Sensor Noise Camera Identification: Countering Counter-Forensics

Miroslav Goljan, Jessica Fridrich, and Mo Chen
Department of Electrical and Computer Engineering
SUNY Binghamton, Binghamton, NY 13902-6000
Ph: (607) 777 5793, Fax: (607) 777 4464
Email: {mgoljan, fridrich}@binghamton.edu, mchen@jadaktech.com

ABSTRACT

In camera identification using sensor noise, the camera that took a given image can be determined with high certainty by establishing the presence of the camera's sensor fingerprint in the image. In this paper, we develop methods to reveal counter-forensic activities in which an attacker estimates the camera fingerprint from a set of images and pastes it onto an image from a different camera with the intent to introduce a false alarm and, in doing so, frame an innocent victim. We start by classifying different scenarios based on the sophistication of the attacker's activity and the means available to her and to the victim, who wishes to defend herself. The key observation is that at least some of the images that were used by the attacker to estimate the fake fingerprint will likely be available to the victim as well. We describe the so-called "triangle test" that helps the victim reveal attacker's malicious activity with high certainty under a wide range of conditions. This test is then extended to the case when none of the images that the attacker used to create the fake fingerprint are available to the victim but the victim has at least two forged images to analyze. We demonstrate the test's performance experimentally and investigate 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.

Keywords: Camera identification, digital forensics, photo-response non-uniformity, sensor fingerprint, counter-forensics.

1. INTRODUCTION

Human fingerprints have been used in forensic science since 1892 when an Argentine police officer Juan Vucetich made the first criminal fingerprint identification. Since then, the science of human fingerprinting has developed to a rigorous discipline that often plays a decisive role in prosecution. Apparently, it is less known that faking a human fingerprint is possible. And in our information age, such counter-forensic techniques are available to the masses after a few clicks on Google, http://www.ehow.com/how_2122642_fake-fingerprints.html. The method uses a few drops of superglue to copy a fingerprint from a smooth surface onto a thin layer of dried wood glue that can be stuck to your thumb and, after gently rubbing your face, printed on any object of your choice, giving you the ability to frame the victim who unknowingly left his/her fingerprint behind. It is certainly conceivable to think of a number of counter counter-forensic methods that would be effective against this type of fingerprint forging. The most obvious one is detecting the mismatch between the DNA in the fingerprint residual with the person's identity or looking at the fine structure of the fingerprint to reveal some tell-tale signs of the faking process. The important message here is that even human fingerprints can be forged given enough skill and resources. The second important message is that such counter-forensic attempts can likely be revealed with more advanced forensic methods.

Digital sensor fingerprints [1] are likewise vulnerable to forging by a skilled individual. The sensor fingerprint is essentially a spread-spectrum watermark and, as such, it is vulnerable to attacks previously developed for robust watermarks, such as the watermark-copy attack [2]. In particular, if an attacker obtains access to images from a given camera, she can estimate its fingerprint and paste it onto an arbitrary image to make it look as if it came from the camera with the stolen fingerprint. An early attempt to investigate such counter-forensic methods appeared in [3]. In this work, we investigate whether it is possible to detect forged camera fingerprints, then develop appropriate countermeasures, and study their effectiveness.

The rest of this introduction defines the preliminary concepts and notation used in this paper. In Section 2, we formulate our assumptions under which we develop our techniques based on the actions of the attacker. The actual techniques are detailed in Section 3, depending on the means available to the attacker and the victim. Each technique is deployed on a real-life example and then analyzed and experimentally tested on a larger data set in Section 4. The paper is concluded in Section 5, where we discuss consequences of the newly obtained results and outline future directions.

1.1 Preliminaries

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

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

For an image **I** obtained by a digital camera, its noise residual is defined as $\mathbf{W_I} = \mathbf{I} - F(\mathbf{I})$, where F is a denoising filter. The major component of the sensor fingerprint is due to photo-response non-uniformity (PRNU), which can be captured using a multiplicative factor **K** (the sensor "fingerprint"). Using the model described in [1], the noise residual can be written as

$$\mathbf{W_I} = \mathbf{aIK} + \mathbf{\Theta}, \tag{1}$$

where $\mathbf{\Theta}$ stands for all other random noise components, such as the shot noise or the readout noise, and **a** is an attenuation factor of the same dimension as **K** that, in general, depends on the image content.

The maximum likelihood estimator of the PRNU factor **K** from images $\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}$ has the form:

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

Under the assumption that no geometric transform was applied to image **J** (e.g., cropping, scaling, and digital zooming), the presence of the camera fingerprint represented by an estimate $\hat{\mathbf{K}}$ is established through the correlation detector:

$$\rho = corr(\mathbf{W_J}, \mathbf{J}\hat{\mathbf{K}}). \tag{3}$$

2. ATTACK SCENARIOS

In this section, we introduce the basic terminology, notation, and scenarios under which the attacker may operate. We assume that Alice owns a digital camera *C* and Eve wants to frame her by forging evidence. Eve takes an image **J** from a different camera *C*' and makes it appear as if it came from *C*. We assume that the best choice for Eve is to first estimate the fingerprint of *C* and then properly superimpose it onto **J**. Eve may or may not attempt to suppress the fingerprint of *C*' from **J** or remove any other artifacts, such as color-interpolation correlations [4]. In this paper, we conservatively assume that camera *C*' is in the full possession of Eve, images from *C*' never appeared in public, and Eve destroys *C*' after framing Alice. Thus, we cannot take advantage of knowing any information about *C*'. Next, we detail Eve's forging activity.

