

Sensor Fingerprint Digests for Fast Camera Identification from Geometrically Distorted Images

Miroslav Goljan and Jessica Fridrich

Department of ECE, SUNY Binghamton, NY, USA, {mgoljan,fridrich}@binghamton.edu

ABSTRACT

In camera identification using sensor fingerprint, it is absolutely essential that the fingerprint and the noise residual from a given test image be synchronized. If the signals are desynchronized due to a geometrical transformation, fingerprint detection becomes significantly more complicated. Besides constructing the detector in an invariant transform domain (which limits the type of the geometrical transformation) a more general approach is to maximize the generalized likelihood ratio with respect to the transform parameters, which requires a potentially expensive search and numerous resamplings of the entire image (or fingerprint). In this paper, we propose a measure that significantly reduces the search complexity by reducing the need to resample the entire image to a much smaller subset of the signal called the fingerprint digest. The technique can be applied to an arbitrary geometrical distortion that does not involve spatial shifts, such as digital zoom and non-linear lens-distortion correction.

1. INTRODUCTION

The concept of identifying a specific camera using sensor fingerprint has been introduced in 2005. Since then, the technique has been substantially improved and generalized to be applicable in a wide range of practical situations.^{1-8,15} The sensor fingerprint is caused by subtle yet consistent deviations from the nominal output of pixels. It needs to be estimated from one or more images positively known to have been taken by the camera. To provide evidence that a given image under investigation was taken by the *exact same* camera (not just the same model), one establishes the presence of the fingerprint in the image. A positive match between an image and a camera fingerprint ties the image with very high certainty to the camera. Applications of this technology include all cases when a crime is committed by taking a picture, such as in child pornography and movie piracy cases, as well as intelligence gathering and criminal investigation.*

The sensor fingerprint is mostly formed by the so-called Photo-Response Non-Uniformity (PRNU)¹³ due to the fact that each pixel consistently outputs voltage that is slightly different from its nominal value. As a result, each image the sensor takes is overlaid with a unique noise-like pattern modulated by the scene light intensity. The presence of a specific known fingerprint in an image can be established using standard deterministic-signal detectors, whose specific form depends on the statistical nature of the modeling noise. With unknown noise parameters, the fingerprint detection is correctly framed as a two-channel detector.¹⁰ Due to the complexity of the Generalized Likelihood Ratio (GLR) detector, practitioners use the simplified Peak to Correlation Energy (PCE) ratio as a convenient test statistic. A large-scale test with nearly 6,900 cameras and over one million images³ indicates that thresholding the PCE with $\tau = 60$ corresponds to probability of 10^{-5} of falsely identifying an image as taken by a specific camera (false alarm).

As with any correlation-based detector, it is important that the fingerprint and the image be synchronized. Due to the random character of the fingerprint, however, even a small desynchronization by a single pixel may lead to a missed detection. In particular, when some geometrical processing was applied to the image under investigation, the detector needs to resynchronize it with the fingerprint. This could be achieved by constructing the detector in a transform-invariant domain.^{12,14} However, since this approach needs to be developed for each

*This forensic method passed the Daubert challenge http://en.wikipedia.org/wiki/Daubert_standard in the State of Alabama in July 2011. In March 2009, M. Goljan testified in Scotland as an expert witness in a high-profile case that involved child abuse crimes by a pedophile ring. See the article "Operation Algebra" at <http://p10.hostingprod.com/@spyblog.org.uk/blog/2009/05/>.

transform type separately, it is not a viable solution if one desires a general solution applicable to an arbitrary transform, including focal-length-dependent non-linear lens distortion correction nowadays commonly used in compact digital cameras.⁷ An alternative approach, also taken in this paper, is to maximize the PCE over all parameters of the geometrical transform. An example of this approach for images subjected to scaling/cropping and lens-distortion correction appeared in Ref. [5] and Ref. [7], respectively.

The search requires resampling either the image or the fingerprint for each value of the parameter vector that describes the geometrical transform. Due to the random character of the fingerprint, the peak in the PCE w.r.t. the parameter is rather narrow, which limits the applications of faster gradient search methods – the bulk of the search needs to be carried out by a simple grid search. In Ref. [7], the authors lowered the complexity of this search by working with resized versions of both signals (resized by a factor of 3). In this paper, we propose a different approach that uses only the outlier values in the fingerprint – the so-called fingerprint digest⁹ – to obtain a much better trade-off between computational complexity and the detection reliability.

In the next section, we introduce some background concepts, procedures, and prior art relevant for this work. Then, in Section 3, we describe the proposed method in detail. Experiments confirming the merit of the proposed approach together with a comparison to prior art appear in Section 4. The paper is concluded in Section 5.

2. PRELIMINARIES

Boldface symbols will be reserved for matrices while the calligraphic font will be used solely for sets. The ij th element of matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ will be denoted as $\mathbf{A}[i, j]$. In general, for a subset of indices $\mathcal{D} \subset \{0, \dots, M - 1\} \times \{0, \dots, N - 1\}$, $\mathbf{A}[\mathcal{D}]$ is a vector of elements of \mathbf{A} in some fixed order. For set \mathcal{X} , its cardinality is denoted as $|\mathcal{X}|$. All arithmetic operations between matrices will be elementwise. Finally, Greek letters will be used to denote noise components while θ is reserved for the geometrical transform parameter.

