Videos versus still images: Asymmetric sensor pattern noise comparison on mobile phones

Chiara Galdi, Frank Hartung, and Jean-Luc Dugelay; Eurecom (France) and FH Aachen (Germany)

Abstract

Nowadays, the most employed devices for recoding videos or capturing images are undoubtedly the smartphones. Our work investigates the application of source camera identification on mobile phones. We present a dataset entirely collected by mobile phones. The dataset contains both still images and videos collected by 67 different smartphones. Part of the images consists in photos of uniform backgrounds, especially collected for the computation of the RSPN. Identifying the source camera given a video is particularly challenging due to the strong video compression. The experiments reported in this paper, show the large variation in performance when testing an highly accurate technique on still images and videos.

Introduction

Source camera identification is one of the most important topics in Image Forensics, considering that it can be applied for associating videos or still images with illegal content to the source camera and possibly to its owner. Nowadays, the most employed devices for recoding videos or capturing images are undoubtedly the smartphones. However, the large variety of imaging sensor and software with very different characteristics (e.g. resolution, image pre-processing, and file format) makes the source camera identification on mobiles very challenging, in particular when dealing with videos subject to strong compression. Our study focus on the source camera identification issue on mobile devices and in analysing the variations in performances when applied on videos and when comparing video versus still images.

It is worth disambiguating between two main categories of source camera identification techniques. They are both based on the analysis of the traces left by the different processing steps in the image acquisition and storage phases. These traces mark the image with some kind of camera fingerprint, which can be used for authentication [1]. The first group of techniques tries to distinguish between different camera models by analysing acquisition artefacts produced by lenses or Color Filter Array (CFA) interpolation. The second, on a more challenging level, aims to distinguish between single devices, even different exemplars of the same camera model. The latter technique is based on the distinctive pattern due to imperfections in the silicon wafer during the sensor manufacturing.

We adopt a well known technique for Sensor Patter Noise extraction, belonging to the second category described above, namely the Enhanced Sensor Patter Noise technique presented by Li in 2010 [10]. The most important aspect of our work is the asymmetric comparison of the SPN extracted from videos and still images. It is known that videos captured by mobile phone are strongly compressed and this has a severe impact on the SPN extraction. The results show the great gap in performances when using videos in place of still images for source camera identification.

Experiments are carried out on a large image database especially collected for source camera identification on mobile devices. Performances are assessed in terms on Receiver Operating Characteristic (ROC) curve, Comulative Match Characteristic (CMC) curve, and Equal Error Rate (EER).

Related Works

As stated before, we adopt a technique, namely the Enhanced Sensor Patter Noise extraction, presented in [10] by Li. This technique is based on the observation that imaging sensors have various defects that produce a noise pattern in the pixel values [13]. The sensor noise is the result of three main components, that are the pixel defects, the fixed pattern noise (FPN), and the Photo Response Non Uniformity (PRNU).

Geradts et al in [14] attempt at reconstructing pixel defects patterns by taking images with 12 black or green background with 12 different cameras. The defect points are then compared showing that each camera has distinct patterns also across the same model. However, not all camera models contain any defective pixels and some cameras eliminate them. Therefore, this method is not applicable to every digital camera [1].

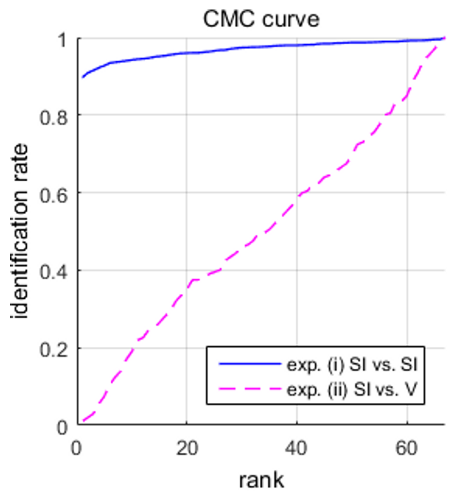
FPN and PNRU are the two components of the so-called pattern noise, and depend on dark currents in the sensor and pixel non-uniformities, respectively [1]. In [6], Lukas et al. propose to analyse the sensor pattern noise (SPN) for camera identification, as it is a unique stochastic characteristic for both CCD and CMOS sensors [1]. They show that the SPN extracted from images taken by the same camera is more correlated than those extracted from different cameras.

The SPN is estimated by computing the difference between an image I and its denoised version:

$$n = DWT(I) - F(DWT(I)) \quad (1)$$

where $DWT()$ is the discrete wavelet transform to be applied on image I and $F()$ is a denoising function applied in the DWT domain. $F()$, is a filter proposed in appendix A of [6].