Fingerprint estimation. Eve takes N images, $\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}$ from camera *C* and estimates its fingerprint using an algorithm Φ:

$$\hat{\mathbf{K}} = \Phi(\mathbf{I}^{(1)}, \ldots, \mathbf{I}^{(N)}; P_K). \tag{4}$$

The symbol P_K stands for the settings of the estimation procedure, such as the choice of the denoising filter used to extract the noise component from images, the parameters of this filter, or the formula for the aggregation of the noise residuals and its parameters. Fundamentally, the estimation procedure is some form of averaging of the noise residuals

$$\hat{\mathbf{K}} = \sum_{i=1}^{N} \mathbf{h}_i \mathbf{W}^{(i)}, \tag{5}$$

where $\mathbf{h}^{(i)} = 1/N$ for simple averaging [5] or $\mathbf{h}^{(i)} = \mathbf{I}^{(i)} / \sum_{i=1}^{N} (\mathbf{I}^{(i)})^2$ for the maximum likelihood estimator (2).

Preprocessing. After estimating the fingerprint, Eve proceeds with preprocessing **J** to suppress the fingerprint of *C*' and/or to remove any artifacts in **J** that are incompatible with *C*. We argue that suppressing the trace of the fingerprint of *C*' is not an easy task [6] and one that Eve may skip altogether. This is because the PRNU component **JK'** in **J** is very weak to be detected per se and *C*' has been destroyed anyway. In fact, we argue that Eve should avoid processing **J** too much as it may introduce artifacts of its own.

A potentially important issue for Eve is the JPEG quantization table of **J**. Eve will likely compress her forgery. If she does so with a quantization table that is incompatible with camera *C*, Alice will know that the image has been manipulated and did not come directly from her camera. If the quantization table of **J** is incompatible with camera *C*, she will necessarily introduce double-compression artifacts into **J**, giving Alice again a starting point of her defense.

The forged image may also contain color-interpolation artifacts of *C*' incompatible with those of *C*. Alice could leverage upon techniques developed for camera brand and model identification [7] and prove that there is a mismatch between the camera models. Eve, in turn, could attempt to remove such artifacts and introduce interpolation artifacts of *C*, for example, using the method described in [8].

The issues discussed above indicate that Eve's job of creating a "perfect" forgery is more difficult than what it appears at first sight. While there certainly exist other potential countermeasures based on inspecting the traces of previous compression or color interpolation artifacts, no attempt is made in this paper to exploit these discrepancies to reveal the forged fingerprint. We simply assume that Eve can do her job well enough to avoid such pitfalls, for example by using *C*' of the same model.

The final step for Eve is to plant the estimated fingerprint in **J**. She achieves this using the procedure

$$\mathbf{J}' = \Psi(\hat{\mathbf{K}}, \mathbf{J}; P_F). \tag{6}$$

Here, again P_F denotes the fingerprint planting settings, including the forgery method and the fingerprint strength α. In this paper, we allow Eve to use either an additive or multiplicative forgery:

Additive forgery. Here, Eve creates **J′** by adding a scaled PRNU factor to **J**

$$\mathbf{J}' = [\mathbf{J} + \alpha \hat{\mathbf{K}}], \tag{7}$$

where $\alpha > 0$ is a scalar and $[x]$ is the operation of rounding x to the set $\{0, 1, ..., 255\}$. Equation (7) should be understood as three equations for each color channel of **J**. Eve may tune the strength α so that the fake fingerprint is not too weak or suspiciously too strong. To find the correct fingerprint strength α, Eve may utilize the predictor of correlation described in [1] or use as a reference another image from Alice's camera that has a similar content, which is the approach we used in this paper. In fact, finding the correct strength is not an easy task. The authors intend to elaborate on this important issue in a follow up work. Finally, **J′** undergoes JPEG compression with the same or similar quantization table as that of the original image **J**.

Multiplicative forgery. In this case, Eve adds a scaled PRNU factor modulated by the content of **J** as this better mimics the process if **J** was indeed taken by *C*:

$$\mathbf{J}' = \mathbf{J}[1 + \alpha \hat{\mathbf{K}}]. \tag{8}$$

Both attacks can, indeed, succeed in fooling the camera identification algorithm in the sense that the detector response (correlation [1] or the PCE [9]) will be high enough to indicate that the forged image **J′** was taken by camera *C*. We reiterate that our task is to develop methods that Alice, the victim, could use to her defense and prove that **J′** is a forgery that did not come from her camera. This is the subject of the next section.