On the most general level, the sensor fingerprint detection can be formalized as follows. Given two noisy estimates of sensor fingerprints, two matrices $\hat{\mathbf{K}}^{(1)}$ and $\hat{\mathbf{K}}^{(2)}$, the task is to decide whether they contain a common signal – the sensor fingerprint \mathbf{K} . It is assumed that $\hat{\mathbf{K}}^{(1)}$ has been estimated from $n_1 \geq 1$ images, all processed using the same geometrical transform, while $\hat{\mathbf{K}}^{(2)}$ was estimated from $n_2 \geq 1$ images, also assumed to have undergone the same geometrical transform but with a potentially different parameter than the one for $\hat{\mathbf{K}}^{(1)}$.

The fingerprint detection can be formulated as the following two-channel hypothesis problem:¹⁰

$$H_0 : \hat{\mathbf{K}}^{(1)} = \mathbf{K} + \xi^{(1)}, \hat{\mathbf{K}}^{(2)} = T_\theta(\mathbf{K}') + \xi^{(2)}, \quad (1)$$

$$H_1 : \hat{\mathbf{K}}^{(1)} = \mathbf{K} + \xi^{(1)}, \hat{\mathbf{K}}^{(2)} = T_\theta(\mathbf{K}) + \xi^{(2)}, \quad (2)$$

where \mathbf{K} and \mathbf{K}' are sensor fingerprints coming from two different cameras, $\xi^{(i)}$, $i = 1, 2$, are i.i.d. Gaussian modeling noise terms with unknown variances, and T_θ is the geometrical transform between $\hat{\mathbf{K}}^{(1)}$ and $\hat{\mathbf{K}}^{(2)}$ parametrized by an unknown (vector) parameter θ . The null hypothesis corresponds to the situation when the observables come from two different devices, while under the alternative hypothesis both observables come from the same device (a positive match).

A simplification of the GLR test for this problem that appears in Ref. [10] to obtain a detection statistic with a reasonable computational complexity, is the Peak Correlation to Energy (PCE) ratio maximized over all possible values of θ :

$$\delta(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)}) \triangleq \max_{\theta} \text{PCE}(T_\theta(\hat{\mathbf{K}}^{(1)}), \hat{\mathbf{K}}^{(2)}), \quad (3)$$

where

$$\text{PCE}(\mathbf{X}, \mathbf{Y}) \triangleq \frac{\rho_{\max}^2}{\frac{1}{MN - |\mathcal{N}_{\max}|} \sum_{(i,j) \notin \mathcal{N}_{\max}} \rho^2[i, j]}, \quad \rho[i, j] \triangleq \frac{\sum_{(k,l)} (\mathbf{X}[k, l] - \bar{\mathbf{X}}) (\mathbf{Y}[k - i, l - j] - \bar{\mathbf{Y}})}{\sqrt{\sum_{(k,l)} (\mathbf{X}[k, l] - \bar{\mathbf{X}})^2} \sqrt{\sum_{(k,l)} (\mathbf{Y}[k, l] - \bar{\mathbf{Y}})^2}}, \quad (4)$$

where $[i_{\max}, j_{\max}] = \arg \max_{i,j} \rho[i, j]$, $\rho_{\max} = \rho[i_{\max}, j_{\max}]$, M, N are the fingerprint dimensions, $0 \leq i, k \leq M - 1$, $0 \leq j, l \leq N - 1$, \mathcal{N}_{\max} is a small neighborhood of $[i_{\max}, j_{\max}]$, and the bar denotes the sample mean. Additionally,

we assume circular symmetry of both signals to resolve the issue when either $k - i$ or $l - j$ get out of their range.[†] This form of the PCE is suitable when the geometrical transform, T_θ , includes a spatial shift (cropping). This is implicit in the definition of the PCE where we take the maximal value of the cross-correlation ρ . If the analyst has a priori knowledge that T_θ does not involve any spatial shift (e.g., for lens-distortion correction or digital zoom), the maximum value of $\rho[i, j]$ (under H_1) occurs as $i = j = 0$, and the PCE attains the form

$$\text{PCE}(\mathbf{X}, \mathbf{Y}) \triangleq \frac{\rho^2[0, 0]}{\frac{1}{MN - |\mathcal{N}_{[0,0]}} \sum_{(i,j) \notin \mathcal{N}_{[0,0]}} \rho^2[i, j]}, \quad (5)$$

where now $\mathcal{N}_{[0,0]}$ is a small neighborhood of $[0, 0]$.

The detection thus involves a numerical maximization of the PCE as a function of θ . As already mentioned above, the random character of the signals involved makes the peak in θ rather narrow, which precludes application of fast gradient-based methods and leaves us with a potentially rather expensive search on a fine grid. The most time consuming parts to compute the PCE (3) for a given θ are the geometrical transform, which will generally require resampling on a non-uniform pixel grid, and evaluating the cross-correlation $\rho[i, j]$ for all i, j .

