downloading results.

Except from the Facebook and YouTube web platforms, WhatsApp mobile application was employed for exchanging both image and video files. All the multimedia files included in this process were exchanged via WhatsApp v2.17.41, using an iPhone7 A1778 device with iOS v10.3.1. The reason for choosing mobile application instead of a desktop one in this case is that the latter does not make compression computations during the exchange process, while the mobile one does. An interesting fact about WhatsApp exchanging processes is that they differ in regards to the device type. For example, iPhone devices obtain a less compressed multimedia file in the exchange process, in comparison to the Android devices. In case of images, this level of compression can be placed somewhere in the range between compression achieved by using low and high quality image uploading through the Facebook platform. Taking that into account, by transferring image files over WhatsApp in addition to the previously two exchanging methods

[1] ClipGrab v3.6.3, available on URL: www.clipgrab.org

which used Facebook as a social media platform, results from a large span of processing methods performed on the same set of images were obtained. Furthermore, as YouTube and WhatsApp do not use the same compression methods, diversity of processing methods was provided for the video files, as well.

6.3 Experiments and results

The introduced VISION dataset was used for multimedia forensics test evaluations in the similar manner as it was performed using the initial dataset created for MOSES mobile application. In other words, source identification based on PRNU fingerprint estimation was performed using obtained multimedia files. One of the differences between the previously conducted experiments and the ones whose execution was provided by VISION dataset is a larger number of differently processed images and videos, in comparison to the dataset used in the experiments described in the previous chapter. Moreover, VISION gives the ability of PRNU factor estimates comparison. As this dataset contains both images and videos acquired from the same source devices, it provides the opportunity to investigate the differences between the estimations produced using video frames and using image files.

Fingerprint computation for each source was conducted using the well-known PRNU method on *flat* multimedia files. Single image and video files were then processed in order to estimate their fingerprints and conduct source identification based on the obtained PCE values. For each device of interest, two fingerprints were computed: one based on 100 images obtained from that device and the other based on the first 100 frames recorded by the same. For each device, all available matching cases (images/videos from the same device) and the same number of mismatching cases (images/videos randomly chosen from other devices) were considered.

6.3.1 Image and video source identification

In the case of images, four experiments were performed for source identification purposes. Each of them employed different types of images: native, WhatsApp exchanged, Facebook high-quality exchanged, and Facebook low-quality exchanged, as shown in Fig. 6.3. The results obtained by the experiments execution are shown in Fig. 6.4, in the form of ROC curve of true

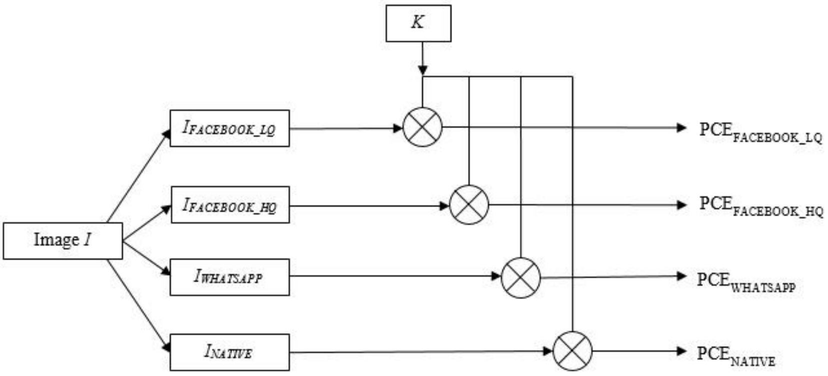


Figure 6.3: Scheme of image source identification using VISION.

positive against false alarm rate. It can be seen from the figure that the worst results were obtained using Facebook low-quality exchanged images, which implies that the compression used in this case of image transfer has a large impact on the ability of reliable source identification. Results for the other three types of images were relatively close and were slightly better for the case of native, natural images, in comparison to the other ones employed.

As it was shown in the previous chapter that digital stabilization highly affects the results of source identification in the case when videos are processed, the experiment was repeated on the video dataset provided by VISION, in order to check the behavior of different videos included in the analysis. Scheme of video source identification using VISION is shown in Fig. 6.5. The results when videos from all the employed devices were used is shown in Fig. 6.6, while Fig. 6.7 provides the results obtained for the case when devices which inherently provide automatic stabilization were excluded from the analysis. The repeated experiment has confirmed the conclusions derived using initial MOSES video dataset: performances strongly drop when digitally stabilized videos are involved.

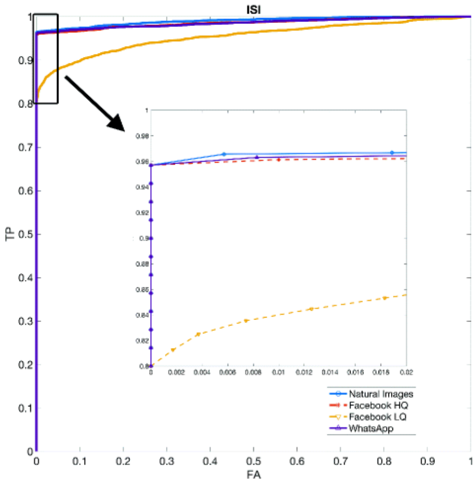


Figure 6.4: ROC curve of image source identification performances on native, WhatsApp exchanged, Facebook high-quality exchanged, and Facebook low-quality exchanged images [6].

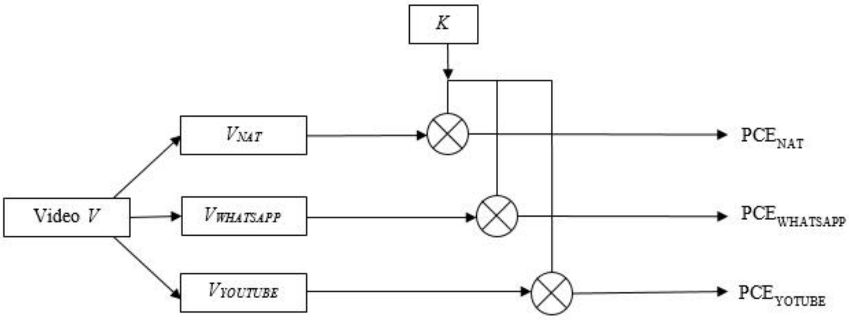


Figure 6.5: Scheme of video source identification using VISION.

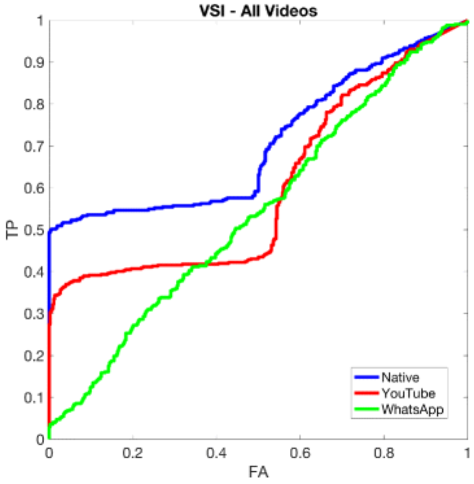


Figure 6.6: ROC curve of video source identification performances on native, YouTube exchanged and WhatsApp exchanged videos [6].

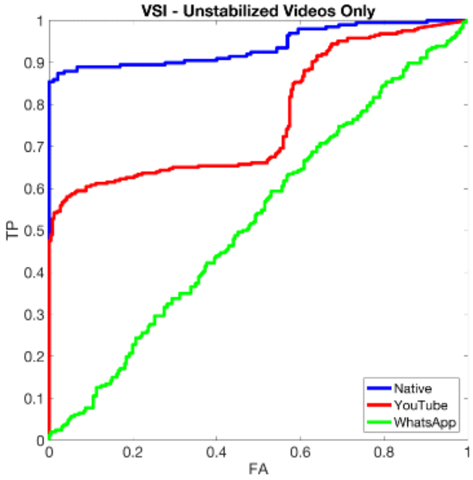


Figure 6.7: ROC curve of video source identification performances on native, YouTube exchanged and WhatsApp exchanged videos, excluding devices with automatic digital stabilization [6].

6.3.2 Source Pattern Noise fingerprint comparison: images vs. videos

Most of the state-of-the-art researches focus only on image or video source identification, but not both, even in cases of available datasets of both of these types of multimedia files. The reason for that is the fact that fingerprints computed using images and videos from the same devices are highly different. Even if the imaging sensor is the same, videos are usually acquired at a much lower resolution than images. While today's smartphones can easily capture 20-megapixel images, 4K video resolution is the highest reachable one. For the comparison purposes, 4K video has 8 megapixels per frame.

Having different maximum possible resolutions for images and videos acquired by the same camera implies having different processes of their acquisition. When recording a video, central crop is first carried to adapt the sensor size to the desired aspect ratio, which is commonly 16:9. Selected pixels are scaled to match the desired video resolution afterwards. This process introduces fingerprint changes, because scaling and other geometrical operations generally affect PRNU-based fingerprint, regardless of multimedia type of interest. Since the process of image acquisition does not require central crop and scaling, it is justified that fingerprints are different for images and videos acquired from the same device.

In case of source identification when both images and videos acquired by the same source device are taken into account, fingerprints of different multimedia need to be adjusted in order to correctly identify source device. For those purposes, image-based and video-based fingerprints are linked by cropping and scaling factors between image and video sensor portion, which usually changes across different device models [6]. The authors in [112] investigated and described the geometrical relation between image and video acquisition processes, which explains this procedure. We invite reader to find more details about it in the aforementioned paper. Procedure described in [112] is used in so-called *Hybrid Source Identification* (HSI) approach, which combines image- and video-based fingerprints.

For the purposes of supporting HSI approach in the analysis conducted in this thesis, cropping and scaling factors for linking the corresponding fingerprints were estimated for several devices. Non-stabilized devices were chosen for this analysis, due to the complexity of devices with the automatic digital stabilization. Estimation was performed on *flat* types of images and videos from VISION dataset. After computation of reference fingerprint, based

Table 6.2: Estimated cropping and scaling factors for non-stabilized videos from VISION dataset [6].

ID	D01	D03	D07	D08	D09	D11	D13	D16	D17	D21	D24	D26	D27	D28	D30	D31	D32	D33	D35	D22
Scaling	0.5	0.48	0.27	1	0.61	0.59	1	0.46	0.59	n.a.	0.5	n.a.	0.5	0.36	0.47	0.46	0.59	0.52	0.39	0.49
Cropping [x y]	[0 228]	[0 372]	[0 7]	[408 354]	[227 411]	[0 307]	[-160 0]	[8 396]	[0 1]	n.a.	[0 240]	n.a.	[0 228]	[0 0]	[39 10]	[9 397]	[0 0]	[464 693]	[0 306]	[0 246]

on 100 images and the same number of video frames, cropping and scaling factors were estimated by brute force search, as suggested in [107]. This approach was used in order to avoid de-synchronization attacks, described as a well-known vulnerability of PRNU-based approach in Section 4.3.

The obtained results are shown in Table 6.2. In case the obtained maximum PCE was lower than the threshold value, the parameter search is considered unsuccessful and denoted as ”n.a.” in the table. Threshold value was accepted to be equal to 45, as it is proposed in [49], due to the obtained empirical results.

It is worth noting that cropping factor is represented in the form of coordinates of corresponding cropping corner, which is the upper-left corner along x and y axes. The reported scaling factors and cropping corners represent the values which cause yielding to the maximum PCE for examined devices. For example, the information given in Table 6.2 can be read as follows: device D07 showed the best performances in terms of PCE values when its fingerprint was scaled for the factor of 0.27 and then cropped on the upper-left side by 7 pixels along the y axis.

Chapter 7

PRNU-based source identification using HDR images

This chapter presents a novel dataset of HDR and SDR images and tests the performances of well-known PRNU-based source identification algorithm. Analysis in terms of image and fingerprint type, as well as reliability of source identification using this method are presented. PCE optimization algorithm is finally proposed in the last section of the chapter.

7.1 Dataset formation

The procedure conducted during the process of creation VISION dataset [6] is adopted for the novel dataset of HDR images. As in VISION, camera that provides the best quality, usually located at the back side of the mobile device, was used for capturing in case of all the used devices.

A novel dataset of HDR and SDR images was created using 23 different portable devices, including Huawei, Apple, Samsung, Xiaomi, Asus, Gionee and One Plus. Devices were configured for capturing in default camera mode. In case of Apple devices, default mode is usually the one that provides the highest quality and resolution available, while that is not necessarily a rule for Android devices. Captures were taken without using flash, in different atmospheres, including both indoor and outdoor scenes. Twenty three models of mobile devices produced by seven different manufacturers were employed. Besides HDR images, created dataset contains standard

SDR images captured by the same devices, in order to enable comparison of PRNU-based source identification between different image types.

Conducted research took into account different possibilities of capturing motions in order to investigate the possible impacts of the pixel artifacts caused by the camera shake on the PRNU estimation and the final source identification. Therefore, the introduced dataset contains images captured using tripod, by a steady hand, and by a shaky hand.

The number of used devices per each manufacturer is as follows:

- 7 Huawei devices,
- 6 Apple devices,
- 4 Samsung devices,
- 3 Xiaomi devices,
- 1 Asus device,
- 1 Gionee device,
- 1 One Plus device.

Seventeen of the employed devices use Android operating system (OS), while the remaining six operate using iOS. Characteristics of the devices can be seen in Table 7.1, which also provides information about the resolution of captured images, their number in accordance to the type (SDR or HDR), and camera movement mode at the time of capturing. Information listed in the table are sorted by brands, and then ordered by models, from the oldest to the newest. Devices are shortly named based on their operating system, e.g. "A" stands for the device that uses Android, while "I" represents the device that uses iOS operating system. Captures were named descriptively, following the format *"device_category_movement_number"*. Abbreviated name of the device model is represented by the *"device"*, *"category"* refers to the image type: HDR or SDR, *"movement"* describes the camera movements, which can be TRIPOD, HAND or SHAKING, while *"number"* represents the ordinal number of the captured image for the device of interest. More information about the meaning of each camera movement type will be provided in the remainder of this section.

Dataset structure is shown in Fig. 7.1. All the captures were first divided into two groups: FLAT and NAT images. The term FLAT refers to the images of flattish, monotonous surfaces, such as walls and skies, which are valuable for PRNU estimation, detection, and the final source identification using PRNU factor. The other category, NAT, consists of images of natural scenes, which can be very detailed, textured, colorful and with

Table 7.1: Characteristics of employed devices and captured images.

Device Class	Device Name	Brand	Model	OS	Image Resolution	SDR Flat	HDR Flat	SDR Hand	HDR Hand	SDR Shaking	HDR Shaking	SDR Tripod	HDR Tripod
A12	Huawei-Honor6plus	Huawei	PE-TL10	Android 6.0	2448 × 3264	50	50	20	20	20	20	20	20
A13	Huawei-Honor6plus	Huawei	PE-TL20	Android4.4.2	2448 × 3264	50 wall	50 wall	20	20	20	20	20	20
A02	Huawei-P8	Huawei	GRA-L09	Android 6.0	4160 × 3120	50 wall	50 wall	24	24	24	24	24	24
A06	Huawei-Y5	Huawei	CUN-L21	Android 5.1	3264× 2448	50(wall)	50(wall)	24	24	24	24	24	24
A04	Huawei-P10	Huawei	VTR-AL00	Android7.0	3968× 2976	51(wall)	50(wall)	15	15	20	20	26	28
A03	Huawei-Honor9	Huawei	STF-AL00	Android7.0	3264× 1840	50(sky)	50(sky)	20	20	20	20	20	20
A05	Huawei-Mate10Pro	Huawei	BLA-L29	Android 8.0	3968× 2976	50(wall)	50(wall)	24	24	24	24	24	24
A09	Galaxy-Note5	Samsung	SM-N920C	Android 7.0	5312× 2988	50(sky)	50(sky)	24	24	24	24	24	24
A07	Galaxy-S7	Samsung	SM-G930F	Android 7.0	4032× 3024	52(wall)	50(wall)	21	21	24	24	21	21
A08	Galaxy-S7	Samsung	SM-G930F	Android 7.0	4032× 2268	50(sky)	50(sky)	24	24	24	24	24	24
A10	Galaxy-J7	Samsung	SM-J730F	Android 7.0	4128× 3096	50(sky)	50(sky)	24	24	24	24	24	24
A15	Xiaom-3	Xiaomi	Redmi Note3	Android 7.1	4608× 2592	50(wall)	50(wall)	24	24	24	24	24	24
A11	Xiaomi5	Xiaomi	MI 5	Android7.0	3456× 4608	50(wall)	87(wall)	21	21	21	21	21	21
A14	Xiaomi-5A	Xiaomi	Note 5A Prime	Android 7.1	4160× 2340	50(sky)	50(sky)	24	24	24	24	24	24
A01	GioneeS55	Gionee	GN9000	Android 4.4	3120× 4208	50(sky)	50(sky)	20	20	20	20	20	20
A17	AsusZenfone-2	Asus	ASUS_Z00ED	Android 6.1	3264× 1836	50(sky)	50(sky)	24	24	24	24	24	24
A16	OnePlus-3t	OnePlus	A3003	Android 8.0	4640× 3480	50(wall)	50(wall)	24	24	24	24	24	24
I06	iPhone 5S	Apple	15A372	iOS 11	3264× 2448	50(wall)	50(wall)	24	24	24	24	24	24
I04	iPad Air	Apple	A1475	iOS 11.0.1	2592× 1936	50(wall)	50(wall)	24	24	24	24	24	24
I05	iPhone 6	Apple	A1586	iOS 11.3	2448× 3264	50(wall)	50(wall)	21	21	21	24	21	21
I02	iPhone se	Apple	A1723	iOS 10.3.3	4032× 3024	54(sky)	54(sky)	19	19	19	19	19	19
I03	iPhone 7	Apple	A1778	iOS 11.3	4032 × 3024	50(wall)	50(wall)	24	24	24	24	24	24
I01	iPhone 8	Apple	A1863	iOS 11.3	3024× 4032	50(sky)	50(sky)	15	15	15	15	15	15

large illumination alternations. FLAT set of images wes created in order to enable PRNU extraction, while NAT images represent a set of real-case scenario images, whose source camera identification could be needed.

As the images were captured in both SDR and HDR mode, they can be further classified in accordance to their type. For the case of NAT set, it is useful to differ images based on camera movements that occurred at the time of capturing. They were therefore sorted as shown in Fig. 7.1. Naming convention used for the movements description is intuitive, where TRIPOD stands for the images captured when camera device is fixed on the tripod, HAND category contains the images taken by hand, while SHAKING category involves images captured by shaky hand. Tripod enables camera steadiness, which minimizes possibility of pixel artifacts. On the other hand, captures taken by steady hand include only small pixel artifacts, mostly invisible to the human eye. Blurring effect becomes visible if the capturing is performed while shaking a camera device, due to the pixel shifting. As HDR images are created as a combination of multiple SDR captures, it is expected that the artifacts will be accentuated and make the source identification procedure more difficult.

The examples of captures are given in Fig. 7.2.

7.2 Experiments

7.2.1 Fingerprint computation

Camera fingerprints were computed based on PRNU estimation for all 23 devices employed in the novel dataset. For the purposes of testing the quality of PRNU estimation, three fingerprints were produced for each device, based on different sets of flat images.

The first group of fingerprints was computed using flat HDR images, where the number of images deviated from 50 to 87 for different devices. As the improved method for PRNU factor estimation requires at least 30 images [49] for successful procedure conduction, the chosen number of images employed in the analysis ensures reliable estimation results. Considering the specifics of HDR images, it is expected that source identification based on PRNU factor would produce better results when both fingerprint of the device and the image taken by the same are of HDR type. Analogy applies to the SDR images. In order to test this assumption, second group of finger-

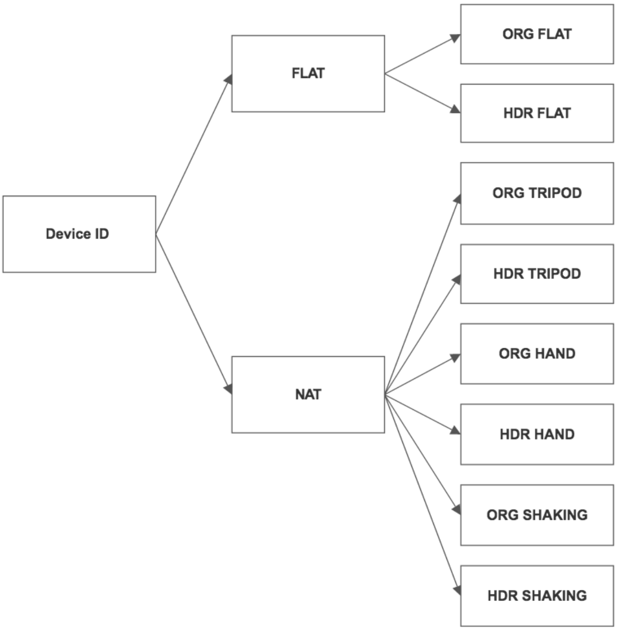


Figure 7.1: The dataset structure.

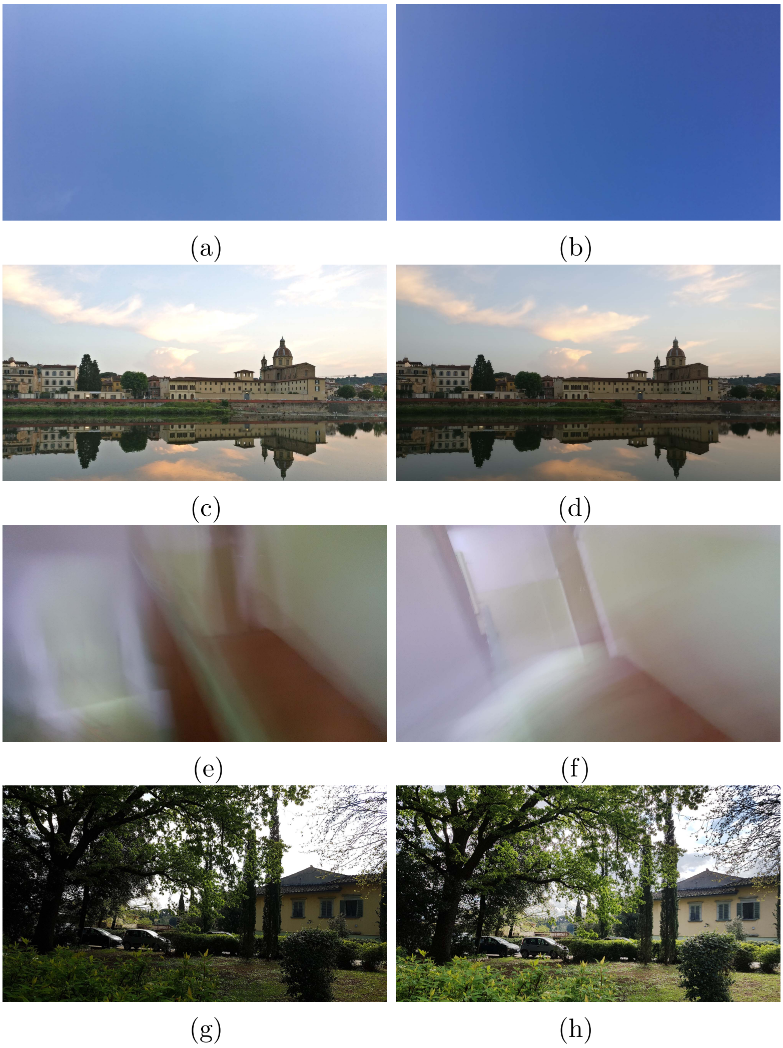

(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)

Figure 7.2: Sample pictures from the Dataset: (a) SDR FLAT, (b) HDR FLAT, (c) SDR TRIPOD, (d) HDR TRIPOD, (e) SDR SHAKING, (f) HDR SHAKING, (g) SDR HAND, (h) HDR HAND.

prints was calculated from the set of 50 to 59 flat SDR images, where the concrete number of images differed between the devices.

Finally, mixing SDR and HDR images, a group of captures called MIX was formed. Including both types of images of interest, MIX category provides the most reliable results for source identification, considering that it is not common in the real-case scenario that users possess an information about the image properties. Fingerprints from MIX set were obtained using 100 to 137 images per device, thus ensuring even higher reliability of the fingerprint estimation results.

7.2.2 Test images processing

Camera photo response non-uniformity detector was obtained using the generalized likelihood test based on cross-correlation maximization. Following the PRNU procedure presented in Chapter 4, test images from the NAT part of the dataset were processed. After the noise extraction was performed, it was correlated with the PRNU factor estimate, in accordance to relation 4.7. Maximum of the normalized correlation ρ_b is considered to be a good approximation of the generalized likelihood ratio test [113] and it was therefore computed.

7.2.3 Parameter of comparison

Peak to Correlation Energy (PCE) ratio was used in the experiments as the measure of relevance of PRNU-based source identification. It was computed over all the images acquired by the device of interest and values for single images were compared to the threshold value that separates acceptance of hypotheses H_0 and H_1. Threshold value was accepted to be equal to 45, as it is proposed in [49], due to the obtained empirical results. For more details about choosing the threshold value, we invite readers to refer to the analysis conducted in [49]. If the value was higher than threshold, hypothesis H_1 (matching image for the source device) was accepted. Otherwise, accepted hypothesis was H_0 (non-matching image for the source device).

It is worth noting that PCE computation was performed for all the analyzed images captured by a certain camera model, and was subsequently averaged. Results have shown that averaged PCE was less prone to result variations and camera movements have had less impact on the results in the case of using this parameter.

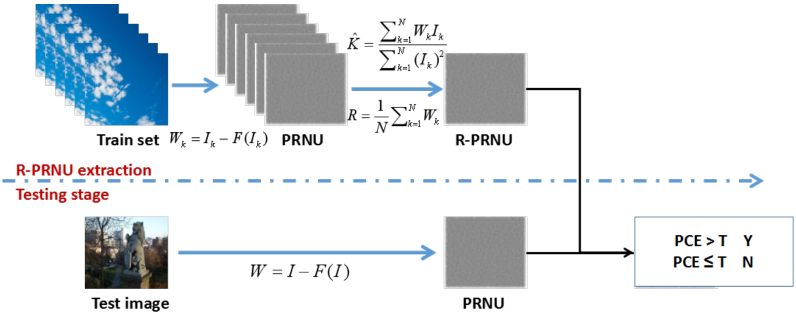


Figure 7.3: The framework of PRNU-based algorithm.

The framework of PRNU-based algorithm is shown in Fig. 7.3.

7.3 Results

This section provides description of multiple stages of analysis during the process of final PRNU-based source identification. Since two type of images were used, the one with the standard dynamic range - SDR, and the one with wider dynamic range - HDR, the first aim was to conclude if there is a noticeable difference between the types of the images. The analysis was further expanded on the fingerprints created from multiple images of the same type, as well as on the groups of images of different types. This step was conducted with the same purpose of revealing the differences between SDR and HDR captures and to finally conclude if their fingerprints converge to the same result, or they differ despite numerous images captured by the same device were included in the analysis. During the analysis, the impact of image and fingerprint types, as well as the impact of motions occurred in the time of image capturing were observed.

After the first step of types and fingerprints difference recognition, there was a need of analyzing PCE values produced for single devices. The aim was to find a reason for low PCE values for single images or a set of images and to find the correlation between the images with low PCE values, if the one exists. The final step of the analysis deals with the problem of defining how reliable were the produced results for source identification and the methods for the reliability increase.

7.3.1 Analysis in terms of image type: SDR vs. HDR

Theoretical introduction to differences between SDR and HDR images is provided in the previous chapters. Following that knowledge, correlation was first computed between the noise extracted from SDR image and fingerprint computed on multiple images of SDR type. Analogously, correlation was computed between HDR-based components. This experiment was conducted in order to define if the complexity of HDR images creation is an aggravating factor in PRNU-based source identification. Furthermore, it aimed to determine the difference between SDR and HDR images in terms of digital image processing procedure.

Correlation of noise, extracted from SDR images, with flat SDR-based fingerprint resulted in generally higher PCE values in comparison to the case when noise from HDR images was employed. The results can be seen in Fig. 7.3.1-7.7. Android devices A01-A06 have shown the biggest difference between PCE values in case when the captures were taken by a tripod. While results obtained for SDRs correlation were characterized with high PCE values in that case, PCEs of most of the HDRs were low. In some cases, they were even below the threshold. Difference in terms of higher PCE value for SDRs in comparison to HDRs is noticeable in the case of captures taken by steady hand, as well. In case of captures made by shaky hand, analogy to the previous two cases cannot be applied. While devices A02, A04 and A05 were shown to have similar PCE values for both cases, when noise is extracted from SDR and HDR images, the other half of the devices was shown to have higher PCE values when SDR components are correlated.

Similar results were obtained with devices A07-A17. Differences between PCE values of SDR and HDR images were not as emphasized as for the previously considered set of devices, rather minor in case of devices A07-A10. On the other hand, A11-A17 followed the same behavior as A01-A06 devices. Images captured by shaky hand did not show to follow any pattern. While PCEs were similar for devices A08-A11 and A15, they were distinctively higher for SDR than HDR images captured by other devices from the analyzed set.

Finally, iPhone devices I01-I06 have shown to obtain similar PCE values for both of the image types, regardless of the camera motion. While PCE values computed during the analysis of SDR images coming from devices I02-I06 were slightly higher than the ones corresponding to the HDR images, I01

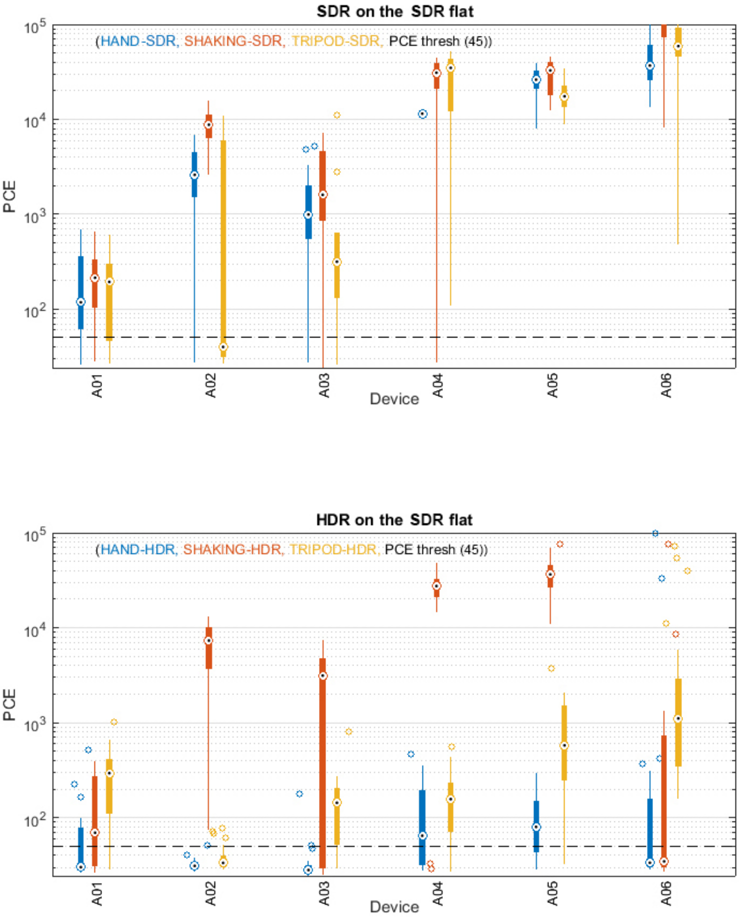


Figure 7.4: PCE values obtained by SDR and HDR images when compared with a flat SDR - based fingerprint (devices A01-A06).

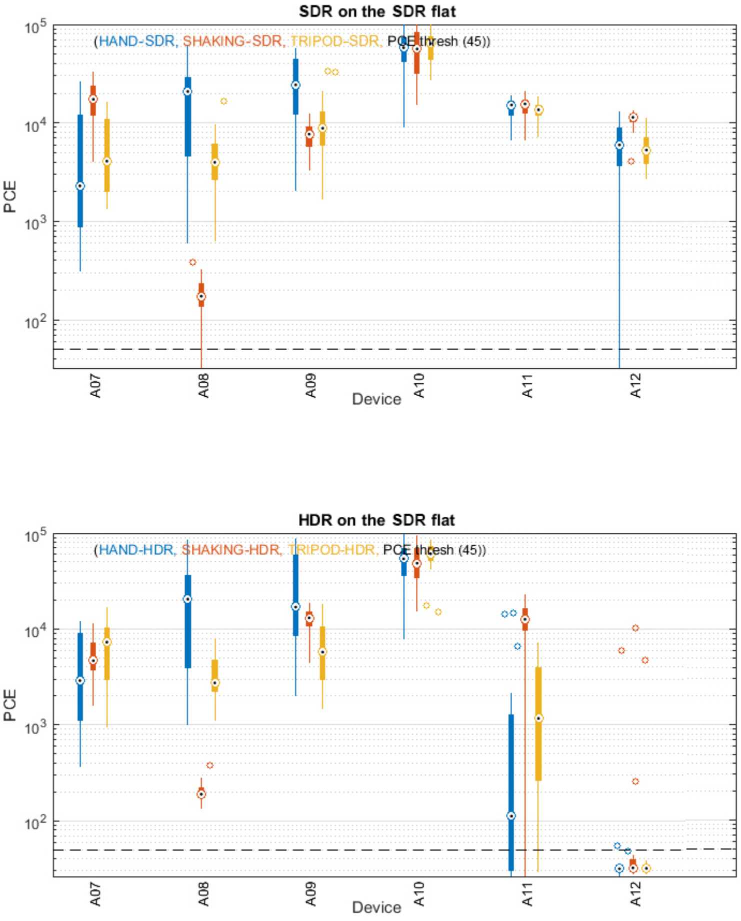


Figure 7.5: PCE values obtained by SDR and HDR images when compared with a flat SDR - based fingerprint (devices A07-A12).

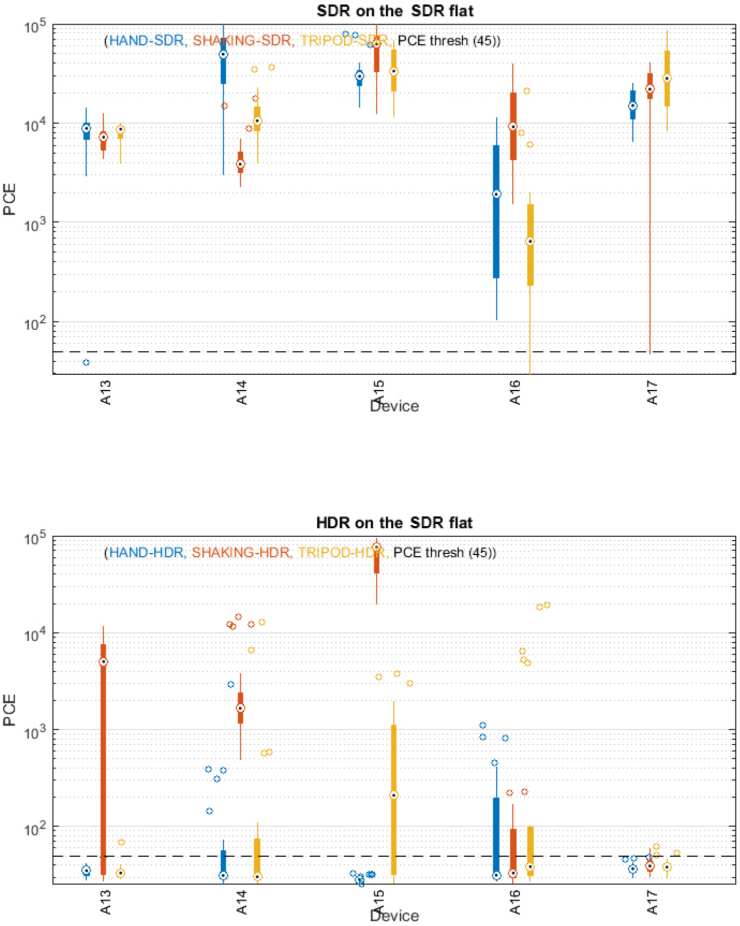
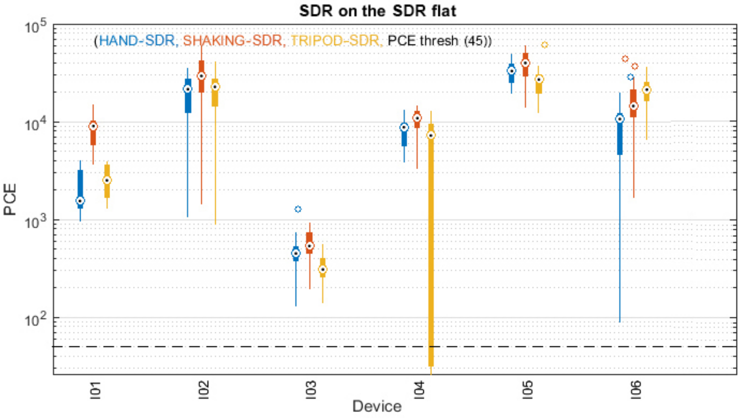


Figure 7.6: PCE values obtained by SDR and HDR images when compared with a flat SDR - based fingerprint (devices A13-A17).



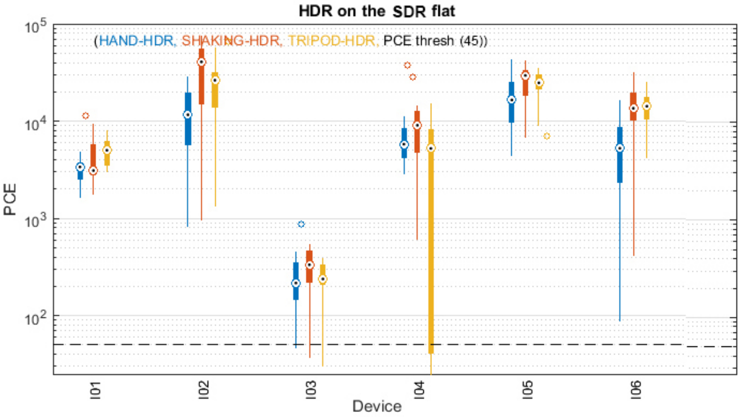


Figure 7.7: PCE values obtained by SDR and HDR images when compared with a flat SDR - based fingerprint (devices I01-I06).

device has shown the unexpected results. With this device, PCE values for images captured by steady and by shaky hand are shown to be higher for HDR images correlated with the SDR-based fingerprint.

At this stage of analysis, it is already noticeable that camera motions have an impact on the image noise. However, the results did not show the best performances in case of complete steadiness during the image capturing, nor that the motions have the same impact on all the devices. Taking into account different possibilities of image creation and different types of imaging sensors, it is expected to obtain results that cannot be generalized to all the camera devices.

The analysis was further conducted by comparison of PCE values when noise from SDR and HDR images was correlated with HDR-based fingerprint. Results are shown in Fig. **??**. All the devices have shown the analogous behavior as the previously described one. It was noted that the majority of PCE values obtained by correlation of two image components (noise and fingerprint) of the same type was higher than the threshold value, while correlation of components of different types results in high or low PCE values, depending on the employed camera device. This fact lead to the conclusion that manufacturers choice of camera hardware has a great impact on possibility and reliability of PRNU-based source identification.

It is worth to show and to discuss the atypical result cases. An example obtained by a single I02 device is shown in Fig. 7.12. It is noticeable that PCE values of HDR images captured by this device model were above the threshold value for all the tested images, when they were correlated with the fingerprint of SDR set of images. This occurred regardless of the camera movements. Similar was obtained by employing SDR images captured by completely different device model, A07, and correlating the relevant noise to fingerprint of HDR images. Obtained result is given in Fig. 7.13. These two examples show that some of the devices can be identified easier than others, and that the correct identification of those devices can be provided regardless of the type (HDR or SDR) of the images. On the other hand, most of the devices have shown significantly different PCE values of images, depending on their type.

7.3.2 Analysis in terms of fingerprint type

Previous section describes the analysis which focuses on the impact of single image type on the obtained results. While fingerprints were the constant, referent components of analysis, sets of images belonging to different devices were variables. On the other hand, this section analyses the impact of different fingerprints on the same sets of images.

Comparing results shown in figures 7.3.1-7.7 with 7.3.1-7.11 for the same set of devices, it is noticeable that SDR images have had higher PCE value when HDR-based fingerprint was employed in case of A01, A05, I01 and I03 devices, for all three motion scenarios. Difference in terms of PCE value enhancement in case of motion change cannot be seen for these devices. On the other hand, images captured by devices A11, A15 and A17 have had significantly higher PCE value when the noise extracted from SDR images was correlated with SDR-based fingerprint. Images captured by all the other devices were shown to have similar PCE values for both of the fingerprints. Therefore, only seven of the employed devices were shown to have a noticeable impact of fingerprint on the obtained results. This lead us to the conclusion that a combination of bigger number of images, regardless of their type, can suppress the anomalies and specific characteristics of different image types to produce a reliable PRNU estimate in case of most of the devices. It is assumed that PRNU estimate converges to the same estimation result in case of both HDR and SDR images for these devices.

To confirm the previous statement, analysis of fingerprint impact on PCE values was performed for HDR images, as well. It was shown that images from devices A11 and A15 have had higher PCE values in correlation to SDR-based fingerprint. While A11 device showed no differences in the amount of PCE improvement for different motion scenarios, images from A15 device have had significantly higher PCE in case of capturing by shaky hand. Improvements were noticeable for images captured using tripod, but there were no differences in case of images captured by steady hand. In contrast to the previously mentioned devices, A01, A17, I01 and I03 have had better performances when the noise extracted from their HDR images was correlated with corresponding HDR-based fingerprint. Differences in terms of camera motions were not noticeable in these cases. All the other employed devices were shown to have similar results for PCE values of HDR images, regardless of the fingerprint type. Considering the fact that the majority of devices did not show a big difference between the fingerprints based on HDR and SDR images, the conclusion derived for previously described SDR images analysis can be confirmed.

Comparing devices that deviate from the conclusion for SDR and HDR images at this stage of analysis, it can be seen that the same devices appear in both of the cases. This makes a total of 7 out of 23 devices which produce a noticeably different fingerprint when different types of images are used in the process of fingerprint computation.

Having conducted the analysis of both image types and fingerprint types separately, it is worth paying attention to the overall results. The ones obtained for devices A13 and A14 are of a special interest because they show a distinctive difference between SDR and HDR images taken by those devices. In those cases, SDR images have had high PCE values, no matter if the correlation was performed using flat SDR- or HDR-based fingerprint, while HDR images have had low PCE value, except from the images captured while shaking the camera device. Therefore, fingerprint type did not show to have a big influence on the results in this case, but the type of the images did. Moreover, camera movements were confirmed to have an impact on the results.

7.3.3 MIX category results analysis

The previous research stages considered separately HDR and SDR images, not only as single objects used for the purposes of source identification testing, but also in the process of fingerprint computation. This section deals with combined sets of HDR and SDR images captured by the same device, which are contained in the MIX category. At this stage, fingerprints were computed based on the relevant MIX set of images for all the employed devices. Results from the previously described analysis served as a motivation for this step, because it was shown that sources are identifiable even in the cases when correlation was computed between the HDR image noise and SDR image fingerprint and vice versa. Considering the fact that the original image properties are rarely available in the real-case scenarios when source identification is needed, MIX category of images can provide the most reliable results for source identification.

The obtained results are shown in Fig. 7.14-7.17. It can be seen that the averaged PCE value of images captured by most of the devices was above the threshold when MIX category of images was used as a reference.

SDR images from devices A01-A06 have shown better performances than their HDR counterparts when they were captured by steady hand or using tripod. On the other side, images taken by shaky hand using devices A02, A04 and A05 have had similar PCE values, regardless of the image type, in correlation to the MIX flat fingerprint. The other half of the devices from this set have shown better results for SDR images in case of shaking motion. In this case, captures taken by shaky hand lead to bigger variations in results, comparing to more steadily captured images. This observation is justified by the fact that camera movement shifts the fingerprint matrices, making different offsets for the analyzed images. The offset depends on the velocity of the camera, which has not been measured during the dataset formation process.

Difference between SDRs and HDRs in terms of PCE value was not significant for devices A07-A10 when the MIX-based fingerprint was employed. SDRs have shown better performances for devices A11 and A12, with the exception of images taken by shaky hand using A11 device. In that case, PCEs were comparable for SDR and HDR images. Similar conclusions can be conducted by analyzing results obtained for devices A13-A17, where only captures taken by devices A15 and A17 in shaking motion have similar PCE values for both SDRs and HDRs, while the rest of the devices and motion scenarios show the advantage of SDR images in source identification process using PRNU method.

Deviation from the previous results occurred in the analysis of iPhone devices. Images captured by I01-I06 in different motions have shown comparable PCE values for both SDR and HDR images. All the values were above the threshold, with the exception of one part of the images taken by I04 device using tripod. These results lead us to the conclusion that iPhone devices are easier to identify than other devices included in this research, no matter of the type of the analyzed image. This conclusion corresponds to the one conducted after analyzing impact of using different types of images and the same SDR- or HDR-based fingerprint for PRNU computation.

7.3.4 Reliability of source identification

The final stage of analysis was determining reliability of PRNU-based source identification. As it is stated in the previous sections, reliability of the procedure results was decided based on calculated PCE values. The threshold PCE value was chosen to be equal to 45.

Results have shown that both SDR and HDR image sources can be detected using this threshold, with the exception of HDR images taken from devices A12, A14 and partially A6 and A17. Considering this fact, it is clear that PRNU method cannot be generally applied, because the devices themselves can introduce variable hidden digital content to the images they produce or affect the procedure in other manner.

The most reliable source identification was provided for devices A07, A09, A10, I01, I02, I03, I05 and I06. Camera movements and usage of flat images were shown to have a minimal effect to PCE value for the previously mentioned devices. On the other hand, devices A06, A12 and A16 were shown to produce higher PCE values for SDR, than HDR images. Furthermore, source identification from SDR images was less prone to camera movements for those devices. Taking the previous statements into account, it can be concluded that complexity of HDR images introduces difficulties in source identification for some devices. This phenomenon requires further analysis of the HDR images creation procedure for the devices of interest.

7.3.5 Analysis of low PCE values

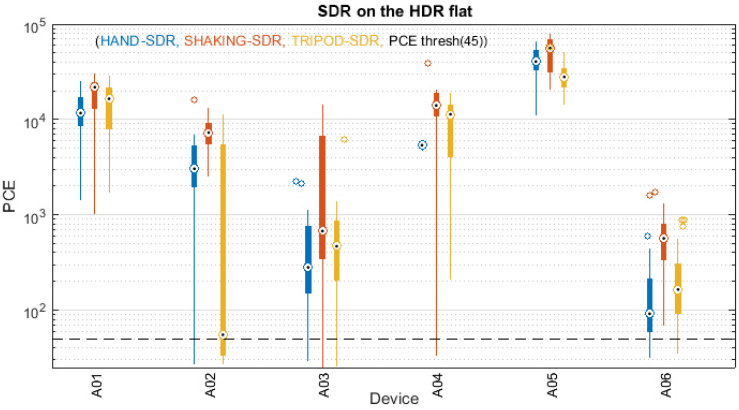
During the image analysis using standard PRNU method, it was noticed that PCE value is unexpectedly low for some of the captures, in comparison to the PCE values of other analyzed images from the same set. Guided by this fact, it was decided to post-process the results in order to determine the reason of poor PCE values for single cases.

The result obtained from A01 model is provided in Fig. 7.18. Twenty groups of images were captured in different motions and modes. Each group was provided the same image content as controlled variable. Considering the differences in acquisition process of SDR and HDR images, it can be concluded, by comparing the PCE values among three different motions, that image alignment has a serious impact on performances of the PRNU-based method. As shown in Fig. 7.18, PCE values of SDR images are higher than the ones of HDR images captured in hand motion. However, situation is opposite in tripod motion. The reason could be that the image alignment operation in hand motion changes positions of pixels, which leads to the mismatch between the noise image extracted from HDR image and R-PRNU. In the case of tripod motion, multiple images with perfect alignment are used to extract noise image. It is well-known that the more images are employed, the more precisely the PRNU is calculated. Therefore, higher PCE values could be obtained for HDR images in this case. In case of shaking motion, depending on the algorithms used in each device, on the one hand, the shift among the images would be too big to align images, which improves the PCE value of HDR images. On the other hand, image alignment is executed reducing the PCE value of HDR images.

In order to further explore the reason behind the change of PCE value between SDR and HDR images, PRNU method based on the pixel patches was applied. Firstly, the images and R-PRNU were cropped into non-overlapping pixel patches with 128×128 size. Then, the PCE values for each pixel patch were calculated and for each image pair (SDR and HDR images), they were mapped into the same scale with log function to obtain the PCE map. PCE maps of SDR and HDR images captured in hand motion (Fig. 7.19) are shown in Fig. 7.20.

