Investigating Inconsistencies in PRNU-Based Camera Identification

INVESTIGATING INCONSISTENCIES IN PRNU-BASED CAMERA IDENTIFICATION

Nabeel Nisar Bhat and Tiziano Bianchi

Politecnico di Torino

ABSTRACT

PRNU (Photo-response non-uniformity) is widely considered a unique and reliable fingerprint for identifying the source of an image. The PRNU patterns of two different sensors, even if belonging to the same camera model, are strongly uncorrelated. Therefore, such a fingerprint is used as evidence by various law enforcement agencies for source identification, manipulation detection, *etc.* However, in recent smartphones, images are subjected to significant in-camera processing associated with computational photography. This heavy processing introduces non-unique artifacts (NUA) in such images and masks the uniqueness of the PRNU fingerprint. In this work, we investigate the robustness of PRNU in modern smartphones. We propose a model that explains the unexpected behavior of PRNU in such smartphones. Finally, we present two methods to identify images suffering from NUA. Our methods achieve high accuracy in identifying such images.

***Index Terms*—** Image forensics, source identification, photo-response non-uniformity, non-unique artifacts.

1. INTRODUCTION

Over the past decade, the photo-response non-uniformity (PRNU) trace has been widely used to identify the origin of an image [1]. The PRNU pattern is unique to a sensor [1]. Moreover, the pattern is relatively constant over time and survives most camera processing [1]. Apart from image-source identification, the PRNU trace is also used to detect manipulations in an image [2], video-source identification from still-images [3] and clustering images based on the source device [4]. Researchers have thoroughly investigated the effectiveness of the PRNU fingerprint by performing tests on many datasets, *e.g.,* the Dresden dataset, the VISION dataset [5], *etc*. The results have been quite convincing; high detection accuracies and very low percentages of false alarms have been associated with the PRNU fingerprint. A considerable number of refinements have also been proposed to improve the robustness of PRNU [6], [7].

Recently computational photography has made it possible to introduce features like Portrait, HDR [8], Night Mode, *etc.* in the modern smartphone cameras. Moreover, acquisition processing like pixel-binning [9] is used to improve the low-light performance. However, the processing involved in such features introduces non-unique artifacts (NUA) in the images, masking the uniqueness of the PRNU fingerprint. Recently, Iuliani and his team [10] highlighted this issue of PRNU in modern smartphones, which employ such processing. The team documented high percentages of false alarms for many modern smartphones. In such a scenario, PRNU will lead to erroneous camera identification and serious consequences. This represents a problem given the usage of PRNU in court evidence and various forensic softwares *e.g.*, PRNU Compare Professional [11] and Amped Authenticate [12]. This problem is under-researched, and no solution to counter the problem has been proposed so far.

In this work, we discuss the problem of fingerprint collision in modern smartphones, *i.e.,* the fact that different smartphones of the same model tend to have similar fingerprints. Our focus is on a subset of smartphone models analyzed by Iuliani, and his team [10]. We propose a model that explains the cause of the problem in these smartphones. Based on this model, we present two methods, SPAM Classifier, and Meta-Data SceneType Classifier, to identify images suffering from NUA. The experiments validate our model and show that images affected by NUA can be reliably identified by the proposed methods.

2. PROBLEM STATEMENT

Before introducing the problem, we recall the standard PRNU model [1]:

$$S = S_0 + S_0 P_N + N \tag{1}$$

where S represents the acquired image, S_0 represents the noise-free image, P_N represents the PRNU fingerprint, and N represents the equivalent noise. According to this model, we can estimate from the image S, the fingerprint P_N or consider the noise residual, $W = S_0 P_N + N$. This PRNU model can be used to solve different problems. In our case, we are interested in the device linking problem, *i.e.,* verifying whether two images have been acquired by the same device. This amounts to computing a similarity metric between the fingerprint estimates or the noise residuals, as follows:

$$\rho = f(W_1, W_2) \tag{2}$$

where $f()$ can be either Normalized Correlation [1] or Peak to Correlation Energy (PCE) [7] and ρ represents the similarity score of the two residuals, W_1 and W_2. If ρ is greater than the threshold, we decide that both pictures come from the same device, otherwise they come from different devices.