In a later study, Li [10] proposes to refine the previous method by enhancing the SPN. Li observed that the SPN can be contaminated by fine details or structures of the depicted scene, since both the image noise and details are located in high frequencies. This deviation might reduce the probabilities of matching with a reference. Li proposes to enhance the SPN estimation by weighting noise components in a way inversely proportional to their magnitude, in order to suppress information derived from non-smooth image parts. As a result, high classification accuracy is obtained also on small-sized image regions [1].



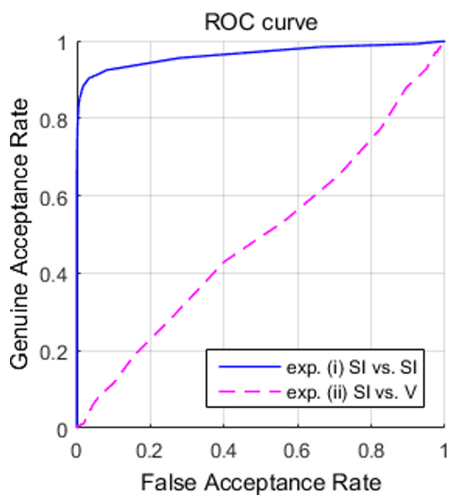


Figure 1. *CMC and ROC curves for experiment (i) still images vs. still images and (ii) still images vs. videos.*

The first large and publicly available image database for benchmarking of source sensor recognition techniques has been proposed in 2010, namely the "Dresden Image Database" [2]. It is composed by more than 14,000 images acquired with 73 cameras of 25 different models. It has been used in a number of works [4][9][3][5]. Another small database for blind source cell-phone model identification has been presented in 2008 by Çeliktutan et al. in [7]. It contains more than 3.000 pictures collected using 17 mobile phones of 15 different models.

Database

In order to perform our experiments, we collected a novel dataset of still images and videos, namely the SOCRatES database. In its current state, the database is made up of about 6.200 images and 680 videos captured with 67 different smartphones of 14 different makes and 42 different models. It also contains several pictures of uniform backgrounds for the RSPN extraction. However, RSPN extraction can be performed also on non-uniform-color images still obtaining optimal performances, as demonstrated in [11].

The acquisition has been performed in uncontrolled conditions. In order to collect the database, many people were involved and asked to use their personal smartphone to collect a set of pictures. The reason behind this choice is, on the one hand, to collect a database of heterogeneous pictures and to maximize the number of devices employed, and, on the other hand, to carefully replicate realistic acquisition conditions.

A total of 90 photos and 10 videos have been collected for each smartphone: 50 are photo of the blue sky, or of another uniform color surface, needed for the RSPN computation; 40 pictures portray random scenes, avoiding privacy and copyright sensitive subjects. Ten short video clips are recorded with each device. Their duration varies from 2 to 5 seconds. Involved persons in the first acquisition session are mostly EURECOM students.

A naming convention has been adopted to distinguish the images/videos captured with different devices, an ID number has been assigned to each different device, and to indicate the type of the acquired item, i.e.: "background picture", "foreground picture", "video".

Along with pictures and videos, annotation files describing the characteristics of the smartphones employed are provided. In particular, they list the smartphone model, the Operating System, the digital camera model, the photo resolution and the video resolution employed during acquisition.

Thanks to this dataset we analyse the advantages and disadvantages of performing source camera recognition on mobile phones and in using videos versus still images. The database and its description will be soon made available the following URL: `http://socrates.eurecom.fr`

Video vs. still images SPN extraction on mobile devices

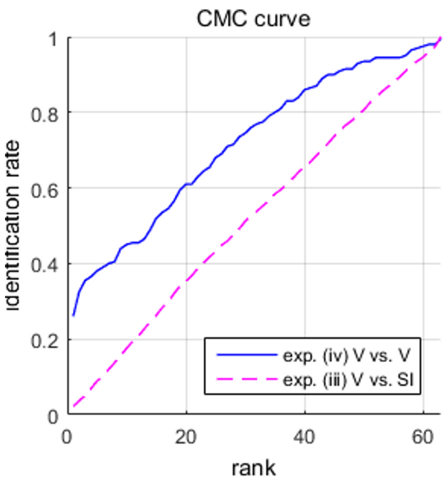
The problem addressed by this work is two-fold: (i) we first assess the performances of Li's technique for source camera identification for the first time on a large database of images captured only by mobile devices; (ii) we analyse the problem of SPN extraction from strongly compressed videos, such as the ones generated by smartphones.

