On the Repudiability of Device Identification and Image Integrity Verification Using Sensor Pattern Noise

Chang-Tsun Li, Chih-Yuan Chang, and Yue Li

Department of Computer Science
University of Warwick
Coventry CV4 7AL
UK
{ctli,cyc,yxl}@dcs.warwick.ac.uk

Abstract. In this work we study the power of the methods for digital device identification and image integrity verification, which rely on sensor pattern noise as device signatures, and the repudiability of the conclusions drawn from the information produced by this type of methods. We prove that the sensor pattern noise existing in the original images can be destroyed so as to confuse the investigators. We also prove that sensor pattern noise of device A can be easily embedded in the images produced by another device B so that the device identifier would mistakenly suggest that the images were produced by device A, rather than by B, and mislead forensic investigations.

Keywords: Digital Device Identification, Digital Forensics, Digital Investigation, Digital Evidence, Sensor Pattern Noise, Integrity Verification.

1 Introduction

While digital imaging devices, such as digital cameras and scanners, bring unquestionable convenience of image acquisition, powerful image processing software also provides means for editing images so as to serve good and malicious purposes. To combat image manipulations for malicious purposes, researchers have proposed ways of verifying the integrity of images based on the detection of local inconsistencies of device attributes or data processing related characteristics, such as sensor pattern noise [1], camera response function [3], resampling artifacts [7], color filter array (CFA) interpolation artifacts [8, 12], and JPEG compression [10]. Similar device attributes and data processing related characteristics have also been exploited to identify and classify the source devices in aiding forensic investigation [2, 5, 6, 9, 11]. While many methods [3, 7, 8, 12] require that specific assumptions be satisfied, methods based on sensor pattern noise have drawn much attention due to the relaxation of the similar assumptions. The deterministic component of pattern noise is mainly caused by imperfections during the sensor manufacturing process and different sensitivity of pixels to light due to the inhomogeneity of silicon wafers [4]. It is because of the inconsistency and the uniqueness of manufacture imperfections and sensitivity to light that even sensors made from the same silicon wafer would possess uncorrelated pattern noise, which can be extracted from the images produced by the devices. This

D. Weerasinghe (Ed.): ISDF 2009, LNICST 41, pp. 19–25, 2010.

property makes sensor pattern noise a robust signature for identifying the origin and verifying the integrity of images.

Although good performance of source identification and integrity verification has been reported [1, 2, 5, 6, 9, 11], we observed that, due to the fact that sensor pattern noise is treated as additive signal to images during its extraction process, sensor pattern noise can be manipulated and substituted to confuse investigators and mislead forensic investigations. Therefore, the conclusions made by these methods can only be useful in aiding forensic investigations by narrowing down cases under investigation, but further research is necessary to enhance the non-repudiability of their findings before they can be accepted as admissible evidence in the court of law.

The rest of this work is organised as follows. We discuss the way digital source identifiers and integrity verifiers using sensor pattern noise work in Section 2 and prove how sensor pattern noise can be manipulated in Section 3. Section 4 concludes this work.

2 Device Identification and Image Integrity Verification Using Sensor Pattern Noise

Although the use of sensor pattern noise, $\boldsymbol{n}_D$, in different methods is slightly different, sensor pattern noise is commonly treated as an additive high-frequency signal to an image, $\boldsymbol{I}_D$, and the way it is extracted is similar to that used in [6], which is formulated as

$$\boldsymbol{n}_D = \boldsymbol{I}_D - F(\boldsymbol{I}_D) \tag{1}$$

where F is a denoising filtering function which filters out the sensor pattern noise. The subscript, D, indicates that $\boldsymbol{I}_D$ is an image taken by device D and $\boldsymbol{n}_D$ is the sensor pattern noise extracted from $\boldsymbol{I}_D$. Although various denoising filters can be used as F, the wavelet-based denoising filter described in [6] has been reported as effective in producing good results and our experiments confirm with the report. Therefore, this filter is used in the current work. We use the *average* of the sensor pattern noise, $\boldsymbol{P}_D$, of a set of images taken by a particular device D to represent that device. In this work, we call this *average* sensor pattern noise, $\boldsymbol{P}_D$, *signature* of sensor D in order to differentiate it from sensor pattern noise, $\boldsymbol{n}_D$, extracted from *individual* photos. The correlation of ρ_{AB}, as formulated in Eq. (2), between the sensor pattern noise $\boldsymbol{n}_A$ of an image $\boldsymbol{I}_A$ and $\boldsymbol{P}_B$ is used to decide whether image $\boldsymbol{I}_A$ is taken by device B.

