Expert Systems With Applications 55 (2016) 59–69

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# Identification of smartphone brand and model via forensic video analysis

Luis Javier García Villalba[a,*], Ana Lucila Sandoval Orozco[a], Raquel Ramos López[a], Julio Hernandez Castro[b]

[a] Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), Faculty of Computer Science and Engineering, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases, 9, Ciudad Universitaria, Madrid 28040, Spain
[b] School of Computing, Office S129A, University of Kent, Cornwallis South Building, Canterbury CT2 7NF, UK

ARTICLE INFO

*Keywords:*
Key frame extraction
Video forensics analysis
Video source acquisition
Sensor pattern noise
PRNU
Smartphone

ABSTRACT

Recording videos on smartphones and other mobile devices, given their enormous popularity, is currently very common. The portability of these devices facilitates their use for recording videos in a wide variety of situations, including while witnessing criminal activities. These videos can be later used as evidence in legal proceedings. Therefore, the forensic analysis of videos taken with mobile device videos is important, and could serve for legal and also investigative purposes. It is necessary, however, to use techniques that are quite specific to this type of devices, given some peculiar features of their cameras. In this paper, we will address the issue of video source acquisition identification by presenting a technique based on sensor noise and wavelet transform extraction from video key frames. These frames are extracted using an efficient algorithm that takes their content into account, improving the selection of frames to be analyzed over past proposals. The scheme presented consists of four stages: (1) Key frames extraction, (2) sensor pattern noise extraction, (3) feature extraction, and (4) classifier training and prediction. We also present experimental results that support the validity of the techniques used and show promising results.



## 1. Introduction

Images captured by electronic devices (i.e. smartphones) are often considered part of evidence in Court, and in a few minutes a video can communicate an enormous amount of information. According to the traffic meter "Alexa, The Web Information Company" (Alexa Internet, Inc., 2014), YouTube is currently the third most visited website in the world, which gives us a clear indication of the online popularity of videos. Video is widely used in everyday life due to the availability of a wide range of mobile devices that can reproduce and/or record it, such as mobile phones, tablets, portable game consoles and digital cameras or camcorders. As for mobile devices, Gartner Inc. Gartner (2014), states that sales of smartphones grew by 36% in the fourth quarter of 2013, and represented 57.6% of the global sales of mobile phones in the fourth quarter, compared to 44% with respect to 2012.

As digital cameras have swept away traditional film cameras in terms of popularity, nowadays mobile devices equipped with cameras have an important role in putting an end to the rapid growth that digital cameras previously experimented. A report by IC Insights (IC Insights, 2014) predicted that by 2016 the market rate of DSCs (Digital Still Camera) will drop from 47% in 2012 to 27%; it also predicts a rise in sales of digital cameras built into smartphones, PCs and tablets, from 31% in 2012 to 42% by 2016.

There has been fierce competition lately between mobile device manufacturers, who strive to integrate higher definition video cameras with every new model, and make them available to the user at all times. A clear example was the presentation of the Galaxy S5 at the Mobile World Congress 2014 held in Barcelona, with the ability to record Ultra HD (4K) video. As a result of this, smartphones have become the first video recording equipment of choice for many of us.

Due to the frequent use of mobile devices, in some cases there exist legal restrictions or limitations to their use in various locations, such as schools, universities, government offices, companies, etc. In parallel, videos are increasingly used, either directly or indirectly, in legal proceedings as evidence for law enforcement (Wen & Yang, 2006). This is despite the fact that manipulation of digital video is becoming increasingly easier, because of the emergence of

* Corresponding author. Tel.: +34 91 394 76 38; fax: +34 91 394 75 47.
*E-mail addresses:* javiergv@fdi.ucm.es (L.J. García Villalba), asandoval@fdi.ucm.es (A.L. Sandoval Orozco), raqram01@ucm.es (R. Ramos López), J.C.Hernandez-Castro@kent.ac.uk (J. Hernandez Castro).
*URL:* http://gass.ucm.es/people/javier/ (L.J. García Villalba)

http://dx.doi.org/10.1016/j.eswa.2016.01.025
0957-4174/

new powerful multimedia processing tools. This is unfortunately even possible for those without a great deal of expertise or experience. A clear example of how easy is to edit video was presented in Jokela, Mäkelä, and Karukka (2007).

Therefore, given the increasing importance of video, digital video forensics are particularly relevant. Their main goal is the acquisition and analysis of digital video in order to find forensically sound evidence, generally while investigating a crime. Within this discipline, *Digital Video Integrity* aims to establish whether a digital video has been tampered with, *Digital Video Steganography* studies if a video contains hidden data and *Video Source Camera Identification* aims to identify which specific camera has been used to capture a video. *Video Source Camera Identification* has many applications in real world scenarios, and its study is especially important and becoming more relevant with every passing day. For example, when a video is presented as evidence in a court of law, identifying the acquisition device of the video could be as important as the video itself. Not doing this in a forensically sound way can lead to legal challenges and render the evidence invalid (Brown, 2014). Additionally, images or videos shared through social networks (Flickr, Instagram, Facebook, Twitter, etc.) or personal email can be authenticated and linked to the device (in this case, the smartphone or digital camera).