3. COUNTER COUNTERMEASURES

Assuming that Alice has her camera C available, we recognize two cases that differ by what data is available to Alice for her forensic investigation and propose the corresponding methods to detect Eve's manipulation. In the first case, analyzed in Section 3.1, Eve has created one forged image $\mathbf{J}$ and Alice has access to at least one of her images, $\mathbf{I}^{(1)},\ldots,\mathbf{I}^{(N)}$, copies of which Eve stole and misused. In the second case (Section 3.2), Eve has produced at least two forgeries, $\mathbf{J}_1'$, $\mathbf{J}_2'$, while Alice has no access to the images used by Eve to estimate the fingerprint.

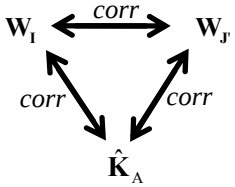
3.1 Stolen images

In this scenario, some of the N images used by Eve to estimate the fingerprint of C are available to Alice but Alice does not know which they are. She has a pool of N_c candidate images that she suspects Eve may have used. This is a plausible scenario because, unless Eve obtains access to Alice's camera and takes images of her own and then erases them from the camera memory card before returning the camera to Alice, Eve will have to use images taken by Alice, such as images posted by Alice on the Internet. It turns out that this case is quite favorable for Alice allowing her to not only prove that the forged image did not originally come from her camera but also determine exactly which images Eve used, thus providing her with a solid defense.

We now explain the key observation based on which Alice can construct her defense. Let $\mathbf{I}$ be one of the images available to Alice that Eve used to forge $\mathbf{J}'$. Because the noise residual $\mathbf{W_I}$ participates in the computation of $\hat{\mathbf{K}}$ through the averaging formula (5), $\mathbf{J}'$ will contain a scaled version of the *entire* noise residual $\mathbf{W_I} = \mathbf{aIK} + \boldsymbol{\theta}$, independently of whether Eve uses an additive or multiplicative forgery. This will make the correlation $c_{\mathbf{I},\mathbf{J}'} = corr(\mathbf{W_I}, \mathbf{W_{J'}})$ larger than what it should 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 significantly fluctuate across images, the test that evaluates the statistical increase must be calibrated. We call this test the *triangle test*.

3.1.1 The triangle test

First, Alice builds her fingerprint $\hat{\mathbf{K}}_{\mathrm{A}}$ from images guaranteed to not have been used by Eve. (She can, for example, take new images with her camera C if needed.) Then, in addition to $c_{\mathbf{I},\mathbf{J}'}$, she computes $c_{\mathbf{I},\hat{\mathbf{K}}_{\mathrm{A}}} = corr(\mathbf{W_I}, \hat{\mathbf{K}}_{\mathrm{A}})$ and $c_{\mathbf{J}',\hat{\mathbf{K}}_{\mathrm{A}}} = corr(\mathbf{W_{J'}}, \hat{\mathbf{K}}_{\mathrm{A}})$ (see the triangle diagram below).



The point is 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 $c_{\mathbf{J},\hat{\mathbf{K}}_{\mathrm{A}}}$ and $c_{\mathbf{J}',\hat{\mathbf{K}}_{\mathrm{A}}}$. On the other hand, when $\mathbf{I}$ *was* used in the forgery, the correlation $c_{\mathbf{I},\mathbf{J}'}$ will be higher than the estimate.

To obtain a more accurate relationship, we will work by blocks of pixels, denoting the signal constrained to block b with subscript b. We adopt the model (1) for the noise residuals and a corresponding model for Alice's fingerprint:

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

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

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

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

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

N_B is the number of blocks, $q \le 1$ is the quality[1] of the fingerprint $\hat{\mathbf{K}}_A$, $q^{-2} = 1 + (SNR_{\hat{\mathbf{K}}_A})^{-1}$, $SNR_{\hat{\mathbf{K}}_A} = \|\mathbf{K}\|^2 / \|\boldsymbol{\xi}\|^2$, and the bar denotes the sample mean. The model coefficients are estimated from the computed block-wise correlations as follows,

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

The dependence between the random variables $c_{\mathbf{I},\mathbf{J}'}$ and $\hat{c}_{\mathbf{I},\mathbf{J}'}$, for images $\mathbf{I}$ *not* used by Eve (innocent images), is reasonably well fit with a straight line $c_{\mathbf{I},\mathbf{J}'} = \lambda \hat{c}_{\mathbf{I},\mathbf{J}'}$ while the deviation from the linear fit seems to be random and independent of $\hat{c}_{\mathbf{I},\mathbf{J}'}$ (see Figure 2a). Thus, we assume that the conditional probability

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

is independent of $\hat{c}_{\mathbf{I},\mathbf{J}'}$ for images $\mathbf{I}$ *not* used by Eve.

Alice runs the following composite hypothesis test

$$\begin{aligned} &\mathrm{H}_0 : c_{\mathbf{I},\mathbf{J}'} - \lambda \hat{c}_{\mathbf{I},\mathbf{J}'} \sim f(x \,|\, \mathbf{I} \text{ not used for forgery of } \mathbf{J}'), \\ &\mathrm{H}_1 : c_{\mathbf{I},\mathbf{J}'} - \lambda \hat{c}_{\mathbf{I},\mathbf{J}'} \sim f(x \,|\, \mathbf{I} \text{ used for forgery of } \mathbf{J}'). \end{aligned} \tag{14}$$

The reason why (14) is not a simple hypothesis test is that the distribution of $c_{\mathbf{I},\mathbf{J}'}$ when $\mathbf{I}$ 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 resort to the Neyman-Pearson test and set our decision threshold t to bound the probability of false alarm, P_{FA}:

$$\Pr(c_{\mathbf{I},\mathbf{J}'} - \lambda \hat{c}_{\mathbf{I},\mathbf{J}'} > t \,|\, \mathrm{H}_0) = P_{\mathrm{FA}}. \tag{15}$$

Alternatively, for each tested image $\mathbf{I}$, Alice can evaluate its p-value and then, on a certain level of statistical significance, identify images that were used by Eve for the fingerprint estimation. The pdf $f(x)$ is often very close to the Gaussian but for some images $\mathbf{J}$, the tails show a hint of a polynomial dependence. Thus, to be conservative, we used Student's t-distribution for the fit. Note that, depending on $\mathbf{J}'$, the constant of proportionality $\lambda > 1$, which suggests the presence of an unknown multiplicative hidden parameter in (10). 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).

3.2 Multiple images with forged fingerprints

Here, we develop forensic techniques for a different, but quite plausible situation. To make her case stronger and convince the judge and the jury, Eve decides to forge more than one image to frame Alice. In particular, she adds *the same* fingerprint to at least two different images from C', $\mathbf{J}_1$ and $\mathbf{J}_2$, and obtains two forged images, $\mathbf{J}_1'$ and $\mathbf{J}_2'$. The new twist here is that Alice can inspect both images (or more pairs if available) but she has no access to images $\mathbf{I}^{(1)},\ldots,\mathbf{I}^{(N)}$ used by Eve to estimate the fingerprint. Interestingly, the same triangle test applies to this case as well because $\mathbf{J}_1'$ and $\mathbf{J}_2'$ will again share another common component besides the PRNU term. The only difference is in replacing the tested image $\mathbf{I}$ with $\mathbf{J}_2'$.

---

[1] The quality of fingerprint $\hat{\mathbf{K}}$ is defined as $q = corr(\mathbf{I},\hat{\mathbf{K}}) / corr(\mathbf{I},\mathbf{K})$ for some flat-field image $\mathbf{I}$ from the same camera and its true PRNU factor $\mathbf{K}$.

4. EXPERIMENTS

In this section, we report all our experiments and implementation details of the forensic methods explained in the previous section.

As in camera identification using sensor fingerprints [1], the signals that enter the triangle test must be preprocessed to remove all common signals called Non-Unique Artifacts (NUAs) [9] introduced by similar in-camera processing, such as demosaicking or on-board signal transfer, and by JPEG compression. Periodic NUAs are suppressed by matrix zero-meaning [1], while artifacts due to JPEG compression and any remaining non-periodic NUAs are suppressed by Wiener filtering in the frequency domain [1].[2]

4.1 Stolen images (experiments)

The experimental setup in general depends on a large number of parameters and choices made by Eve. In presenting our test results, we opted for what we consider (after many initial experiments) to be the most advantageous setting for Eve and the hardest one for Alice:

1. To estimate the fingerprint, Eve uses the most accurate estimator (2) she can find in the literature implemented using the denoising filter described in [10] with parameter $\sigma = 3$.
2. Then, Eve slightly denoises the 24-bit color image **J** using the same denoising filter ($\sigma = 1$) to suppress the fingerprint from the camera *C*' that took **J** and other artifacts introduced by *C*'.
3. Eve determines the fingerprint strength factor α so that other images from Alice's camera that have a similar content also have a similar correlation detector (3) response.
4. Finally, Eve applies the multiplicative modulation (8) because it better mimics the real impact of PRNU and saves the output image as a high-quality JPEG image.

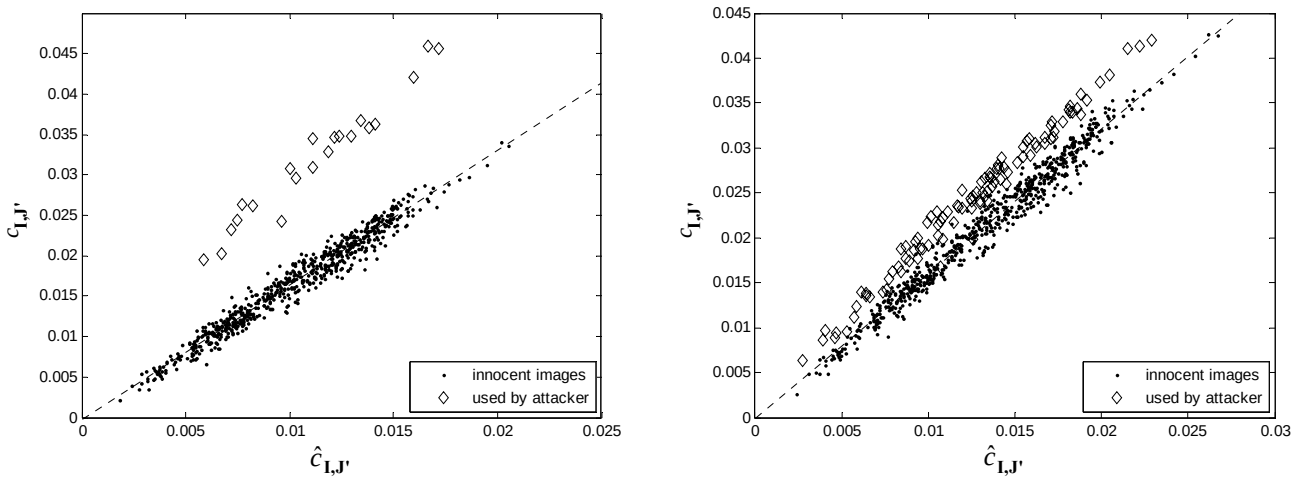
On the defense side, Alice estimates her fingerprint $\hat{\mathbf{K}}_A$ from N_A flat-field images. Surprisingly, the quality of $\hat{\mathbf{K}}_A$ has little impact on the triangle test. Tests with $N_A = 15$ and $N_A = 100$ produced essentially identical results. The experiments below were run for $N_A = 15$ with the quality of $\hat{\mathbf{K}}_A$, $q = 0.806$. The number of blocks N_B, the parameter input in (11), was found to have little impact as well, as long as the block size stayed within some "reasonable range," e.g., $100 \le N_B \le 400$ for a four-megapixel image. In all our experiments, $N_B = 20 \times 20 = 400$, which would correspond to 9,690 pixels in each block.

In order to demonstrate the power of the proposed triangle test, we started with $N = 20$ and then increased it to $N = 50$ and $N = 100$. The six test images **J** (see Figure 1) that Eve attacked were chosen randomly from a set of 250 images downloaded from the image-sharing website Flickr. The source camera *C*' is the Canon PS A520 while the camera *C* is Canon PS G2, which has the same native resolution. We do not show images after their fingerprints are forged as they are visually identical to the originals. Instead, we report the Peak Signal-to-Noise Ratio (PSNR) between **J**′ before it is JPEG compressed and **J**. This distortion includes the denoising Step 2 and quantization to 24-bit colors after adding the fingerprint. The output JPEG quality factor is $Q = 90$, which is slightly smaller than the quality Q_0 of the original JPEG images (see the first column in Table 1). To accurately estimate the pdf $f(x|H_0)$, we used 669 images from camera *C* that were for sure not used by Eve. The images were all taken within a period of about four years. In practice, depending on the situation, sufficient accuracy may be obtained using a much smaller sample.

Figure 2 presents a typical plot of $c_{\mathbf{I},\mathbf{J}'}$ vs. $\hat{c}_{\mathbf{I},\mathbf{J}'}$ for $N = 20$ and $N = 100$. Quite understandably, the separation between innocent images and those used by Eve deteriorates with increasing *N*. The probability of correct detection P_D in the hypothesis test (14) for all six test images is shown in Table 1. Each value of P_D was obtained by running the entire experiment as depicted in Figure 2 and evaluating the p-values for all images $\mathbf{I}^{(1)},\ldots,\mathbf{I}^{(N)}$ used by Eve.

[2] We skipped filtering of $\mathbf{W}_\mathbf{I}$ for all tested images (both those used and not used by Eve) to save computation time when producing experimental data for this paper.

Figure 1. Original images (Canon PS A520, source Flickr.com), (#1, #2, #3); (#4, #5, #6).



(a) $N = 20$ (b) $N = 100$

Figure 2. Correlations $c_{\mathbf{I,J'}}$ vs. $\hat{c}_{\mathbf{I,J'}}$ for image **J** #5 and 669 "innocent" images. Eve estimated the fingerprint from N images.

The lower detection rate for image #3 is likely 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 involves identifying pixels that were originally saturated and setting their attenuation factors to zero.

Table 1. Detection rate P_D [%] and PSNR for the case analyzed in Section 4.1, JPEG quality $Q = 90$.

Image #,(Q_0,Q)	PSNR($\mathbf{J}$, $\mathbf{J}'$) [dB]			P_D [%] for $P_{FA} = 0.001$			P_D [%] for $P_{FA} = 0.0001$		
	$N = 20$	$N = 50$	$N = 100$	$N = 20$	$N = 50$	$N = 100$	$N = 20$	$N = 50$	$N = 100$
1 (93,90)	52.58	52.56	52.57	100	98	76	100	98	61
2 (93,90)	52.77	52.72	52.71	100	98	48	100	96	29
3 (93,90)	51.67	51.63	51.59	100	60	20	95	40	6
4 (93,90)	51.01	51.02	51.00	100	94	25	100	70	6
5 (93,90)	51.50	51.47	51.45	100	94	48	100	90	11
6 (97,90)	52.61	52.56	52.53	100	100	67	100	94	54

The success of the triangle test largely depends on the number of images, N, used by Eve to estimate the fingerprint of camera C. The triangle test in our study is quite reliable when N is small, e.g., $N < 30$. If Eve has enough resources and can obtain a large number of images, say $N > 200$, the reliability of the triangle test applied to each image one by one will likely be quite low. However, the situation is not completely hopeless for Alice if she has a set of N_c candidate images while a good portion of them are among those N used by Eve. After testing all N_c images one by one, Alice can pool the results and test whether the differences from the model $c_{\mathbf{I},\mathbf{J}'} - \lambda \hat{c}_{\mathbf{I},\mathbf{J}'}$ indeed follow $f(x|\mathrm{H}_0)$. Since the differences are independent across different images, the scaled log-likelihood of all N_c observations

$$L_{N_c} = \frac{1}{\sqrt{N_c}} \sum_{i=1}^{N_c} \log\left(f\left(c_{\mathbf{I},\mathbf{J}'}^{(i)} - \lambda \hat{c}_{\mathbf{I},\mathbf{J}'}^{(i)} \mid \mathrm{H}_0\right)\right), \tag{16}$$

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

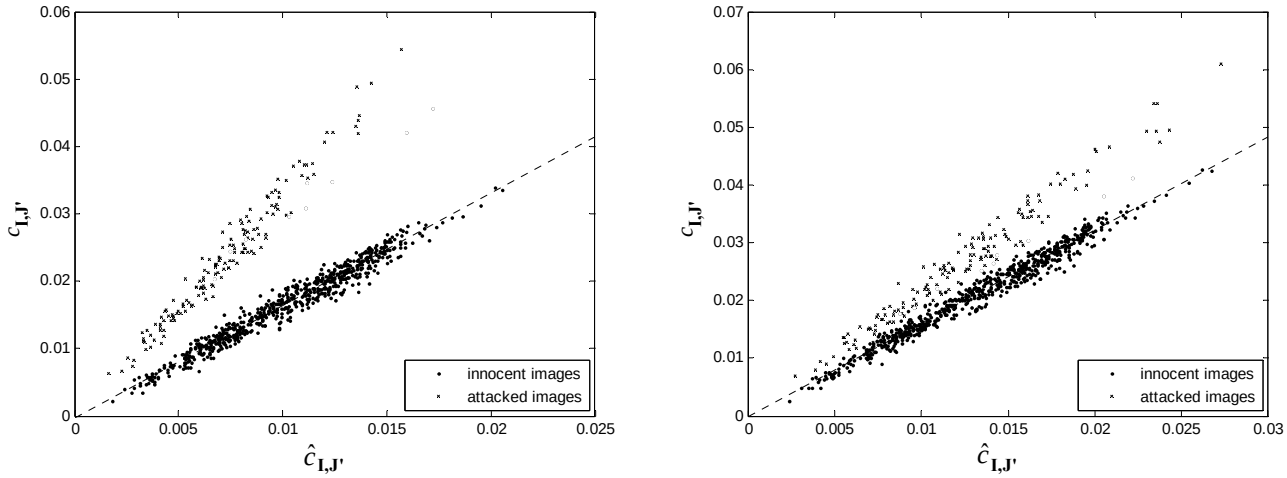
To demonstrate this pooling approach, we provide a sample result with the test image #5 for $N = 100$ and $N_c = 200$. If Alice randomly selects, say, $k = 40$ images out of N_c candidates, approximately $k/2$ of them were used by Eve to estimate her fingerprint. With $P_{FA} = 0.001$ (or $P_{FA} = 0.0001$), by testing single images, $P_D \approx 48\%$ (or 11%) (see Table 1). After pooling all 40 images, however, $P_D \approx 99\%$ (or 82%). These probabilities were obtained by repeating the pooling test using bootstrapping 500 times.

4.2 Multiple images with forged fingerprints (experiments)

In this section, we put to test the forensic method proposed in Section 3.2. Here, Alice has no access to any of the images used by Eve to estimate the fingerprint, however, Eve has forged two images, $\mathbf{J}_1'$ and $\mathbf{J}_2'$, and Alice can inspect them both. The method is the same triangle test applied to $\mathbf{J}_1'$ and $\mathbf{J}_2'$.

To enlarge the experiments, we downloaded additional 147 images $\mathbf{J}$ from camera C' from Flicker and created 3×147 (147 for each choice of N) forgeries by adding Eve's fingerprint to them using the same parameters as described in the beginning of Section 4.1. All other images are reused from the previous section. Figure 3 and Table 2 are the equivalents of Figure 2 and Table 1. Because the same fingerprint was added to each forgery, the PSNR between $\mathbf{J}$ and $\mathbf{J}'$ was the same as indicated in Table 1. This experiment was run separately for $N = 20, 50, 100$.

In general, the larger the correlation between $\mathbf{J}$ and $\mathbf{J}_2'$, the more reliably the triangle test can decide between the hypotheses (14). The decision becomes less reliable when the correlation is low, a situation induced by our rather conservative choice of the fingerprint strength α. This is apparent in Table 2 in the column $N = 20$ that shows some missed detections. On the other hand, in this test, P_D does not drop with N as fast as in the "stolen images" case studied in Section 4.1 and reported in Table 1.



(a) $N = 20$ (b) $N = 100$

Figure 3. Correlations $c_{\mathbf{I},\mathbf{J}'}$ vs. $\hat{c}_{\mathbf{I},\mathbf{J}'}$ for image **J** #5 and 669 "innocent" images **I**. For the 147 forged images, $\mathbf{I} = \mathbf{J}_2'$ ran through all 147 images and $\mathbf{J}' = \mathbf{J}_1'$ was fixed.

Table 2. Detection rate P_D [%] for the method reported in Section 4.2. The JPEG quality factor $Q = 90$, $N_A = 15$.

P_D [%]	$P_{FA} = 0.001$			$P_{FA} = 0.0001$		
Image #	$N = 20$	$N = 50$	$N = 100$	$N = 20$	$N = 50$	$N = 100$
1	97	95	85	95	92	79
2	95	93	77	92	91	66
3	82	63	40	73	52	28
4	97	89	71	95	83	61
5	97	95	77	95	90	59
6	99	92	80	95	87	70

5. CONCLUSION

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. We designed a test, the so-called "triangle test," using which Alice can identify which images 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. 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.

Finally, we would like to mention that in our investigation, we have explored many diverse ideas, most of which ended up in a dead end. Given a forged image **J**, a seemingly promising idea was to acquire the same scene using Alice's camera (the so-called baseline image) and then compare the fingerprint presence in the two images for consistency, for example, on small blocks. The baseline image could be obtained by displaying **J** on a computer monitor and taking picture of it using Alice's camera. Alternatively, we may print out **J** on a high-quality printer and again take a picture of it. The biggest problem with this approach is the fact that the fine-grain details of the baseline image and those of **J** will likely be quite different, resulting in a very different noise term $\boldsymbol{\Theta}$ (see (1)). Additional complications are caused by a different gamma factor, distorted colors, or the interference between the discrete nature of the display and the camera sensor.

6. ACKNOWLEDGEMENTS

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

REFERENCES

[1] 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.

[2] Kutter, M., Voloshynovskiy, S., and Herrigel, A.: "The Watermark Copy Attack." *Proc. SPIE, Security and watermarking of multimedia contents II*, vol. 3971, San Jose, CA, January 24–26, pp. 371–380, 2000.

[3] Gloe, T., Kirchner, M., Winkler, A., and Böhme, R.: "Can we Trust Digital Image Forensics?" In *ACM Multimedia (ACMMM)*. ACM, Augsburg, Germany, pp. 78–86, 2007.

[4] Popescu, A. and Farid, H.: "Exposing Digital Forgeries in Color Filter Array Interpolated Images." *IEEE Transactions on Signal Processing* **53**(10), 3948–3959, 2005.

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

[6] 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**.**

[7] 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.

[8] 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**.**

[9] 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.

[10] 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.

[11] Fridrich, J., Chen, M., Goljan, M., and Lukáš, J.: "Digital Imaging Sensor Identification (Further Study)." *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, vol. 6505, San Jose, CA, January 28–February 2, pp. 0P–0Q, 2007.

7. APPENDIX

We assume the model (9) for the noise residual $\mathbf{W_I}$ from image **I** and the camera fingerprint **K**. Furthermore, we adopt the following simplifying assumptions.

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

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

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 $\mathbf{K}_b$ is a realization of a stationary random variable with finite variance:

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

where we denoted $\mathbf{p}_k = \frac{1}{\| \mathbf{K}_b \|^2} \sum_{\mathbf{I}_b[i]\mathbf{J}_b[i]=k} \mathbf{K}_b^2[i]$. To explain the last approximate equality, realize that for a stationary signal $\mathbf{K}_b[i]$, $i = 1, \ldots, N_b$, 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

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

which can be simplified by introducing

$$SNR_\mathbf{I} = \frac{\sum_b a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2}{\sum_b \| \mathbf{\Theta}_{\mathbf{I},b} \|^2} , \; SNR_{\hat{\mathbf{K}}_\mathrm{A}} = \frac{\sum_b \| \mathbf{K}_b \|^2}{\sum_b \| \xi_b \|^2} \tag{A.5}$$

$$corr(\mathbf{W}_\mathbf{I}, \hat{\mathbf{K}}_\mathrm{A}) = \frac{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b} \| \mathbf{K}_b \|^2}{\sqrt{\sum_b a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2} \sqrt{1 + \frac{1}{SNR_\mathbf{I}}} \sqrt{\sum_b \| \mathbf{K}_b \|^2} \sqrt{1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}}} = \frac{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b} \| \mathbf{K}_b \|^2}{\sqrt{\sum_b a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2} \sqrt{1 + \frac{1}{SNR_\mathbf{I}}} \sqrt{1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}}} \| \mathbf{K} \|^{-1} \tag{A.6}$$

Similarly, we now write for

$$corr(\mathbf{W}_\mathbf{I}, \mathbf{W}_\mathbf{J}) = \frac{\sum_b a_{\mathbf{I},b}a_{\mathbf{J},b}\overline{\mathbf{I}_b\mathbf{J}_b} \| \mathbf{K}_b \|^2}{\sqrt{\sum_b a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2} \sqrt{\sum_b a_{\mathbf{J},b}^2 \overline{\mathbf{J}_b^2} \| \mathbf{K}_b \|^2} \sqrt{1 + \frac{1}{SNR_\mathbf{I}}} \sqrt{1 + \frac{1}{SNR_\mathbf{J}}}} . \tag{A.7}$$

By comparing (A.6) and (A.7), we see that

$$corr(\mathbf{W}_\mathbf{I}, \mathbf{W}_\mathbf{J}) = corr(\mathbf{W}_\mathbf{I}, \hat{\mathbf{K}}_\mathrm{A}) corr(\mathbf{W}_\mathbf{J}, \hat{\mathbf{K}}_\mathrm{A}) \frac{\sum_b a_{\mathbf{I},b}a_{\mathbf{J},b}\overline{\mathbf{I}_b\mathbf{J}_b} \| \mathbf{K}_b \|^2}{\sum_b a_{\mathbf{I},b}\overline{\mathbf{I}_b} \| \mathbf{K}_b \|^2 \cdot \sum_b a_{\mathbf{J},b}\overline{\mathbf{J}_b} \| \mathbf{K}_b \|^2} \| \mathbf{K} \|^2 \left(1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}} \right) . \tag{A.8}$$

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:

$$\| \mathbf{K}_b \|^2 = \frac{\| \hat{\mathbf{K}}_{\mathrm{A},b} \|^2}{1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}}} \tag{A.9}$$

and the formula for $corr(\mathbf{W_I}, \mathbf{W_J})$ can be expressed using the norm of $\hat{\mathbf{K}}_\mathrm{A}$ rather than the unknown norm of $\mathbf{K}$:

$$corr(\mathbf{W_I}, \mathbf{W_J}) = corr(\mathbf{W_I}, \hat{\mathbf{K}}_\mathrm{A}) corr(\mathbf{W_J}, \hat{\mathbf{K}}_\mathrm{A}) \frac{\sum_b a_{\mathbf{I},b} a_{\mathbf{J},b} \overline{\mathbf{I}_b \mathbf{J}_b} \| \hat{\mathbf{K}}_{\mathrm{A},b} \|^2}{\sum_b a_{\mathbf{I},b} \overline{\mathbf{I}_b} \| \hat{\mathbf{K}}_{\mathrm{A},b} \|^2 \cdot \sum_b a_{\mathbf{J},b} \overline{\mathbf{J}_b} \| \hat{\mathbf{K}}_{\mathrm{A},b} \|^2} \| \hat{\mathbf{K}}_\mathrm{A} \|^2 \left(1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}} \right). \tag{A.10}$$

Determining the factors $a_{\mathbf{I},b}$:

$$\mathbf{c}_{\mathbf{I},b} = corr(\mathbf{W}_{\mathbf{I},b}, \mathbf{I}_b \hat{\mathbf{K}}_{\mathrm{A},b}) = \frac{a_{\mathbf{I},b} \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2}{\sqrt{a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2} \sqrt{1 + \frac{\| \mathbf{\Theta}_{I,b} \|^2}{a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2}} \sqrt{\overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2} \sqrt{1 + \frac{\| \xi_b \|^2}{\| \mathbf{K}_b \|^2}}} = \tag{A.11}$$

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

From here after some simple algebra,

$$\frac{\| \mathbf{\Theta}_{\mathbf{I},b} \|^2}{a_{\mathbf{I},b}^2} = \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2 \left\{ \mathbf{c}_{\mathbf{I},b}^{-2} \left(1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}} \right)^{-1} - 1 \right\}. \tag{A.13}$$

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

$$\| \mathbf{W}_{\mathbf{I},b} \|^2 = a_{\mathbf{I},b}^2 \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2 + \| \mathbf{\Theta}_{\mathbf{I},b} \|^2. \tag{A.14}$$

Because the solution to the system

$$\begin{array}{ll} y / x = R & x = S/(d+R) \\ xd + y = S & \text{is} \quad y = SR/(d+R) \end{array} \tag{A.15}$$

and because in our case R is the r.h.s. of (A.13), $d = \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2$, $S = \| \mathbf{W}_{\mathbf{I},b} \|^2$, $x \triangleq a_{\mathbf{I},b}^2$, we obtain

$$a_{\mathbf{I},b}^2 = \frac{\| \mathbf{W}_{\mathbf{I},b} \|^2}{\overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2 + \overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2 \left\{ \mathbf{c}_{\mathbf{I},b}^{-2} \left(1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}} \right)^{-1} - 1 \right\}} = \frac{\| \mathbf{W}_{\mathbf{I},b} \|^2}{\overline{\mathbf{I}_b^2} \| \mathbf{K}_b \|^2} \mathbf{c}_{\mathbf{I},b}^2 \left(1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}} \right), \tag{A.16}$$

which can be written using (A.9) to involve the norm of $\hat{\mathbf{K}}_{\mathrm{A},b}$ rather than $\mathbf{K}$:

$$a_{\mathbf{I},b}^2 = \frac{\| \mathbf{W}_{\mathbf{I},b} \|^2}{\overline{\mathbf{I}_b^2} \| \hat{\mathbf{K}}_{\mathrm{A},b} \|^2} \mathbf{c}_{\mathbf{I},b}^2 \left(1 + \frac{1}{SNR_{\hat{\mathbf{K}}_\mathrm{A}}} \right)^2. \tag{A.17}$$