$$\rho_{AB} = \frac{(\boldsymbol{n}_A - \overline{\boldsymbol{n}}_A)\cdot(\boldsymbol{P}_B - \overline{\boldsymbol{P}_B})}{\left\|\boldsymbol{n}_A - \overline{\boldsymbol{n}}_A\right\|\cdot\left\|\boldsymbol{P}_B - \overline{\boldsymbol{P}_B}\right\|} \tag{2}$$

where $\overline{\boldsymbol{n}}_A$ and $\overline{\boldsymbol{P}_D}$ are the means of $\boldsymbol{n}_A$ and $\boldsymbol{P}_D$, respectively. A large value of ρ_{AB} indicates high likelihood that $\boldsymbol{I}_A$ is taken by device B. We could expect that $\rho_{AA} > \rho_{AB}$ and $\rho_{BB} > \rho_{AB}$ if A and B are two different devices because ρ_{AA} and ρ_{BB} are *intra-class* metrics of similarity while ρ_{AB} indicates *inter-class* similarity.

On the other hand, methods for image integrity verification using sensor pattern noise, such as [1], are based on detecting local inconsistencies of sensor pattern noise blocks introduced into the forged images. The basic way of making use of sensor pattern noise is the same as that of device identifiers except that sensor pattern noise is extracted from each individual image *blocks* and compared to the corresponding blocks of the device signature, rather than the *whole* signature.

3 Potential Attacks

From the presentation in Section 2 and Eq. (1), we can see that digital device identifiers and image integrity verifiers using sensor pattern noise consider the sensor pattern noise as an additive signal to images. However, it is this additive nature of the sensor pattern noise that opens gap for potential attacks and manipulations. We present some experiments in the next two sub-sections to demonstrate how attacks with two different malicious intentions could be launched on photos taken by digital cameras. We call the first attack *Signature Removal*, which could be maliciously carried out to confuse investigators while the other attack –*Signature Substitution*, as we call it, could be applied to mislead forensic investigations.

3.1 Signature Removal

We define *Signature Removal* as an attack of removing the sensor signature $\boldsymbol{P}_A$ of image $\boldsymbol{I}_A$, taken by device A, using Eq. (3) so that the device identifiers cannot extract the sensor pattern noise.

$$\boldsymbol{I}_{A'} = \boldsymbol{I}_A - \alpha \boldsymbol{P}_A \tag{3}$$

where α is a positive strength factor determining the amount of sensor signature to be removed.