Source identification for modern smartphones turns out to be challenging. Pair-wise correlations between PRNU patterns corresponding to the images of different devices of the same smartphone model result in unexpected distributions. Ideally, these correlation values should be less than the detection threshold. Instead, for modern smartphones, many correlation values surpass the threshold. In other words, the PRNU fingerprints of different devices collide with each other. This inconsistent behavior is verified for both correlation metrics, PCE and Normalized Correlation. As far as Normalized Correlation is concerned, the correlation values do not approximate the expected zero-mean Gaussian distribution.

3. NUA MODELING

We believe that the problem of fingerprint collision occurs due to the introduction of NUA in images of modern smartphones. The imaging pipeline and the processing associated with computational photography are customized by the manufacturer. Therefore, it is likely that a manufacturer uses similar processing in different devices of the same smartphone model. This common processing introduces NUA shared among different devices of the same model. We assume that NUA may affect only a fraction of images acquired by a certain device. The images that are not affected by NUA are called *good images* and can be modeled by standard PRNU model (1). On the other hand, the images that are affected by NUA are called *bad images* and can be modeled by:

$$S = S_0 + S_0 P_N + N + \epsilon \tag{3}$$

where ϵ refers to the perturbation (a sort of additional fingerprint) due to NUA. This is similar to the concept of noiseprint introduced in [13]. However, noiseprint jointly considers the PRNU and NUA and is therefore aimed at camera model identification, not at the device level. Instead, here we are trying to separate the two components, so that device identification can be performed reliably.

PRNU patterns and noise residuals of all bad images corresponding to different devices of the same smartphone model suffer from the same non-unique perturbation. On the other hand, the fingerprints and residuals of good images do not have a contribution due to NUA. Therefore, according to our model, we expect that correlations between two good images or one good and one bad image coming from different devices result in values less than the threshold. Only correlations between two bad images from different devices of the same smartphone model result in unexpected values greater than the threshold due to the presence of a common NUA term ϵ. Thus, if we compute correlations between pairs of images coming from different devices, we expect that those correlations will follow a bimodal distribution. One mode will have the usual zero-mean distribution of non-matching fingerprints, while the other mode will have an unexpected positive mean. This second mode can be related to the subset of bad images only.

4. PROPOSED METHODS

We propose two methods to counter the anomalous behavior of PRNU. Both methods aim at identifying good and bad images. To validate the performance of our methods, we identify good and bad images according to our model, which serves as ground truth.

4.1. Ground Truth Generation

We consider a dataset of images from known smartphones models and devices. We use two strategies to expose the problem of fingerprint collision. Under strategy 1, we perform pair-wise correlations between the noise residuals corresponding to the images of different devices of the same model. Under strategy 2, we instead perform pair-wise correlations between the fingerprints. Then we divide the correlations into good and bad scores according to a threshold. For Normalized Correlation, the threshold is calculated based on the sensor size for a fixed probability of false alarm equal to 10^{-6}. While for PCE, the threshold is set to the commonly agreed value of 60. The subset of images that contribute to bad correlation scores (greater than the threshold) is labeled as bad, while the remaining images are labeled as good.

4.2. SPAM Classifier

The first method, SPAM Classifier, is based on SPAM features [14]. SPAM features capture dependencies between pixels of an image. In bad images, the dependencies are modified due to NUA. Therefore, SPAM features should be distinct for good and bad images. First, we extract 2^{nd} order SPAM features from labeled ground truth images. Since the SPAM features for an image consist of 686 dimensions, we reduce their dimensionality using principal component analysis (PCA) to 30 features or dimensions. The dataset is then divided into training and test data using a 70:30 split. The final step involves training a classifier with the training data to learn the features of good and bad images. Finally, we estimate the accuracy of the classifier on the test data.

