Expert Systems with Applications 42 (2015) 1927–1940

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Smartphone image clustering

Luis Javier García Villalba *, Ana Lucila Sandoval Orozco, Jocelin Rosales Corripio

Group of Analysis, Security and Systems (GASS), Department of Software Engineering and Artificial Intelligence (DISIA), School of Computer Science, Office 431, Universidad Complutense de Madrid (UCM), Calle Profesor José García Santesmases s/n, Ciudad Universitaria, 28040 Madrid, Spain

A R T I C L E I N F O

Article history:
Available online 25 October 2014

Keywords:
Image clustering
Image forensics analysis
PRNU
Sensor Pattern Noise

A B S T R A C T

Every day the use of images from mobile devices as evidence in legal proceedings is more usual and common. Therefore, forensic analysis of mobile device images takes on special importance. This paper explores the branch of forensic analysis which is based on the identification of the source, specifically on the grouping or clustering of images according to their source acquisition. In contrast with other state of the art techniques for source identification, hierarchical clustering does not involve a priori knowledge of the number of images or devices to be identified or training data for a future classification stage. That is, a grouping by classes with all the input images is performed. The proposal is based on the combination of hierarchical and flat clustering and the use of *Sensor Pattern Noise* (SPN). There has been a series of experiments which emulate similar situations to those that may occur in reality to test the robustness and reliability of the results of the technique. The results are satisfactory in all the experiments, obtaining high rates of success.

1. Introduction

At present, the number of cameras integrated into mobile devices has proliferated, allowing millions of consumers to take photographs and even easily share captured content. The mobile industry has developed the technology to reduce costs and thus make them very accessible to the public.

The accessibility and easy use of mobile cameras has the consequence that a large number of photos are taken with them generating more evidence presented before the law on crimes such as credit card fraud, child pornography, industrial espionage, public safety, street violence, etc. Therefore, forensic analysis of such images is particularly important in criminal investigations. There are two main branches within digital image forensic analysis: image source acquisition identification and malicious tampering detection. This work focuses on the first branch. Also, since mobile device cameras have some characteristics that make them different from the rest, this work focuses on images from this type of devices.

There are two major approaches regarding source acquisition identification: closed scenarios and open scenarios. A closed scenario is one in which the image source identification is performed on a specific and known beforehand set of cameras. For this approach a set of images from each camera is normally used to train a classifier and later the image source acquisition under investigation is predicted. The most commonly used technique for the digital imaging classification task is *Support Vector Machine* (SVM), although there are other options, such as the use of neural networks. This work focuses on image source acquisition identification in open scenarios, i.e., the forensic analyst does not know a priori the camera set to which images whose source identification will be identified belong. Obviously, in this type of classification in which data from cameras are not known beforehand, the objective is not to identify the make and model of the images, but to be able to group the different images into disjoint sets in which all their images belong to the same device. This approach is very close to real-life situations, since in many cases the set of cameras to which a set of images may belong is completely unknown to the analyst. In addition, it is virtually impossible to have a set of images to train a classifier with all mobile device cameras existing in the world. In this case, being able to group images into sets that belong to the same device is very useful, as this can provide very valuable and in some cases conclusive information to judicial investigators.

In this paper a clustering algorithm based on Caldelli, Amerini, Picchioni, and Innocenti (2010) is proposed. As elements for classification we use a set of features obtained from SPN noise. Broadly speaking, the main difference is that our proposal takes into account the evolutionary process of cluster formation when

* 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), jocerosa@ucm.es (J.R. Corripio).
URL: http://gass.ucm.es/en/people/javier/ (L.J. García Villalba).

http://dx.doi.org/10.1016/j.eswa.2014.10.018
0957-4174/

calculating the coefficient that determines the cohesion between the elements of the same cluster and separation between different clusters that are being generated.

This work is divided into five sections, the first being this introduction. Section 2 briefly presents previous work related to forensic techniques for mobile device image source acquisition identification. The proposed technique is presented in Section 3. The experiments and their results are presented in Section 4. Finally, in Section 5 the conclusions drawn from this work are presented.

2. Related works

Most research on image source acquisition identification focuses on traditional digital cameras or *Digital Still Camera* (DSC); most of these techniques are not valid for mobile device images. The main reason is that most of the techniques are based on directly or indirectly use of sensor features or in the lens of the digital camera. Regarding the sensor, it is the component that is responsible for capturing the light and generate a digital signal according to its intensity. There are currently two types of sensor technologies that meet this latter purpose in digital cameras: glsCCD y *Complementary Metal Oxide Semiconductor* (CMOS). Both types of sensors essentially consist of metal-oxide (*Metal Oxide Semiconductor* (MOS)) distributed in a matrix and they work in a similar way. However the key difference is in the way in which pixels are scanned and the way in which the reading of the charges is carried out. *Charge Coupled Device* (CCD) sensors need an additional chip to process the sensor 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, as they themselves perform the digitalization, offering speed and reducing the size and cost of the systems that make up a digital camera. Another difference between these two types of sensors is that the pixels in a CCD array capture light simultaneously, which promotes a more uniform output. CMOS sensors generally perform the reading as progressive scan (avoiding the *blooming* effect). CCD sensors are far superior to the CMOS in terms of noise and dynamic range; on the other hand, CMOS sensors are more sensitive to light and behave better in low light conditions. Early CMOS sensors were somewhat worse than CCDs, but nowadays this has been practically corrected. The CCD technology has reached its limit and nowadays the CMOS technology is developing and gradually overcoming their shortcomings. Most of DSCs use CCD sensors, in mobile devices is more common to use sensors CMOS. Even day by day, reducing the quality differences between CCD and CMOS sensors, in the great majority of cases DSCs sensors notably exceed in quality to sensors in mobile devices digital cameras, and this is a strong reason to require specific techniques for image source acquisition source. Likewise to the case of sensor, mobile devices digital camera lenses, in general, are lower of quality than DSCs lenses.

