

Sensor-Fingerprint Based Identification of Images Corrected for Lens Distortion

Miroslav Goljan and Jessica Fridrich

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

ABSTRACT

Computational photography is quickly making its way from research labs to the market. Recently, camera manufacturers started using in-camera lens-distortion correction of the captured image to give users more powerful range of zoom in compact and affordable cameras. Since the distortion correction (barrel/pincushion) depends on the zoom, it desynchronizes the pixel-to-pixel correspondence between images taken at two different focal lengths. This poses a serious problem for digital forensic methods that utilize the concept of sensor fingerprint (photo-response non-uniformity), such as “image ballistic” techniques that can match an image to a specific camera. Such techniques may completely fail. This paper presents an extension of sensor-based camera identification to images corrected for lens distortion. To reestablish synchronization between an image and the fingerprint, we adopt a barrel distortion model and search for its parameter to maximize the detection statistic, which is the peak to correlation energy ratio. The proposed method is tested on hundreds of images from three compact cameras to prove the viability of the approach and demonstrate its efficiency.

1. MOTIVATION

Camera identification based on sensor fingerprint is nowadays a mature area of research^{2,3,5,11} that has already found its way to courtrooms.* The method works by first estimating the sensor fingerprint from a set of images positively known to have been taken by the camera. Then, to prove 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.

The basis of the sensor fingerprint is the so-called Photo-Response Non-Uniformity (PRNU),⁹ which quantifies the fact that each pixel on the sensor consistently outputs a photon count (charge) that very slightly but consistently differs from its nominal value. Consequently, each image the sensor takes is overlaid with a unique noise-like pattern modulated by the scene light intensity. The presence of the fingerprint in an image can be established using standard deterministic-signal detectors, whose form depends on the statistical nature of the modeling noise. An especially convenient detection statistic is the Peak to Correlation Energy (PCE) ratio. A large-scale test with nearly 6900 cameras and over one million images⁵ 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, which typically requires searching over all possible parameter values of the transformation (see, e.g., the camera identification from scaled and cropped images⁶).

*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/>.

Recently, camera manufacturers started using in-camera lens-distortion correction to compensate for picture distortion as well as chromatic aberration. They were forced to cover the ever increasing demand for powerful zoom even in compact cameras where it is not feasible or economical to use a complex lens assembly with high quality rendering under a wide range of focal lengths.[†] Due to the ever-increasing power of processors incorporated in modern digital cameras, this convenient picture correction is quickly becoming quite wide-spread. Indeed, it is more economical to use a cheap lens and correct the distortion later through software than to try to manufacture a high-quality lens. The natural selection went the same way – although the human eyes contain fairly poor optics, the picture we perceive is high quality due to powerful post-processing inside the brain.

The lens distortion can be quite complicated. The most common flaw is the pincushion and barrel distortion (Figure 4). The strength of the distortion is, in general, dependent on the camera focal length. Thus, when identifying a given image using sensor fingerprint, one encounters a problem – an image can only be positively identified with a fingerprint estimated from images taken with the exact same focal length as fingerprints estimated from images taken at other focal lengths will be desynchronized.

The easiest solution that would not require any modification of the existing camera identification method³ is to read the focal length from the EXIF header of the image under investigation and estimate the fingerprint from images taken at the same focal length. However, even when the EXIF header can be trusted, the focal length may be only approximate or “quantized” and it may be rather hard to focus the camera to the exact same focal length to take images from which the fingerprint can be estimated. Moreover, this approach cannot be taken if the EXIF header is removed by post-processing or when the camera is not available to the investigator. What is needed is an automatic search for the distortion parameters to resynchronize the fingerprint with the tested image. This is the subject of this paper.

In Section 2, we introduce basic notation and briefly recapitulate the original camera identification method. The purpose of Section 3 is to demonstrate that the lens-distortion correction constitutes a problem and assess its severity. In Section 4, we describe the lens-distortion correction model and validate it by experiments in which defective pixels are used as anchors. The search algorithm for finding the parameter of the lens distortion is described in Section 5. This section also contains error analysis and a detailed pseudo-code for the final detection algorithm. Finally, the method is put to test on three compact cameras. We report the missed detection rates by testing random pairs of fingerprints and images taken from the entire range of focal lengths offered by the camera. A summary of the paper together with a list of possible improvements and extensions appear in Section 7.