SPN extraction from videos is a well known issue [12]. The sensor pattern noise is strongly impacted by compression and also, compared to photos captured by the same sensor and with the same resolution, the recorded scene is somehow cropped. It is observed that the resulting SPN comparison lead to a much lower correlation when comparing videos recorded by the same sensor. One possible way to mitigate the problem, is to pre-select the video frames to be used in the SPN computation, by taking into account mainly the I-frames [12], on which the impact of video compression is weaker.

As mentioned before, Li proposes to enhance the weakest SPN components and to suppress the strongest ones that are more likely to correspond to scene details [10]. Different models are proposed in [10] to compute the Enhanced SPN (ESPN), we adopt the following:

$$n_e(i,j) = \begin{cases} e^{-0.5n^2(i,j)/\alpha^2}, & \text{if } 0 <= n(i,j) \\ -e^{-0.5n^2(i,j)/\alpha^2}, & \text{otherwise} \end{cases} \tag{2}$$

where n_e is the ESPN, n is the SPN, i and j are the indices



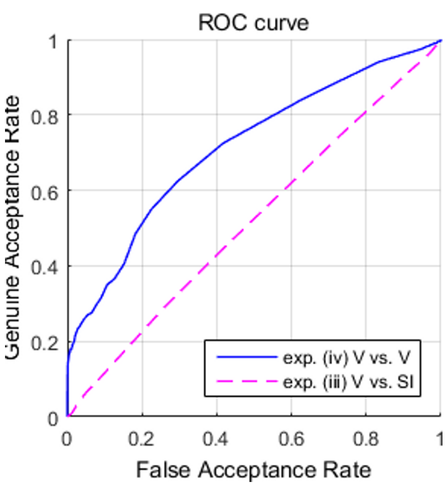


Figure 2. *CMC and ROC curves for experiment (iii) videos vs. still images and (iv) videos vs. videos.*

of the components of n and n_e, and α is a parameter that is set to 7, as indicated in [10].

To know if a given picture/video frame belongs to a specific sensor, the extracted ESPN is compared with the Reference SPN of the sensor (RSPN). The RSPN n_r corresponds to the average SPN computed over N images:

$$n_r = \frac{1}{N} \times \sum_{k=1}^{N} n_k \tag{3}$$

The process to compute RSPN and ESPN from still images is trivial. It is only necessary to have enough images for the RSPN extraction, 50 is the number of images employed both by Lukas et al. [6] and by Li [10]. For videos, it is required to first extract the single video frames. In our experiments, we extract the first 100 frames from one (or more videos in case the first one is too short) video for computing the RSPN, and we use the first frame of each video for the ESPN extraction.

Experimental results

The following experiments have been performed:

i RSPN still images vs. ESPN still images;
ii RSPN still images vs. ESPN videos;
iii RSPN videos vs. ESPN still images;
iv RSPN videos vs. ESPN videos.

For each experiment, the performances are assessed in terms of Receiver Operating Characteristic (ROC) curve, Area Under ROC curve (AUC), Comulative Match Characteristic (CMC) curve, Recognition Rate (RR = CMC(1) - value of the CMC at rank 1), and Equal Error Rate (EER). These perfomances are computed from a distance matrix made up of the correlation scores obtained comparing the RSPNs of the sensors against the ESPNs extraxted from the images/video frames.

As expected, the performances for experiment (i) *still images vs. still images*, are very good: RR = 0.90, EER = 0.08, and AUC = 0.96. Meaning that 2789 images have been matched to the corresponding 67 sensors with a rate of correct classification of the 96%. The corresponding performance graphs are illustrated in figure 1. Li's technique assure a high rate of correct matching even when using small parts of the image for the comparison. In our tests, RSPN and ESPN are extracted from a window of 1024×1024 pixels centred in the image/video frame.

Less predictable are the results obtained by experiments (ii) *still images vs. videos* and (iii) *videos vs. still images*. As stated before, it is known that the SPN extracted from videos has a slightly different resolution compared with pictures captured by the same device and that the SPN is affected by the strong video compression. Nonetheless, from the graphs presented in figures 1 and 2, it is observable that the noise pattern is completely uncorrelated. The corresponding performance values are AUC = 0.49, RR = 0.01, EER = 0.51 and AUC = 0.52, RR = 0.02, EER = 0.48 for experiment (ii) and (iii), respectively.