4.3. Meta-data SceneType Classifier

The second method, Meta-Data SceneType Classifier, is based on the screening of meta-data. Analyzing the standard meta-data settings (Exposure Time, ISO, and Aperture), we could not find any link between the settings and the unexpected behavior of correlations. However, one of the

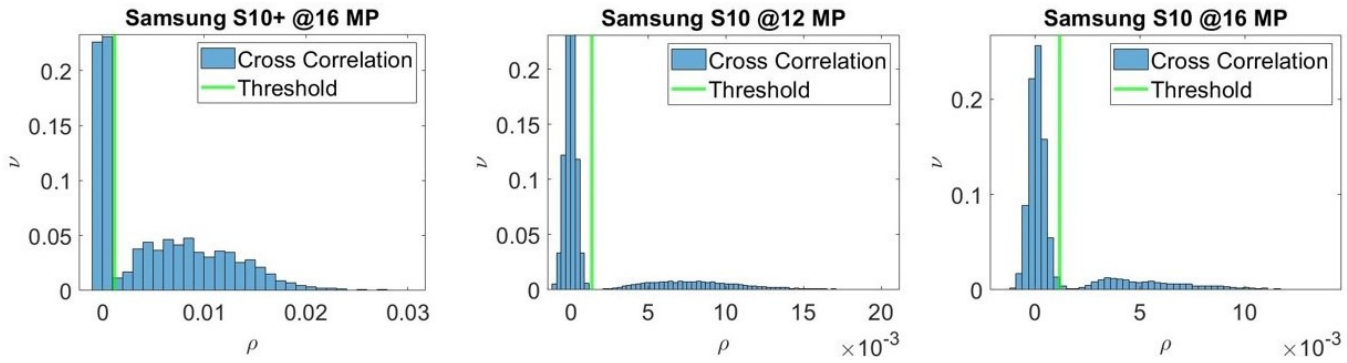

Fig. 1. Distribution of Normalized Correlation values.

meta-data tags, *SceneType: A directly photographed image*, correlates highly with the expected behavior. Those images which contain this tag resemble good images introduced earlier *i.e.*, the corresponding pair-wise correlations result in values less than the threshold. Moreover, those images which do not contain this tag resemble bad images. The absence of this tag implies additional post-processing. The classifier simply labels all images showing this tag as good images.

5. EXPERIMENTS

We collect images from the Flickr database introduced in [10] corresponding to Huawei (Mate 20 Pro and P30 Pro), Samsung (S10, S10+, S10e and A50) and Xiaomi (Redmi Note 7) smartphones. We collect a total of 4643 images from 67 different Flickr users (devices). The details of the dataset are presented in Table 1. In [10], the genuinness of the data has been validated by looking at the *Model*, *Width, Height* and *GPSInfo* tags in meta-data. Therefore, it is safe to assume that different Flickr users correspond to different devices of the same smartphone model. For the sake of space constraints, we report results corresponding to strategy 1 and Normalized Correlation for a subset of models. However, the same holds for strategy 2 and PCE. Figure 1 shows the distribution of correlation values for Samsung S10+ (16 MP), S10 (12 MP) and S10 (16 MP), from left to right respectively. The X-axis depicts the ρ values and Y-axis depicts frequency (ν). A green line indicates a detection threshold, which is calculated based on a fixed probability of false alarm (10^{-6}). From the histograms, we can see that correlation values follow a bimodal distribution. Though most of the correlation values show the expected behavior *i.e.*, less than the threshold, a considerable number of correlation values unexpectedly surpass the threshold. The same behavior is seen for S10+ (12 MP), A50, and S10e. On the other hand, the problem is not that significant for Huawei Models and Redmi Note 7. Notably, for the highest resolutions of Huawei Models, Mate 20 Pro (40 MP) and P30 Pro (40 MP), none of the pair-wise correlations exceed the threshold, and hence the problem does not exist.

Table 1. Details of the Dataset.

Model	Resolution	#Users	#Images
Mate 20 Pro	10 MP	10	700
Mate 20 Pro	20 MP	7	490
Mate 20 Pro	40 MP	2	140
P30 Pro	10 MP	10	700
P30 Pro	40 MP	2	139
S10	12 MP	6	378
S10	16 MP	4	279
S10+	12 MP	8	558
S10+	16 MP	2	140
A50	12 MP	3	209
S10e	12 MP	5	350
Redmi Note 7	12 MP	8	560

Table 2. Validation of the Model.