2. NOTATION AND PRELIMINARIES

Everywhere in this chapter, boldface font will denote vectors (or matrices) of length specified in the text. For \mathbf{X} an $M \times N$ matrix, \mathbf{X}_{ij} denotes the i, j th element, while \mathbf{X}_i is the i th component of \mathbf{X} when converted to a vector by columns. Unless mentioned otherwise, all operations among vectors or matrices, such as product, ratio, raising to a power, etc., are *element-wise*. The Euclidean dot product of vectors is denoted as $\mathbf{X} \cdot \mathbf{Y}$ with $\|\mathbf{X}\| = \sqrt{\mathbf{X} \cdot \mathbf{X}}$ being the L_2 (Euclidean) norm of \mathbf{X} . Denoting the sample mean with a bar, the normalized correlation and the PCE are defined as

$$\rho(\mathbf{X}, \mathbf{Y}) = \frac{(\mathbf{X} - \bar{\mathbf{X}}) \cdot (\mathbf{Y} - \bar{\mathbf{Y}})}{\|\mathbf{X} - \bar{\mathbf{X}}\| \|\mathbf{Y} - \bar{\mathbf{Y}}\|}, \quad \text{PCE}(\mathbf{X}, \mathbf{Y}) = \frac{\rho^2}{\frac{1}{MN - |\mathcal{N}|} \sum_{\mathbf{s} \in \mathcal{I} - \mathcal{N}} \rho^2(\mathbf{s})}, \quad (1)$$

where $\mathcal{I} = \{0, \dots, M-1\} \times \{0, \dots, N-1\}$ and $\mathcal{N} = \{\mathbf{s} = (s_1, s_2); (s_1 \leq 5 \vee s_1 \geq M-5) \wedge (s_2 \leq 5 \vee s_2 \geq N-5)\}$ is a small neighborhood of $\mathbf{s} = (0, 0)$. For brevity, we use

$$\rho(\mathbf{s}) = \rho(\mathbf{X}_{i+s_1, j+s_2}, \mathbf{Y}_{i, j}), \quad (2)$$

with the remark that the indices wrap up cyclically whenever they get out of their original ranges (e.g., $i + s_1$ is understood as $\text{mod}(i + s_1, M)$).

[†]<http://www.zdnet.com/blog/digitalcameras/is-in-camera-lens-distortion-correction-cheating-or-not/3130>, <http://www.dpreview.com/articles/distortion/>

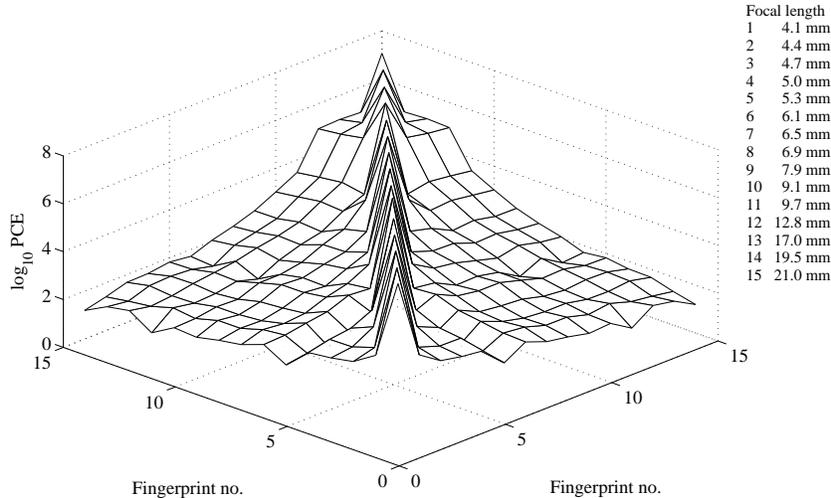


Figure 1. Logarithm of PCE($\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)}$) for all pairs of fingerprints estimated from images taken at 15 different focal lengths.