An interesting phenomenon occurs at smooth image regions with low luminocity, such as ground with low brightness. PCE values of HDR images have had higher values than their SDR counterparts in that case. The same results were obtained for both over- and under-exposed image regions. On contrary, PCE values were decreased for the pixel patches with smooth and high luminance, such as the blue sky. The reason could be that HDR images keep balance between the dark and bright areas and the PRNU-based method performs better for the images with much smoother and higher luminance. According to the previously described analysis, it can be concluded that, for the smooth pixel patches with higher luminance, but not saturation, HDR and SDR images both have high PCE values. Moreover, image regions with over/under-exposure usually lead to low PCE value. In addition, the images captured with strong amount of noise, such as the night scene shown in the last column of Fig. 7.20, also have low PCE value.

The above presented analysis is more specific, rather general, due to the fact that each device has its own specifics which directly influence the results of PCE values obtained on images acquired by them. Considering that, the further analysis in terms of image acquisition [114] and sensor pattern noise specifics [80] is required. Proposed dataset provides the ability for this and wider researches, such as estimation of displacement fields from pairs of digital images [115] and characterization of the dynamic behaviour of a mechanical chain tensioner by functional tolerating [116].



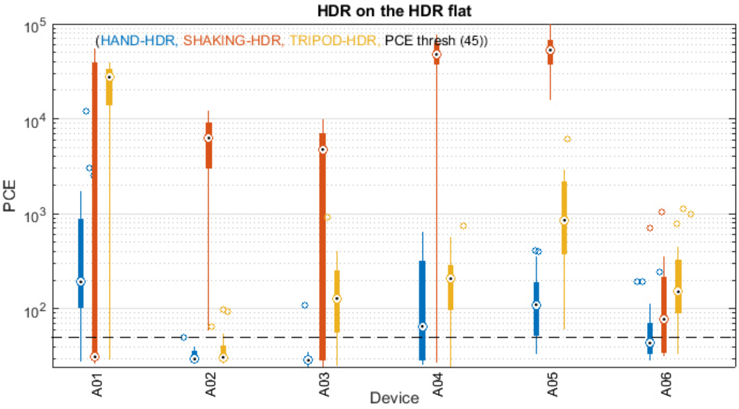


Figure 7.8: PCE values obtained by SDR and HDR images when compared with a flat HDR - based fingerprint (devices A01-A06).

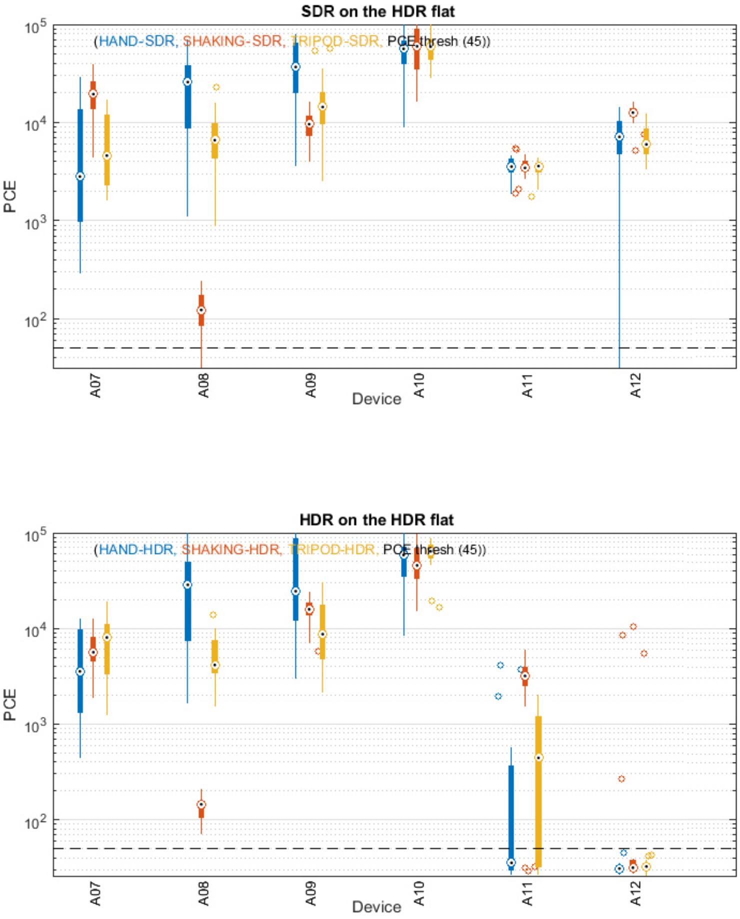
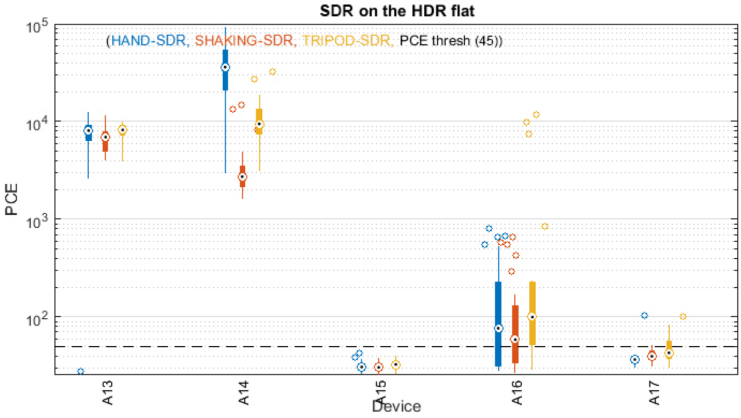


Figure 7.9: PCE values obtained by SDR and HDR images when compared with a flat HDR - based fingerprint (devices A07-A12).



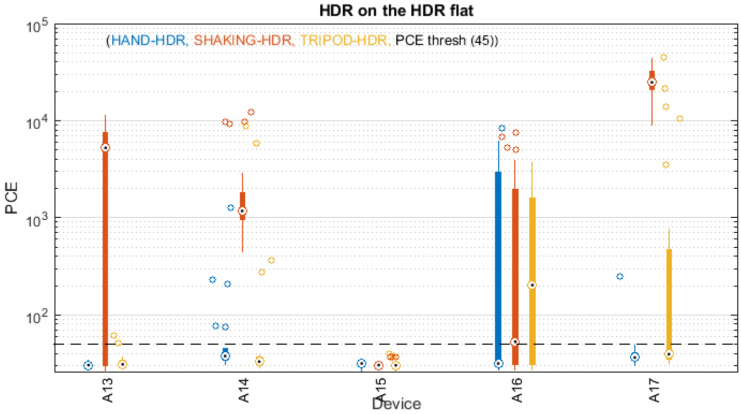
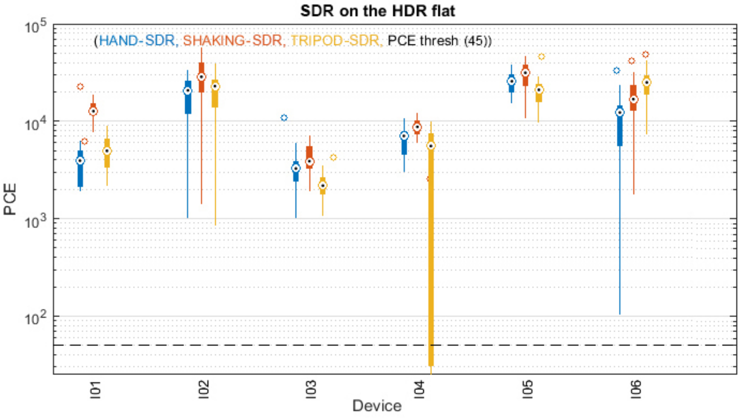


Figure 7.10: PCE values obtained by SDR and HDR images when compared with a flat HDR - based fingerprint (devices A13-A17).



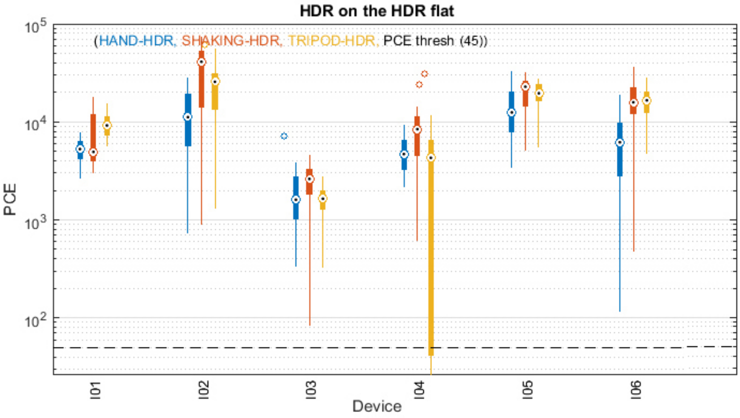


Figure 7.11: PCE values obtained by SDR and HDR images when compared with a flat HDR - based fingerprint (devices I01-I06).

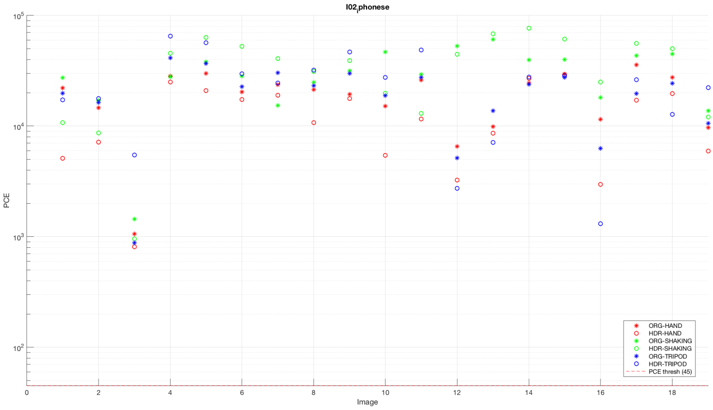


Figure 7.12: Example of result obtained correlating noise from HDR images captured by I02 model with SDR images fingerprint.

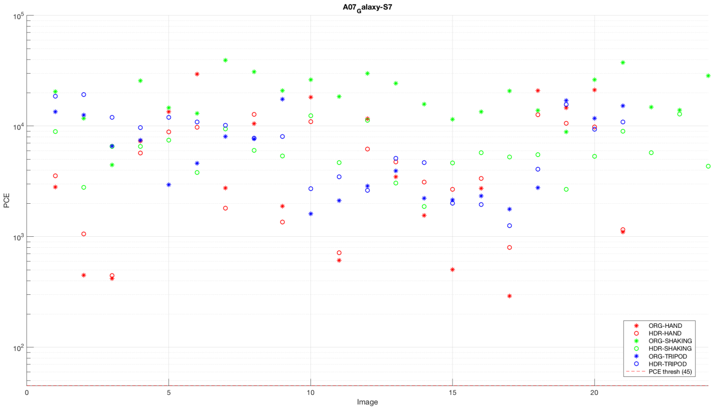


Figure 7.13: Example of result obtained correlating noise from SDR images captured by A07 model with HDR images fingerprint.

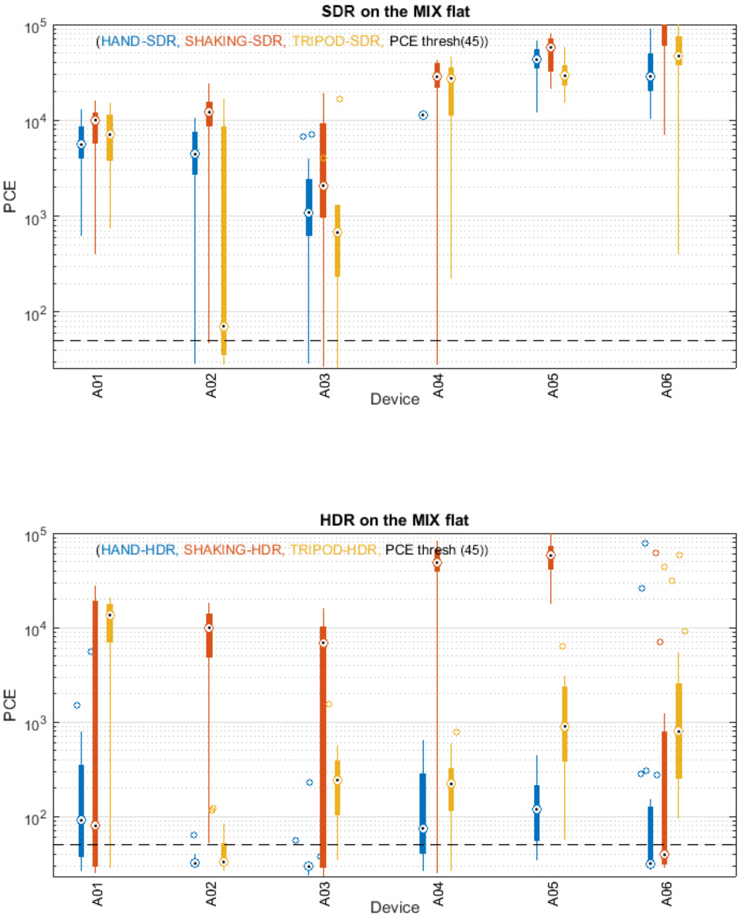
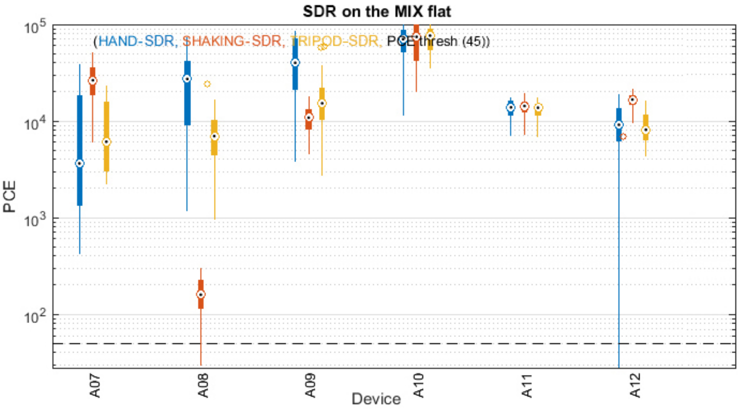


Figure 7.14: PCE values obtained by SDR and HDR images when compared with a flat MIX- based fingerprint (devices A01-A06).