Research in this field studies techniques to identify both the maker and model of the devices used to generate digital videos. It is analogous to ballistics, that try to relate a gun with its bullets, in that it tries to identify the link between videos and the digital camera which has generated them (Wang, Guo, Kong, & Meng, 2009).

This paper presents a combination of forensic analysis techniques for the identification of a video source device, but focusing on videos generated by mobile devices, mostly smartphones.

The paper is divided into six sections, the first being this introduction. Section 2 presents the differences between the pipeline in the creation of an image and a video. Section 3 introduces a state of the art for the forensic analysis of images and videos, regarding the issue of source acquisition identification. The proposed technique is presented in detail in Section 4. The supporting experiments are presented in Section 5. Finally, Section 6 shows the conclusions drawn from this work.

## 2. Video capturing process

It is important to understand the basics of the procedure employed by digital cameras to generate an image; this is shown schematically in Fig. 1. The process is similar for the generation of a video, although in video there is an additional stage in which the sequence of images is encoded over time.

Firstly, the lens system captures light from the scene by controlling the exposure, focus, and image stabilization. Next, the light passes through a set of filters that improve the visual quality of the image, and then the light gets to the image sensor called Color Filter Array; this is an array of light sensitive elements called pixels. Note that the choice of the CFA can greatly influence the sharpness and the final appearance of the image since there are quite different CFA patterns. The most commonly used model is the Green-Red-Green-Blue (GRGB) Bayer pattern; other models are: Red-Green-Blue-Emerald (RGBE), Cyan-Yellow-Yellow-Magenta (CYYM), Cyan-Yellow-Green-Magenta (CYGM) or Red-Green-Blue-White (RGBW). The incident light on the colored filters gets to a sensor which is responsible for generating an analogue signal proportional to the intensity of received light, keeping these values in an internal array. There are currently two types of sensor technologies that meet this latter purpose in digital cameras: CCD (Charge Coupled Device) and CMOS (Complementary Metal Oxide Semiconductor). Both types of sensors essentially consist of Metal Oxide Semiconductors (MOS) and they work in a similar way, although the key difference is in the way in which pixels are scanned and the way in which the reading of the charges is performed. CCD sensors need an additional chip to process the sensor's output information; this causes the manufacture of devices to be more costly and the sensors to be bigger. In contrast, CMOS sensors have independent active pixels and, as they are able to perform the digitalization themselves, offer extra speed and a reduction on size and cost. Another difference between these two types of sensors is that pixels in a CCD array capture light simultaneously, which generates a more uniform output. CMOS sensors generally perform the reading as progressive scan (avoiding the blooming effect). CCD sensors are far superior to CMOS in terms of noise reduction and dynamic range; on the other hand, CMOS sensors are more sensitive to light and generally behave better in low light conditions. Early CMOS sensors were somewhat worse than CCDs, but nowadays this gap has been practically corrected. CCD technology has reached its limit, and it is now when CMOS is being developed and improved so that its weaknesses are being slowly solved, so much so that the majority of smartphones in the market use CMOS sensors. Data stored by the CCD/CMOS sensor are then converted into a digital signal and transmitted to the image processor. Once the image processor receives the digital signal, it eliminates noise and other anomalies. Some other processes applied to the signal are color interpolation, gamma correction, and color correction.

Only in the case of video generation, there is an additional final step that encodes the resulting frames to create a final single video file. This encoding aims to transform all the frames captured into a sequence over a period of time. It also seeks to achieve the most optimal final file size, since in a video there are typically many captured frames that are redundant. That is, sometimes it is possible to share scene features from one frame to another that facilitate the reduction of the final video size without losing visual contents. For example, in MPEG encoding, there is a structure called GOP (Group of Pictures) which specifies the order in which images are sorted and solves the problem of redundancy encoding. A GOP can contain different types of images: type I, B and P. I (intra-coded) images are reference images that represent fixed contents that are independent of other image types. P (prediction encoding) images contain information about motion compensation of the previous image, either type P or I. B (bi-directional prediction encoding) images contain different information from both the previous and the next picture. A GOP always starts with a type I picture, followed by several type P images. The remaining gaps are

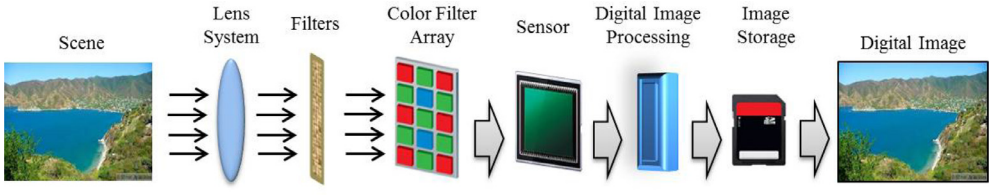


**Fig. 1.** Image acquisition process in digital cameras.