We reserve the symbols \mathbf{I} , $\mathbf{W}(\mathbf{I})$, and \mathbf{K} for a grayscale image, its noise residual, and for camera fingerprint; all three are $M \times N \triangleq n$ matrices of integers, $\mathbf{I} \in \{0, \dots, 255\}^n$, and real numbers, $\mathbf{K}, \mathbf{W} \in \mathbb{R}^n$. The noise residual is obtained using a denoising filter F , $\mathbf{W} = \mathbf{I} - F(\mathbf{I})$. In this paper, as well as in our entire previous work, we use the wavelet denoising filter¹² with $\sigma^2 = 3$. The denoising filter is applied to each color channel and then the three color components are combined using the formula for conversion from RGB to grayscale as described in Ref. 3.

The fingerprint estimation and post-processing follows the same citation. In particular, we use the maximum likelihood estimator of \mathbf{K} and apply the zero-meaning to each of the four sub-matrices of $\hat{\mathbf{K}}$ that correspond to red, blue, green, and second green pixels in the layout of the Bayer filter array. We then extract the noise from the magnitude of $\hat{\mathbf{K}}$ in the Fourier domain using the Wiener filter, $W(\cdot, \sigma^2)$, with noise variance $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{K}}_i^2$ (all operations elementwise):

$$\mathbf{F} = \mathcal{F}(\hat{\mathbf{K}}), \quad \hat{\mathbf{K}} \leftarrow \text{Real} \left[\mathcal{F}^{-1} \left(\mathbf{F} \cdot \frac{|\mathbf{F}| - W(|\mathbf{F}|, \sigma^2)}{|\mathbf{F}|} \right) \right], \quad (3)$$

where \mathcal{F} is the orthonormal Fourier transform. Both the zero-meaning and the filtering are necessary to remove Non-Unique Artifacts (NUAs) of in-camera signal and image processing that may otherwise increase the false alarm probability. In the presence of lens-distortion correction, however, the correct removal of NUAs is quite challenging. More on this topic appears in Section 6.2. The presence of fingerprint \mathbf{K} in noise residual \mathbf{W} is established using the test statistic in the form of $\text{PCE}(\mathbf{W}, \hat{\mathbf{K}})$.

3. WHEN THE ORIGINAL APPROACH FAILS

The purpose of this section is to assess the severity of the problem and quantify how much of the distinguishing power is lost due to desynchronization introduced by lens distortion correction.

For our initial experiments, we chose a Panasonic Lumix DMC-ZS7 camera equipped with a 14.5 Mpixel sensor (3000×4000 pixels effective output) and $12\times$ optical zoom. We estimated 15 fingerprints for this camera taken at 15 different focal lengths by taking 10 snapshots of cloudy sky[‡] with no zoom, zooming a little, taking

[‡]A gray scene with an approximately uniform content, such as cloudy sky, has all three color components in balance and allows for easy separation of noise from content.

another 10 snapshots, until we collected 15×10 images.[§] In total, we covered the focal lengths in the range of 4.1 – 49.2 mm as determined from the EXIF headers (also see Figure 1). We also turned off the digital zoom.

Figure 1 shows $\log_{10} \text{PCE}(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)})$ for every pair of fingerprints estimated above. Note that the PCE drops rather sharply once the focal lengths between fingerprints do not match. This sensitivity is the largest for short focal lengths, where the PCE drops below 100, while in the telephoto range the effect of desynchronization disappears. It is shown in the appendix that the PCE between high-quality fingerprints (the values on the diagonal) is approximately equal to $n\rho^2(\mathbf{K}^{(1)}, \mathbf{K}^{(2)})$. Realize that when testing a single image w.r.t. a fingerprint rather than two high quality fingerprint estimates, the drop of PCE is even larger depending on the SNR between the fingerprint and the noise residual (23) and often prevents successful detection of the original camera identification method as demonstrated in Section 6.

4. LENS DISTORTION MODEL

In this paper, we adopt a simple model for radially symmetric barrel/pincushion distortion.^{8,10,13} Denoting the original pixel coordinates before and after distortion as (x, y) and (x', y') , the geometrical transformation is a one-to-one mapping $T_a : (x, y) \rightarrow (x', y')$

$$x' = x_p + (x - x_p)(1 + ar^2) \quad (4)$$

$$y' = y_p + (y - y_p)(1 + ar^2) \quad (5)$$

where $r^2 = (x - x_p)^2 + (y - y_p)^2$ is the squared radial distance to the optical center of the image at (x_p, y_p) and a is a parameter. The distance is scaled so that $r = 1$ corresponds to one half of the image diagonal. When $a < 0$, we speak of barrel distortion, while $a > 0$ corresponds to the pincushion distortion.