The most surprising results are those relative to the the third experiment *videos vs. videos*, see figure 2 for reference. With AUC = 0.71, RR = 0.26, and EER = 0.33, it is possible to realize how strongly is the SPN impacted by video compression. The test has been performed over 200 videos. Thus, the performances are expected to decrease when testing the recognition rate on a larger number of videos.

Conclusions

A database of still images and videos recorded with 67 different smartphones of 42 different models has been collected. On this database, a highly accurate technique for source sensor recognition has been tested, namely Li's Enhanced Sensor Pattern Noise. This method is able to associate a given picture to the correct sensor even when comparing cameras of the same model. The results showed above, demonstrate, as expected, that the ESPN works very well on still images, with a rate of correct classification of 96% on a large set of images. On the other hand, performances drop when using the same technique on videos. This is the first time that source sensor recognition is tested on a large set of pictures captured by mobile devices and the first time that the impact of video compression on the ESPN technique is analysed over a large number of videos. Performances are reported to drop from the 96% of correct classification to the 71% when testing still images vs. still images and videos vs. videos, respectively. Performances drop aroung 50% when performing asymmetric SPN comparison between still images and videos.

Acknowledgments

This work was partly supported by the European Union's Horizon 2020 - Marie Skłodowska-Curie Actions - Research and Innovation Staff Exchange. Project name: IDENTITY: Computer Vision Enabled Multimedia Forensics and People Identification.

References

[1] Judith A. Redi, Wiem Taktak, and Jean-Luc Dugelay, Digital Image Forensics: a booklet for beginners, Multimed Tools Appl (2011) 51: 133. doi:10.1007/s11042-010-0620-1

[2] Thomas Gloe and Rainer Böhme, The Dresden Image Database for Benchmarking Digital Image Forensics, Journal of Digital Forensic Practice, vol. 3 (2-4), pp. 150-159, 2010. doi: 10.1080/15567281.2010.531500

[3] Francesco Marra, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, A study of co-occurrence based local features for camera model identification, Multimed Tools Appl (2016). doi:10.1007/s11042-016-3663-0

[4] Zhonghai Deng, Arjan Gijsenij, and Jingyuan Zhang, Source camera identification using Auto-White Balance approximation, 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 57-64. doi: 10.1109/ICCV.2011.6126225

[5] Thomas Gloe, Stefan Pfennig, and Matthias Kirchner, Unexpected artefacts in PRNU-based camera identification: A 'Dresden Image Database' case-study, Proc. 14th ACM Workshop Multimedia Security, pp. 109-114, 2012.

[6] Jan Lukás, Jessica Fridrich, and Miroslav Goljan, Digital camera identification from sensor pattern noise, IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 205-214, Jun.

[7] Oya Çeliktutan, Bülent Sankur, and İsmail Avcibaş, Blind identification of source cell-phone model, IEEE Transactions on Information Forensics and Security 3(3), 553-566 (2008).

.

[8] Chiara Galdi, Michele Nappi, and Jean-Luc Dugelay, Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity, Pattern Recognition Letters, 0167-8655, 2015. doi: 10.1016/j.patrec.2015.09.009

[9] Giovanni M. Farinella, Mario V. Giuffrida, Vincenzo Digiacomo, and Sebastiano Battiato, On Blind Source Camera Identification, Advanced Concepts for Intelligent Vision Systems LNCS, Volume 9386, pp. 464-473, 2015.

[10] Chang-Tsun Li, Source camera identification using enhanced sensor pattern noise, IEEE Trans-actions on Information Forensics and Security 5(2): pp. 280-287, 2010.

[11] Chiara Galdi, Michele Nappi, and Jean-Luc Dugelay, Multimodal authentication on smartphones: Combining iris and sensor recognition for a double check of user identity, Pattern Recognition Letters, 0167-8655, 2015. doi: 10.1016/j.patrec.2015.09.009

[12] Wei-Hong Chuang, Hui Su, and Min Wu, Exploring compression effects for improved source camera identification using strongly compressed video, 2011 18th IEEE International Conference on Image Processing, Brussels, 2011, pp. 1953-1956. doi: 10.1109/ICIP.2011.6115855

[13] Gerald C. Holst (1998) CCD Arrays, Cameras, and Displays, Second Edition. JCD Publishing & SPIE Press, USA

[14] Zeno J. Geradts, Jurrien Bijhold, Martijn Kieft, Kenji Kurosawa, Kenro Kuroki, Naoki Saitoh (2001) Methods for identification of Images Acquired with Digital Cameras. Proc. of SPIE, Enabling Technologies for Law Enforcement and Security, vol. 4232, pp. 505512