Given n images $\mathbf{I}^{(1)}, \dots, \mathbf{I}^{(n)}$ obtained by the same camera, one of the most commonly used procedures for estimating the sensor fingerprint is the estimator derived in Ref. [3]:

$$\hat{\mathbf{K}} = \frac{\sum_{i=1}^n \mathbf{W}^{(i)} \mathbf{I}^{(i)}}{\sum_{i=1}^n (\mathbf{I}^{(i)})^2}, \quad (6)$$

where $\mathbf{W}^{(i)}$ is the noise residual of $\mathbf{I}^{(i)}$ obtained using a denoising filter F , $\mathbf{W}^{(i)} = \mathbf{I}^{(i)} - F(\mathbf{I}^{(i)})$, $i = 1, \dots, n$.

By far the most common case in practice is when an analysts is matching a single image \mathbf{I} to a camera with a fingerprint estimate $\hat{\mathbf{K}}^{(2)}$. In this case, $n_1 = 1$, $\mathbf{I}^{(1)} = \mathbf{I}$, $\mathbf{W}^{(1)} = \mathbf{W}$, and the estimator (6) simplifies into $\hat{\mathbf{K}}^{(1)} = \mathbf{W}/\mathbf{I}$. Following the reasoning in Ref [3], we will use a modified detector in the form

$$\delta_1(\mathbf{W}, \hat{\mathbf{K}}^{(2)}) \triangleq \max_{\theta} \text{PCE}(T_\theta(\mathbf{W}), T_\theta(\mathbf{I})\hat{\mathbf{K}}^{(2)}), \quad (7)$$

for this situation. A disadvantage of this detector is that whenever T_θ is not an identity, while searching for θ we need to resample both \mathbf{W} and \mathbf{I} , which increases the computational complexity. A possible remedy exists when T_θ is close to identity (e.g., for lens-distortion correction) assuming the image intensity (content) does not change significantly over small neighborhoods. In this case, the complexity can be lowered by approximating the term $T_\theta(\mathbf{I})$ with the image intensity itself:

$$\delta_2(\mathbf{W}, \hat{\mathbf{K}}^{(2)}) \cong \max_{\theta} \text{PCE}(T_\theta(\mathbf{W}), \hat{\mathbf{K}}^{(2)}). \quad (8)$$

3. PCE WITH SIGNAL DIGESTS

As stated in the introduction, the goal of this article is to speed up the search for the nuisance parameter θ . To this end, we propose to work with a suitably chosen subset of both arguments of the PCE.[‡] This will decrease the complexity of the resampling operation T_θ and speed up the computation of the test statistics (3), (5), (7), and (8).

Given a signal \mathbf{S} , imagine that we leave its $k \ll MN$ largest (in absolute value) elements unchanged and set the remaining elements in \mathbf{S} to zero. When this operation is performed on both arguments entering (any form of) the PCE statistic, we are essentially introducing noise into the signals, which has the effect of decreasing the PCE. We already, however, achieve a speed up as, for each θ in our search we only need to resample k values

[†]Note the slight difference between the PCE used in our previous work,^{1-8, 15} where the peak was determined by the maximum of $\rho^2[i, j]$ rather than $\rho[i, j]$. This is to prevent false identification due to strong anticorrelation (negative ρ) between the signals, assuming such a case can only happen due to strong non-unique artifacts present in the fingerprints.

[‡]The idea of speeding up the camera identification using sensor fingerprints by restraining the computations to the largest components in the fingerprint was independently proposed by Hu et al.¹¹ and by Goljan et al.⁹

instead of MN . Note that if T_θ involves a spatial shift, one needs to compute the full 2D cross-correlation $\rho[i, j]$ and search for its maximum value. In this case, no savings in complexity is obtained as the 2D cross-correlation is best implemented as a convolution using Fast Fourier Transform (FFT), which “does not see” the signal sparseness. On the other hand, if we know that there is no spatial shift, we can work only with the k outlier values, order them into a k -dimensional vector and perform a 1D cross-correlation, achieving thus additional savings in complexity.[§]

For the remainder of this article, we will constrain ourselves to transformations T_θ that do not involve any spatial shift. This means that both δ_1 (7) and δ_2 (8) will be computed using the PCE in the form (5). For transformations with a spatial shift, the authors suggest the previously proposed downsampling to speed up the search for the geometrical transform parameter.⁷

By ordering the indices (positions) of the k largest outliers of \mathbf{S} into an ordered set $\mathcal{D} \subset \{0, \dots, M-1\} \times \{0, \dots, N-1\}$ of cardinality k , we obtain the so-called *digest* of \mathbf{S} . Sometimes, we will refer to $\mathbf{S}[\mathcal{D}]$ as the digest hoping that confusion will not arise as to whether we are talking about values of \mathbf{S} or its indices.