At this point, it is important to realize that our task is not really to model the true lens distortion but to invert the transformation performed inside the camera. The shift between the center of the image and the optical center is mostly caused by a mechanical misalignment between the lens and the sensor. As such discrepancy will almost certainly be different from camera to camera, it is unlikely that manufacturers correct for it as that would require an algorithm individualized for each camera made. Thus, we will assume $(x_p, y_p) = (0, 0)$ and use a simplified model, which now can be written only using the radial distance to the image center:

$$r' = r(1 + ar^2). \quad (6)$$

Using Taylor expansion, it is routine to show that the mapping between two images, one distorted with parameter a and the other with b , $r'' = r(1 + br^2)$, requires additional higher-order terms of even power:

$$r'' = r' (1 + (a - b)r'^2 + 3a(a - b)r'^4 + \dots). \quad (7)$$

With each additional parameter, however, the complexity of the parameter search significantly increases. Fortunately, the coefficient at the quartic term is of higher order when compared to the quadratic one. Thus, we adopted (6) as a single-parameter model between distorted images (fingerprints) and verified its accuracy using defective pixels as anchor points (since for most consumer-end cameras the raw sensor output before any processing is applied is not available). We performed this model validation for the Panasonic Lumix camera.

First, we identified all dead and hot pixels whose distance to the image center was at least 1/4 of the diagonal as defects closer to the center move only by a small amount and taint the estimation of the transform. The defects were identified in fingerprints from the lower half of zoom range covered by 15 fingerprints (see the previous section for details). First, we confirmed that the optical center is in the geometrical center of the image by fitting straight lines to the defect positions as they change with zoom (see Figure 2). The optical center appears to coincide with the image center within about ten pixels (see the center closeup on the right).

[§]To make sure that 10 images are sufficient for a reasonably good estimate of the fingerprint, we set the exposure manually in an attempt to maintain the pixel intensities high but not saturated.

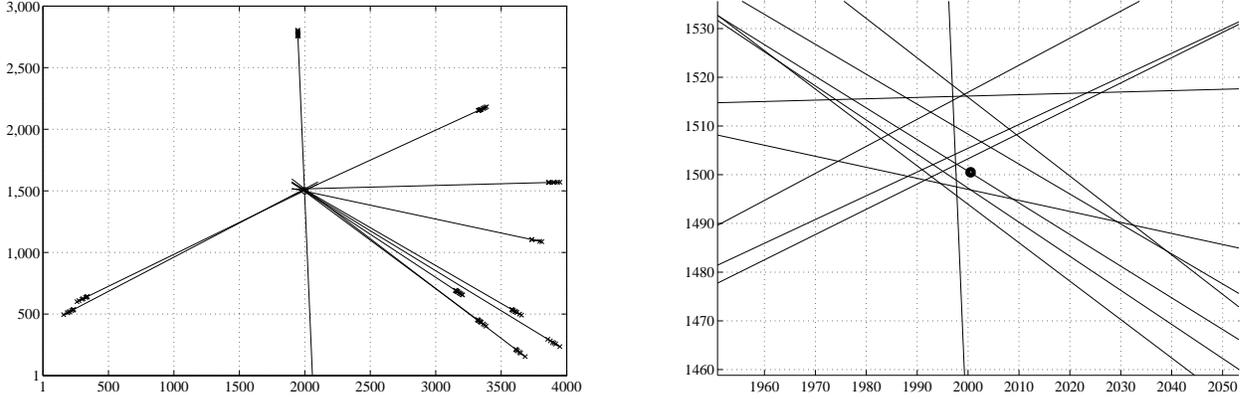


Figure 2. Left: optical center assessment by fitting straight lines through defects (marked by crosses). Right: central detail with image center marked by a black circle at coordinates (1500.5, 2000.5).

For each focal length, we fit the quadratic model 6 to the pixels' radial distance using least squares. The results are shown in Figure 3. The deviation of the defect position from the model fit was always within one pixel, which confirms the validity of the quadratic model. Notice that even for two relatively close focal length values of $f = 7.9$ and $f = 4.1$ the pixels located far from the center are displaced by hundreds of pixels. One defect positioned at $r = 2250$ was moved by 250 pixels (or by 11%).