For any type of image classification, either in open or closed scenarios, it is necessary to obtain certain features that allow classification techniques to perform their task. According to Van Lanh, Chong, Emmanuel, and Kankanhalli (2007), four groups of techniques can be established for this purpose: based on lens aberration (Choi, 2006; Choi, Lam, & Wong, 2006; Van, Emmanuel, & Kankanhalli, 2007), based on the *Color Filter Array* (CFA) matrix interpolation (Bayram, Sencar, & Memon, 2006, 2008; Long & Huang, 2006), based on the sensor imperfections (Chen, Fridrich, Goljan, & Lukás, 2008; Costa, Eckmann, Scheirer, & Rocha, 2012; Kang, Li, Qu, & Huang, 2012; Lukas, Fridrich, & Goljan, 2006) and based on the use of image features (Hu, Li, & Zhou, 2010; Mckay, Swaminathan, Gou, & Wu, 2008; Meng, Kong, & You, 2008; Liu et al., 2012; Ozparlak & Avcibas, 2011). Within the latter group a subdivision can be made based on color features, quality features, and wavelet domain statistics. In Sandoval Orozco, Arenas González, Rosales Corripio, García Villalbas, and Hernandez-Castro (2013) an overview of this research can be seen.

This work uses techniques based on sensor imperfections, particularly those based on the SPN. The main components of image noise are the *Fixed Pattern Noise* (FPN) and the *Photo Response Non Uniformity* (PRNU). There are several sources of imperfections and noise introduced at different stages of the creating pipeline of an image in a digital camera. Even if a uniform and fully lighted picture is taken it is possible to see small changes in the intensity between pixels. This is due to the shot noise is random and, in large part, the pattern noise is deterministic and is kept approximately equal if several pictures of the same scene are taken.

The noise pattern of an image refers to any spatial pattern that does not change from one image to another. It is composed for the spatial noise which is independent of the signal (FPN) and for the spatial noise due to the difference in the response of each pixel to the incident signal (PRNU). The noise pattern structure is shown in Fig. 1.

Noise FPN is generated by the dark current and it also depends on exposure and temperature. Since the FPN is an independent additive noise, some cameras automatically removed by subtracting a dark frame to generated images.

Noise PRNU is the dominant part of the Sensor Pattern Noise of an image and it is a multiplicative noise dependent. Noise PRNU is mainly formed by noise *Pixel Non-Uniformity* (PNU) and by the low frequency defects as *zoom* settings and light refraction in the dust particles and lenses. Noise PNU is the light sensitivity difference between pixels of the sensor array. It is generated by the lack of homogeneity of the silicon wafers and by the imperfections during the sensor manufacturing process. Due to the nature and origin, it is very unlikely that even the sensors from the same wafer have PNU correlated patterns. This noise is not affected by ambient temperature nor by humidity. Noise PNU is usually more common, complex and significant in CMOS sensors, due to the complexity of pixel array circuitry.

Once you have the features to be used for classification of images we will focus on issues relating to the classification by clustering. The analysis of clusters, or clustering, aims to group a collection of objects into representative classes called clusters, without a priori information, in such a way that the objects belonging to each cluster keep a greater similarity to objects from other clusters.

Image grouping can be performed using supervised or unsupervised learning techniques. In the first case it is essential to know the device information a priori, i.e., it is clearly identified with the classification in closed scenarios which requires a training stage with the features extracted from the images and a second

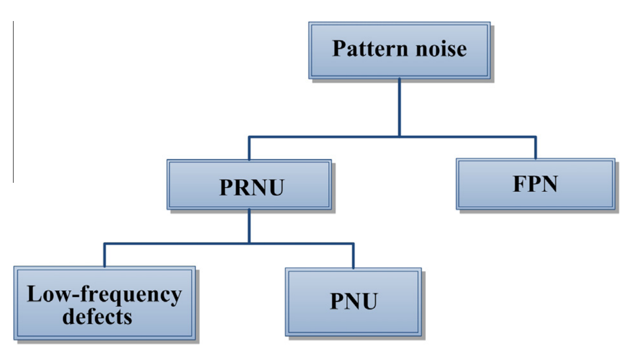

Fig. 1. Sensor Pattern Noise.