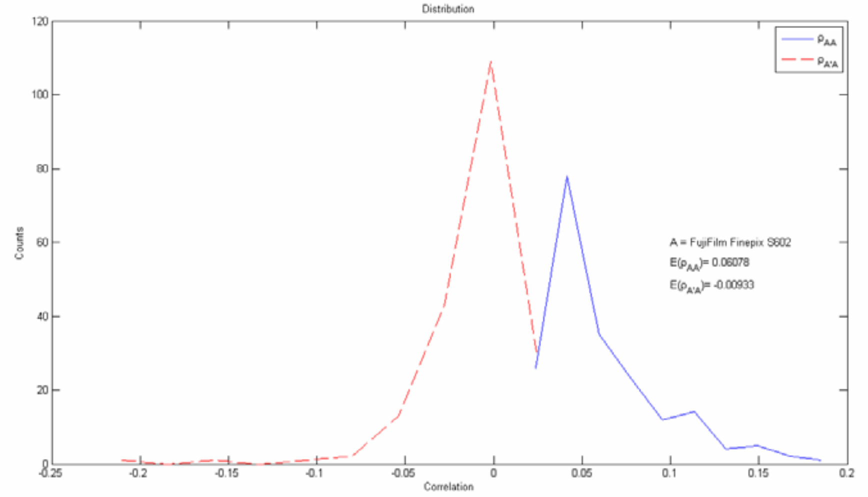


Fig. 1. Demonstration of *Signature Removal* attack

To evaluate the effect of *Signature Removal* attack, we apply Eq. (3) to 200 photos of 1536 × 2048 pixels taken by camera *A* (FujiFilm Finepix S602) in our experiment and demonstrate the results in Fig. 1. The blue solid curve is the distribution of the correlations between the sensor pattern noise of the 200 photos and the signature, $\boldsymbol{P}_A$, of camera *A*, with a mean $E(\rho_{AA})$ equal to 0.06078. After manipulating those 200 photos we extract the fake sensor pattern noise $\boldsymbol{n}_{A'}$ from each of them using Eq. (1) and then calculate the correlations $\rho_{A'A}$ according to Eq. (3), with $\alpha =$ 0.75. We can see from Fig. 1 that $\rho_{A'A}$, illustrated with red dashed curve, distribute around $E(\rho_{A'A})$ = -0.009334 (≈ 0), which far less than $E(\rho_{AA})$ (= 0.06078) of ρ_{AA}. This is a clear indication that the original sensor pattern noises, $\boldsymbol{n}_A$, have been removed from their corresponding host photos. This may well lead the forensic investigators to think that the photos in question were taken by some unknown cameras, but not by camera *A*.

3.2 Signature Substitution

We define *Signature Substitution* as an attack of removing signature, $\boldsymbol{P}_A$, of device *A* from image $\boldsymbol{I}_A$ taken by device *A* and then embedding the signature, $\boldsymbol{P}_B$, of device *B* so that device identifiers would mistakenly suggest that the manipulated image, $\boldsymbol{I}_{A'}$, were taken by device *B*, rather than by device *A*. The operation is as follows.

$$\boldsymbol{I}_{A'} = \boldsymbol{I}_A - \alpha \boldsymbol{P}_A + \beta \boldsymbol{P}_B \tag{4}$$

where α, like in Eq. (3), is a positive strength factor determining the amount of sensor signature of device *A* to be removed and β is also a positive strength factor determining the amount of sensor signature of device *B* to be embedded.

Fig. 2 demonstrates our experimental results when *Signature Substitution* (i.e., Eq. (4)) is applied, with $\alpha = \beta = 1$. In this experiment, we use the same 200 photos taken by camera *A* (FujiFilm Finepix S602) and another 200 photos of 1536 × 2048 pixels taken by camera *B* (Olympus C730UZ). From Fig. 2 we could clearly see that ρ_{AB}, which is a metric of *inter-class* similarity, distribute around a mean close to 0 ($E(\rho_{AB})$ = -0.00027) while ρ_{AA} and ρ_{BB}, which are both metrics of *intra-class* similarity, distribute around relatively greater means ($E(\rho_{AA})$ = 0.06078 and $E(\rho_{BB})$ = 0.02823, respectively). These are good indications that sensor pattern noise can indeed effectively identify the source devices. However, after attacking the 200 photos taken by camera *A* (FujiFilm Finepix S602) according to Eq. (4), we can see that the correlations, $\rho_{A'B}$, between the fake sensor pattern noise, $\boldsymbol{n}_{A'}$, (i = 1, 2, 3,…, 200), extracted from the manipulated photos, $\boldsymbol{I}_{A'}$, and $\boldsymbol{P}_B$ distribute around a significant mean ($E(\rho_{A'B})$ = 0.06723). Based on the significant $\rho_{A'B}$, the identifier would mistakenly conclude that images $\boldsymbol{I}_{A'}$ were taken by camera *B* (Olympus C730UZ), which may well mislead forensic investigations.

By the same token, to avoid detection by image integrity verifiers, an attacker could remove the device signature of *K* photos, $\boldsymbol{I}_{A_k}$ (k = 1, 2, 3, …, *K*), taken by the same or different devices using Eq. (3) to create $\boldsymbol{I}_{A_k}^{'}$ before carrying out a forgery