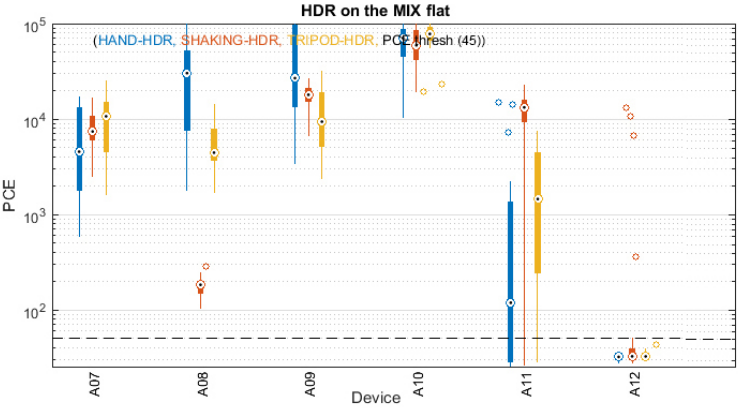
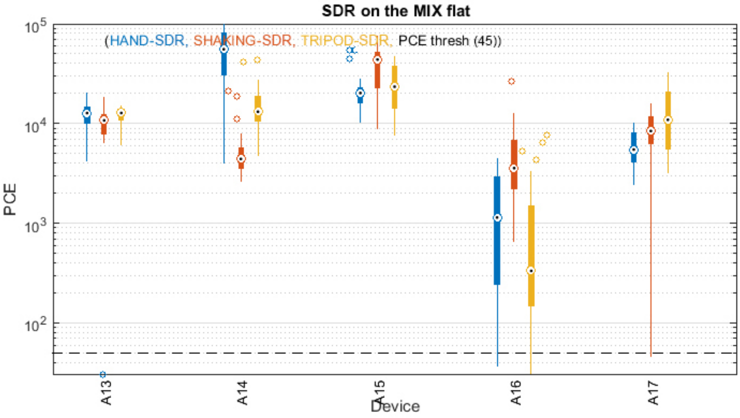


Figure 7.15: PCE values obtained by SDR and HDR images when compared with a flat MIX- based fingerprint (devices A07-A12).



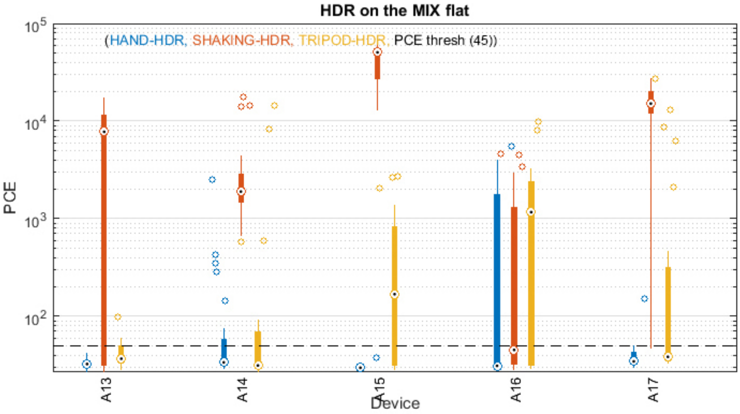
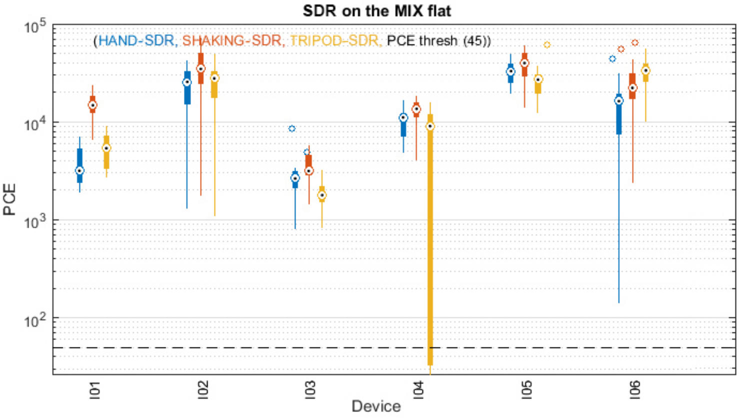


Figure 7.16: PCE values obtained by SDR and HDR images when compared with a flat MIX- based fingerprint (devices A13-A17).



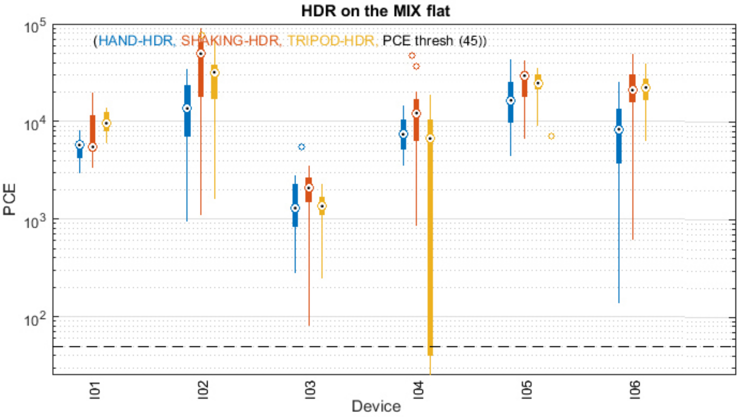


Figure 7.17: PCE values obtained by SDR and HDR images when compared with a flat MIX- based fingerprint (devices I01-I06).

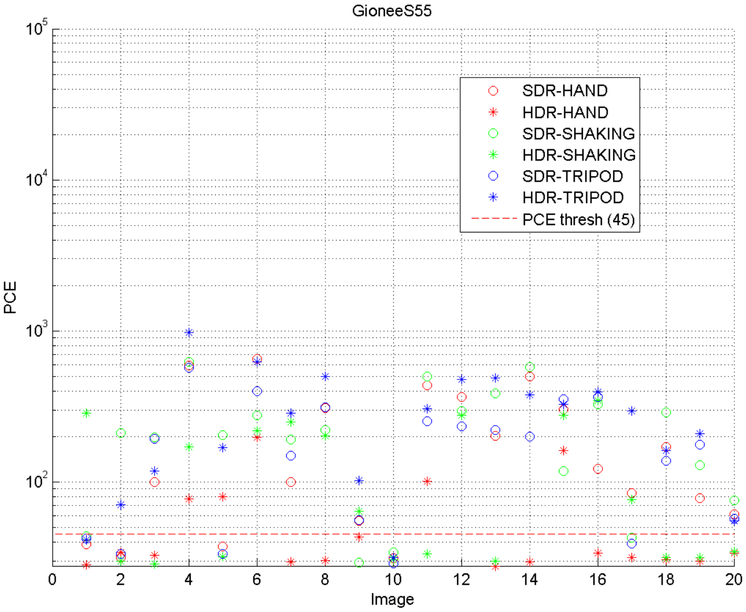
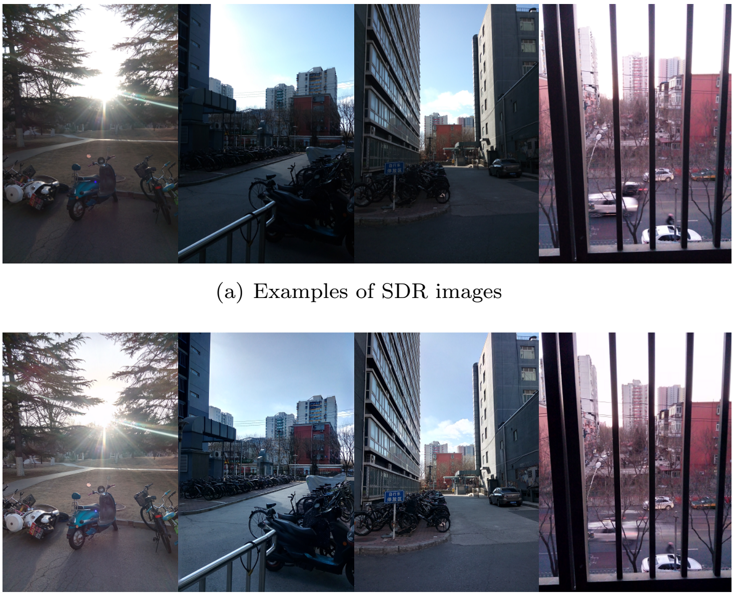


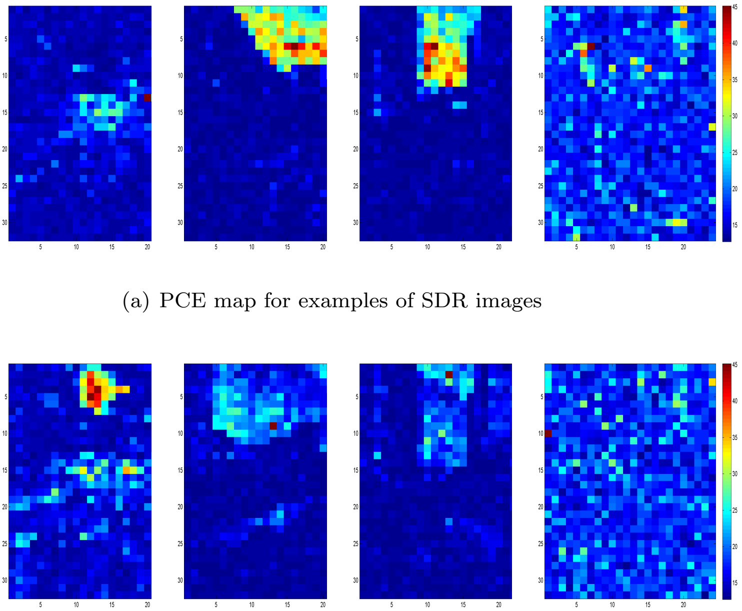
Figure 7.18: Example of result obtained correlating noise from HDR images captured by A01 model with SDR images fingerprint.



(a) Examples of SDR images

(b) Examples of HDR images

Figure 7.19: Examples of (a) SDR and (b) HDR images.



(a) PCE map for examples of SDR images

(b) PCE map for examples of HDR images

Figure 7.20: PCE maps for examples of SDR and HDR images.

Chapter 8

Conclusion

In order to obtain reliable results in multimedia forensics investigations, having a properly formed dataset with all the information needed is of a crucial importance. During the research activities conducted within this thesis, three novel datasets of images and videos were introduced. Datasets were further used in source identification procedure based on the well-known algorithms, in order to investigate the impact of acquisition processes of modern smartphone devices on the procedure. Different camera motions, multimedia types and compression algorithms used in several social media platforms were taken into account.

As most of the available databases face the problem of non-expandability, they suffer from becoming out-dated and inadequate for investigations in the field of multimedia forensics. MOSES was introduced as a solution in the form of mobile application able to record and store videos in the dataset already containing a large number of video files. Being easily accessible and simple for using, MOSES has a good potential for creating a large, continuously updated and expanded video dataset. As it is relatively novel application, its popularity and success should be traced in order to make improvements and optionally include possibility for uploading images alongside of videos. Moreover, for enabling further researches on the created dataset, it will become publicly available on-line in the near future.

VISION was introduced as the second proposed dataset, containing both images and videos. PRNU-based source identification was tested on this dataset and PRNU estimates computed on the basis of videos and images acquired by the same devices were compared. Comparison has shown that the estimates have significant differences and that some manipulations have to be performed in order to match them. Considering the adequate algorithm has not been developed up to this date, VISION can serve as a suitable testing dataset in the researching process, due to the large span of multimedia and information it contains.

Furthermore, previously mentioned datasets include images and videos exchanged through the most popular social media platforms, thus enabling investigation of the exchanging impact on the original files. The research conducted in this thesis has shown that this procedure introduces difficulties in source identification based on PRNU estimation, lowering the correlation between the exchanged multimedia and its acquiring device. The thesis introduced an analysis of ROC curves produced for original multimedia and multimedia exchanged through WhatsApp, Facebook and YouTube, providing comparison of source identification reliability. However, deeper analysis of the impact of different compression algorithms, factors and the number of compression times on source identification procedure is required.

PRNU-based source identification algorithm was shown to have obstacles in case of MOSES and VISION datasets, although they contained multimedia files which were mostly obtained using standard capturing profile. For the purposes of investigating other possible obstacles introduced with modern smartphone devices, HDR images were analyzed as well, due to their complexity and wider dynamic range. It was shown that HDR images, in most cases, are harder to correlate with the source device, comparing to SDR images. As their popularity increases, this can become a burning problem and research topic not only for HDR images, but also for HDR videos.

Different camera motions and device characteristics were also shown to affect the results. As the obtained results could not be generalized to all the capturing devices, the need for deeper analysis in terms of acquisition processes of the employed devices is imposed. Further researches on this topic can also include photographic devices, which use different capturing procedures than smartphones. This investigation can help in obtaining more generalized results in terms of reference pattern noise-based source identification possibilities.

Considering the scope and content of datasets presented in this thesis, they can be employed in various multimedia forensics investigations, but also in other scientific fields related to image and video files.

Appendix A

Publications

This research activity has led to several publications in international journals and conferences. These are summarized below.[1]

International Journals

1. **Al Shaya, O**, Yang. P, Ni. R, Zhao. Y, and Piva. A. “A new dataset for source identification of High Dynamic Range images”, *Sensors*, 2018

1. Shullani. D, Fontani. M, Iuliani. M, **Al Shaya, O** and Piva. A.“VISION: a video and image dataset for source identification”, *EURASIP Journal on Information Security*, 2017

National Conferences

1. Shullani, D., **Al Shaya. O**, Iuliani. M, Fontani. M, and Piva. A. “A Dataset for Forensic Analysis of Videos in the Wild”, in International Tyrrhenian Workshop on Digital Communication (pp. 84-94) *Springer,Cham*, Italy, September 2017

[1] The author’s bibliometric indices are the following: H-index = 1, total number of citations = 14 (source: Google Scholar on Month 1, 2019).

Bibliography

[1] T.-T. Ng, S.-F. Chang, C.-Y. Lin, and Q. Sun, “Passive-blind image forensics,” in *Multimedia security technologies for digital rights management*, pp. 383–412, Elsevier, 2006.

[2] K. Bahrami, A. C. Kot, L. Li, and H. Li, “Blurred image splicing localization by exposing blur type inconsistency,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 5, pp. 999–1009, 2015.

[3] J. Nakamura, *Image sensors and signal processing for digital still cameras.* CRC press, 2016.

[4] E. Reinhard, W. Heidrich, P. Debevec, S. Pattanaik, G. Ward, and K. Myszkowski, *High dynamic range imaging: acquisition, display, and image-based lighting.* Morgan Kaufmann, 2010.

[5] D. Shullani, O. Al Shaya, M. Iuliani, M. Fontani, and A. Piva, “A dataset for forensic analysis of videos in the wild,” in *International Tyrrhenian Workshop on Digital Communication*, pp. 84–94, Springer, 2017.

[6] D. Shullani, M. Fontani, M. Iuliani, O. Al Shaya, and A. Piva, “Vision: a video and image dataset for source identification,” *EURASIP Journal on Information Security*, vol. 2017, no. 1, p. 15, 2017.

[7] R. YANG, W. LUO, and J. HUANG, “Multimedia forensics,” *SCIENTIA SINICA Informationis*, vol. 43, no. 12, pp. 1654–1672, 2013.

[8] A. Piva, “An overview on image forensics,” *ISRN Signal Processing*, vol. 2013, 2013.

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

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

[11] E. Casey, *Digital evidence and computer crime: Forensic science, computers, and the internet.* Academic press, 2011.

[12] R. Böhme, F. C. Freiling, T. Gloe, and M. Kirchner, "Multimedia forensics is not computer forensics," in *International Workshop on Computational Forensics*, pp. 90–103, Springer, 2009.

[13] A. E. Hassanien, M. Fouad, A. A. Manaf, M. Zamani, R. Ahmad, J. Kacprzyk, *et al.*, *Multimedia Forensics and Security.* Springer, 2017.

[14] J. Crabtree and A. Sellers, "Rating organization cybersecurity using active and passive external reconnaissance," Aug. 2 2018. US Patent App. 15/823,363.

[15] R. C. Pandey, S. K. Singh, and K. K. Shukla, "Passive forensics in image and video using noise features: a review," *Digital Investigation*, vol. 19, pp. 1–28, 2016.

[16] F. Y. Shih, *Digital watermarking and steganography: fundamentals and techniques.* CRC press, 2017.

[17] S. K. Tripathi and B. Gupta, "An efficient digital signature scheme by using integer factorization and discrete logarithm problem," in *Advances in Computing, Communications and Informatics (ICACCI), 2017 International Conference on*, pp. 1261–1266, IEEE, 2017.

[18] M. Yin, "Multimedia authentication for copyright protection," in *IOP conference series: earth and environmental science*, vol. 69, p. 012160, IOP Publishing, 2017.

[19] O. Tayan, M. N. Kabir, and Y. M. Alginahi, "A hybrid digital-signature and zero-watermarking approach for authentication and protection of sensitive electronic documents," *The Scientific World Journal*, vol. 2014, 2014.

[20] T. Lakshmanan and M. Muthusamy, “A novel secure hash algorithm for public key digital signature schemes.,” *Int. Arab J. Inf. Technol.*, vol. 9, no. 3, pp. 262–267, 2012.

[21] X. Liu and Y. Zha, “Copyright protection of digital movies using the coalition of technology and law in china,” *Chinese Studies*, vol. 7, no. 04, p. 259, 2018.

[22] J. S. Hendricks, M. L. Asmussen, and J. S. McCoskey, “Electronic book security and copyright protection system,” Mar. 10 2016. US Patent App. 14/845,106.

[23] P.-H. Vo, T.-S. Nguyen, V.-T. Huynh, and T.-N. Do, “A robust hybrid watermarking scheme based on dct and svd for copyright protection of stereo images,” in *Information and Computer Science, 2017 4th NAFOSTED Conference on*, pp. 331–335, IEEE, 2017.

[24] Y. Xiang, I. Natgunanathan, Y. Rong, and S. Guo, “Spread spectrum-based high embedding capacity watermarking method for audio signals,” *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 23, no. 12, pp. 2228–2237, 2015.

[25] S. Jindal and N. Kaur, “Digital image steganography survey and analysis of current methods,” *International Journal of Computer Science and Information Technology & Security*, vol. 6, pp. 10–13, 2016.

[26] T. Pevny, J. Fridrich, and A. D. Ker, “From blind to quantitative steganalysis,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 445–454, 2012.

[27] Z. Li, D. Gong, F. Liu, and A. G. Bors, “3d steganalysis using the extended local feature set,” in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 1683–1687, IEEE, 2018.

[28] Z. Xia, X. Wang, X. Sun, and B. Wang, “Steganalysis of least significant bit matching using multi-order differences,” *Security and Communication Networks*, vol. 7, no. 8, pp. 1283–1291, 2014.

[29] V. Holub and J. Fridrich, “Low-complexity features for jpeg steganalysis using undecimated dct,” *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 2, pp. 219–228, 2015.

[30] N. Meghanathan and L. Nayak, "Steganalysis algorithms for detecting the hidden information in image, audio and video cover media," *international journal of Network Security & Its application (IJNSA)*, vol. 2, no. 1, pp. 43–55, 2010.

[31] A. D. Ker, "Steganalysis of lsb matching in grayscale images," *IEEE signal processing letters*, vol. 12, no. 6, pp. 441–444, 2005.

[32] S. Lyu and H. Farid, "Steganalysis using higher-order image statistics," *IEEE transactions on Information Forensics and Security*, vol. 1, no. 1, pp. 111–119, 2006.

[33] D. Zou, Y. Q. Shi, W. Su, and G. Xuan, "Steganalysis based on markov model of thresholded prediction-error image," in *Multimedia and Expo, 2006 IEEE International Conference on*, pp. 1365–1368, IEEE, 2006.

[34] M. Saini and R. Chhikara, "Dwt feature based blind image steganalysis using neural network classifier," *International Journal of Engineering Research & Technology (IJERT)*, vol. 4, no. 04, pp. 776–782, 2015.

[35] K. Wang, J. Han, and H. Wang, "Digital video steganalysis by subtractive prediction error adjacency matrix," *Multimedia tools and applications*, vol. 72, no. 1, pp. 313–330, 2014.

[36] Y. Ren, M. Wang, Y. Zhao, L. Wang, and T. Cai, "Steganalysis of msu stego video based on block matching of interframe collusion and motion detection," *Wuhan University Journal of Natural Sciences*, vol. 17, no. 5, pp. 441–446, 2012.

[37] Y. QIN and B. XU, "Video steganalysis method based on temporal and spatial redundancies," *Optical Instruments*, vol. 5, p. 006, 2011.

[38] Y. Yang, R. Pintus, H. Rushmeier, and I. Ivrissimtzis, "A 3d steganalytic algorithm and steganalysis-resistant watermarking," *IEEE transactions on visualization and computer graphics*, vol. 23, no. 2, pp. 1002–1013, 2017.

[39] K. Pathak and M. Bansal, "A fpga based steganographic system implementing a modern steganalysis resistant lsb algorithm," *Defence Science Journal*, vol. 67, no. 5, p. 551, 2017.

[40] Y.-T. Lin, C.-M. Wang, W.-S. Chen, F.-P. Lin, and W. Lin, "A novel data hiding algorithm for high dynamic range images," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 196–211, 2017.

[41] S. F. Mare, M. Vladutiu, and L. Prodan, "Hdr based steganographic algorithm," in *2011 IEEE 17th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, pp. 333–338, IEEE, 2011.

[42] M.-T. Li, N.-C. Huang, and C.-M. Wang, "A data hiding scheme for high dynamic range images," *International Journal of Innovative Computing Information and Control (IJICIC)*, vol. 7, no. 5A, pp. 2021–2035, 2011.

[43] B. Martini and K.-K. R. Choo, "An integrated conceptual digital forensic framework for cloud computing," *Digital Investigation*, vol. 9, no. 2, pp. 71–80, 2012.

[44] H. Farid, "Image forgery detection," *IEEE Signal processing magazine*, vol. 26, no. 2, pp. 16–25, 2009.

[45] H. Farid, "Digital image ballistics from jpeg quantization: A followup study," *Department of Computer Science, Dartmouth College, Tech. Rep. TR2008-638*, vol. 7, pp. 1–28, 2008.

[46] 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*, pp. 404–407, IEEE, 2011.

[47] X. Kang, Y. Li, Z. Qu, J. Huang, *et al.*, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 393–402, 2012.

[48] J. Fridrich, "Digital image forensics," *IEEE Signal Processing Magazine*, vol. 26, no. 2, 2009.

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

[50] A. Rocha, W. Scheirer, T. Boult, and S. Goldenstein, "Vision of the unseen: Current trends and challenges in digital image and video forensics," *ACM Computing Surveys (CSUR)*, vol. 43, no. 4, p. 26, 2011.

[51] R. Köhler, M. Hirsch, B. Mohler, B. Schölkopf, and S. Harmeling, "Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database," in *European Conference on Computer Vision*, pp. 27–40, Springer, 2012.

[52] L. Xu, S. Zheng, and J. Jia, "Unnatural l0 sparse representation for natural image deblurring," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1107–1114, 2013.

[53] T. Grosch, "Fast and robust high dynamic range image generation with camera and object movement," *Vision, Modeling and Visualization, RWTH Aachen*, pp. 277–284, 2006.