Model (Resolution)	μ	#Bad Images	Validation
Mate 20 Pro (10 MP)	0	200	10%
Mate 20 Pro (20 MP)	0	413	10.5%
P30 Pro (10 MP)	0	66	30%
S10 (12 MP)	0	185	93.98%
S10 (12 MP)	7	180	99.5%
S10 (16 MP)	0	197	39.5%
S10 (16 MP)	7	136	88.84%
S10+ (12 MP)	0	312	71.74%
S10+ (12 MP)	7	282	87.43%
S10+ (16 MP)	0	113	82.2%
S10+ (16 MP)	7	103	97.96%
A50 (12 MP)	0	84	96.61%
A50 (12 MP)	7	83	100%
S10e (12 MP)	0	118	99.81%
S10e (12 MP)	7	118	99.81%

Table 2 shows the validation of our model based on PCE stats. By validation, we mean the number of times correlation

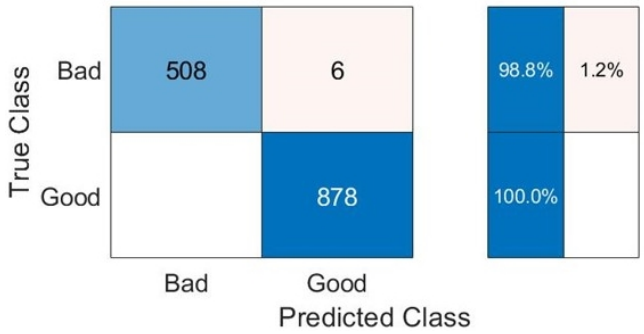

Fig. 2. SPAM Classifier on test data.

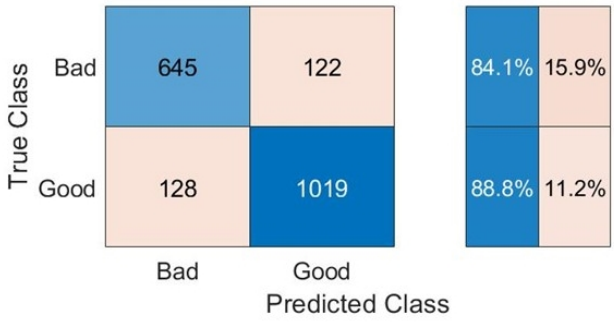

Fig. 3. SceneType Classifier on Samsung models.

between bad images results in values greater than the threshold. For some models, we introduce a margin ($\mu = 7$), *i.e.,* we consider an image as bad only if it is included in at least μ bad correlation values. The rationale is that those images do not have significant NUA; therefore, they can be considered as good images for the validation of the model. We leave the Redmi Note 7 and highest resolutions of Huawei out, owing to 2,0 and 0 number of bad images, respectively. As we can see from Table 2, our model makes sense for Samsung smartphones where the subset of identified bad images is responsible for the unexpected correlation values. Moreover, our proposed model is relatively simple. We might have more than one class of bad images suffering from different NUA due to different processing (Portrait, HDR *etc.*), and this explains why in some cases, *e.g.,* the lowest resolutions of Huawei models, the validation percentages are low.

5.1. Performance of Methods

For the SPAM Classifier, the training set consists of 1171 bad images and 2080 good images, a total of 3251 training examples. We use 20-fold cross-validation to train a KNN-based classifier. The test data consists of 1392 examples, 514 bad images, and 878 good images. The classifier achieves very high accuracy, 100%, and 98.8%, in correctly identifying good and bad images, respectively, from the test data, as shown in the confusion matrix in Figure 2. On the other hand, Meta-data SceneType Classifier, shown in Figure 3, achieves decent accuracy of 84.1% and 88.8% in identifying bad and good images corresponding to Samsung models. For the other models like Huawei and Redmi note 7, the SceneType tag can not be used to identify images suffering from NUA.

For the SPAM Classifier, we do an additional validation on the test images. We perform pair-wise correlations between images predicted as good and bad, respectively, corresponding to different devices of the same smartphone model. Figure 4 reveals the distribution of correlation scores for the images predicted as bad on the left and images predicted as good on the right, for Samsung S10e. For the bad images, we can see that almost all of the pair-wise correlations between classifier-predicted bad images are greater than the threshold. On the other hand, the pair-wise correlations between good images and good and bad images are less than the threshold. Thus, the SPAM Classifier achieves high accuracy in correctly identifying images suffering from NUA.