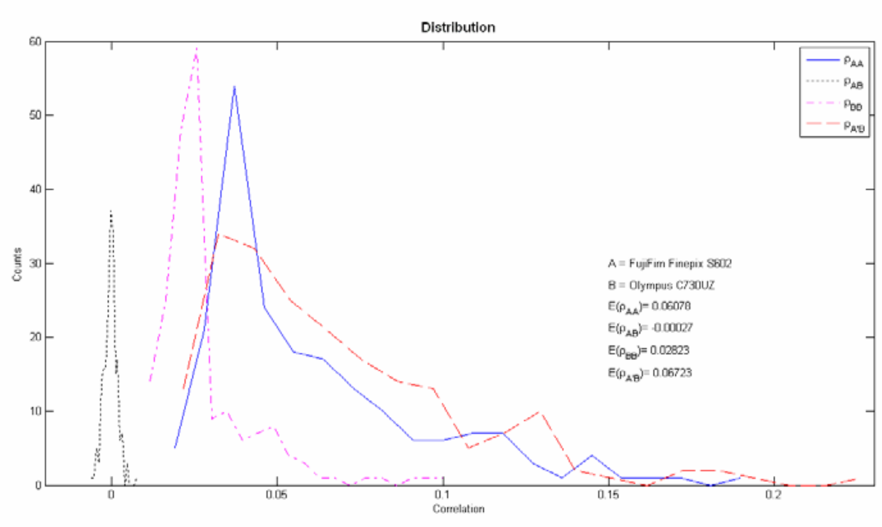
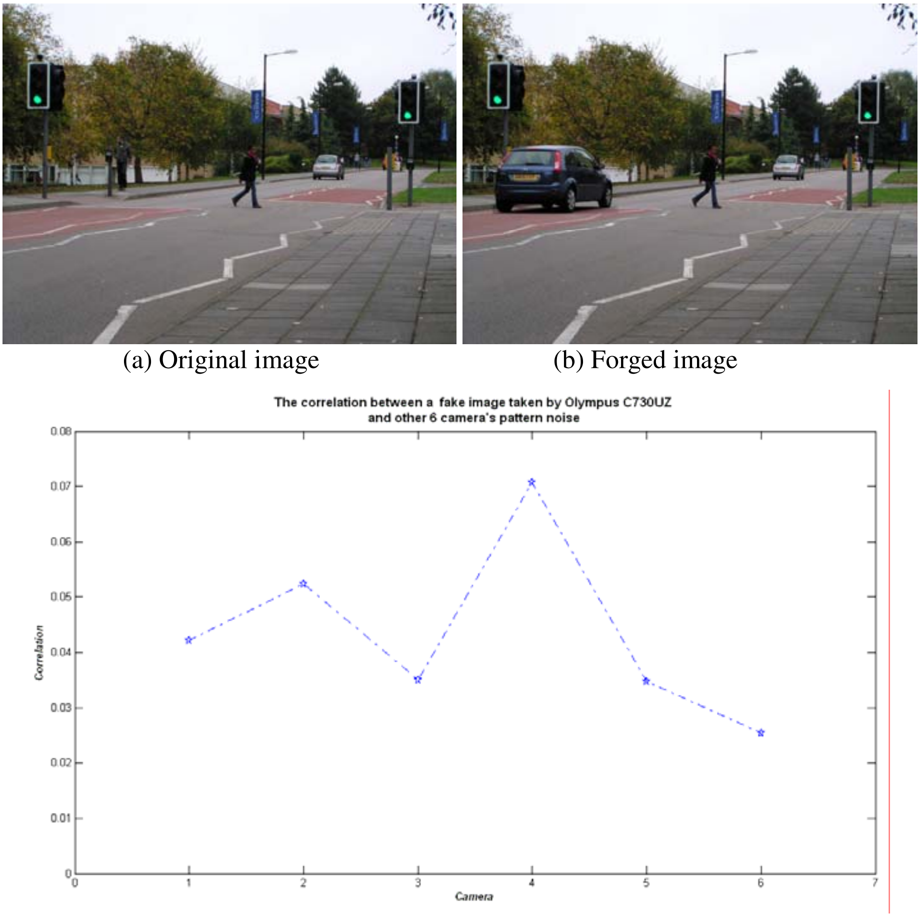


Fig. 2. Demonstration of *Signature Substitution* attack

operation to create a fake image $\boldsymbol{I}_{A'}$. Finally, the attacker could embed the signature, $\boldsymbol{P}_{B,}$ of a specific device *B* to make the fake image $\boldsymbol{I}_{A''}$ appear to be created by device *B*. The attack can be formulated as

$$\boldsymbol{I}_{A''} = \sum_{k=1}^{K} \boldsymbol{I}_{A'_k} + \beta \boldsymbol{P}_B \tag{5}$$