To evaluate the PCE statistic on digests in practice, we work with the *inverse* mapping, T_θ^{-1} , as this can be implemented much easier. For example, to compute, $\text{PCE}(T_\theta(\mathbf{W}), \mathbf{I}\hat{\mathbf{K}}^{(2)})$ for a digest \mathcal{D} of $\hat{\mathbf{K}}^{(2)}$ (the fingerprint digest), we first compute the k *locations* where the digest pixels are mapped under T_θ^{-1} and then interpolate \mathbf{W} *only* around these k points (we use a linear interpolation using the four closest neighbors) and denote the interpolated signal $\mathbf{W}[T_\theta^{-1}(\mathcal{D})]$. Once the signals are reduced to digests, they become vectors of length k , $\mathbf{x} = \mathbf{W}[T_\theta^{-1}(\mathcal{D})]$, $\mathbf{y} = \mathbf{I}[\mathcal{D}]\hat{\mathbf{K}}^{(2)}[\mathcal{D}]$, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^k$. Evaluating the normalized cross-correlation ρ in the PCE thus involves only a 1D sum over the digest rather than a 2D sum. To be absolutely precise, to evaluate the PCE (8) at digest \mathcal{D} ,

$$\text{PCE}\left(T_\theta(\mathbf{W}), \mathbf{I}\hat{\mathbf{K}}^{(2)}\right) = \text{PCE}(\mathbf{x}, \mathbf{y}) = \frac{\rho^2[0]}{\frac{1}{k-1} \sum_{i \neq 0} \rho^2[i]}, \quad (9)$$

where now

$$\rho[i] \triangleq \frac{\sum_{d=0}^{k-1} (\mathbf{x}[d] - \bar{\mathbf{x}})(\mathbf{y}[d-i] - \bar{\mathbf{y}})}{\sqrt{\sum_{d=0}^{k-1} (\mathbf{x}[d] - \bar{\mathbf{x}})^2} \sqrt{\sum_{d=0}^{k-1} (\mathbf{y}[d] - \bar{\mathbf{y}})^2}}, \quad i = 0, \dots, k-1. \quad (10)$$

Thus, computing PCE on a digest with k elements reduces the computational complexity by the factor of MN/k .

Notice that instead of working with the fingerprint digest, we could work with the digest of the *product* $\mathbf{I}\hat{\mathbf{K}}^{(2)}$ taking thus into consideration the image intensities. In fact, later in this paper, we will explore the possibility to further improve the performance of the proposed methods exactly in this manner. As yet another note, we remark that to compute (7) on a fingerprint digest, we map the fingerprint digest $T_\theta^{-1}(\hat{\mathbf{K}}^{(2)}[\mathcal{D}])$ and interpolate both \mathbf{W} and \mathbf{I} around these k mapped values before computing the PCE in a similar manner as above.

As already pointed out above, working with signal subsets introduces noise in the signals and thus reduces the PCE. Figure 1 shows that this loss is much smaller when we use digests than when the signal length is reduced by trimming (cropping). The dependence of the PCE (9) on the digest length was derived in Ref. [9] and is marked as “Digest – model” in the figure legend. Thus, using fingerprint digests for reducing the grid search complexity is likely to be very effective as well. It is important to stress here that the digest will be used only to estimate the parameter θ during the grid search. The final detection statistic will always be evaluated using either (5) or (7) for the full fingerprint size with θ obtained by the grid search. This way, we ensure that the false alarm probability is not affected by the method used to estimate θ .

For some transformations, it may be important to consider not only the values of the fingerprint elements but also their position w.r.t. the geometrical transform. For example, the radial lens distortion correction⁷ is more accurately estimated from digest pixels on the perimeter of the image, where the distortion is the largest. Thus, it might be better to select the digest based not only on the outlier fingerprint values but also consider their distance to the image center. However, our investigative experiments that attempted to prefer pixels with larger distance from the image center failed to give more accurate estimates of θ . On the positive side, we confirmed

[§]Replacing PCE with the simple correlation coefficient, ρ , in the objective function during the grid search phase is also possible for additional computation speed-up.

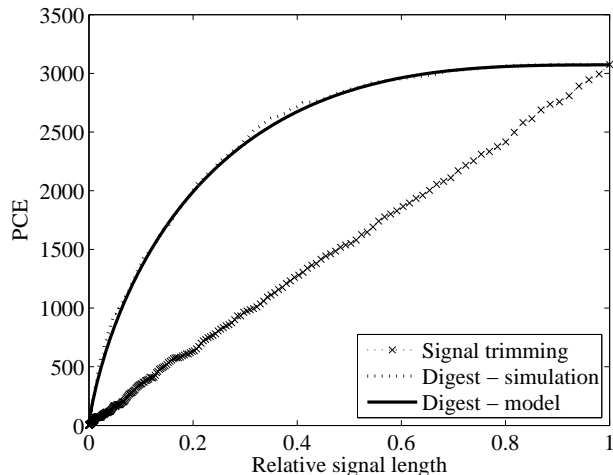


Figure 1. PCE $\delta(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)})$ (3) as a function the relative signal length for two fingerprint estimates, $\hat{\mathbf{K}}^{(1)} = \mathbf{K} + \xi^{(1)}$, $\hat{\mathbf{K}}^{(2)} = \mathbf{K} + \xi^{(2)}$, with \mathbf{K} simulated as a Gaussian signal with the noise terms leading to $\rho_{\max} = 0.04$. With decreasing signal length, the PCE decreases linearly for trimming while it stays significantly higher for digests.

that restricting the digest \mathcal{D} to pixels that are not in the proximity of the image center does eliminate occasional missed detection cases caused by inaccurate or wrong estimation of the distortion parameter θ . In particular, by omitting 1/12 of all pixels lying within a small circle around the center, we observed a $\sim 3\%$ increase in the final PCE (5) ($\sim 97\%$ unchanged).

