An Improved Statistic for the Pooled Triangle Test against PRNU-Copy Attack

*Mauro Barni, *Fellow, IEEE*, Héctor Santoyo García, Benedetta Tondi, *Member, IEEE*

Abstract

We propose a new statistic to improve the pooled version of the triangle test used to combat the fingerprint-copy counter-forensic attack against PRNU-based camera identification [1]. As opposed to the original version of the test, the new statistic exploits the one-tail nature of the test, weighting differently positive and negative deviations from the expected value of the correlation between the image under analysis and the candidate images, i.e., those image suspected to have been used during the attack. The experimental results confirm the superior performance of the new test, especially when the conditions of the test are challenging ones, that is when the number of images used for the fingerprint-copy attack is large and the size of the image under test is small.

Index Terms

Forensics and counter-forensics, sensor-based camera identification, camera fingerprint, adversarial signal processing, triangle test.

arXiv:1805.02899v1 [cs.CR] 8 May 2018

An Improved Statistic for the Pooled Triangle Test against PRNU-Copy Attack

I. INTRODUCTION

Photo-Response Non Uniformity (PRNU) noise [2] has been successfully used for forensic camera identification [3] and image forgery detection [4], [5]. Techniques based on PRNU are prone to the so-called fingerprint-copy (or PRNU-copy) attack [6], according to which, a forger, usually referred to as Eve, estimates the PRNU from a set of publicly available images acquired by the camera of a victim, say Alice, and implant the estimated PRNU into an image shot by a different camera. An effective countermeasure against the fingerprint-copy attack is the *triangle test* proposed in [1]. The test exploits the fact that an image forged with the fingerprint-copy attack shares with the images used by Eve to estimate the PRNU other noise components in addition to the PRNU, hence resulting in an unnaturally high correlation between the forged image and the images used to create the forgery. In its simplest version, the triangle test allows Alice to understand *which* images, in a set of publicly available images acquired by her camera, have been used to produce the forgery. In other cases, Alice's goal is *just* to prove that the image under analysis has been forged by means of a fingerprint-copy attack, without the need to identify the exact subset of images used to produce the forgery. To do so, Alice can resort to the *pooled* version of the test [1]. The pooled test is generally very powerful and the effectiveness of the counter-forensic methods proposed so far against the single-image triangle test, e.g. [7], [8], [9], is dramatically reduced when the pooled triangle test is considered.

In this paper, we propose a refined statistic for the pooled triangle test, that allows to improve the performance of the test with particular reference to those situations where the test is less reliable, namely when the number of images Eve has access to is large and when the size of the analysed image is small. The improved statistic relies on the observation that the original pooled test treats in the same way both images exhibiting an unnaturally high correlation with the image under test and those for which this correlation is lower than expected. In this way, the analysis somewhat neglects the one-tail nature of the test[1] according to which the images used for the PRNU-copy attack are expected to exhibit a larger correlation with respect to those that have not been used to create the forgery. The new statistic, on the contrary, accumulates the deviations from the expected correlation by considering their sign. The resulting test, then, decides that the image under analysis has been subject to a PRNU-copy attack only in the presence of positive deviations. The superior performance of the proposed statistic are assessed in a wide variety of cases, by varying the parameters that impact most on the performance of the test, that is, the number N of images used by Eve to estimate the PRNU, the overall number N_c of public images available, and the size of the images.

---

[1]We remark that such an observation does not apply to the single-image version of the test (see eq. (16) in [1]).

The paper is organized as follows. In Section II, we review the PRNU-copy attack and the pooled triangle test. The proposed improved statistic is described in Section III. The results of the experimental validation are presented and thoroughly discussed in Section IV. Eventually, we draw our conclusions and present some directions for future work in Section V.

II. PRNU-COPY ATTACK AND POOLED TRIANGLE TEST

Let us denote with $\mathcal{C}_{1,pub}$ a public dataset of N_c images acquired by Alice's camera C_1. Eve's goal is to take an image J coming from another camera C_2 and modify it in such a way that it looks like as if it was generated by C_1. To do so, Eve estimates the PRNU of C_1 from a subset of N images, I_i, $i = 1, .., N$, belonging to $\mathcal{C}_{1,pub}$, as follows:

$$\hat{K}_E = \frac{\sum_{i=1}^N W_{I_i} I_i}{\sum_{i=1}^N I_i^2}, \tag{1}$$

where $\hat{K}_E$ is the PRNU estimate obtained by Eve, $W_{I_i} = I_i - F(I_i)$ is the noise residual of I_i, and F is a denoising filter, e.g. the one in [10]. The noise residual has the form $W_{I_i} = I_i K + \theta$, where K is the true PRNU of C_1 and θ collects the non-PRNU noise components of the residual [2]. Then, Eve superimposes the estimated PRNU onto J, obtaining the forged image

$$J' = [J(1 + \alpha \hat{K}_E)], \tag{2}$$

where $[\cdot]$ indicates rounding to integers and α is the fingerprint strength. The value of α must be sufficiently large to pass the threshold-based correlation test (see below), but, at the same time, as small as possible to make the forgery undetectable.

On the analyst side, camera attribution is carried out by relying on a threshold-based correlation test, that is by computing $\rho = \text{corr}(W_I, I\hat{K}_A)$, where I is the image under test, and $\hat{K}_A$ is Alice's estimation of the PRNU fingerprint of C_1, which can be reliably obtained from a limited number of flat-field images. Image I is attributed to C_1, if ρ is above a threshold, set by fixing the false alarm probability. The forged image J' can easily pass the correlation test [6], thus being wrongly attributed to C_1.

As a countermeasure, Alice can apply the *triangle test* [1] to the images attributed to C_1, to determine if they are genuine images shot by C_1, or they are the result of a PRNU-copy attack. The idea behind the triangle test is the following: each image I_i used by Eve to estimate K, shares with the forged image J' not only the PRNU term (as it happens for a genuine - non forged - image), but also the other terms of the noise residual W_{I_i}; then, the correlation of the residual of J' with the one of I_i, namely $c_{I_i,J'} = \text{corr}(W_{I_i}, W_{J'})$, is typically larger when J' is a forgery and image I_i has been used to estimate the fingerprint implanted in J'.

By following [1], given a non-forged image J and an image I from C_1, it is possible to compute the expected value of $c_{I,J}$, named $\hat{c}_{I,J}$. The dependence between the real value of $c_{I,J}$ and $\hat{c}_{I,J}$ when I has not been used by Eve to forge J, is well fit by a straight line, hereafter referred to as *inference line*, $c_{I,J} = \lambda \hat{c}_{I,J} + \eta$, for some slope λ and intercept η. On the contrary, if I has been used by Eve to forge J', the correlation $c_{I,J'}$ takes much
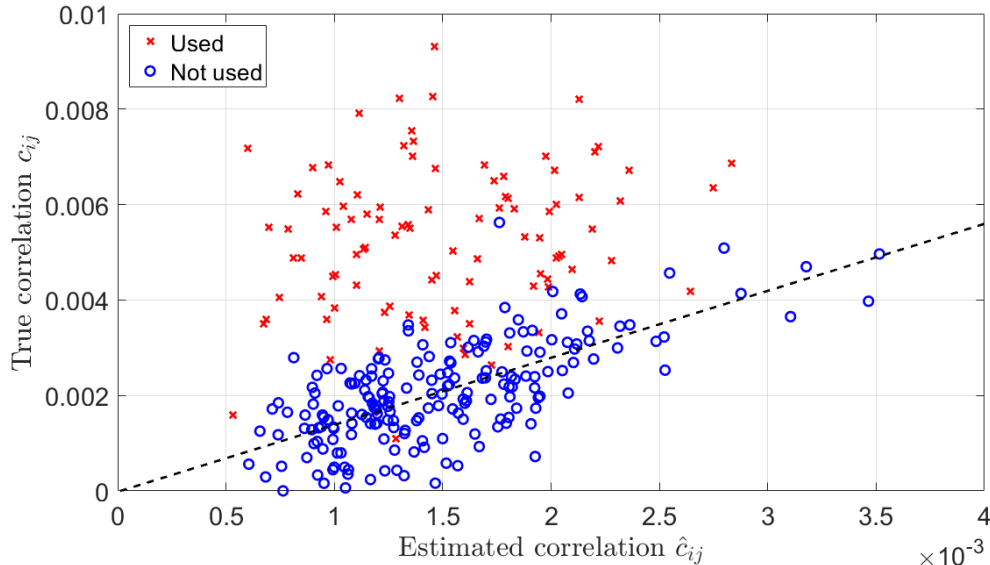
Fig. 1: True correlation $c_{I,J'}$ as a function of the estimated correlation $\hat{c}_{I,J'}$ for an image J' forged by Eve with $N = 100$ ($N_c = 300$).

larger values. Figure 1 illustrates a typical plot of $c_{I,J'}$ as a function of $\hat{c}_{I,J'}$ for a forged image J' for $N = 100$, when $N_c = 300$.

For notational simplicity, in the following, given a test image J and a candidate image I_i, we let $d_{J,i} = c_{I_i,J} - \lambda \hat{c}_{I_i,J} - \eta$. In [1] it is shown that the distribution of $d_{J,i}$ is approximately constant with I_i (and $\hat{c}_{I_i,J}$), so we can write:

$$Pr\{d_{J,i} = x | \hat{c}_{I_i,J}\} \approx f_J(x), \tag{3}$$

for some f_J, independent of I_i and $\hat{c}_{I,J}$. Let, μ_J and σ_J denote the mean and variance of $d_{J,i}$ when I_i is not used by Eve to create the forgery J^2 (expectedly, μ_J is very close to 0). In [1], it is argued that f_J is often close to a Gaussian distribution, that is $f_J \sim \mathcal{N}(\mu_J, \sigma_J)$, even if for some images a Student's t-distribution may be a more conservative choice. For sake of brevity, in the following, we stick to the Gaussian model, the difference with respect to the Student's t-model being very small based on our experiments.

A. The pooled triangle test

Let J be the to-be-tested image and let H_0 be the hypothesis that J has not been forged, or, equivalently in our scenario, that no image in $\mathcal{C}_{1,pub}$ has been used by Eve to forge J. Let H_1 be the opposite hypothesis that some of the images in $\mathcal{C}_{1,pub}$ have been used to forge J. Let k be the number of candidate images considered by Alice to carry out the test (we have $k = N_c$ when the entire public set is used for the test). We denote with $\mathcal{C}_{1,pub}^k$ the corresponding subset. The pooled triangle test described in [1] uses the following statistic to decide if some of the

---

[2]This may either correspond to a situation in which J is a forgery but I_i has not been used to create it, or to a case in which J is not a forgery.

images in $\mathcal{C}_{1,pub}^k$ have been used to forge J:

$$L_k^J = \sum_{i \in \mathcal{C}_{1,pub}^k} \log\left(f_J(d_{J,i})\right). \tag{4}$$

When f_J is a Gaussian, testing L_k^J is very similar in spirit to base the test on the sum of the squared distances. In fact, in such a case, we have

$$L_k^J = -k\log(\sqrt{2\pi\sigma^2}) - \sum_{i \in \mathcal{C}_{1,pub}^k} \left(\frac{d_{J,i} - \mu_J}{\sqrt{2}\sigma_J}\right)^2. \tag{5}$$

By observing that L_k^J corresponds to the log-likelyhood of the deviations $d_{J,i}$ under H_0, the image J is said to be a forgery if $L_k^J < T$, where T is set by fixing the false alarm probability.

III. An improved statistic for the pooled test

A limit of a test based on L_k^J is that such a statistic considers (the log of) the probability of observing the deviations $d_{J,i}$'s under H_0 without exploiting the knowledge we have about the distribution of $d_{J,i}$ under H_1. In fact, even if the exact distribution of $d_{J,i}$ under H_1 is not known, we know that when the image I_i has been used to forge J, the measured correlation $c_{I,J}$ tends to be larger than expected, hence resulting in a larger, positive, value of $d_{J,i}$. More precisely, by assuming (w.l.o.g.) that μ_J is 0, we know that (see also Figure 1):

$$Pr\{d_{J,i} < 0 | \ I_i \text{ used to forge } J\} < Pr\{d_{J,i} < 0 | \ I_i \text{ not used}\}. \tag{6}$$

This is the typical example of one-tailed statistical test, for which the sign of the deviation from the expected value should be taken into account in addition to the magnitude of the deviation. Such one-tailed nature of the test is discarded with the statistic in (5), which, by looking at the quadratic distances $d_{J,i}$, implicitly assumes that a large positive and a large negative value of $d_{J,i}$ are equally probable when I_i is used by Eve for the PRNU-copy attack. Note that, even if we exemplified this problem by assuming a Gaussian distribution for $d_{J,i}$, the above observations are generally true for any distribution f_J. Based on the above observation, we propose to replace L_k^J with a new statistic that takes into account the sign of the deviation $d_{J,i}$, with the understanding that only positive values contribute to form the evidence that J has been forged by Eve. Specifically, we suggest to replace L_k^J with the following:

$$V_k^J = \sum_{i \in \mathcal{C}_{1,pub}^k} \text{sign}(d_{J,i} - \mu_J) \left(\frac{d_{J,i} - \mu_J}{\sigma_J}\right)^2, \tag{7}$$

where, as before, μ_J and σ_J are the mean and variance of $d_{J,i}$ under the hypothesis that I_i has not been used by Eve to forge J[3]. With reference to (5), it is evident that the main difference between L_k^J and V_k^J is the dependence of V_k^J on the sign of $d_{J,i} - \mu_J$. In this way, V_k^J exploits the knowledge that $Pr\{d_{J,i} < \mu_J | H_0\} > Pr\{d_{J,i} < \mu_J | H_1\}$, thus resulting in a more accurate test. An additional advantage of directly considering the distances from the inference line rather than the probability values, is that we do not need to make any assumption on the distribution of

---

[3]Following [1], the pooled test is implemented by replacing μ_J and σ_J with their sample estimates.

$d_{J,i}$ for the images not used by Eve (f_J). In general, other n-powers could be considered for the distance term $(d_{J,i} - \mu_J)/\sigma_J$ in (7). For instance, we run some experiments by accumulating linear rather than quadratic distances obtaining similar results. In this paper, we chose the square distances to ease the comparison with the statistic L_k^J, which in fact results in the accumulation of quadratic distances when f_J is a Gaussian (see (5)).

Eventually, the test decides in favour of H_1 if $V_k^J > T'$, where the threshold T' is fixed by imposing a constraint on the false alarm probability. On this regard, we observe that, as for L_k^J, there are two sources of randomness in V_k^J, namely J and $\mathcal{C}_{1,pub}^k$[4]. Then, the false alarm probability can be evaluated by varying either $\mathcal{C}_{1,pub}^k$ or J. In the former case (which is the approach followed in [1] to test the performance of L_k^J), J is fixed, and the distribution of V_k^J under H_0 can be theoretically approximated to a Gaussian. The terms of the sum in (7), in fact, are independent under H_0, although they are not identically distributed because of the presence of the sign. The central limit theorem can then be applied (the Lindeberg condition [11] is satisfied), and V_k^J assumed to be normally distributed, thus allowing to set the threshold T' theoretically.

IV. EXPERIMENTS

We run our tests by considering the Nikon D7000 camera (C_1) and the Nikon D90 camera (C_2) in the RAISE dataset [12]. We split the images from C_1 as follows: a total number of 1000 images were used to build the public set $\mathcal{C}_{1,pub}$ (in some experiments only a subset of 600 images was used as $\mathcal{C}_{1,pub}$); 300 images were used to build the private set $\mathcal{C}_{1,priv}^{(1)}$, used by Alice to estimate the parameters of the triangle test, that is, to estimate λ and η and build the inference line; another set $\mathcal{C}_{1,priv}^{(2)}$ of 300 images was used to establish the decision threshold of the correlation detector (with a true positive rate set to 0.9). Other 300 images, passing the correlation test, formed a third set $\mathcal{C}_{1,priv}^{(3)}$ used in the experiments to simulate H_0. Eventually, all the 100 flat-field images available in the RAISE dataset for the camera C_1 were used to estimate the PRNU. A number of 300 images coming from a camera Nikon D90 were used to build Eve's set $\mathcal{C}_2$. The original sizes of the images from Eve's and Alice's cameras C_1 and C_2 are different. In our experiments, we considered image sizes of 1936×1296 (medium size) and 1024×1024 (small size) pixels, obtained by cropping the central parts of the images from C_1 and C_2. With regard to the fingerprint-copy attack performed by Eve, for simplicity, we considered the minimum strength α resulting in a positive identification in the correlation test. This is a worst case assumption for Alice, since in practice Eve can not reproduce exactly Alice's test, and then she will apply an α which is larger than such a minimum value to be sure to pass the test.

We run our experiments by considering two slightly different versions of the pooled test, corresponding to two different interpretations of the error probability and, in particular, the false alarm probability. The two resulting settings correspond to the following testing conditions:

a) Given a test image J, the error probabilities are computed by varying the subset of k images used to compute

---

[4]Strictly speaking, L and V depend on the set $\mathcal{C}_{1,pub}^k$. With a slight abuse of notation, we simply denote such a dependence with the letter k in the pedex.

V_k^J (res. L_k^J). In this setting, the false alarm probability corresponds to the probability that, given J, k images at random taken from $\mathcal{C}_{1,pub}$ result in a value of V_k^J (res. L_k^J) larger (res. lower), than the detection threshold;

b) Given k images in $\mathcal{C}_{1,pub}$, the error probabilities are computed by varying the to-be-tested image J. In particular, the false alarm probability corresponds to the probability that C_1 produces an image for which V_k^J (res. L_k^J) is larger (res. lower), than the detection threshold.

Two considerations are in order. The setup a) is equal to the one used in [1]. As we have already noticed, in this case both V_k^J and L_k^J can be assumed to be normally distributed, hence the detection threshold can be determined theoretically by fixing the false alarm probability and estimating the mean and variance of the test statistic by resorting to bootstrapping (as in [1]). With regard to b), the distribution of the statistics V_k^J and L_k^J under H_0 is not known, so it is not possible to set the detection threshold theoretically by fixing the false alarm probability. In this case, then, we evaluated the performance of the test by plotting the ROC curve of the test any evaluating the missed detection probability for a given false alarm probability set by choosing a suitable operating point on the ROC curve.

A. Performance of the test for the setup a)

To test the performance in this case, we fixed the forged image J, obtained by taking an image in $\mathcal{C}_2$ and applying the attack in (2). Then, we picked a random set of k images out of the N_c images in $\mathcal{C}_{1,pub}$, and we computed the statistics V_k^J and L_k^J. We repeated this procedure by changing the random selection of the k images, thus getting a number of observations for both statistics under H_1. Finally, we measured the correct detection probability P_d, for a fixed theoretical target P_{fa}. Specifically, we computed the p-value corresponding to the observed statistics and the image J is said to be forged if the p-value of the observation is lower than P_{fa}. From the discussion in the previous section, the p-value is computed by considering the Gaussian model for V_k^J (res. L_k^J) under H_0. As in [1], we let $k = 60$, then we evaluated P_d by bootstrapping, i.e., by repeating the process 30000 times, each time changing the random selection of k images in $\mathcal{C}_{1,pub}$. Figure 2 shows the results of the tests carried out on 2 randomly chosen images in $\mathcal{C}_2$. The tests were run for various values of N, with $N_c = 600$, and $P_{fa} = 10^{-3}$. For each N, the to-be-implanted PRNU $\hat{K}_E$ is estimated from N randomly chosen images in the candidate set. The size of the images is 1936×1296. We can see that the use of the improved statistic V_k^J brings a significant advantage when $N/N_c > 0.5$, while for small values of the ratio N/N_c, the new and the old statistics behave similarly. A similar behaviour is observed for different values of N_c. In general, the difference between V_k^J and L_k^J can be better appreciated when N_c is large (say $N_c > 300$), since when N_c is small the pooled test is very powerful and both statistics works very well.

B. Performance of the test for the setup b)

In this case, we fixed $\mathcal{C}_{1,pub}^k$ and run the pooled test by varying the test image J. We computed the statistics V_k^J and L_k^J by forging the images in $\mathcal{C}_2$, whereas the values under H_0 were obtained by considering the images in $\mathcal{C}_{1,priv}^3$. Throughout these these experiments we let $k = N_c$. This is a reasonable assumption that corresponds to assuming that Alice knows the entire public set available to Eve.
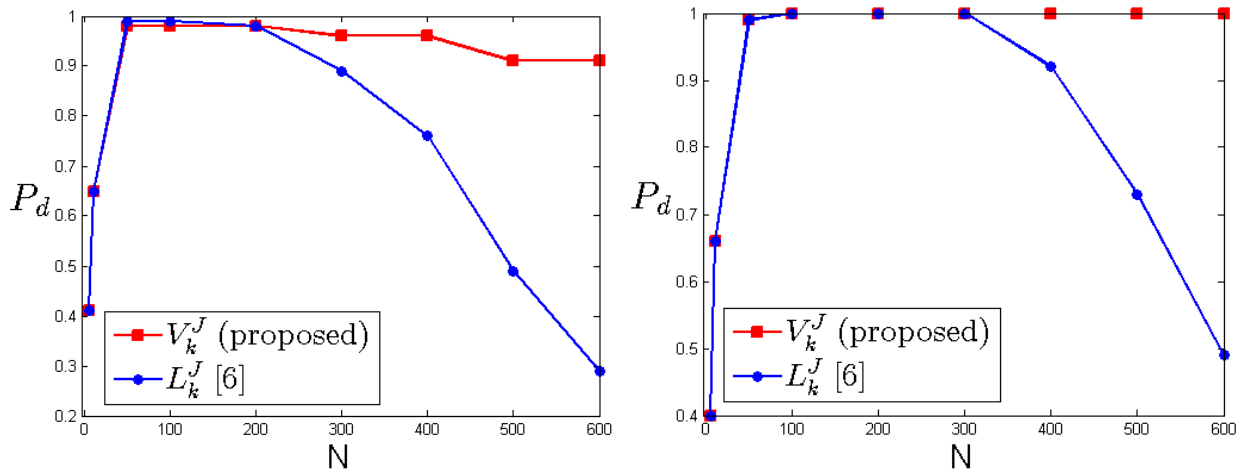
Fig. 2: P_d as a function of N for 2 images in $\mathcal{C}_2$; $P_{fa} = 10^{-3}$, $N_c = 600$. The minimum N considered is 4.

The values of P_d obtained from the ROC curve by fixing the false alarm probability to 0.03 are reported in Figure 3 for various values of N ($N_c = 600$), for both small and medium size images. The advantage of the improved statistic increases with N. Expectedly, with small images the performance of the pooled test are lower and the difference between the two statistics is more evident. We observe that the test achieves perfect results also when N/N_c is very low. This is a consequence of the fact that $k = N_c$ (or, more in general, that k is comparable to N_c), since with this choice the pooled test is very reliable especially when N/N_c is small. A similar behaviour holds for other values of N_c. Figure 4 shows the results we have got with $N_c = 1000$ in the least favorable case of small size images. We see that the test with V_k^J is still reliable with such a large N_c: in particular, at $N = 1000$, we get $P_d = 0.95$, while, for the test with L_k^J, P_d is 0.0767. We verified that for the case of medium size images we still get very close-to-ideal performance with $N_c = 1000$ (in the most difficult case with $N = 1000$, we get $P_d = 0.99$ with V_k^J, and $P_d = 0.15$ with L_k^J).

V. CONCLUSIONS

We have proposed a new statistic for the pooled triangle test originally introduced in [1]. The improved statistic is based on the observation that the statistic proposed in [1] somewhat neglects the one-tailed nature of the test. Experiments show that the proposed statistic achieves better results, especially in the most challenging case when the number of images N used by Eve for the fingerprint-copy attack is large (and comparable to N_c). Further tests could be carried out to investigate the limit values of N (and N_c) for which the test based on the new statistic is still reliable. As a further work, we plan to evaluate the performance of the pooled test based on the improved statistic in the presence of targeted attacks like those introduced in [7], [9].
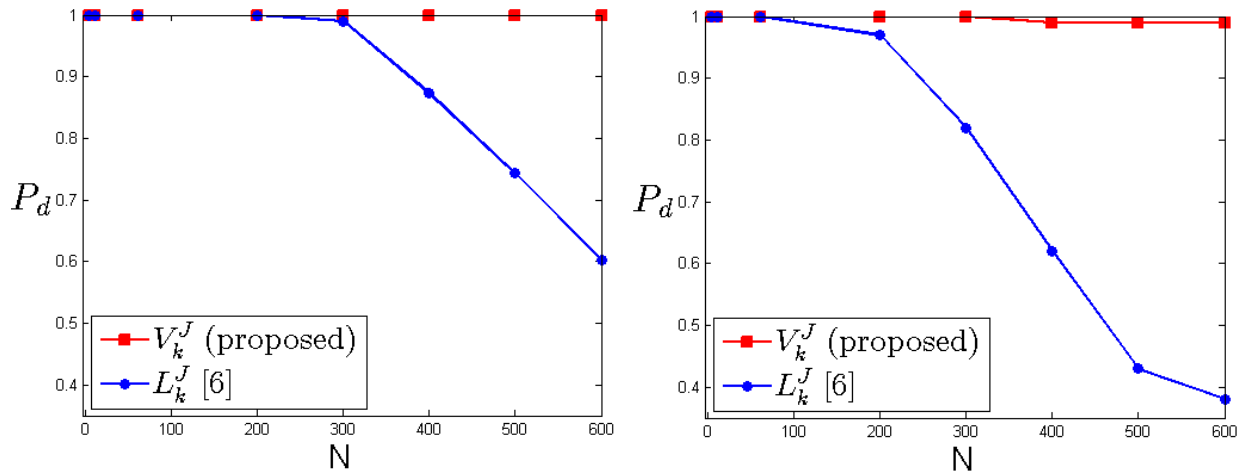
Fig. 3: P_d values obtained from the ROC curve by letting $P_{fa} = 0.03$, $N_c = 600$. Image size: 1936×1296 (left) and 1024×1024 (right). The minimum N considered is $N = 4$.
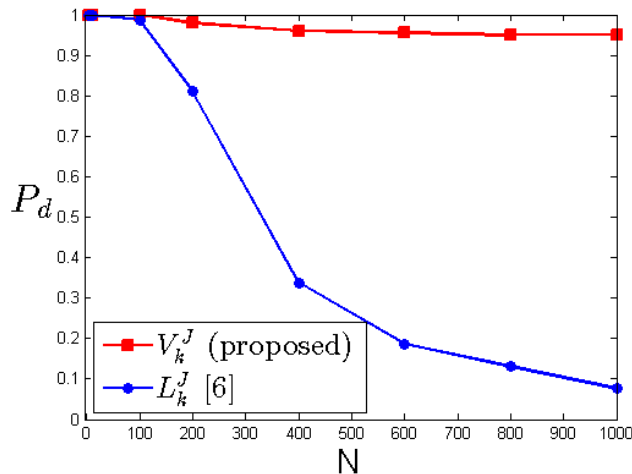

Fig. 4: P_d values obtained from the ROC curve by letting $P_{fa} = 0.03$, $N_c = 1000$. Image size: 1024×1024.

REFERENCES

[1] M. Goljan, J. Fridrich, and M. Chen, "Defending against fingerprint-copy attack in sensor-based camera identification," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 1, pp. 227–236, March 2011.

[2] M. Chen, J. Fridrich, M. Goljan, and J. Lukas, "Determining image origin and integrity using sensor noise," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 74–90, March 2008.

[3] C. T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, June 2010.

[4] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A bayesian-mrf approach for prnu-based image forgery detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 554–567, April 2014.

[5] J. Lukáš, J. Fridrich, and M. Goljan, "Detecting digital image forgeries using sensor pattern noise," in *Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072.   International Society for Optics and Photonics, 2006, p. 60720Y.

[6] T. Gloe, M. Kirchner, A. Winkler, and R. Böhme, "Can we trust digital image forensics?" in *Proceedings of the 15th ACM International Conference on Multimedia*, ser. MM '07.   New York, NY, USA: ACM, 2007, pp. 78–86. [Online]. Available: http://doi.acm.org/10.1145/1291233.1291252

[7] Q. Rao, H. Li, W. Luo, and J. Huang, "Anti-forensics of the triangle test by random fingerprint-copy attack," in *Computational Visual Media Conference*, 2013, pp. 1–6.

[8] F. Marra, F. Roli, D. Cozzolino, C. Sansone, and L. Verdoliva, "Attacking the triangle test in sensor-based camera identification," in *2014 IEEE International Conference on Image Processing (ICIP)*, Oct 2014, pp. 5307–5311.

[9] R. Caldelli, I. Amerini, and A. Novi, "An analysis on attacker actions in fingerprint-copy attack in source camera identification," in *2011 IEEE International Workshop on Information Forensics and Security*, Nov 2011, pp. 1–6.

[10] M. K. Mihcak, I. Kozintsev, and K. Ramchandran, "Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising," in *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No.99CH36258)*, vol. 6, Mar 1999, pp. 3253–3256 vol.6.

[11] P. Billingsley, *Probability and Measure*, 2nd ed.   John Wiley and Sons, 1986.

[12] D.-T. Dang-Nguyen, C. Pasquini, V. Conotter, and G. Boato, "Raise: A raw images dataset for digital image forensics," in *Proceedings of the 6th ACM Multimedia Systems Conference*, ser. MMSys '15.   New York, NY, USA: ACM, 2015, pp. 219–224. [Online]. Available: http://doi.acm.org/10.1145/2713168.2713194