5. METHODOLOGY

In this section, we describe a method for detecting whether noisy estimates of two fingerprints, $\hat{\mathbf{K}}^{(1)}$, $\hat{\mathbf{K}}^{(2)}$, contain a common signal – the sensor fingerprint \mathbf{K} . It is assumed that $\hat{\mathbf{K}}^{(i)}$ were estimated from $N_i \geq 1$ images taken at focal length (zoom setting) f_i and that, in general, $f_1 \neq f_2$. When $N_1 = N_2 = 1$, we speak of image pairing,⁴ which corresponds to the most challenging problem when only two images are available and our task is to determine whether or not they were taken using the same device. The most common application, and one we focus on in this paper, is when one observable is a noise residual and the other a fingerprint estimate.

Formally, our task is a two-channel problem:

$$H_0 : \hat{\mathbf{K}}^{(1)} = \mathbf{K} + \xi^{(1)}, \hat{\mathbf{K}}^{(2)} = T_a(\mathbf{K}') + \xi^{(2)} \quad (8)$$

$$H_1 : \hat{\mathbf{K}}^{(1)} = \mathbf{K} + \xi^{(1)}, \hat{\mathbf{K}}^{(2)} = T_a(\mathbf{K}) + \xi^{(2)}, \quad (9)$$

where \mathbf{K} and \mathbf{K}' are two sensor fingerprints, $\xi^{(i)}$, $i = 1, 2$, are i.i.d. Gaussian modeling noise terms with unknown variances, and T_a is the lens distortion mapping (6) parametrized by a scalar parameter a . 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 generalized likelihood ratio detector for this problem, which appears in Ref. 7, is standardly applied to camera identification from fingerprints.^{3–6} The detection statistic is the PCE maximized over all possible values of a , $\max_a \text{PCE}(\hat{\mathbf{K}}^{(1)}, T_a^{-1}(\hat{\mathbf{K}}^{(2)}))$. The detection thus involves numerical maximization of the PCE as a function of a . As explained in Section 5.1 below, it is easier to implement the inverse mapping, T_a^{-1} , between discrete lattices than the forward transform T_a .

First, a range for the parameter a is adopted, $a \in [-A, A]$. The interval is symmetrical to cover both pincushion and barrel distortions. The proposed method consists of two parts: a grid search on a progressively refined grid and then a golden section search is run on a neighborhood of the grid point with the maximal PCE. The first hierarchical grid search starts with the initial grid $\mathcal{A}_1 = \{-A, 0, A\}$. Given a union of grids $\cup_{m=1}^k \mathcal{A}_m = \{a_1^{(k)}, \dots, a_{n_k}^{(k)}\}$, the next grid is $\mathcal{A}_{k+1} = \{\frac{1}{2}(a_1^{(k)} + a_2^{(k)}), \frac{1}{2}(a_2^{(k)} + a_3^{(k)}), \dots, \frac{1}{2}(a_{n_k-1}^{(k)} + a_{n_k}^{(k)})\}$. The

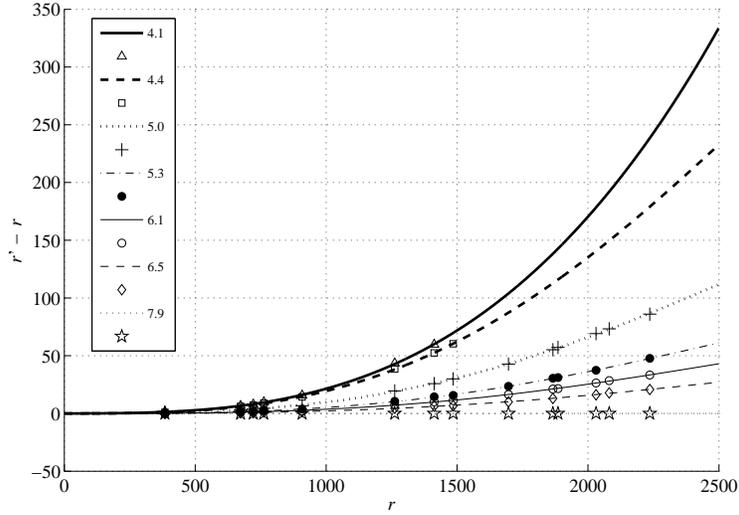


Figure 3. Model (6) fit to defective pixels' shifts.