4. EXPERIMENTS

We demonstrate the merit of the proposed idea on the example of an in-camera lens-distortion correction transform T_θ with a scalar parameter θ , which controls how the original distance of each pixel to the image center, r , changes after the transform is applied:

$$r' = r(1 + \theta r^2). \quad (11)$$

We note that this equation is commonly used for modeling barrel and pincushion optical distortion.¹⁶

4.1. Prior art

The proposed method of digests will be compared with the method described in Ref. [7]. In this section, we briefly describe this approach while referring to the original publication for details.

The method consists of a two-phase search for θ on downsampled signals after which the final PCE is evaluated for the full size signals. The first phase of the search is a simple grid search on progressively refined grids. If this coarse search succeeds to find a PCE larger than a certain threshold, the procedure continues with the second, refining phase, which is a golden search on a small neighborhood of the parameter identified in the first phase. Otherwise, the algorithm announces a failure to match the camera. The searches in both phases are carried out for both the fingerprint and the image noise residual subsampled by a factor of three to speed up the search.[¶] If the golden search manages to find θ for which the PCE value exceeds a certain threshold, the algorithm computes the final detection statistic PCE (7) for the *full size* signals and the estimated θ .

The method of digests proposed here essentially follows the same procedure except it works with fingerprint digests instead of downsampled signals. We show experimentally that the same error rates (final PCE values)

[¶]The downsampling function was implemented by subselecting every third row and column while keeping the center of the image (and fingerprint) exactly in the geometrical center to avoid introducing a shift before the radial transform is applied.

can be achieved with substantially shorter digests than the resized signals. We first illustrate this in the next section to prove the concept and then subject the method to a more rigorous test in Section 4.3.

4.2. Sample experiment

The camera was the Canon SX230 HS equipped with a CMOS sensor with native resolution of 4000×3000 pixels. We estimated the sensor fingerprint $\hat{\mathbf{K}}^{(2)}$ from only ten images of a white wall illuminated by natural light. Since the camera zoom was not engaged, the focal length was 5.0 mm (the widest angle). The first three test images (shown in Figure 3) were randomly chosen among the total 275 test images from this camera and their noise residuals were assigned to $\hat{\mathbf{K}}^{(1)}$.

In Figure 2, we compare the detection statistic δ (5) evaluated for $\hat{\mathbf{K}}^{(1)}$ and $\hat{\mathbf{K}}^{(2)}$ resized by the factor of $r = 1/m$, $m = 3, 3.5, 4, 4.5, 5, 5.5, 6$ with the same statistic implemented using a digest of $\hat{\mathbf{K}}^{(2)}$ with cardinality k equal to that of the resized signals. For brevity, we do not include similar comparative tests for non-matching camera fingerprints (hypothesis H_0) as the distribution of δ is unaffected by the choice of the signal subset. The exact dimension and size in pixels after downsampling included in this experiment are shown in Table 1.

Table 1. Dimensions of signals downsampled by factor m . The native resolution of the images was 3000×4000 pixels.

m	3	3.5	4	4.5	5	5.5	6
Dimensions	1000×1334	856×1142	750×1000	666×888	600×800	544×726	498×666
No. of pixels	1,334,000	977,552	750,000	591,408	480,000	394,944	331,668

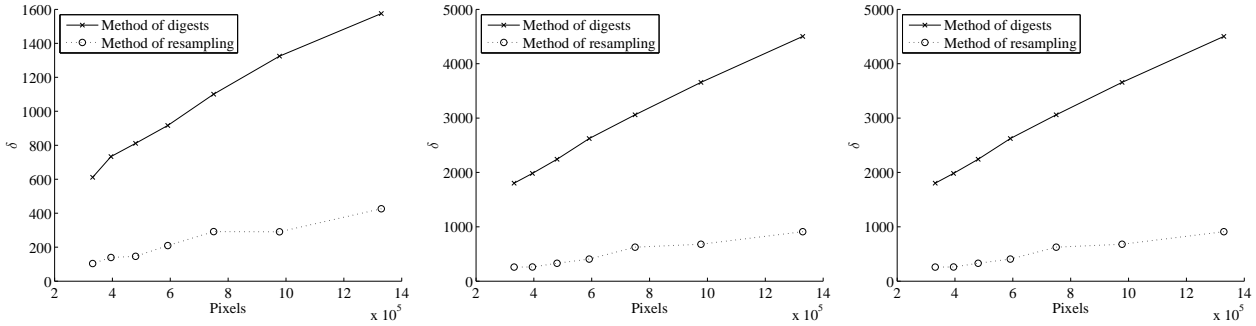


Figure 2. The gain in the PCE when using the method of digests over the prior art⁷ that used resampling by the factor of three for images shown in Figure 3.