[54] G. Cao, Y. Zhao, R. Ni, and H. Tian, "Anti-forensics of contrast enhancement in digital images," in *Proceedings of the 12th ACM Workshop on Multimedia and Security*, pp. 25–34, ACM, 2010.

[55] C. Rajalakshmi, M. G. Alex, and R. Balasubramanian, "Study of image tampering and review of tampering detection techniques," *International Journal of Advanced Research in Computer Science*, vol. 8, no. 7, 2017.

[56] A. Makrushin, C. Kraetzer, T. Neubert, and J. Dittmann, "Generalized benford's law for blind detection of morphed face images," in *Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security*, pp. 49–54, ACM, 2018.

[57] T. Kunkel, E. Reinhard, G. Damberg, and A. Ballestad, "Light detection, color appearance models, and modifying dynamic range for image display," July 9 2013. US Patent 8,483,479.

[58] T. Bianchi, A. De Rosa, and A. Piva, "Improved dct coefficient analysis for forgery localization in jpeg images," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pp. 2444–2447, IEEE, 2011.

[59] Y. Huang, W. Lu, W. Sun, and D. Long, “Improved dct-based detection of copy-move forgery in images,” *Forensic science international*, vol. 206, no. 1-3, pp. 178–184, 2011.

[60] J. Zhao and J. Guo, “Passive forensics for copy-move image forgery using a method based on dct and svd,” *Forensic science international*, vol. 233, no. 1-3, pp. 158–166, 2013.

[61] Z. He, W. Lu, W. Sun, and J. Huang, “Digital image splicing detection based on markov features in dct and dwt domain,” *Pattern Recognition*, vol. 45, no. 12, pp. 4292–4299, 2012.

[62] T. Mahmood, Z. Mehmood, M. Shah, and T. Saba, “A robust technique for copy-move forgery detection and localization in digital images via stationary wavelet and discrete cosine transform,” *Journal of Visual Communication and Image Representation*, vol. 53, pp. 202–214, 2018.

[63] I. A. Ansari, M. Pant, and C. W. Ahn, “Svd based fragile watermarking scheme for tamper localization and self-recovery,” *International Journal of Machine Learning and Cybernetics*, vol. 7, no. 6, pp. 1225–1239, 2016.

[64] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, “A sift-based forensic method for copy–move attack detection and transformation recovery,” *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099–1110, 2011.

[65] X. Bo, W. Junwen, L. Guangjie, and D. Yuewei, “Image copy-move forgery detection based on surf,” in *Multimedia information networking and security (MINES), 2010 international conference on*, pp. 889–892, IEEE, 2010.

[66] H. J. Kim, S. Lim, B. Kim, and E. S. Jung, “A new approach to photography forensics using 3d analysis for correcting perception errors: a case study,” in *Proceedings of the 2010 ACM workshop on Surreal media and virtual cloning*, pp. 27–30, ACM, 2010.

[67] J. Yin and Y. Fang, “Markov-based image forensics for photographic copying from printed picture,” in *Proceedings of the 20th ACM international conference on Multimedia*, pp. 1113–1116, ACM, 2012.

[68] A. De Rosa, F. Uccheddu, A. Costanzo, A. Piva, and M. Barni, "Exploring image dependencies: a new challenge in image forensics," in *Media Forensics and Security II*, vol. 7541, p. 75410X, International Society for Optics and Photonics, 2010.

[69] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," in *Proceedings of the 9th workshop on Multimedia & security*, pp. 35–42, ACM, 2007.

[70] A. Popescu and H. Farid, "Exposing digital forgeries by detecting duplicated image regions. department computer science, dartmouth college, technology report tr2004-515," 2004.

[71] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting double mpeg compression," in *Proceedings of the 8th workshop on Multimedia and security*, pp. 37–47, ACM, 2006.

[72] T. Shanableh, "Detection of frame deletion for digital video forensics," *Digital Investigation*, vol. 10, no. 4, pp. 350–360, 2013.

[73] C. Feng, Z. Xu, W. Zhang, and Y. Xu, "Automatic location of frame deletion point for digital video forensics," in *Proceedings of the 2nd ACM workshop on Information hiding and multimedia security*, pp. 171–179, ACM, 2014.

[74] A. Gironi, M. Fontani, T. Bianchi, A. Piva, and M. Barni, "A video forensic technique for detecting frame deletion and insertion.," in *ICASSP*, pp. 6226–6230, 2014.

[75] Q. Dong, G. Yang, and N. Zhu, "A mcea based passive forensics scheme for detecting frame-based video tampering," *Digital Investigation*, vol. 9, no. 2, pp. 151–159, 2012.

[76] M. J. Sorrell, "Digital camera source identification through jpeg quantisation," in *Multimedia forensics and security*, pp. 291–313, IGI Global, 2009.

[77] S. Bayram, H. T. Sencar, and N. Memon, "Efficient sensor fingerprint matching through fingerprint binarization," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 4, pp. 1404–1413, 2012.

[78] P. J. Bateman, A. T. Ho, and J. A. Briffa, "Image forensics of high dynamic range imaging," in *International Workshop on Digital Watermarking*, pp. 336–348, Springer, 2011.

[79] I. Yerushalmy and H. Hel-Or, "Digital image forgery detection based on lens and sensor aberration," *International journal of computer vision*, vol. 92, no. 1, pp. 71–91, 2011.

[80] C.-T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, 2010.

[81] S. Agarwal and S. Chand, "Image tampering detection using local phase based operator," in *Emerging Trends in Electrical Electronics & Sustainable Energy Systems (ICETEESES), International Conference on*, pp. 355–360, IEEE, 2016.

[82] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Correlation clustering for prnu-based blind image source identification," in *Information Forensics and Security (WIFS), 2016 IEEE International Workshop on*, pp. 1–6, IEEE, 2016.

[83] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Blind prnu-based image clustering for source identification," *IEEE Trans. Inf. Forensics Secur*, vol. 12, no. 9, pp. 2197–2211, 2017.

[84] A. Lawgaly and F. Khelifi, "Sensor pattern noise estimation based on improved locally adaptive dct filtering and weighted averaging for source camera identification and verification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 2, pp. 392–404, 2017.

[85] I. Amerini, R. Caldelli, A. Del Mastio, A. Di Fuccia, C. Molinari, and A. P. Rizzo, "Dealing with video source identification in social networks," *Signal Processing: Image Communication*, vol. 57, pp. 1–7, 2017.

[86] W.-H. Chuang, H. Su, and M. Wu, "Exploring compression effects for improved source camera identification using strongly compressed video," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pp. 1953–1956, IEEE, 2011.

[87] B. Bayar and M. C. Stamm, "Towards open set camera model identification using a deep learning framework," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2007–2011, IEEE, 2018.

[88] L. Baroffio, L. Bondi, P. Bestagini, and S. Tubaro, "Camera identification with deep convolutional networks," *arXiv preprint arXiv:1603.01068*, 2016.

[89] D. Freire-Obregón, F. Narducci, S. Barra, and M. Castrillón-Santana, "Deep learning for source camera identification on mobile devices," *Pattern Recognition Letters*, 2018.

[90] I. Amerini, T. Uricchio, and R. Caldelli, "Tracing images back to their social network of origin: A cnn-based approach," in *Information Forensics and Security (WIFS), 2017 IEEE Workshop on*, pp. 1–6, IEEE, 2017.

[91] R. Caldelli, R. Becarelli, and I. Amerini, "Image origin classification based on social network provenance," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 6, pp. 1299–1308, 2017.

[92] R. Caldelli, I. Amerini, and C. T. Li, "Prnu-based image classification of origin social network with cnn," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1357–1361, IEEE, 2018.

[93] M. Klíma, K. Fliegel, P. Pata, S. Vitek, M. Blažek, P. Dostal, L. Krasula, T. Kratochvíl, V. Rícnỳ, M. Slanina, *et al.*, "Deimos–an open source image database.," *Radioengineering*, vol. 20, no. 4, 2011.

[94] T. Gloe and R. Böhme, "The'dresden image database'for benchmarking digital image forensics," in *Proceedings of the 2010 ACM Symposium on Applied Computing*, pp. 1584–1590, ACM, 2010.

[95] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, "Hmdb: a large video database for human motion recognition," in *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 2556–2563, IEEE, 2011.

[96] M. D. Fairchild, "The hdr photographic survey," in *Color and Imaging Conference*, vol. 2007, pp. 233–238, Society for Imaging Science and Technology, 2007.

[97] P. Korshunov, P. Hanhart, T. Richter, A. Artusi, R. Mantiuk, and T. Ebrahimi, "Subjective quality assessment database of hdr images compressed with jpeg xt," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*, pp. 1–6, IEEE, 2015.

[98] B. Funt and L. Shi, "The rehabilitation of maxrgb," in *Color and imaging conference*, vol. 2010, pp. 256–259, Society for Imaging Science and Technology, 2010.

[99] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "Raise: a raw images dataset for digital image forensics," in *Proceedings of the 6th ACM Multimedia Systems Conference*, pp. 219–224, ACM, 2015.

[100] "Digital media lab high dynamic range (dml-hdr) video dataset created at the university of british columbia." `http://dml.ece.ubc.ca`.

[101] M. Azimi, A. Banitalebi-Dehkordi, Y. Dong, M. T. Pourazad, and P. Nasiopoulos, "Evaluating the performance of existing full-reference quality metrics on high dynamic range (hdr) video content," *arXiv preprint arXiv:1803.04815*, 2018.

[102] J. Froehlich, S. Grandinetti, B. Eberhardt, S. Walter, A. Schilling, and H. Brendel, "Creating cinematic wide gamut hdr-video for the evaluation of tone mapping operators and hdr-displays," in *Digital Photography X*, vol. 9023, p. 90230X, International Society for Optics and Photonics, 2014.

[103] 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.

[104] Z. Yu, R. Abma, J. Etgen, and C. Sullivan, "Attenuation of noise and simultaneous source interference using wavelet denoising," *Geophysics*, vol. 82, no. 3, pp. V179–V190, 2017.

[105] M. Srivastava, C. L. Anderson, and J. H. Freed, "A new wavelet denoising method for selecting decomposition levels and noise thresholds," *IEEE Access*, vol. 4, pp. 3862–3877, 2016.

[106] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," in *Media Forensics and Security*,

vol. 7254, p. 72540I, International Society for Optics and Photonics, 2009.

[107] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, vol. 6819, p. 68190E, International Society for Optics and Photonics, 2008.

[108] C.-H. Wei, *Modern Library Technologies for Data Storage, Retrieval, and Use.* IGI Global, 2013.

[109] D.-K. Hyun, S.-J. Ryu, M.-J. Lee, J.-H. Lee, H.-Y. Lee, and H.-K. Lee, "Source camcorder identification from cropped and scaled videos," in *Media Watermarking, Security, and Forensics 2012*, vol. 8303, p. 83030E, International Society for Optics and Photonics, 2012.

[110] C. Whitelam, E. Taborsky, A. Blanton, B. Maze, J. C. Adams, T. Miller, N. D. Kalka, A. K. Jain, J. A. Duncan, K. Allen, *et al.*, "Iarpa janus benchmark-b face dataset.," in *CVPR Workshops*, vol. 2, p. 6, 2017.

[111] M. Moltisanti, A. Paratore, S. Battiato, and L. Saravo, "Image manipulation on facebook for forensics evidence," in *International Conference on Image Analysis and Processing*, pp. 506–517, Springer, 2015.

[112] M. Iuliani, M. Fontani, D. Shullani, and A. Piva, "A hybrid approach to video source identification," *arXiv preprint arXiv:1705.01854*, 2017.

[113] M. Goljan, M. Chen, and J. Fridrich, "Identifying common source digital camera from image pairs," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 6, pp. VI–125, IEEE, 2007.

[114] H. T. Sencar and N. Memon, "Digital image forensics: There is more to a picture than meets the eye," *Counter-Forensics: Attacking Image Forensics*, pp. 327–366, 2013.

[115] G. Besnard, F. Hild, and S. Roux, "âfinite-elementâ displacement fields analysis from digital images: application to portevin–le châtelier bands," *Experimental Mechanics*, vol. 46, no. 6, pp. 789–803, 2006.

[116] M. Calì, S. M. Oliveri, R. Ambu, and G. Fichera, “An integrated approach to characterize the dynamic behaviour of a mechanical chain tensioner by functional tolerancing.,” *Strojniski Vestnik/Journal of Mechanical Engineering*, vol. 64, no. 4, 2018.