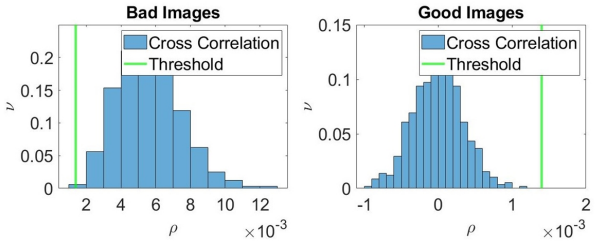

Fig. 4. Normalized Correlation for predicted images.

6. CONCLUSION

In this paper, we verified the problem of PRNU fingerprint collision in modern smartphones. We saw that the problem is significant for Samsung smartphone models, while it is less evident for Huawei models and Redmi Note 7. The highest resolutions of Huawei models did not suffer from NUA. However, in the case of Samsung models, the highest resolutions suffered heavily from the issue of NUA. Therefore, the problem can not be linked to a specific resolution. Moreover, we did not find any link between the standard meta-data settings and the unexpected behavior. We proposed a model to explain the cause of fingerprint collision and validated it on our dataset. Finally, we presented two methods to tackle the problem. The SPAM Classifier achieves high accuracy in correctly identifying images affected by NUA but needs an exact camera model for training. On the other hand, a Meta-data-based Classifier applies to Samsung models only. In practice, the proposed approach could be used to single out images affected by NUA, for example, excluding them from the subset used to compute a reference fingerprint or giving low confidence to tests involving them.

7. REFERENCES

[1] Jessica Fridrich, Jan Lukas, and M Goljan, "Digital camera identification from sensor noise," *IEEE Transactions on Information Security and Forensics*, vol. 1, no. 2, pp. 205–214, 2006.

[2] Mo Chen, Jessica Fridrich, Miroslav Goljan, and Jan Lukás, "Determining image origin and integrity using sensor noise," *IEEE Transactions on information forensics and security*, vol. 3, no. 1, pp. 74–90, 2008.

[3] Massimo Iuliani, Marco Fontani, Dasara Shullani, and Alessandro Piva, "Hybrid reference-based video source identification," *Sensors*, vol. 19, no. 3, pp. 649, 2019.

[4] Francesco Marra, Giovanni Poggi, Carlo Sansone, and Luisa Verdoliva, "Blind PRNU-based image clustering for source identification," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 9, pp. 2197–2211, 2017.

[5] Dasara Shullani, Marco Fontani, Massimo Iuliani, Omar Al Shaya, and Alessandro Piva, "Vision: a video and image dataset for source identification," *EURASIP Journal on Information Security*, vol. 2017, no. 1, pp. 1–16, 2017.

[6] Miroslav Goljan and Jessica Fridrich, "Camera identification from cropped and scaled images," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*. International Society for Optics and Photonics, 2008, vol. 6819, p. 68190E.

[7] Miroslav Goljan, Jessica Fridrich, and Tomáš Filler, "Large scale test of sensor fingerprint camera identification," in *Media forensics and security*. International Society for Optics and Photonics, 2009, vol. 7254, p. 72540I.

[8] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K Mantiuk, and Jonas Unger, "HDR image reconstruction from a single exposure using deep CNNS," *ACM transactions on graphics (TOG)*, vol. 36, no. 6, pp. 1–15, 2017.

[9] Nikolai E Bock, "Apparatus and method for pixel binning in an image sensor," Aug. 15 2006, US Patent 7,091,466.

[10] Massimo Iuliani, Marco Fontani, and Alessandro Piva, "A leak in PRNU based source identification—questioning fingerprint uniqueness," *IEEE Access*, vol. 9, pp. 52455–52463, 2021.

[11] Netherlands Forensic Institute, "PRNU Compare Professional," https://www.forensicinstitute.nl/products-and-services/forensic-products.

[12] Amped Software, "Amped Authenticate," https://ampedsoftware.com/authenticate.

[13] Davide Cozzolino and Luisa Verdoliva, "Noiseprint: a CNN-based camera model fingerprint," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 144–159, 2019.

[14] Tomáš Pevny, Patrick Bas, and Jessica Fridrich, "Steganalysis by subtractive pixel adjacency matrix," *IEEE Transactions on information Forensics and Security*, vol. 5, no. 2, pp. 215–224, 2010.