Figure 3. Three test images taken with from Canon SX230 HS with focal lengths 11.73, 26.19, and 5.0 mm, respectively.

4.3. Experiments on three cameras

The purpose of this section is to show that the method of digests needs significantly fewer pixels to achieve the same error rate in camera identification when compared to the downsampling method (Section 4.1). We

investigate two types of digests – one computed directly from the fingerprint $\hat{\mathbf{K}}^{(2)}$ and one from the product $\mathbf{I}\hat{\mathbf{K}}^{(2)}$.

We compare the PCE (5) obtained for the downsampled signals with (7) and (8), which we evaluate for fingerprint digests of lengths from $k = 60,000$ to $300,000$ in steps of $20,000$.

4.3.1. Data acquisition

The following experiments include three compact cameras: Canon SX230 HS, Panasonic Lumix ZS7, and Nikon CoolPix S9100. All images used for the camera fingerprint estimation as well as all test images were taken at the full 3000×4000 resolution. These cameras do not apply any resizing operation during the acquisition. To prevent the digital zoom from being engaged accidentally, we turned it off. The cameras offer a large optical zoom from which we sampled the fingerprints and test images. Since the zooming proceeds in small discrete steps, the focal length can only attain certain fixed values from the available zoom range.

For each camera, we prepared a series of images in groups of ten (taking advantage of the camera’s burst mode) while adjusting the zoom before proceeding to the next group. For the Canon and Nikon cameras, we were taking snapshots of a well illuminated white wall while cloudy sky shots were used for the Panasonic camera. We used the manual mode to ensure high luminance and no saturation in these images. Each camera fingerprint was estimated from ten such images all taken at the same focal length. The fingerprints cover the entire range of focal lengths the cameras offer. The test images were mostly acquired outdoors. The scenes were often zoomed in to supply images with many possible focal length values.

Table 2. Cameras, fingerprints, and test images included in the experiments.

	Focal length (mm)	Image sensor	JPEG compression	# of fingerprints	# of test images
Canon SX230 HS	5.0 – 70.0	1/2.3" CMOS	3 bit/pixel	50	275
Panasonic ZS7	4.1 – 49.2	1/2.33" CCD	4 bit/pixel	36	183
Nikon CoolPix S9100	4.5 – 81.0	1/2.3" CMOS	2 bit/pixel	20	514

4.3.2. Experiment setup and outcome

Each tested image-fingerprint pair was chosen to have unequal focal lengths. Figure 4 depicts the distribution of these focal lengths for all 100 pairs and three cameras. The fingerprints were taken one by one ordered by the focal length, and repeated as necessary, while the test images were chosen randomly among those with focal length different from the images the fingerprint was estimated from.

The lens-distortion parameter θ was searched for in the interval $[-0.22, 0.22]$, which is sufficient to cover all possible lens distortion correction found in any of the three tested cameras. In general, the search interval may have to be adjusted for other camera models.

The size of the digest \mathcal{D} that is sufficient to meet a fixed detection threshold on PCE (a threshold between 60 to 100 typically leads to a low level of false alarm³) varies largely across the cameras, fingerprint quality, image content, and camera settings (resolution, ISO, JPEG compression, etc.). The digest-based method, however, consistently needed significantly fewer pixels during the search for the unknown parameter in all experiments. On a few examples, Figure 5 depicts how the detection statistics depend on the size of the digest. The points where the plots cross the horizontal dashed line determine the digest length at which the corresponding PCE is equal to the one obtained by downsampling by a factor of three.

The first two digest-based methods use the digest of the fingerprint $\hat{\mathbf{K}}^{(2)}$ and of the product $\mathbf{I}\hat{\mathbf{K}}^{(2)}$. The third method starts the first phase of the grid search with the digest of $\mathbf{I}\hat{\mathbf{K}}^{(2)}$ but recomputes it from $T_\theta(\mathbf{I})\hat{\mathbf{K}}^{(2)}$ before the start of the second, refining phase with the parameter θ found during the first phase. As expected, this last option delivers the highest PCE values at the price of having to perform an additional resampling step to re-estimate the digest. The difference between the last two methods is, however, rather negligible. This experiment clearly shows the benefit of computing the digest from the product $\mathbf{I}\hat{\mathbf{K}}^{(2)}$ rather than just the fingerprint.

