International Journal of Information Security and Privacy
Volume 12 • Issue 3 • July-September 2018

Blind Image Source Device Identification: Practicality and Challenges

Udaya Sameer Venkata, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

Ruchira Naskar, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India

ABSTRACT

This article describes how digital forensic techniques for source investigation and identification enable forensic analysts to map an image under question to its source device, in a completely blind way, with no a-priori information about the storage and processing. Such techniques operate based on blind image fingerprinting or machine learning based modelling using appropriate image features. Although researchers till date have succeeded to achieve extremely high accuracy, more than 99% with 10-12 candidate cameras, as far as source device prediction is concerned, the practical application of the existing techniques is still doubtful. This is due to the existence of some critical open challenges in this domain, such as exact device linking, open-set challenge, classifier overfitting and counter forensics. In this article, the authors identify those open challenges, with an insight into possible solution strategies.

KEYWORDS

Counter-Forensics, Digital Forensics, Exact Device Linking, Open-Set Challenge, Overfitting, Source Camera Identification

INTRODUCTION

Images play a major role in domains such as the legal industry, by acting as the primary sources of evidence towards any event in the court of law. Under such circumstances, it becomes crucial to verify the authenticity of images before their usage. Traditional techniques for protection and verification of the integrity of digital images, such as digital watermarking and steganography, fall under the purview of active protection mechanisms. Such techniques rely on data pre-processing in some form or the other, such as watermark computation, data embedding etc. On the contrary, the rapidly evolving domain of digital forensics provides image security and authentication measures which are completely post--processing based, hence called passive techniques. Image forensics in particular deals majorly with two problems namely Copy Move Forgery detection which is identifying forgeries in images, Source Camera Identification which is identifying the source camera which has captured the image under question.

In this paper, we delve into one of the most important image forensic problem in today's date, known as the Source Camera Identification (SCI) problem. The problem is to map a suspect's camera to an illegal image repository like child--pornography, to settle copy--right cases, to ascertain the validity and authenticity of whistle-blower information and many other sensitive scenarios.

To identify the source camera of an image, the Meta Data which stores the camera information in an image, in the image headers, can be well-exploited. However, wide availability of efficient image

DOI: 10.4018/IJISP.2018070105

editing tools today; it would take minimal effort to modify such headers. Thus, such pre-processed information added to the images cannot be trusted or treated as reliable. This makes the forensic expert to completely rely on post--processed information to evaluate the authenticity of images. The image processing pipeline in any digital camera adds information particular to the imaging sensor, due to the hardware and software artefacts.

As the source camera identification techniques are used to provide legal evidences, the false alarm rate in this problem domain has to be kept minimal, given its sensitive context. With more than 99% accuracy achieved by recent researchers in source camera identification, there lie a number of critical open challenges which would hinder the practical usage of such techniques in most real--life contexts. In this paper, we present the underlying challenges in this domain, and provide insights into possible solution strategies of each.

The rest of the paper is organized as follows. In the next section, we present an overview of the existing source attribution techniques. In the third section, we present and discuss the pragmatic challenges in source identification, along with individual possible solution strategies. In the fourth section, we propose a generalized solution strategy to overcome the challenges faced in forensic investigation of image sources. Finally, we conclude with related future research directions.

RELATED WORK

Source camera identification has been solved following two primary approaches. First, using camera fingerprints (Lukas, 2006), and second, through machine learning based model (Kharrazi et al., 2004). In the camera fingerprinting based techniques, Photo Response Non-Uniformity (PRNU) (Lukas, 2006) noise, a unique fingerprint formed on the camera's sensor while an image is captured, acts as the primary attribute to map an image to its source. Every camera manufacturer uses different sensors for different devices. The photo--electronic conversion of incident light to digital form, generates a noise at each pixel location of the sensor, hence producing a noise pattern, completely unique to the underlying sensor and thus the camera device.

Sensor Fingerprint Based Techniques

To extract the camera's fingerprint, also called the Sensor Pattern Noise (SPN), the PRNU noise of many images taken by the camera is averaged. The forensic expert having physical access to a finite number of cameras extracts the sensor pattern noises of each camera and stores those. To map an unknown test image to one of those finite cameras, PRNU of the image is extracted and a correlation-based mechanism is employed against the available sensor pattern noises. Depending on the correlation values, the forensic expert can determine the possible source of the image.

A digital camera imaging output can be written as:

$$P_x = P_0 + (P_0 F + \phi_1) \tag{1}$$

where P_x is the image output, P_0 is the amount of incident light, F is the PRNU factor and ϕ_1 is the collection of other noises such as dark current, shot noise etc. The Noise Residual or the PRNU component of a single (i^{th}) image I_i can be calculated as:

$$PRNU_i = P_x^i - DF(P_x^i) \tag{2}$$

where, the original image P_x^i is passed through a *Denoising Filter* (DF). The denoised image is then subtracted from the original image to generate the noise residual $PRNU_i$. The Sensor Pattern Noise (SPN) of a camera model C_j can then be calculated as: