Defending Against Fingerprint-Copy Attack in Sensor-Based Camera Identification

Miroslav Goljan, *Member*, *IEEE*, Jessica Fridrich, *Member*, *IEEE*, and Mo Chen

Abstract—**Sensor photo-response non-uniformity has been proposed as a unique identifier (fingerprint) for various forensic tasks, including digital-camera ballistics in which an image is matched to a specific camera that took it. The problem investigated here concerns the situation when an adversary estimates the sensor fingerprint from a set of images and superimposes it onto an image from a different camera to frame an innocent victim. This paper proposes a reliable method for detecting such fake fingerprints under rather mild and general assumptions about the adversary's activity and the means available to the victim. The proposed method is subjected to experiments to evaluate its reliability as well as its limitations. The conclusion that can be made from this study is that planting a sensor fingerprint in an image without leaving a trace is significantly more difficult than previously thought.**

Index Terms—**Camera identification, digital forensics, photo-response non-uniformity, sensor fingerprint, counter-forensics.**

I. Introduction

Photo-response non-uniformity (PRNU) of imaging sensors can be used as a unique fingerprint to address various forensic tasks involving digital images and video [1]. One of the most important applications of this technology is matching an image or a video clip to the camera that took it, which is a task similar in spirit to matching a bullet to a gun barrel. Since every image taken by a given sensor contains its PRNU signal, an image can be matched to the sensor (camera) by proving that its noise residual contains the same PRNU. The sensor fingerprint can be estimated by averaging noise components of natural images [2][3]. Because the fingerprint is essentially a random spread-spectrum signal, it can be detected using some form of a matched filter [2]–[4].

Since the inception of this technology in 2005, researchers have realized that the fingerprint can be copied onto an image that did not come from the camera and thus frame an innocent victim. In the most typical and quite plausible scenario, Alice, the victim, posts her images on the Internet. Eve, the adversary, estimates the fingerprint of Alice's camera and properly superimposes it onto another image. Indeed, as already shown in the original publication [2] and, more recently, in [5][6], threshold-based correlation detectors cannot distinguish between a genuine fingerprint and a fake one.

The goal of this paper is to develop a countermeasure against this fingerprint-copy attack. In the next section, we introduce the notation and include the basics of sensor-based camera identification needed in this work. Section III discusses the options available to the adversary and explains the assumptions under which we will work. The actual techniques are described in Section IV, which is divided into two subsections, each addressing a different scenario depending on the means available to the adversary and the victim. Experimental validation of all proposed techniques and analysis of their limitations appear in Section V. The last experimental Section VI contains the results of experiments whose aim is to evaluate how the strength with which the adversary adds the fake fingerprint influences the reliability of the proposed methods. The paper is concluded in Section VII, where we summarize the paper and discuss the consequences of the newly obtained results.

II. Background

A. Notation

Everywhere in this article, boldface symbols represent either vectors or matrices. At times, it will be convenient to index a matrix using a one-dimensional index instead of an index pair. It is hoped that switching between these two index types will cause no confusion. For two matrices of the same dimensions, $\mathbf{X}$ and $\mathbf{Y}$, their element-wise product (or element-wise division) is a matrix $\mathbf{Z}$ of the same dimensions, $\mathbf{Z}[i,j] = \mathbf{X}[i,j]\,\mathbf{Y}[i,j]$ (or $\mathbf{Z}[i,j] = \mathbf{X}[i,j] / \mathbf{Y}[i,j]$), and we simply write $\mathbf{Z} = \mathbf{XY}$ (or $\mathbf{Z} = \mathbf{X}/\mathbf{Y}$). The dot product is denoted as $\mathbf{X} \odot \mathbf{Y} = \sum_{i=1} \mathbf{X}[i]\mathbf{Y}[i]$ with $\|\mathbf{X}\| = \sqrt{\mathbf{X} \odot \mathbf{X}}$ being the L_2 norm of $\mathbf{X}$. Denoting the sample mean with a bar, the normalized correlation is defined as

$$corr(\mathbf{X},\mathbf{Y}) = \frac{(\mathbf{X}-\bar{\mathbf{X}}) \odot (\mathbf{Y}-\bar{\mathbf{Y}})}{\|\mathbf{X}-\bar{\mathbf{X}}\| \cdot \|\mathbf{Y}-\bar{\mathbf{Y}}\|}. \quad (1)$$

B. Fingerprint estimation and detection

For any digital image $\mathbf{I}$ and a denoising filter F, the noise residual of $\mathbf{I}$ is defined as $\mathbf{W_I} = \mathbf{I} - \mathrm{F}(\mathbf{I})$. The PRNU signal can be captured using a multiplicative factor $\mathbf{K}$, which plays the role of a sensor "fingerprint." Adopting the model described in [7], the noise residual has the form:

$$\mathbf{W_I} = \mathbf{aIK} + \mathbf{\Theta}, \quad (2)$$

where $\mathbf{\Theta}$ stands for all other noise components, such as the shot noise or the readout noise, and $\mathbf{a}$ is an attenuation factor of the same dimension as $\mathbf{K}$. In general, $\mathbf{a}$ depends on the image content and the processing to which $\mathbf{I}$ was subjected to.

When modeling $\mathbf{\Theta}$ in (2) as an i.i.d. Gaussian, the maximum likelihood estimator of the PRNU factor $\mathbf{K}$ from N noise residuals $\mathbf{W}^{(i)} = \mathbf{W}_{\mathbf{I}^{(i)}}$, $i = 1, \ldots, N$, has the form [7]:

$$\hat{\mathbf{K}} = \frac{\sum_{i=1}^{N} \mathbf{W}^{(i)} \mathbf{I}^{(i)}}{\sum_{i=1}^{N} \left(\mathbf{I}^{(i)}\right)^2}. \tag{3}$$

The quality of the fingerprint estimate $\hat{\mathbf{K}}$ is defined as

$$q = corr(\mathbf{K}, \hat{\mathbf{K}}). \tag{4}$$

For camera identification, it is important that the fingerprint not contain any other artifacts (called Non-Unique Artifacts or NUAs in [8]) that might be common across sensors/cameras of the same make because such artifacts are not unique to each particular sensor and would increase the false alarm. Since most of these artifacts are due to demosaicking algorithms that depend on the Color Filter Array (CFA) and are periodic in nature, they can be removed by zero-meaning the rows and columns of $\hat{\mathbf{K}}$ separately for each pixel type as defined by the CFA. Assuming $\hat{\mathbf{K}}$ is an $m \times n$ matrix, for the Bayer CFA there are four types of pixels forming four interleaved submatrices $\hat{\mathbf{K}}^{\mathrm{T}}$, $\mathrm{T} \in \{\mathrm{G1}, \mathrm{G2}, \mathrm{R}, \mathrm{B}\}$, where $\hat{\mathbf{K}}^{\mathrm{T}}$ is of dimension $(m/2) \times (n/2)$. The operation of zero-meaning is described using the following pseudo-code executed for all four T:

```
$r_i = 2/n \sum_{j=1}^{n/2} \hat{\mathbf{K}}^{\mathrm{T}}[i,j]$
for $i$=1 to $m/2$ { $\hat{\mathbf{K}}^{\mathrm{T}}[i,j] \leftarrow \hat{\mathbf{K}}^{\mathrm{T}}[i,j] - r_i$ for $j$=1, …, $n/2$}
$c_j = 2/m \sum_{i=1}^{m/2} \hat{\mathbf{K}}^{\mathrm{T}}[i,j]$
for $j$=1 to $n/2$ { $\hat{\mathbf{K}}^{\mathrm{T}}[i,j] \leftarrow \hat{\mathbf{K}}^{\mathrm{T}}[i,j] - c_j$ for $i$=1, …, $m/2$}
```

Reassembling the four submatrices into one matrix, the final fingerprint estimate is further processed in the DFT domain using a Wiener filter W: $\hat{\mathbf{K}} \leftarrow \hat{\mathbf{K}} - \mathrm{W}(\hat{\mathbf{K}})$ to further suppress any remaining NUAs, such as non-periodic artifacts [8].

If no geometric transform was applied to image **J** (e.g., cropping, scaling, and digital zoom), the presence of the camera fingerprint in **J** is established through the correlation detector:

$$\rho = corr(\mathbf{W}_{\mathbf{J}}, \mathbf{J}\hat{\mathbf{K}}), \tag{5}$$

where $\hat{\mathbf{K}}$ is a fingerprint estimate. Alternative statistics proposed for the detection based on different modeling assumptions include the generalized matched filter (GMF) [7] and the peak to correlation energy (PCE) ratio [4].

III. The Fingerprint-Copy Attack

In this section, we position ourselves into the role of the adversary Eve and analyze the impact of her actions on how difficult it will be for Alice to reveal the copied fingerprint. We assume that Alice owns a digital camera *C*. Eve takes an image **J** from a different camera *C*' with fingerprint $\mathbf{K}' \neq \mathbf{K}$ and makes it appear as if it was taken by *C*. She does so by first estimating the fingerprint of *C* from some set of Alice's images and then properly superimposes it onto **J**. Next, we detail Eve's options in her forging activity.

[Fingerprint estimation] Eve has access to *N* images, $\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}$, from *C* and estimates its fingerprint using an algorithm Φ:

$$\hat{\mathbf{K}}_{\mathrm{E}} = \Phi(\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}; P_{\mathrm{K}}). \tag{6}$$

The symbol P_{K} stands for the settings of the estimation procedure, which might include the choice of the denoising filter used to extract the noise component from images, the parameters of this filter, or the formula for aggregation of the noise residuals. Fundamentally, any estimation procedure will be some form of averaging of the noise residuals:

$$\hat{\mathbf{K}}_{\mathrm{E}} = \sum_{i=1}^{N} \mathbf{h}^{(i)} \mathbf{W}^{(i)}, \tag{7}$$

where $\mathbf{h}^{(i)} = 1/N$ for simple averaging [9] or $\mathbf{h}^{(i)} = \mathbf{I}^{(i)} / \sum_{i=1}^{N} (\mathbf{I}^{(i)})^2$ for the estimator (3).

[Preprocessing] Having estimated the fingerprint, Eve may preprocess **J** to suppress the PRNU term $\mathbf{J}\mathbf{K}'$ introduced by the sensor in *C*' and/or to remove any artifacts in **J** that are incompatible with *C*. Because suppressing the PRNU term is not an easy task [10], quite likely the best option for Eve is to skip this step altogether. This is because the PRNU component $\mathbf{J}\mathbf{K}'$ in **J** is very weak[1] to be detected per se and because the chances that Alice will gain access to *C*' may be rather small. In fact, Eve should avoid processing **J** too much as it may introduce artifacts of its own.

If Eve compresses her forgery using JPEG, she needs to make sure that the quantization table is compatible with camera *C*, otherwise Alice will know that the image has been manipulated and did not come directly from her camera. If camera *C*' uses different quantization tables than *C*, Eve will inevitably introduce double-compression artifacts into **J**, giving Alice again a starting point of her defense.

Unless *C* and *C*' are of the same model, the forged image may contain color-interpolation artifacts of *C*' incompatible with those of *C*. Alice could leverage upon techniques developed for camera brand/model identification [11] and prove that there is a mismatch between the camera model and the color interpolation artifacts. A knowledgeable adversary may, in turn, attempt to remove such artifacts of *C*' and introduce interpolation artifacts of *C*, for example, using the method described in [12]. It is now apparent that it is far from easy to create a "perfect" forgery.

While it is certainly possible for Alice to utilize traces of previous compression or color interpolation artifacts, no attempt is made in this paper to exploit these discrepancies to reveal the forged fingerprint. Our goal is to develop techniques capable of identifying images forged by Eve even in the most difficult scenario for Alice when *C*' is of exactly the same model as *C* to avoid any incompatibility issues discussed above. Thus, we do not allow Alice to take advantage of knowing any a priori information about *C*'.

[1] The energy of the PRNU component is around 51dB depending on the camera model and the average luminance of **J**.

[Forging] The final step for Eve is to plant the estimated fingerprint in $\mathbf{J}$, creating thus the forged image $\mathbf{J}'$. In her attempt to mimic the acquisition process, and in accordance with (2), Eve superimposes the fake fingerprint multiplicatively, which is what would happen if $\mathbf{J}$ was indeed taken by C:

$$\mathbf{J}' = [\mathbf{J}(1+\alpha\hat{\mathbf{K}}_\mathrm{E})], \tag{8}$$

where $\alpha > 0$ is a scalar fingerprint strength and $[x]$ is the operation of rounding x to integers forming the dynamic range of $\mathbf{J}$. Finally, Eve saves $\mathbf{J}'$ as JPEG with the same or similar quantization table as that of the original image $\mathbf{J}$. Formula (8) should be understood as three equations for each color channel of $\mathbf{J}$.

As already reported in [2][5][6], this attack succeeds in fooling the camera identification algorithm in the sense that the response of the fingerprint detector on $\mathbf{J}'$ (either the correlation (5), the GMF [7], or the PCE [13]) will be high enough to indicate that $\mathbf{J}'$ was indeed taken by camera C.

A very important issue for Eve is the choice of the strength α. We call α the "natural" strength if $\mathbf{J}'$ elicits the same response of the GMF fingerprint detector [7] implemented with the true fingerprint $\mathbf{K}$ as when $\mathbf{J}'$ was indeed taken by C. By selecting the natural strength, Eve essentially creates the most natural-looking forgery in which the fingerprint is not suspiciously weak or strong as this would likely give Alice additional avenues for her defense.[2]

The natural strength can be estimated using a predictor of the detector response Pred($\mathbf{J}'$,$\mathbf{K}$), such as the one described in [7]. While it is certainly true that Eve cannot easily construct the predictor because she does not have access to the true fingerprint $\mathbf{K}$, she may select the natural strength α by pure luck. To be more precise here, we grant Eve the ability to guess the natural strength instead of giving her access to $\mathbf{K}$. We note that similar assumptions postulating a clairvoyant adversary are commonly made in many branches of information security.

The main bulk of experiments described in Section V is carried out with the natural fingerprint strength. In Section VI, we investigate how the performance of the proposed methods changes when Eve uses other values of α.

IV. Detecting fake fingerprints

Here, we describe a test using which Alice can decide whether an image indeed came from her camera or whether it was forged by Eve as described in Section III. We separately discuss two cases that differ by what data is available to Alice for her forensic investigation and by the actions of Eve. The first and perhaps more plausible case is analyzed in Section IV.A. It is assumed that Eve created one forged image $\mathbf{J}'$ and Alice has access to at least one of the images, $\mathbf{I}^{(1)},\ldots,\mathbf{I}^{(N)}$, used by Eve to estimate $\hat{\mathbf{K}}_\mathrm{E}$. In the second case detailed in Section IV.B, Eve has forged two or more images, $\mathbf{J}_1'$ and $\mathbf{J}_2'$.

A. Single forged image

In this scenario, some of the N images used by Eve to estimate $\hat{\mathbf{K}}_\mathrm{E}$ are available to Alice but Alice does not know which they are. She has a set of $N_\mathrm{c} \geq N$ candidate images that Eve may have possibly used. This is a very plausible scenario because, unless Eve gains access to Alice's camera and takes images of her own and then removes them from the camera before returning the camera to Alice, Eve will have little choice but to use images taken by Alice, such as images posted by Alice on the Internet. In this case, Alice can prove that the forged image did not originally come from her camera by identifying among her candidate images those used by Eve.

We now explain the key observation based on which Alice can construct her defense. Let $\mathbf{I}$ be one of the N images available to Alice that Eve used to forge $\mathbf{J}'$. Since the noise residual $\mathbf{W}_\mathbf{I}$ participates in the computation of $\hat{\mathbf{K}}_\mathrm{E}$ through the averaging formula (7), $\mathbf{J}'$ will contain a scaled version of the *entire* noise residual $\mathbf{W}_\mathbf{I} = \mathbf{aIK} + \boldsymbol{\theta}_\mathbf{I}$. Thus, besides the PRNU term, $\mathbf{W}_\mathbf{I}$ and $\mathbf{W}_{\mathbf{J}'}$ will share another signal – the noise $\boldsymbol{\theta}_\mathbf{I}$. Consequently, the correlation $c_{\mathbf{I},\mathbf{J}'} = corr(\mathbf{W}_\mathbf{I}, \mathbf{W}_{\mathbf{J}'})$ will be larger than what it would be if the only common signal between $\mathbf{I}$ and $\mathbf{J}'$ was the PRNU component (which would be the case if $\mathbf{J}'$ was not forged). As this increase may be quite small and the correlation itself may fluctuate significantly across images, the test that evaluates the statistical increase must be calibrated. We call this test the *triangle test*.

1) The triangle test

Alice starts her defense by computing an estimate $\hat{\mathbf{K}}_\mathrm{A}$ of the fingerprint of her camera from images guaranteed to not have been used by Eve. For instance, she can take new images with her camera C. Then, for a candidate image $\mathbf{I}$, she computes $c_{\mathbf{I},\mathbf{J}'}$, $c_{\mathbf{I},\hat{\mathbf{K}}_\mathrm{A}} = corr(\mathbf{W}_\mathbf{I}, \hat{\mathbf{K}}_\mathrm{A})$, and $c_{\mathbf{J}',\hat{\mathbf{K}}_\mathrm{A}} = corr(\mathbf{W}_{\mathbf{J}'}, \hat{\mathbf{K}}_\mathrm{A})$ (follow the diagram below).

$$\begin{tikzcd} \mathbf{W}_\mathbf{I} \arrow[rr, leftrightarrow, "corr"] \arrow[dr, leftrightarrow, "corr"'] & & \mathbf{W}_{\mathbf{J}'} \arrow[dl, leftrightarrow, "corr"] \\ & \hat{\mathbf{K}}_\mathrm{A} & \end{tikzcd}$$

The test is based on the fact that for images $\mathbf{I}$ that *were not* used to forge $\mathbf{J}'$, the value of $c_{\mathbf{I},\mathbf{J}'}$ can be estimated from $c_{\mathbf{I},\hat{\mathbf{K}}_\mathrm{A}}$ and $c_{\mathbf{J}',\hat{\mathbf{K}}_\mathrm{A}}$, while when $\mathbf{I}$ *was* used in the forgery, the correlation $c_{\mathbf{I},\mathbf{J}'}$ will be higher than the estimate.

In order to obtain a more accurate relationship, we will work by blocks of pixels, denoting the signals constrained to block b with subscript b. We adopt the model (2) for the noise residuals and a similar model for Alice's fingerprint:

[2] This strategy may not be the most advantageous to an adversary who is aware of the methods presented in this paper. More on this issue appears in Section VI.

$$\mathbf{W}_{\mathbf{I},b} = a_{\mathbf{I},b}\mathbf{I}_b\mathbf{K}_b + \mathbf{\Theta}_{\mathbf{I},b},\ \ \mathbf{W}_{\mathbf{J}',b} = a_{\mathbf{J}',b}\mathbf{J}'_b\mathbf{K}_b + \mathbf{\Theta}_{\mathbf{J}',b},$$
$$\hat{\mathbf{K}}_{\mathrm{A},b} = \mathbf{K}_b + \xi_b. \tag{9}$$

In (9), we assume that the attenuation factor $\mathbf{a}_{\mathbf{I},b} \equiv a_{\mathbf{I},b}$ is constant on each block. When **I** was not used by Eve, under some fairly mild assumptions about the noise terms in the models (9), the following estimate of $c_{\mathbf{I},\mathbf{J}'}$ is derived in the appendix:

$$\hat{c}_{\mathbf{I},\mathbf{J}'} = corr(\mathbf{W}_{\mathbf{I}}, \hat{\mathbf{K}}_{\mathrm{A}})\ corr(\mathbf{W}_{\mathbf{J}'}, \hat{\mathbf{K}}_{\mathrm{A}})\mu(\mathbf{I},\mathbf{J}')q^{-2}, \tag{10}$$

where $\mu\,(\mathbf{I},\mathbf{J}')$ is the "mutual-content factor,"

$$\mu(\mathbf{I},\mathbf{J}') = \frac{\sum_b a_{\mathbf{I},b}a_{\mathbf{J}',b}\overline{\mathbf{I}_b\mathbf{J}'_b}}{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b} \cdot \sum_b a_{\mathbf{J}',b}\overline{\mathbf{J}'_b}} N_{\mathrm{B}}, \tag{11}$$

and the bar denotes the sample mean. The integer N_{B} is the number of blocks and $q \le 1$ is the quality of $\hat{\mathbf{K}}_{\mathrm{A}}$, $q^{-2} = 1 + (SNR_{\hat{\mathbf{K}}_{\mathrm{A}}})^{-1}$, $SNR_{\hat{\mathbf{K}}_{\mathrm{A}}} = \|\mathbf{K}\|^2 / \|\xi\|^2$. The attenuation factors can be estimated by computing the following block-wise correlations:[3]

$$a_{\mathbf{I},b} = \frac{\|\mathbf{W}_{\mathbf{I},b}\|}{\sqrt{\overline{\mathbf{I}_b^2}}\,\|\hat{\mathbf{K}}_{\mathrm{A},b}\|} corr(\mathbf{W}_{\mathbf{I},b}, \mathbf{I}_b\hat{\mathbf{K}}_{\mathrm{A},b})q^{-2}. \tag{12}$$

Continuing the analysis of the case when **I** was *not* used by Eve, we consider $c_{\mathbf{I},\mathbf{J}'}$ and $\hat{c}_{\mathbf{I},\mathbf{J}'}$ as random variables over different images **I** for a fixed **J**′. The dependence between these two random variables is well fit with a straight line $c_{\mathbf{I},\mathbf{J}'} = \lambda\,\hat{c}_{\mathbf{I},\mathbf{J}'} + \eta$. Because the distribution of the deviation from the linear fit does not seem to vary with $\hat{c}_{\mathbf{I},\mathbf{J}'}$ (see Fig. 2), we make a simplifying assumption that the conditional probability

$$\Pr(c_{\mathbf{I},\mathbf{J}'} - \lambda\hat{c}_{\mathbf{I},\mathbf{J}'} - \eta = x \mid \hat{c}_{\mathbf{I},\mathbf{J}'}) \approx f_{\mathbf{J}'}(x), \tag{13}$$

is independent of $\hat{c}_{\mathbf{I},\mathbf{J}'}$.

When **I** *was* used by Eve in the multiplicative forgery, due to the additional common signal $\mathbf{\theta}_{\mathbf{I}}$, the correlation $c_{\mathbf{I},\mathbf{J}'} = corr(\mathbf{W}_{\mathbf{I}}, \mathbf{W}_{\mathbf{J}'})$ increases to $\beta c_{\mathbf{I},\mathbf{J}'}$, where β is the following multiplicative factor derived in the appendix:

$$\beta = 1 + \frac{\alpha}{N}\frac{\sum_b a_{\mathbf{J}',b}\overline{\mathbf{J}_b{}'}\,\|\mathbf{\Theta}_{\mathbf{I},b}\|^2}{\sum_b a_{\mathbf{I},b}a_{\mathbf{J}',b}\overline{\mathbf{I}_b\mathbf{J}'_b}\,\|\mathbf{K}_b\|^2}. \tag{14}$$

Notice that the percentual increase is proportional to the fingerprint strength α and the energy of the common noise component $\mathbf{\theta}_{\mathbf{I}}$, and it is inversely proportional to N.

Alice now runs the following composite binary hypothesis test for every candidate image **I** from her set of N_{c} candidate images:

$$\mathrm{H}_0 : c_{\mathbf{I},\mathbf{J}'} - \lambda\hat{c}_{\mathbf{I},\mathbf{J}'} - \eta \sim f_{\mathbf{J}'}(x),$$
$$\mathrm{H}_1 : c_{\mathbf{I},\mathbf{J}'} - \lambda\hat{c}_{\mathbf{I},\mathbf{J}'} - \eta \nsim f_{\mathbf{J}'}(x). \tag{15}$$

The reason why (15) cannot be turned into a simple hypothesis test is that the distribution of $c_{\mathbf{I},\mathbf{J}'}$ when **I** is used for forgery is not available to Alice and it cannot be determined experimentally because Alice does not know what strategy Eve used. Thus, we set our decision threshold t to bound the probability of false alarm, P_{FA}:

$$\Pr(c_{\mathbf{I},\mathbf{J}'} - \lambda\hat{c}_{\mathbf{I},\mathbf{J}'} - \eta > t \mid \mathrm{H}_0) = P_{\mathrm{FA}}. \tag{16}$$

Note that, depending on **J**′, the constant of proportionality $\lambda > 1$, which suggests the presence of an unknown multiplicative hidden parameter in (10) most likely due to some non-periodic NUAs that were not removed using zero-meaning and Wiener filtering as described in Section II.B. The quality of Alice's fingerprint, q, can be considered unknown (or simply set to 1) as different q will just correspond to a different λ (scaling of the x axis in the diagram of $c_{\mathbf{I},\mathbf{J}'}$ vs. $\lambda\hat{c}_{\mathbf{I},\mathbf{J}'}$).

Alice now has two options. She can test each candidate image **I** separately by evaluating its p-value and thus, on a certain level of statistical significance, identify those images that were used by Eve for estimating her fingerprint. Alternatively, Alice can test for N_{c} candidate images **I** all at once whether $c_{\mathbf{I},\mathbf{J}'} - \lambda\hat{c}_{\mathbf{I},\mathbf{J}'} - \eta \sim f_{\mathbf{J}'}(x)$. This "pooled test" will be a better choice for her for large N when the reliability of the triangle test for individual images becomes low.

B. *Multiple images with forged fingerprints*

In this paper, we consider another plausible scenario in which Eve forges more than one image and presents them as evidence to the judge. She may do so with the hope to make her case against Alice stronger. To be more precise, Eve superimposes *the same* fingerprint estimate $\hat{\mathbf{K}}_{\mathrm{E}}$ to at least two different images, $\mathbf{J}_1$ and $\mathbf{J}_2$, and obtains two forged images, $\mathbf{J}_1'$ and $\mathbf{J}_2'$. Interestingly, in this case, Alice will have another option for her defense – she can run the triangle test for the triple $\mathbf{J}_1'$, $\mathbf{J}_2'$, $\hat{\mathbf{K}}_{\mathrm{A}}$. Indeed, the test should work as the common component between the noise residuals of $\mathbf{J}_1'$ and $\mathbf{J}_2'$ will include the modeling noise of the estimated fingerprint $\hat{\mathbf{K}}_{\mathrm{E}}$. Note that Alice can run this test even when she has no access to images $\mathbf{I}^{(1)},\ldots,\mathbf{I}^{(N)}$ used by Eve to estimate $\hat{\mathbf{K}}_{\mathrm{E}}$!

V. Experiments

This section contains experimental evaluation of the triangle test under the assumption that Eve selects the natural strength α for the fake fingerprint as explained in Section III. The results for other values of α appear in Section VI.

In all tests, the signals entering the triangle test were preprocessed by zero-meaning. Wiener filtering, as described in Section II.B to suppress the NUAs, was only applied to $\mathbf{W}_{\mathbf{J}'}$ and not to $\mathbf{W}_{\mathbf{I}}$ to save the computation time. The camera *C*' is the 4-megapixel Canon PS A520 while *C* is Canon PS G2, which has the same native resolution. Both cameras were set to take images at the highest quality JPEG

[3] Eq. (12) holds independently of whether or not **I** was used by Eve. It is derived in the appendix.

compression and the largest resolution. The picture-taking mode was set to “auto.”

As explained in Section III, the fingerprint estimation algorithm Φ and the forging algorithm depend on many parameters. To obtain a compact yet comprehensive report on the performance of the triangle test, the experiments were designed to show the effect of only the most influential parameters – the number of images used by Eve, N, the number of candidate images, N_c, the target false alarm rate P_{FA}, and the fingerprint strength α. To improve the readability, we summarize the range of these parameters for each particular experiment in a table.

Eve estimates the fingerprint using the most accurate estimator she can find in the literature (3) implemented using the denoising filter F described in [14] with the wavelet-domain Wiener filter parameter $\sigma = 3$ (valid for 8-bit per channel color images). From our experiments, the reliability of the triangle test is insensitive to the denoising filter or the mismatch between the filters used by Eve and Alice.

Then, Eve forges a 24-bit color image **J** from camera C’ to make it look as if it came from camera C. She first slightly denoises **J** using the same denoising filter F (with its Wiener filter parameter $\sigma = 1$) to suppress the fingerprint from camera C’ and possibly other artifacts introduced by C’. The filter is applied to each color channel separately. Then, Eve adds the fingerprint to **J**, $\mathbf{J}' = \mathbf{J} + \alpha\mathbf{J}\mathbf{K}_E$, and saves the result as JPEG with quality factor $Q = 90$, which is slightly smaller than the typical qualities of the tested JPEG images **J**.

The fingerprint strength α is determined so that the response of the GMF (Equation (11) in [7]) matches its prediction Pred(**J**,**K**). The predictor was constructed exactly as described in [7]; it is a mapping that assigns a predicted value of the GMF to the pair consisting of a JPEG 90 image **J** and the true fingerprint **K**. The function Pred(·,·) was implemented as a linear combination of intensity, texture, and flattening features, and their second-order terms (total 15 terms). The coefficients of the linear fit were determined from 20 images of natural scenes using the least square fit.

Note that because Eve adjusts α so that the *JPEG-compressed* **J′** elicits the same GMF value as the prediction Pred(**J**,**K**), the proper value of α must be found, e.g., using a binary search.

The true fingerprint **K** was estimated from 300 JPEG images of natural scenes.

On the defense side, Alice estimates her fingerprint $\hat{\mathbf{K}}_A$ from $N_A = 15$ blue-sky raw images (fingerprint quality was $q = 0.56$). Surprisingly, the quality of $\hat{\mathbf{K}}_A$ has little impact on the triangle test. Tests with $N_A = 70$ produced essentially identical results. In particular, it is not necessary for Alice to work with a better quality fingerprint than Eve!

In all experiments, the block size was 128×128 pixels. The triangle test performed equally well for blocks as small as 64×64 and as large as 256×256.

A. Single forged image (individual test)

We first tested how reliably Alice can identify which images were used by Eve (the triangle test applied to each candidate image individually). By far the most influential element is the number of images used by Eve, N, and the content of the forged image **J**. We used $N \in \{20,50,100,200\}$ and six randomly selected test images **J** shown in Fig. 1. To give the reader a sense of the extent of Eve’s forging activity, in Table I we report the PSNR between **J′** before it is JPEG compressed and **J**. The PSNR between **J′** and **J** measures the total distortion that includes the slight denoising, F(**J**), and quantization to 24-bit colors after adding the fingerprint. The PSNR between **J′** and the slightly denoised F(**J**) measures the energy of the PRNU term only.

TABLE I. PSNR BETWEEN THE ORIGINAL IMAGE **J** AND THE FORGERY **J′** BEFORE JPEG COMPRESSION FOR SIX TEST IMAGES (SEE THE EXPERIMENT IN SECTION V.A).

#/N	PSNR(F(**J**), **J′**) [dB]				PSNR(**J**, **J′**) [dB]			
	20	50	100	200	20	50	100	200
1	48.8	51.8	53.2	53.9	47.6	49.5	50.3	50.7
2	49.0	51.8	53.1	53.8	47.8	49.8	50.6	50.9
3	50.1	51.8	52.9	53.4	48.7	49.8	50.5	50.8
4	54.5	56.4	57.5	58.7	49.5	50.0	50.2	50.4
5	49.5	52.2	53.2	54.0	47.7	49.3	49.7	50.1
6	50.8	53.2	54.3	55.1	49.3	51.0	51.6	52.1

In all our experiments, the pdf $f_{\mathbf{J}'}(x)$ (13) was often very close to a Gaussian but for some images **J**, the tails exhibited a hint of a polynomial dependence. Thus, to be conservative, we used Student’s t-distribution for the fit. The distribution was estimated from 358 images from camera C that were not used by Eve. All these images were taken within a period of about four years. In practice, depending on the situation, statistically significant conclusions may be obtained using a much smaller sample.

For a given probability of false alarm, P_{FA}, the Student’s t-fit was used to set a threshold on the test statistic using (16). To verify the threshold, we computed the triangle-test statistic from all images from C available to us (total of 937 images). These images were taken within the time period of seven years; a portion of them were stored as JPEG with varying quality factors, while some were in the raw TIFF format. All these images were recompressed to JPEG quality 90 and converted to grayscale before running the test. For the threshold computed for $P_{FA} = 10^{-3}$ and 10^{-4}, we observed two and zero false alarms in the right tail, respectively. Although these results are compatible with the theoretical false alarm rates, this experiment does not confirm the Student’s t-model as one needs to observe at least 30 false alarms for a reliable estimate (Doddington’s rule of 30 [16]). Unfortunately, an unreasonably large amount of images (30,000) would have to be taken with camera C to confirm the false alarms experimentally.

Table II summarizes the parameters investigated in the experiment in this section.

TABLE II. PARAMETERS FOR THE SINGLE FORGED IMAGE EXPERIMENT (INDIVIDUAL TEST).

Images	6
N	20, 50, 100, 200
N_c	N/A
α	natural
P_{FA}	10^{-3}, 10^{-4}

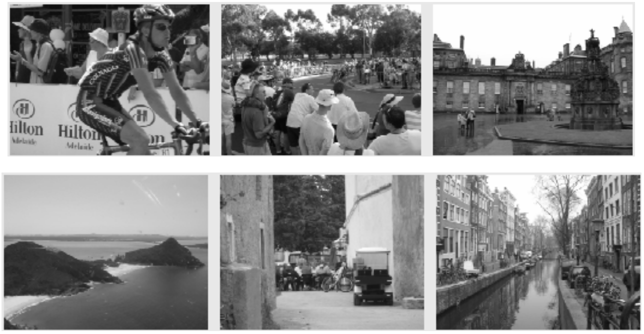


Fig. 1: Six original images **J** from a Canon PS A520 numbered by rows (#1, #2, #3); (#4, #5, #6).

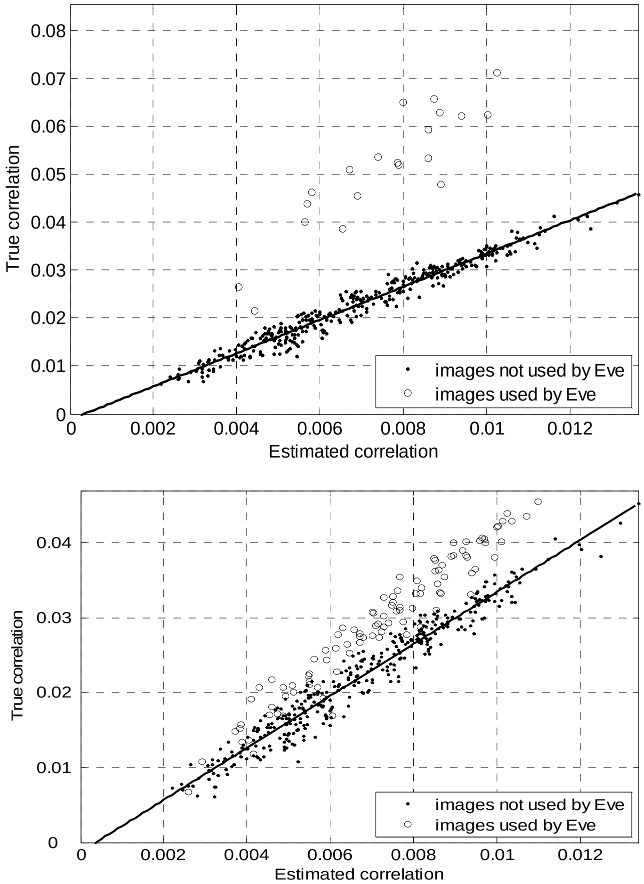


Fig. 2: True correlation $c_{\mathbf{I},\mathbf{J}'}$ vs. the estimate $\hat{c}_{\mathbf{I},\mathbf{J}'}$ for image no. 5. Eve's fingerprint was estimated from $N = 20$ images (top) and $N = 100$ (bottom).

Fig. 2 presents a typical plot of $c_{\mathbf{I},\mathbf{J}'}$ vs. $\hat{c}_{\mathbf{I},\mathbf{J}'}$ for $N = 20$ and $N = 100$. As expected, the separation between images used by Eve and those not used deteriorates with increasing N. When applying the triangle test individually to each candidate image, after setting the decision threshold using (16) to satisfy a desired probability of false alarm, P_{FA}, the probability of correct detection P_D in the hypothesis test (15) is shown in Table III. Each value of P_D was obtained by running the entire experiment as explained in Section IV.A and counting how many images $\mathbf{I}^{(1)},\ldots,\mathbf{I}^{(N)}$ used by Eve were correctly identified by the triangle test.

The lower detection rate for image no. 4 is due to the low energy of the fingerprint (see the corresponding row in Table I) dictated by the predictor. Because the image has a smooth content, which is further smoothened by the denoising filter, the fingerprint PSNR in the noise residual **W** is higher than for other images. Consequently, a low fingerprint energy is sufficient to match the predicted correlation.

Image no. 3 also produced lower detection rates, mostly due to the fact that 27.6% of the image content is overexposed (the entire sky) with fully saturated pixels. The attenuation factor **a** in (9) is thus effectively equal to zero for such pixels, while it is estimated in (12) under H_1 as being relatively large due to the absence of the noise term $\mathbf{\Theta}_{\mathbf{J}',b}$. A possible remedy is to apply the triangle test only to the non-saturated part of the image. However, then we experience a lower accuracy again due to a smaller number of pixels in the image. At this point, we note that if the adversary makes the forgery using (8) without attenuating the PRNU in saturated areas, the fingerprint there will be too strong, which could be used by Alice to argue that the fingerprint has been artificially added and the image did not come from her camera.

TABLE III. DETECTION RATE P_D [%] FOR SIX TEST IMAGES FOR THE EXPERIMENT IN SECTION V.A.

#/N	P_D [%] for $P_{FA} = 10^{-3}$				P_D [%] for $P_{FA} = 10^{-4}$			
	20	50	100	200	20	50	100	200
1	100	92	63	15	100	80	44	6
2	100	84	40	5	100	74	26	0
3	95	78	35	4	95	66	14	0
4	95	64	21	3	95	42	8	1
5	100	90	56	11	100	82	41	2
6	100	94	59	14	100	90	40	2

B. *Single forged image (pooled test)*

Table III shows that the triangle test for individual images is quite reliable for small N. If Eve has enough resources and obtains a large number of images, say $N > 200$, the reliability of the triangle test applied to each image one by one may become quite low depending on the content of the forged image. As discussed in Section IV.B, Alice's goal is to prove that $\mathbf{J}'$ was forged rather than identify which images Eve used. This hypothesis test can be decided more reliably by testing for all N_c candidate images **I** whether $c_{\mathbf{I},\mathbf{J}'} - \lambda\,\hat{c}_{\mathbf{I},\mathbf{J}'} - \eta \sim f_{\mathbf{J}'}(x)$. Since the differences are independent across different images, the scaled log-likelihood of all N_c observations,

$$L_{N_c} = \frac{1}{\sqrt{N_c}} \sum_{i=1}^{N_c} \log\left(f_{\mathbf{J}'}(c_{\mathbf{I},\mathbf{J}'}^{(i)} - \lambda \hat{c}_{\mathbf{I},\mathbf{J}'}^{(i)} - \eta) \right), \qquad (17)$$

is asymptotically Gaussian under H_0. Here, $c^{(i)}_{\mathbf{I},\mathbf{J}'}$ and $\hat{c}^{(i)}_{\mathbf{I},\mathbf{J}'}$ are the correlations for the ith image, $i = 1, \ldots, N_c$.

To determine the limits of the pooled triangle test in practice, we evaluated how its reliability depends on the image content, the number of images used by Eve, N, and the ratio N_c/N. Table IV summarizes the parameters investigated in the experiment in this section.

TABLE IV. PARAMETERS FOR THE SINGLE FORGED IMAGE EXPERIMENT (POOLED TEST).

Images	3
N	100, 150, 200, 250, 300
N/N_c	[0, 1]
α	natural
P_{FA}	10^{-3}, 10^{-4}

The following test was run for images no. 2, 3, and 5. As before, for each tested image $\mathbf{J}'$ the Gaussian fit for the log-likelihood (17) was estimated from 358 90% quality JPEG images $\mathbf{I}$ not used by Eve. For the pooled test, a random set of $k = 60$ images was selected out of N_c candidate images and its p-value was computed. A successful detection was declared when $p < P_{FA}$. The final value of the probability of detection P_D was obtained by bootstrapping (repeating this process over the random selection of k images) 30,000 times.

The results of the pooled test are shown in Fig. 3 for $P_{FA} = 10^{-4}$. Here, we increased the number of images used by Eve to 300. The pooled triangle test can correctly detect forged images for $N/N_c > 0.5$ for N up to 200. There exists a strong correlation between the test performance and the detection statistic (5). In general, the test performs better for images exhibiting larger correlation (smooth and high-luminance images). The detection is still possible when $N = 300$ as long as $N/N_c > 0.5$.

C. Multiple images with forged fingerprints

In this section, we experimentally study the scenario when Eve has forged two images, $\mathbf{J}_1'$ and $\mathbf{J}_2'$. As explained in Section IV.B, the triangle test can be applied to $\mathbf{J}_1'$ and $\mathbf{J}_2'$ to reveal the forgery. This has the advantage that Alice does not need access to the images from which Eve estimated her fingerprint $\hat{\mathbf{K}}_E$.

We used additional 147 images $\mathbf{J}$ from camera C' and, for each choice of N, we created 147 forgeries by adding Eve's fingerprint to them as described at the beginning of Section V. All other images were reused from the previous section. Fig. 4 and Table VI are the equivalents of Fig. 2 and Table III. Because the same fingerprint was added to each forgery, the PSNR between $\mathbf{J}$ and $\mathbf{J}'$ was the same as indicated in Table I. This experiment was run separately for $N = 20, 50, 100, 200, 300$.

Table V summarizes the parameters investigated in the entire experiment.

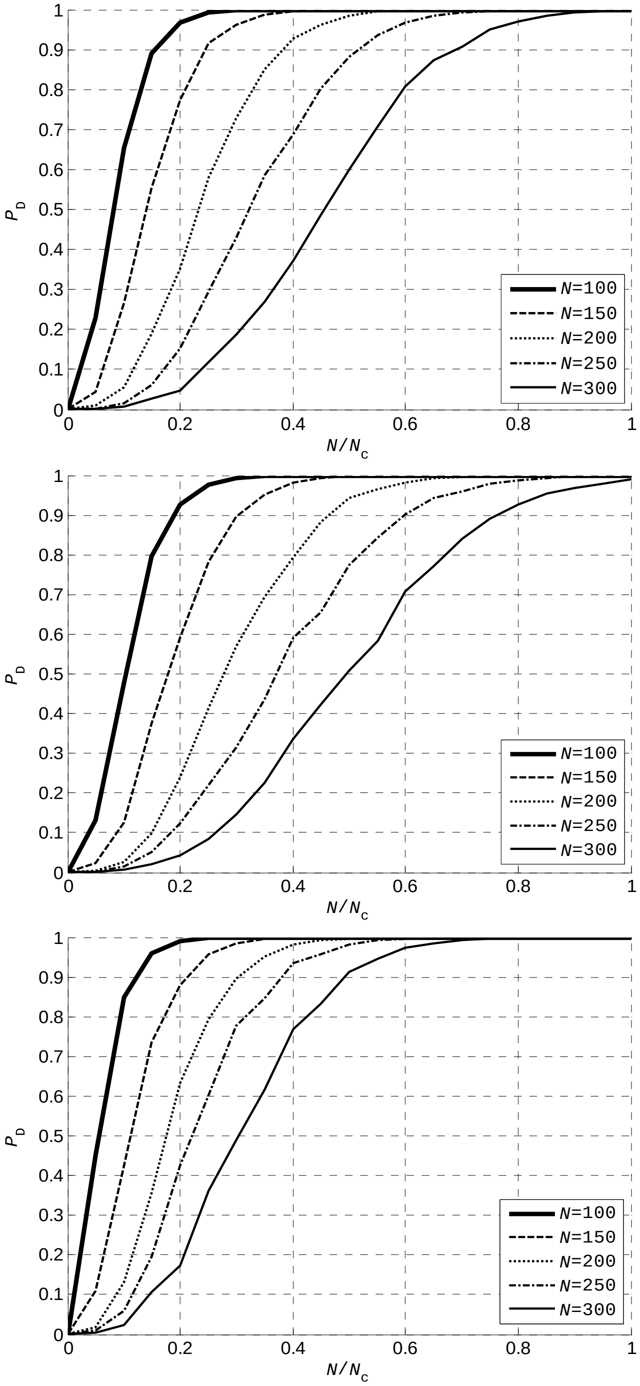


Fig. 3: Probability of revealing the forgery as a function of the ratio N/N_c for the pooled test for images no. 2, 3, and 5 for $P_{FA}=10^{-4}$; N is the number of images used by Eve to estimate the fake fingerprint and N_c is the number of candidate images.

TABLE V. PARAMETERS FOR THE MULTIPLE FORGED IMAGE EXPERIMENT.

Images	147 pairs
N	20, 50, 100, 200, 300
N/N_c	N/A
α	natural
P_{FA}	10^{-3}, 10^{-4}

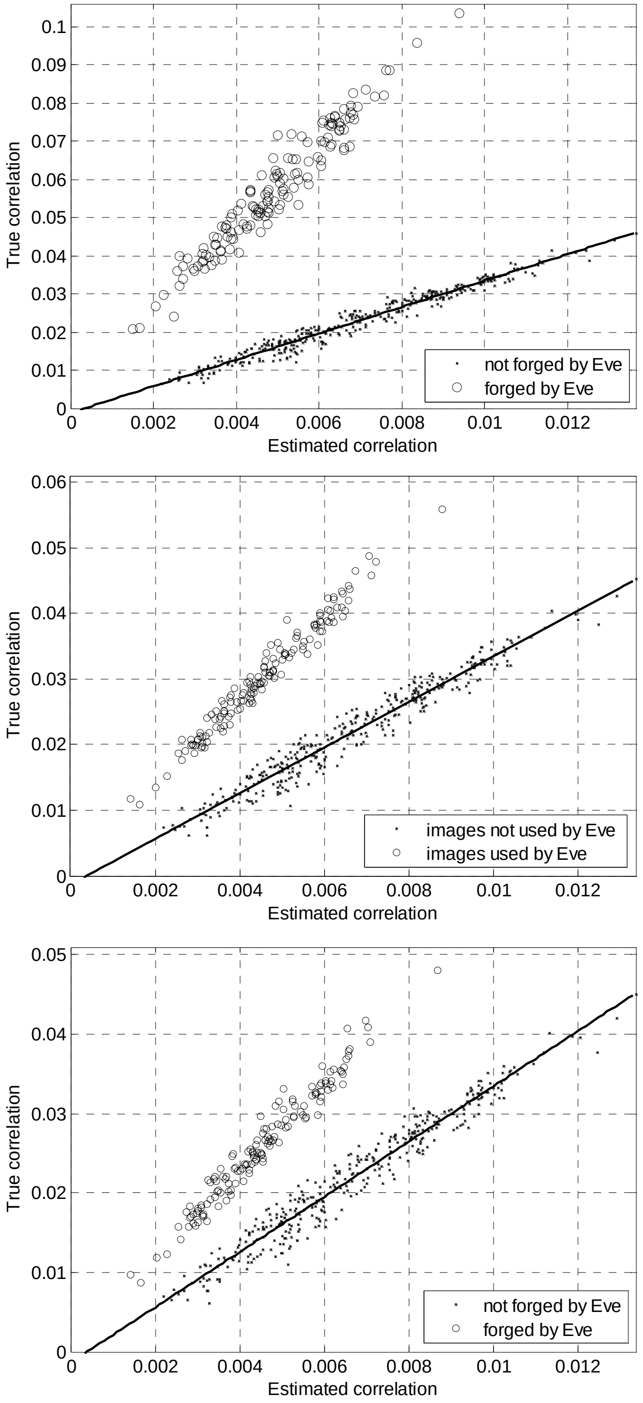


Fig. 4: True correlation $c_{\mathbf{I},\mathbf{J}'}$ vs. estimated $\hat{c}_{\mathbf{I},\mathbf{J}'}$ for the multiple-image test. $\mathbf{I} = \mathbf{J}_2'$ ran through 147 forged images, while $\mathbf{J} = \mathbf{J}_1'$ was fixed to image no. 5. Top to bottom: $N = 20$, 100, and 300.

In general, the larger the correlation between $\mathbf{J}_1'$ and $\mathbf{J}_2'$, the more reliably the triangle test can decide between the hypotheses (15). In this test, P_D does not drop with N as fast as when applying the triangle test to each image individually (see Section V.A and Table III). Reliable detection is possible even for $N = 300$.

TABLE VI: DETECTION RATE P_D [%] FOR THE MULTIPLE FORGED IMAGE TEST FOR $P_{FA} = 10^{-4}$.

#/N	P_D [%]				
	20	50	100	200	300
1	100.0	100.0	100.0	99.3	99.3
2	100.0	100.0	98.6	97.3	95.9
3	100.0	99.3	98.0	93.2	86.4
4	100.0	96.6	76.9	55.8	40.8
5	100.0	100.0	100.0	98.0	97.3
6	100.0	100.0	98.6	98.6	94.6

TABLE VII: DETECTION RATE P_D [%] OF THE TRIANGLE TEST FOR A SINGLE FORGED IMAGE NO. 2 WHEN SCALING THE NATURAL FINGERPRINT STRENGTH BY FACTOR r FOR TWO FALSE ALARM RATES AND FOUR VALUES OF N (SECTION VI).

P_D[%] for $P_{FA} = 10^{-3}$				
r\\N	20	50	100	200
0.10	80	26	1	0
0.25	100	54	30	0
0.50	100	76	32	3
0.75	100	84	37	5.5
1.00	100	84	40	4.5
2.00	100	88	52	9.5
P_D[%] for $P_{FA} = 10^{-4}$				
0.10	70	8	0	0
0.25	95	14	2	0
0.50	100	58	14	0
0.75	100	72	22	0
1.00	100	74	26	0
2.00	100	88	37	1

VI. EFFECT OF FINGERPRINT STRENGTH

All tests reported in Section V were carried out under the assumption that Eve chooses the natural strength for the fake fingerprint to avoid creating a forgery with a suspiciously weak or strong fingerprint as that could be the starting point of other attacks. However, if Eve is aware of the triangle test, she may try to adjust the strength to minimize the chances of being caught by this particular test. We are obviously facing the typical cat-and-mouse scenario when an attack elicits a counter-measure, which in turn elicits an attack, etc.

Inspecting the expression for the triangle test statistic (14) and its dependence on α, it may seem that the best option for Eve to minimize the chances of triggering the triangle test is to use the smallest α that still elicits positive identification in a threshold-based correlation detector. However, it will not be easy for Eve to do this as she does not have access to the true fingerprint (Alice does), and thus Eve cannot really be sure if the strength is large enough for positive identification. Another problem is that a suspiciously low correlation that is incompatible with the predicted correlation would imply that the image was subject to processing, which, in turn, points to violation of the chain of evidence (evidence integrity).

Instead of delving into arguments about the best strategy for Eve, we use the remaining space left in this paper to inform the reader about the performance limitations of the triangle test when Eve uses a strength for her fingerprint that is different from the natural value. The results indicate that the performance decreases rather slowly with decreased strength, and that Eve would need to decrease the natural strength by a rather large margin to escape the triangle test.

Next, we repeat the individual test, the pooled test, and the multiple-forgery test for six different fingerprint strengths obtained by scaling the natural strength by the factor of $r \in \{0.10, 0.25, 0.50, 0.75, 1.00, 2.00\}$.

A. Single forged image

Due to limited space left in this paper, this experiment was carried out only for image No. 2.[4] The detection rates for the false alarm rates 10^{-3} and 10^{-4} are displayed in Table VII. As expected, the triangle test becomes less reliable with decreasing strength α. Fortunately, the decrease is rather slow and does not necessarily invalidate the triangle test, depending on the number of images N used by Eve.

B. Single forged image (pooled test)

Next, we repeated the pooled test for image No. 2 for the same range of six scaled values of the fingerprint strength. The results are depicted in a graphical form in Fig. 5 for N = 100 and 200 for two values of the false alarm (total of four plots). Although the success rate of the triangle test understandably decreases with decreasing α and N/N_c, it is rather interesting that the pooled test is still 80% reliable even when $N = 100$, $N/N_c = 1/2$, and $P_{FA} = 10^{-4}$.

C. Multiple images with forged fingerprints

The last experiment of this section is the multiple image forgery test as described in Section V.C. Again, the test was repeated with pairs of images whose fingerprint strength was scaled by six different factors. One of the images, $\mathbf{J}'$, (image no. 2) was fixed while the other forged image $\mathbf{I}$ in the pair was obtained from 147 images. Table VIII shows the percentage of successfully revealed forgeries out of the 147 tested pairs for two false alarm rates and four different values of the number of images N used by Eve to create both forgeries.

VII. CONCLUSIONS

Camera identification using sensor noise works by establishing the presence of the camera's sensor fingerprint in the image under investigation. An adversary (Eve) may attempt to fool the identification algorithm by pasting a camera fingerprint onto an image that did not come from the camera. In doing so, an innocent victim (Alice) would be framed. In this paper, we investigate techniques that the victim may use to prove that the fingerprint was not inserted during the image acquisition, as the adversary claims, but was later maliciously added.

The crucial breakthrough we experienced in our study came from positioning ourselves into the role of the adversary and realizing what information and data will be available to both Eve and Alice. In her activity, Eve will likely have to rely on images taken by Alice that she decided to share with others, for example on her Facebook site. However, the estimation error of the camera fingerprint estimated from such images will contain remnants of the entire noise residual from all images used by Eve. This fact is the basis of the test we proposed (the "triangle test") using which Alice can identify the images that Eve used for her forgery and, in doing so, prove her innocence. This test was then extended to the case when none of the images is available to the victim but the victim has at least two forged images to analyze. We demonstrated the test's performance experimentally and investigated its limitations. In particular, the test can be applied when Eve uses a high-quality fingerprint estimated from 300 images. The conclusion that can be made from this study is that planting a sensor fingerprint in an image without leaving a trace is significantly more difficult than previously thought.

TABLE VIII: DETECTION RATE P_D [%] OF THE TRIANGLE TEST FOR THE MULTIPLE IMAGE TEST WHEN SCALING THE NATURAL FINGERPRINT STRENGTH BY FACTOR r FOR TWO FALSE ALARM RATES AND FOUR VALUES OF N.

P_D [%] for $P_{FA} = 10^{-3}$				
$r\backslash N$	20	50	100	200
0.10	55.8	6.1	1.4	1.4
0.25	98.6	71.4	44.9	17.0
0.50	100	98.6	91.8	78.9
0.75	100	100	99.3	97.3
1.00	100	100	100	99.3
2.00	100	100	100	100
P_D [%] for $P_{FA} = 10^{-4}$				
0.10	24.5	1.4	1.4	1.4
0.25	89.8	47.6	19.1	1.4
0.50	100	93.9	80.3	58.5
0.75	100	100	95.2	87.8
1.00	100	100	98.6	97.3
2.00	100	100	100	100

VIII. ACKNOWLEDGMENT

This research was supported by an NSF award CNF-0830528.

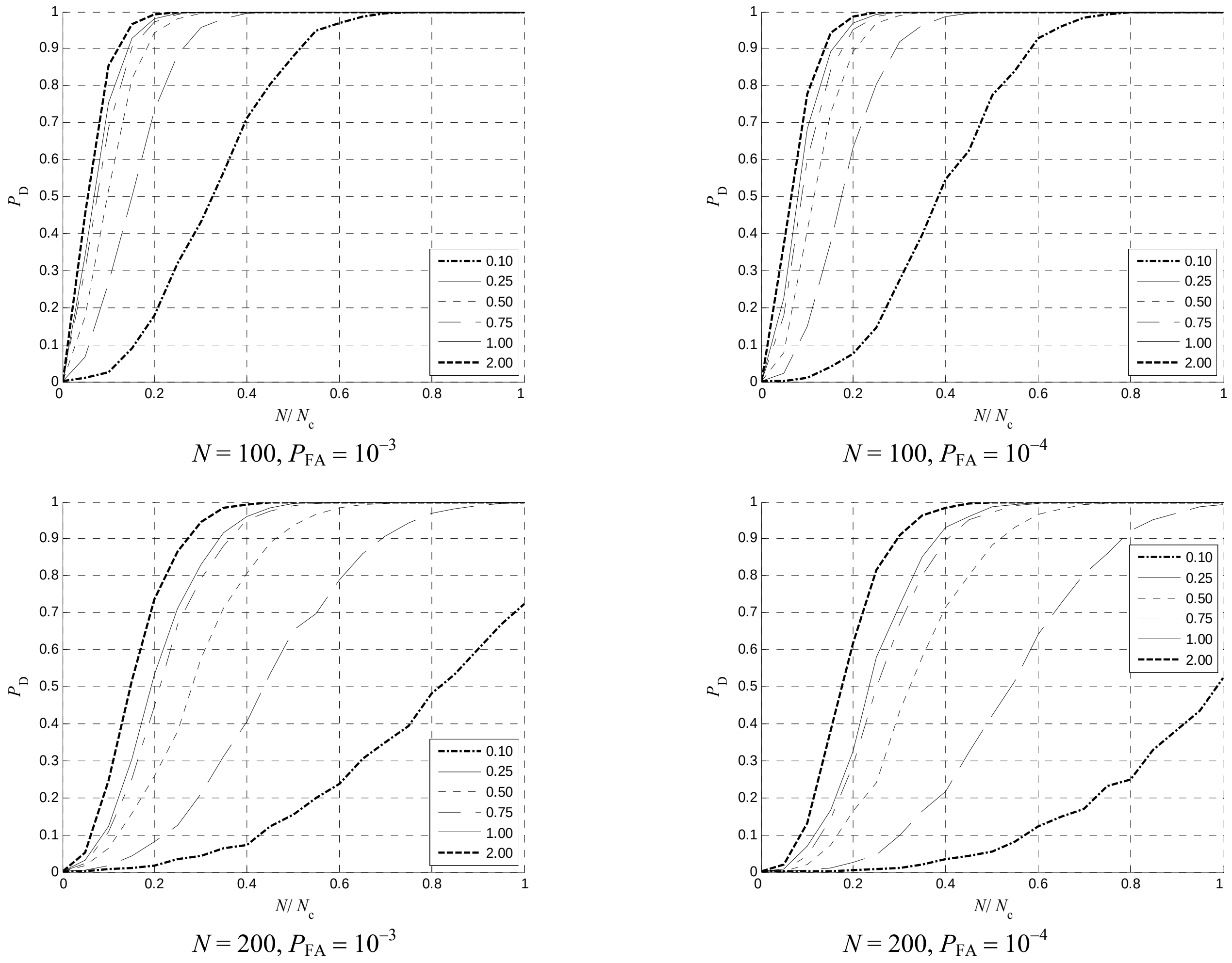
IX. REFERENCES

[1] Lukáš J., Fridrich J., and Goljan M.: "Determining Digital Image Origin Using Sensor Imperfections", *Proc. SPIE Electronic Imaging, Image and Video Communication and Processing*, vol. 5685, San Jose, CA, pp. 249–260, January 16–20, 2005.

[2] Lukáš, J., Fridrich, J., Goljan, M.: "Detecting Digital Image Forgeries Using Sensor Pattern Noise." *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VIII*, vol. 6072, San Jose, CA, January 16–19, pp. 362–372, 2006.