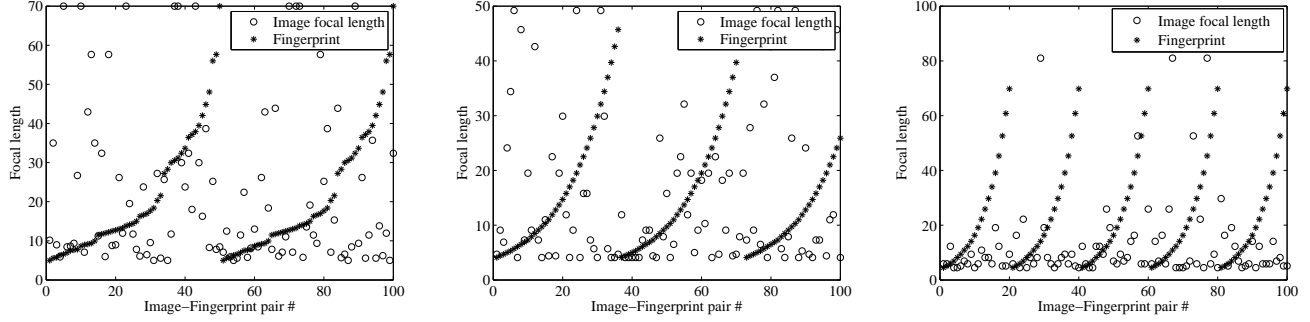


Figure 4. Focal length of all 100 image–fingerprint pairs used in our tests. From left to right: Canon, Panasonic, and Nikon.

Table 3. Median improvement factor – the ratio between the number of pixels obtained by resampling by a factor of three to the digest length that produced the same final PCE value. The success rate counts the cases when $\delta \geq 100$ for digest length $k = 300,000$.

Camera model	Digest from \mathbf{K}	Digest from \mathbf{IK}	Digest from $T(\mathbf{I})\mathbf{K}$	Success rate
Canon SX230 HS	10.08	18.96	19.28	100%
Panasonic ZS7	8.17	15.87	14.77	45%
Nikon CoolPix S9100	14.92	22.31	22.35	71%

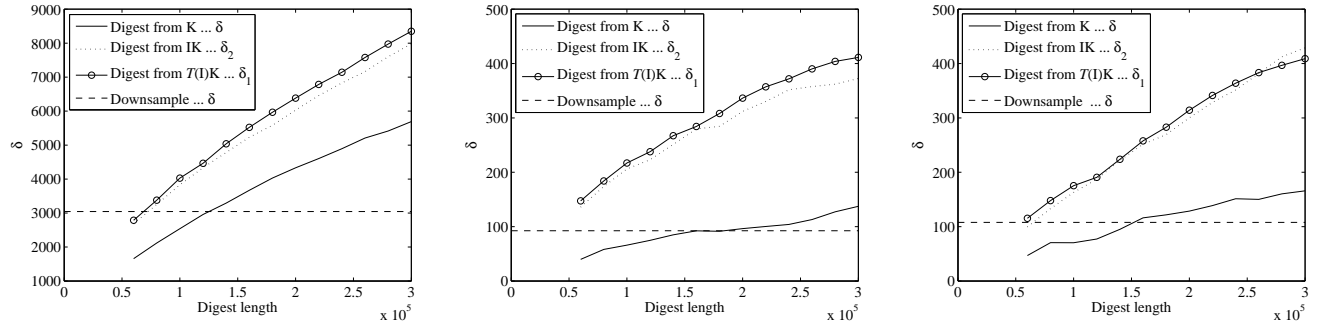


Figure 5. The test statistics (5), (7), and (8) for three versions of the method of digests and for downsampling by the factor of three (dashed line). Left to right: image–fingerprint test pairs from Canon SX230 HS, Panasonic ZS7, and Nikon CoolPix S9100.

To better portray the computational complexity savings, we next computed the ratio between the length of the signal downsampled by the factor of three and the digest length needed to match the performance (the expected error rates) of both methods. The results are shown in Table 3. Depending on the tested image–fingerprint pair, the speed-up factor was typically of the order of 10–20, which is quite significant.

The table confirms what was already observed above on a mere three image–fingerprint pairs:

1. It is clearly beneficial to compute the digest from the product $\hat{\mathbf{IK}}^{(2)}$ instead of just the fingerprint itself.
2. Resampling the intensity after the first phase improves the results only negligibly.

The low success rate for Panasonic may be caused by image stabilization, which is not purely optical for this camera.

5. CONCLUSIONS

Camera identification using sensor fingerprint is a detection problem with the generalized likelihood ratio (GLR) detector in the form of a correlator. Due to the stochastic character of the fingerprint, it is imperative that the test image under investigation be synchronized with the fingerprint or else a missed detection results. Desynchronization may occur when an unknown geometrical transformation, such as scaling, rotation, or lens-distortion correction, has been applied to either the image or the fingerprint.

One possibility to resynchronize the signals is to construct the fingerprint detector in a transform-invariant domain. Such methods, however, need to be worked out for each transform separately, may not be easily derived for general complex transforms, and may be associated with a performance loss. Alternatively, one can keep the GLR detector, whose form now includes a maximization over the parameter space. This, however, leads to a potentially expensive search. One way to speed up this search is to work with subsets of the signals, saving thus on the complexity of resampling and the test statistic evaluation itself.