where $\boldsymbol{I}_{A'_k} = \boldsymbol{I}_{A_k} - \alpha_k \boldsymbol{P}_{A_k}$ (see Eq. (3)) and $\sum_{k=1}^{K}$, not to be taken as summation, is an arbitrary forgery operation, such as splicing and cut-and-paste, involving K images. Fig. 3 illustrates an original photo (Fig. 3(a)) and a forged image (Fig. 3(b)) with a car inserted in the foreground of the scene. Two photos (i.e., K = 2) taken by camera *A* (Olympus C730UZ) are used for forging Fig. 3(b). To avoid detection by the image integrity identifier, we applied Eq. (5) to make the device identifier believe that the forged photo was taken by another camera *B* after the forgery is done. Fig.3(c) illustrates the correlation $\rho_{A''B}$ when the signatures of six different cameras (Canon IXUS850IS, Canon PowerShot A400, Canon IXY DIGITAL 500, FujiFilm Finepix S602, FujiFilm Finepix A920 and Olympus FE210) are embedded. The significant correlations, which are close to the mean correlations between the signatures of the cameras and the photos they take suggest that the attack has been successful. Note that in our experiments involving the seven cameras mentioned in this work, the means of the intra-class correlation are all greater than 0.02, indicating that any correlation greater than 0.02 are significant.



(a) Original image

(b) Forged image

(c) correlation

Fig. 3. Demonstration of image forgery and Signature Substitution attack. a) The original image taken by an Olympus C730UZ camera. b) A forged image using photos taken by the same Olympus C730UZ camera with the sensor signature of FujiFilm Finepix S602 camera embedded. c) The correlations between the sensor pattern noises $n_{A''}$ extracted from Fig. 3(b) and the signatures of six cameras.

4 Conclusions

Sensor pattern noise has been reported in many articles as a robust signature of images produced by various digital devices, with great potential for device identification and integrity verification in the context of digital forensics. However, in this work, we have demonstrated that, due to its additive nature, the sensor pattern noise of an image can be manipulated to confuse forensic investigators or mislead investigations by applying the two attacks that we devised and presented in Section 3. We conclude that device identifiers and image integrity verifiers using sensor pattern noise are effective tools in aiding forensic investigations. Nevertheless, without further research to make the conclusions drawn by these methods non-repudiatable, it is unlikely that their conclusions will be accepted in the court of law as admissible evidence.

References

1. Chen, M., Fridrich, J., Goljan, M., Lukáš, J.: Determining Image Origin and Integrity Using Sensor Noise. IEEE Transactions on Information Forensics and Security 3(1), 74–90 (2008)
2. Gloe, T., Franz, E., Winkler, A.: Forensics for Flatbed Scanners. In: Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, January 29-February 1, vol. 6505, pp. 1I–1J (2007)
3. Hsu, Y.F., Chang, S.F.: Image Splicing Detection Using Camera Response Function Consistency and Automatic Segmentation. In: Proc. IEEEInternational Conference on Multimedia and Expo., Beijin, China, July 2-5 (2007)
4. Janesick, J.R.: Scientific Charge-Coupled Devices, vol. PM83. SPIE, Bellingham (2001)
5. Khanna, N., Chiu, G.T.C., Allebach, J.P., Delp, E.J.: Forensic techniques for Classifiying Scanner, Computer Generated and Digital Camera Images. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Las Vegas, USA, 30 March-4 April, pp. 1653–1656 (2008)
6. Lukáš, J., Fridrich, J., Goljan, M.: Digital Camera Identification from Sensor Pattern Noise. IEEE Transactions on Information Forensics and Security 1(2), 205–214 (2006)
7. Popescu, A.C., Farid, H.: Exposing Digital Forgeries by Detecting Traces of Resampling. IEEE Transactions on Signal Processing 53(2), 758–767 (2005)
8. Popescu, A.C., Farid, H.: Exposing Digital Forgeries in Color Filter Array Interpolated Images. IEEE Transactions on Signal Processing 53(10), 3948–3959 (2005)
9. Sankur, B., Celiktutan, O., Avcibas, I.: Blind Identification of Cell Phone Cameras. In: Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX, San Jose, CA, January 29-February 1, vol. 6505, pp. 1H–1I (2007)
10. Sorell, M.J.: Digital Camera Source Identification through JPEG Quantisation. In: Li, C.-T. (ed.) Multimedia Forensics and Security. Information Science Publishing, Hershey (2008)
11. Sutcu, Y., Batram, S., Sencar, H.T., Memon, N.: Improvements on Sensor Noise based Source camera Identification. In: Proceeding of IEEE International Conference on Multimedia and Expo., Beijing, China, July 2-5, pp. 24–27 (2007)
12. Swaminathan, A., Wu, M., Liu, K.J.R.: Nonintrusive Component Forensics of Visual Sensors Using Output Images. IEEE Transactions on Information Forensics and Security 2(1), 91–106 (2007)