[3] Chen, M., Fridrich, J., Goljan, M.: "Digital Imaging Sensor Identification (Further Study)." *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, vol. 6505, San Jose, CA, Jan 29 – Feb 1, pp. 0P–0Q, 2007.

[4] As can be seen in Table III, image No. 2 constitutes the "middle ground" when it comes to the difficulty of being identified as forged using the triangle test.



$N = 100, P_{FA} = 10^{-3}$

$N = 100, P_{FA} = 10^{-4}$

$N = 200, P_{FA} = 10^{-3}$

$N = 200, P_{FA} = 10^{-4}$

Fig. 5: Probability of revealing the forgery as a function of the ratio N/N_c for the pooled test for image no. 2 and $r \in \{0.10, 0.25, 0.50, 0.75, 1.00, 2.00\}$. The parameters N and P_{FA} are below each plot; N is the number of images used by Eve to estimate the fake fingerprint and N_c is the number of candidate images.

[4] Goljan, M., Chen, M., Fridrich, J.: "Identifying Common Source Digital Camera From Image Pairs," *Proc. IEEE International Conference on Image Processing (ICIP)*, San Antonio, Texas, 2007.

[5] Gloe, T., Kirchner, M., Winkler, A. and Böhme, R.: "Can we trust digital image forensics?" *Proc. The 15th International conference on Multimedia*, Multimedia '07, ACM, Augsburg, Germany, pp. 78–86, 2007.

[6] Steinebach, M., Liu, H., Fan, P., Katzenbeisser, S.: "Cell phone camera ballistics: attacks and countermeasures," *Proc. SPIE, Multimedia on Mobile Devices 2010*, SPIE, vol. 7542, pp. 0B–0C, 2010.

[7] Chen, M., Fridrich, J., Goljan, M., and Lukáš, J.: "Determining Image Origin and Integrity Using Sensor Noise." *IEEE Transactions on Information Security and Forensics* **1**(1), pp. 74–90, March 2008.

[8] Goljan, M.: "Digital Camera Identification from Images – Estimating False Acceptance Probability," *Lecture Notes in Computer Science, Digital Watermarking*, vol. 5450, Proc. IWDW08, Busan, South Korea, pp. 454–468, 2009.

[9] Bloy, G. J.: "Blind Camera Fingerprinting and Image Clustering." *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(3), pp. 532–534, March, 2008.

[10] Rosenfeld, K. and Sencar, H. T.: "A Study of the Robustness of PRNU-Based Camera Identification." *Proc. SPIE, Media Forensics and Security XI*, vol. 7254, San Jose, CA, January 18–22, pp. 0M–0N, 2009**.**

[11] Celiktutan, O., B. Sankur, B., and Avcibas, I.: "Blind Identification of Source Cell-Phone Model." *IEEE Transactions on Information Forensics and Security* **3**(3), pp. 553–566, 2008.

[12] Böhme, R. and Kirchner, M.: "Synthesis of Color Filter Array Pattern in Digital Images." *Proc. SPIE, Media Forensics and Security XI*, vol. 7254, San Jose, CA, January 18–22, pp. 0K–0L, 2009**.**

[13] Goljan, M., Fridrich, J., and Filler, T.: "Large Scale Test of Sensor Fingerprint Camera Identification." *Proc. SPIE, Media Forensics and Security XI*, vol. 7254, San Jose, CA, January 18–22, pp. 0I–0J, 2009.

[14] Mihcak, M. K., Kozintsev, I., and Ramchandran, K.: "Spatially Adaptive Statistical Modeling of Wavelet Image Coefficients and its Application to Denoising." *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 6, Phoenix, Arizona, pp. 3253–3256, March 1999.

[15] Porter, J. E.: "On the "30 error" criterion." In: *National Biometric Test Center* – Collected Works, 1997–2000, San Jose State University, http://www.engr.sjsu.edu/biometrics/nbtccw.pdf.

APPENDIX

We model the noise residuals $\mathbf{W}_\mathbf{I}$ and $\mathbf{W}_\mathbf{J}$ and Alice's fingerprint $\hat{\mathbf{K}}_\mathrm{A}$ using (9). Furthermore, we adopt the following simplifying assumptions.

Uncorrelatedness. For any two images **I** and **J**, not necessarily different, and for every block *b*

$$\mathbf{I}_b\mathbf{K}_b \odot \mathbf{\Theta}_{\mathbf{J},b} = 0\,,\ \mathbf{I}_b\mathbf{K}_b \odot \xi_b = 0\,,$$
$$\mathbf{K}_b \odot \xi_b = 0\,,\ \mathbf{\Theta}_{\mathbf{J},b} \odot \xi_b = 0\,. \tag{A.1}$$

In reality, these assumptions really mean that the dot products are small compared to other quantities in the derivations below.

Dot product. We will need the following approximate equality valid whenever $\mathbf{K}_b[i]$ are i.i.d. realizations of a scalar random variable with finite variance:

$$\begin{aligned}\mathbf{I}_b\mathbf{K}_b \odot \mathbf{J}_b\mathbf{K}_b &= \sum_k k \sum_{\mathbf{I}_b[i]\mathbf{J}_b[i]=k} \mathbf{K}_b^2[i] \\ &= \|\mathbf{K}_b\|^2 \sum_k k\mathbf{p}_k \doteq \overline{\mathbf{I}_b\mathbf{J}_b}\,\|\mathbf{K}_b\|^2,\end{aligned} \tag{A.2}$$

where we denoted $\mathbf{p}_k = \dfrac{1}{\|\mathbf{K}_b\|^2} \sum_{\mathbf{I}_b[i]\mathbf{J}_b[i]=k} \mathbf{K}_b^2[i]$. To explain the last approximate equality in (A.2), realize that for an i.i.d. signal $\mathbf{K}_b[i]$, i = 1, …, N_b, with variance $\sigma_\mathbf{K}^2$:

$$\mathbf{p}_k \doteq \frac{|\{i\,|\,\mathbf{I}_b[i]\mathbf{J}_b[i]=k\}|\,\sigma_\mathbf{K}^2}{N_b\sigma_\mathbf{K}^2}\,, \tag{A.3}$$

which is the sample pmf of the signal $\mathbf{I}_b[i]\mathbf{J}_b[i]$.

The actual derivations.

Our goal is to find a relationship between $corr(\mathbf{W}_\mathbf{I},\hat{\mathbf{K}}_\mathrm{A})$, $corr(\mathbf{W}_\mathbf{J},\hat{\mathbf{K}}_\mathrm{A})$, and $corr(\mathbf{W}_\mathbf{I},\mathbf{W}_\mathbf{J})$, the quantities we can compute for any two images **I**, **J**, and an estimated fingerprint $\hat{\mathbf{K}}_\mathrm{A}$. Using (A.1) and (A.2), we first express

$$\begin{aligned}corr(\mathbf{W}_\mathbf{I},\hat{\mathbf{K}}_\mathrm{A}) &= \frac{\sum_b (a_{\mathbf{I},b}\mathbf{I}_b\mathbf{K}_b + \mathbf{\Theta}_{\mathbf{I},b}) \odot (\mathbf{K}_b + \xi_b)}{\sqrt{\sum_b \|a_{\mathbf{I},b}\mathbf{I}_b\mathbf{K}_b + \mathbf{\Theta}_{\mathbf{I},b}\|^2}\sqrt{\sum_b \|\mathbf{K}_b + \xi_b\|^2}} \\ &\doteq \frac{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b}\,\|\mathbf{K}_b\|^2}{\sqrt{\sum_b a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2 + \|\mathbf{\Theta}_{\mathbf{I},b}\|^2}\sqrt{\sum_b \|\mathbf{K}_b\|^2 + \|\xi_b\|^2}},\end{aligned} \tag{A.4}$$

which can be simplified to

$$\begin{aligned}&= \frac{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b}\,\|\mathbf{K}_b\|^2}{\sqrt{\sum_b a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}\sqrt{1+\frac{1}{SNR_\mathbf{I}}}\sqrt{\sum_b \|\mathbf{K}_b\|^2}\sqrt{1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}}} \\ &= \frac{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b}\,\|\mathbf{K}_b\|^2}{\sqrt{\sum_b a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}\sqrt{1+\frac{1}{SNR_\mathbf{I}}}\sqrt{1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}}}\,\|\mathbf{K}\|^{-1}\end{aligned} \tag{A.5}$$

by introducing

$$SNR_\mathbf{I} = \frac{\sum_b a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}{\sum_b \|\mathbf{\Theta}_{\mathbf{I},b}\|^2}\,,\ SNR_{\hat{\mathbf{K}}_\mathrm{A}} = \frac{\sum_b \|\mathbf{K}_b\|^2}{\sum_b \|\xi_b\|^2}\,. \tag{A.6}$$

To determine the factors $a_{\mathbf{I},b}$, we write

$$\begin{aligned}\mathbf{c}_{\mathbf{I},b} &= corr(\mathbf{W}_{\mathbf{I},b},\mathbf{I}_b\hat{\mathbf{K}}_{\mathrm{A},b}) \\ &= \frac{a_{\mathbf{I},b}\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}{\sqrt{a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}\sqrt{1+\frac{\|\mathbf{\Theta}_{\mathbf{I},b}\|^2}{a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}}\sqrt{\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}} \\ &\times \frac{1}{\sqrt{1+\frac{\|\xi_b\|^2}{\|\mathbf{K}_b\|^2}}} = \frac{1}{\sqrt{1+\frac{\|\mathbf{\Theta}_{\mathbf{I},b}\|^2}{a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}}\sqrt{1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}}}.\end{aligned} \tag{A.7}$$

From here after some simple algebra,

$$\frac{\|\mathbf{\Theta}_{\mathbf{I},b}\|^2}{a_{\mathbf{I},b}^2} = \overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2 \left\{\mathbf{c}_{\mathbf{I},b}^{-2}\left(1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}\right)^{-1} - 1\right\}. \tag{A.8}$$