In this paper, we propose to select the signal subset based on the outlier values in the sensor fingerprint – the so-called fingerprint digest. The advantage of working with a digest over a simple downsampling (or trimming the signal) is that the test statistic for the digest decreases substantially slower with decreasing signal length and thus one obtains a much better trade off between the gain in speed and the detection accuracy. Compared to the savings that can be achieved by resizing the signals to a smaller size, the method of digests leads to a significantly better performance in terms of a lower missed detection probability (larger PCE) or faster processing or both.

We prove the concept for a specific type of a geometrical distortion – the lens distortion correction nowadays commonly applied in most compact digital cameras. According to the tests executed on three cameras, the best overall method is to extract the digest from the product \mathbf{IK} of the test image intensity, \mathbf{I} , and the fingerprint, \mathbf{K} . The digest length can be conveniently chosen to achieve a desired trade-off between processing speed and performance. Depending on the fingerprint quality, image content, camera model and its settings, the method of digests achieved a 10–20 fold speed-up over reducing the signal size via downsampling (at the same level of detection accuracy).

The method of digests is applicable only to geometrical transformations that do not include spatial shift, such as radial lens-distortion correction or central cropping (e.g., digital zoom). For transformations involving a spatial shift, signal size reduction by downsampling is still a viable option.

REFERENCES

1. M. Chen, J. Fridrich, M. Goljan, and J. Lukáš. Determining image origin and integrity using sensor noise. *IEEE Transactions on Information Forensics and Security*, 3(1):70–94, March 2008.
2. J. Fridrich. Digital image forensic using sensor noise. *IEEE Signal Processing Magazine*, 26(2):26–37, 2009.
3. M. Goljan, T. Filler, and J. Fridrich. Large scale test of sensor fingerprint camera identification. In N.D. Memon, E.J. Delp, P.W. Wong, and J. Dittmann, editors, *Proc. SPIE, Electronic Imaging, Media Forensics and Security XI*, volume 7254, pages 0I 01–12, January 19–21, 2009.
4. M. Goljan and J. Fridrich. Camera identification from printed images. In E.J. Delp, P.W. Wong, J. Dittmann, and N.D. Memon, editors, *Proceedings of SPIE Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages 0I 1–12, January 28–30, 2008.
5. M. Goljan and J. Fridrich. Camera identification from scaled and cropped images. In E.J. Delp, P.W. Wong, J. Dittmann, and N.D. Memon, editors, *Proceedings of SPIE Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages 0E 1–13, January 28–30, 2008.
6. M. Goljan and J. Fridrich. Determining approximate age of digital images using sensor defects. In N. Memon, J. Dittmann, A. Alattar, and E. Delp, editors, *Proc. SPIE Electronic Imaging, Steganography, Security, and Watermarking of Multimedia Contents XIII*, volume 7541, pages 08 1–12, January 17–21, 2010.
7. M. Goljan and J. Fridrich. Sensor-fingerprint based identification of images corrected for lens distortion. In N. Memon, A. Alattar, and E. Delp III, editors, *Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics 2012*, volume 8303, pages 0H 1–13, San Francisco, CA, January 23–25, 2012.

8. M. Goljan, J. Fridrich, and M. Chen. Defending against fingerprint-copy attack in sensor-based camera identification. *IEEE Transactions on Information Security and Forensics*, 6(1):227–236, 2011.
9. M. Goljan, J. Fridrich, and T. Filler. Managing a large database of camera fingerprints. In N. Memon, J. Dittmann, and A. Alattar, editors, *Proc. SPIE Electronic Imaging, Steganography, Security, and Watermarking of Multimedia Contents XII*, volume 7541, pages 08 1–12, January 17–21, 2010.
10. C. R. Holt. Two-channel detectors for arbitrary linear channel distortion. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(3):267–273, March 1987.
11. Yongjian Hu, Binghua Yu, and Chao Jian. Source camera identification using large components of sensor pattern noise. In *Computer Science and its Applications (CSA '09). 2nd International Conference on*, pages 1–5, Jeju, South Korea, December 10–12, 2009.
12. Dai-Kyung Hyun, Seung-Jin Ryu, Min-Jeong Lee, Jun-Hee Lee, Hae-yeoun Lee, and Heung-Kyu Lee. Source camcorder identification from cropped and scaled videos. In A. Alattar, N. D. Memon, and E. J. Delp, editors, *Proc. SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV*, volume 8303, pages 0E 1–8, January 23–26, 2012.
13. J.R. Janesick. *Scientific Charge-Coupled Devices*, volume PM83. SPIE Press Monograph, 2001.
14. Bum-Soo Kim, Jae-Gark Choi, Chul-Hyun Park, Jong-Un Won, Dong-Min Kwak, Sang-Keun Oh, Chang-Rim Koh, and Kil-Houm Park. Robust digital image watermarking method against geometrical attacks. *Real-Time Imaging*, 9:139–149, 2003.
15. J. Lukáš, J. Fridrich, and M. Goljan. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security*, 1(2):205–214, June 2006.
16. H. Wolfgang. Correcting lens distortions in digital photographs. 2010. http://www.imagemagick.org/Usage/lens/correcting_lens_distortions.pdf.