From (1), we obtain another equation for both unknowns $\mathbf{a}_{\mathbf{I},b}, \|\mathbf{\Theta}_{\mathbf{I},b}\|^2$:

$$\|\mathbf{W}_{\mathbf{I},b}\|^2 = a_{\mathbf{I},b}^2\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2 + \|\mathbf{\Theta}_{\mathbf{I},b}\|^2. \tag{A.9}$$

Because the solution to the system

$$\begin{matrix} y/x = R \\ xd + y = S \end{matrix}\quad \text{is} \quad \begin{matrix} x = S/(d+R) \\ y = SR/(d+R) \end{matrix} \tag{A.10}$$

and because in our case R is the r.h.s. of (A.8), $d = \overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2$, $S = \|\mathbf{W}_{\mathbf{I},b}\|^2$, $x \triangleq a_{\mathbf{I},b}^2$, we obtain

$$\begin{aligned}a_{\mathbf{I},b}^2 &= \frac{\|\mathbf{W}_{\mathbf{I},b}\|^2}{\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2 + \overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2 \left\{\mathbf{c}_{\mathbf{I},b}^{-2}\left(1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}\right)^{-1} - 1\right\}} \\ &= \frac{\|\mathbf{W}_{\mathbf{I},b}\|^2}{\overline{\mathbf{I}_b^2}\,\|\mathbf{K}_b\|^2}\mathbf{c}_{\mathbf{I},b}^2\left(1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}\right).\end{aligned} \tag{A.11}$$

Assuming the SNR $\|\mathbf{K}_b\|^2/\|\xi_b\|^2 = SNR_{\hat{\mathbf{K}}_\mathrm{A}}$ is independent of *b*, because $\|\hat{\mathbf{K}}_{\mathrm{A},b}\|^2 = \|\mathbf{K}_b\|^2 + \|\xi_b\|^2$, we obtain after some simple algebra:

$$\|\mathbf{K}_b\|^2 = \frac{\|\hat{\mathbf{K}}_{\mathrm{A},b}\|^2}{1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}}. \tag{A.12}$$

Thus, (A.11) can finally be written to involve the norm of $\hat{\mathbf{K}}_{\mathrm{A},b}$ rather than $\mathbf{K}_b$:

$$a^2_{\mathbf{I},b} = \frac{\|\mathbf{W}_{\mathbf{I},b}\|^2}{\overline{\mathbf{I}^2_b}\,\|\hat{\mathbf{K}}_{\mathrm{A},b}\|^2}\mathbf{c}^2_{\mathbf{I},b}\left(1+\frac{1}{SNR_{\hat{\mathbf{K}}_{\mathrm{A}}}}\right)^2. \qquad \text{(A.13)}$$

Now, we derive the expression for $corr(\mathbf{W_I}, \mathbf{W_J})$ when **I** was not used to forge image **J**. Here, the only common signal between $\mathbf{W_I}$ and $\mathbf{W_J}$ is the PRNU:

$$corr(\mathbf{W_I}, \mathbf{W_J}) = \frac{\sum_b a_{\mathbf{I},b} a_{\mathbf{J},b} \overline{\mathbf{I}_b \mathbf{J}_b}\,\|\mathbf{K}_b\|^2}{\sqrt{\sum_b a^2_{\mathbf{I},b}\overline{\mathbf{I}^2_b}\,\|\mathbf{K}_b\|^2}\sqrt{\sum_b a^2_{\mathbf{J},b}\overline{\mathbf{J}^2_b}\,\|\mathbf{K}_b\|^2}} \times \frac{1}{\sqrt{1+\frac{1}{SNR_\mathbf{I}}}\sqrt{1+\frac{1}{SNR_\mathbf{J}}}}. \qquad \text{(A.14)}$$

By comparing (A.5) and (A.14), we see that

$$corr(\mathbf{W_I}, \mathbf{W_J}) = corr(\mathbf{W_I}, \hat{\mathbf{K}}_\mathrm{A}) corr(\mathbf{W_J}, \hat{\mathbf{K}}_\mathrm{A}) \times \frac{\sum_b a_{\mathbf{I},b} a_{\mathbf{J},b} \overline{\mathbf{I}_b \mathbf{J}_b}\,\|\mathbf{K}_b\|^2}{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b}\,\|\mathbf{K}_b\|^2 \cdot \sum_b a_{\mathbf{J},b}\overline{\mathbf{J}_b}\,\|\mathbf{K}_b\|^2}\|\mathbf{K}\|^2\left(1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}\right). \qquad \text{(A.15)}$$

Rewriting $corr(\mathbf{W_I}, \mathbf{W_J})$ using the norm of $\hat{\mathbf{K}}_\mathrm{A}$ rather than the unknown norm of **K**,

$$corr(\mathbf{W_I}, \mathbf{W_J}) = corr(\mathbf{W_I}, \hat{\mathbf{K}}_\mathrm{A}) corr(\mathbf{W_J}, \hat{\mathbf{K}}_\mathrm{A}) \times \frac{\sum_b a_{\mathbf{I},b} a_{\mathbf{J},b} \overline{\mathbf{I}_b \mathbf{J}_b}\,\|\hat{\mathbf{K}}_{\mathrm{A},b}\|^2}{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b}\,\|\hat{\mathbf{K}}_{\mathrm{A},b}\|^2 \cdot \sum_b a_{\mathbf{J},b}\overline{\mathbf{J}_b}\,\|\hat{\mathbf{K}}_{\mathrm{A},b}\|^2} \times \|\hat{\mathbf{K}}_\mathrm{A}\|^2\left(1+\frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}\right). \qquad \text{(A.16)}$$

For the case when **I** was used to forge image **J**, the noise residuals $\mathbf{W_I}$ and $\mathbf{W_J}$ will have another common component besides the PRNU. It is the non-PRNU noise $\mathbf{\Theta_I}$. Assuming for simplicity that Eve estimates her fingerprint $\hat{\mathbf{K}}_\mathrm{E}$ simply by averaging the noise residuals of $\mathbf{I}_1, \ldots, \mathbf{I}_N$ in the form (2)

$$\hat{\mathbf{K}}_\mathrm{E} = \frac{1}{N}\sum_{i=1}^{N}(a_i\mathbf{I}_i\mathbf{K} + \mathbf{\Theta}_i) \doteq a_\mathrm{E}\mathbf{K} + \frac{1}{N}\sum_{i=1}^{N}\mathbf{\Theta}_i. \qquad \text{(A.17)}$$

In (A.17), we made a simplifying assumption that $1/N\sum a_i\mathbf{I}_i\mathbf{K} \doteq a_\mathrm{E}\mathbf{K}$, which is reasonable for all but small N. Eve superimposes $\hat{\mathbf{K}}_\mathrm{E}$ using (8), obtaining

$$\mathbf{J}' = \mathbf{J} + \alpha\mathbf{J}\hat{\mathbf{K}}_\mathrm{E} = \mathbf{J} + \alpha a_\mathrm{E}\mathbf{J}\mathbf{K} + \frac{\alpha}{N}\mathbf{J}\mathbf{\Theta_I} + \frac{\alpha}{N}\mathbf{J}\sum_{i=2}^{N}\mathbf{\Theta}_i, \qquad \text{(A.18)}$$

where, without loss of generality, we assumed that $\mathbf{I}_1 = \mathbf{I}$. In this paper, we assume that Eve creates the perfect forgery in the sense that **J′** elicits the same detector response as if it was not forged. In other words, $\alpha a_\mathrm{E} = 1$. Using this and the model for the noise residual on each block (2),

$$\mathbf{W}_{\mathbf{J}',b} = a_{\mathbf{J}',b}\mathbf{J}_b\mathbf{K}_b + \frac{\alpha a_{\mathbf{J}',b}}{N}\mathbf{J}_b\mathbf{\Theta}_{\mathbf{I},b} + \frac{\alpha a_{\mathbf{J}',b}}{N}\mathbf{J}_b\sum_{i=2}^{N}\mathbf{\Theta}_{i,b} + \mathbf{\Theta}'_{\mathbf{J}',b}, \qquad \text{(A.19)}$$

Note that this expression is of the form $\mathbf{W}_{\mathbf{J}',b} = a_{\mathbf{J}',b}\mathbf{J}_b\mathbf{K}_b + \delta_b\mathbf{J}_b\mathbf{\Theta}_{\mathbf{I},b} + \mathbf{\Theta}_{\mathbf{J}',b}$, for some appropriately defined noise term $\mathbf{\Theta}_{\mathbf{J}',b}$ and $\delta_b = \frac{\alpha a_{\mathbf{J}',b}}{N}$. The derivation of the expression for $corr(\mathbf{W_I}, \mathbf{W_{J'}})$ now follows exactly the same steps as above with one exception. The dot product

$$\mathbf{W_I} \odot \mathbf{W_{J'}} = \sum_b a_{\mathbf{I},b} a_{\mathbf{J}',b}\overline{\mathbf{I}_b\mathbf{J}'_b}\,\|\mathbf{K}_b\|^2 + \frac{\alpha}{N}\sum_b a_{\mathbf{J}',b}\overline{\mathbf{J}_b}\,\|\mathbf{\Theta}_{\mathbf{I},b}\|^2 \qquad \text{(A.20)}$$

now has an additional component because the two noise residuals share more than the PRNU signal. The rest of the expression coincides with that of $\rho(\mathbf{W_I}, \mathbf{W_J})$ in (A.5). Because the fingerprint is zero mean, $\overline{\mathbf{J}_b} = \overline{\mathbf{J}_b'}$ and the increase in correlation can thus be expressed via the ratio

$$\frac{\rho(\mathbf{W_I}, \mathbf{W_{J'}})}{\rho(\mathbf{W_I}, \mathbf{W_J})} = 1 + \frac{\alpha}{N}\frac{\sum_b a_{\mathbf{J}',b}\overline{\mathbf{J}_b'}\,\|\mathbf{\Theta}_{\mathbf{I},b}\|^2}{\sum_b a_{\mathbf{I},b} a_{\mathbf{J}',b}\overline{\mathbf{I}_b\mathbf{J}'_b}\,\|\mathbf{K}_b\|^2}. \qquad \text{(A.21)}$$

in terms of the forged image **J′** only.