algorithm never checks further than k_{\max} , which corresponds to the total number of grid points, $|\mathcal{A}| = 2^{k_{\max}} + 1$, where $\mathcal{A} = \cup_{m=1}^{k_{\max}} \mathcal{A}_m$. If the grid search goes through all $2^{k_{\max}} + 1$ points while the maximal PCE* (found at grid point $a_{i^*}^{(k^*)} \in \mathcal{A}_{k^*}$) is below some threshold τ_1 , the entire algorithm stops and mismatch is proclaimed (hypothesis H_0). If $\text{PCE}^* > \tau_1$, the algorithm continues with the golden search refinement. To speed up the grid search, we stop it and proceed to the golden search once the PCE is sufficiently large ($\text{PCE} > \tau_2 \wedge k > 4$) instead of going through all points in \mathcal{A} as in this case it would be too slow and unnecessary to try all grid points.

Having found $a_{i^*}^{(k^*)}$, the second, refining, step is a golden section search on the interval $[a_{i^*}^{(k^*)-1}, a_{i^*}^{(k^*)+1}]$ of width $A/2^{k^*-1}$ defined by two grid points in \mathcal{A}_{k^*} closest to $a_{i^*}^{(k^*)}$. The golden section search stops when the width of the search interval reaches

$$A/2^{k^*-1} \approx \frac{1}{4d}, \quad (10)$$

where d is the image diagonal in pixels.[¶] The parameter value outputted by the golden section search is denoted \hat{a} .

Since we do not know which of the two observables was corrected more strongly, it is not clear whether in the objective function we should be applying T_a^{-1} to $\hat{\mathbf{K}}^{(2)}$ or T_a to $\hat{\mathbf{K}}^{(1)}$. It is intuitively clear that one option could give significantly higher PCE than the other due to cumulative artifacts of repeated resampling. We take care of this dilemma by taking as the final value of the PCE the maximum:

$$\text{PCE}_{\text{final}} = \max \left\{ \text{PCE} \left(T_{\hat{a}}(\hat{\mathbf{K}}^{(1)}), \hat{\mathbf{K}}^{(2)} \right), \text{PCE} \left(\hat{\mathbf{K}}^{(1)}, T_{\hat{a}}^{-1}(\hat{\mathbf{K}}^{(2)}) \right) \right\}. \quad (11)$$

For practical implementation, it is useful to know that the inverse to the model (6) is

$$r = r'(1 - ar'^2 + 3a^2r'^4 + O(r'^6)). \quad (12)$$

The entire algorithm proclaims a match (hypothesis H_1) if and only if $\text{PCE}_{\text{final}} > \tau_3$. In all other cases, including the one when the golden search is entirely skipped because PCE at $a_{i^*}^{(k^*)}$ is less than τ_1 , we decide H_0 .

[¶]In other words, the search stops when the changes to a are so small that T_a^{-1} at the boundary differs by at most a quarter of a pixel.

Algorithm 1 Detection algorithm for image identification.

```
1: Set  $A = 0.22, \tau_1 = 15, \tau_2 = 75, \tau_3 = 75$ 
2: for  $k = 1$  to  $k_{\max}$  do
3:    $\mathcal{A}_k = \{a_1^{(k)}, \dots, a_{n_k}^{(k)}\}$ 
4:   for  $i = 1$  to  $n_k$  do
5:     if  $\text{PCE}(\hat{\mathbf{K}}^{(1)}, T_{a_i}^{-1}(\hat{\mathbf{K}}^{(2)})) > \text{PCE}^*$  then
6:        $\text{PCE}^* = \text{PCE}(\hat{\mathbf{K}}^{(1)}, T_{a_i}^{-1}(\hat{\mathbf{K}}^{(2)})), i^* = i, k^* = k, a^* = a_i^{(k)}$ 
7:       if  $\text{PCE}^* > \tau_2$  then
8:         go to 16
9:       end if
10:    end if
11:  end for $\{i\}$ 
12: end for $\{k\}$ 
13: if  $\text{PCE}^* < \tau_1$  then
14:   Decide  $H_0$ 
15: end if
16:  $\hat{a} = \text{GOLDEN\_SEARCH}([a_{i^*-1}^{(k^*)}, a_{i^*+1}^{(k^*)}], N_{\text{gold}})$ 
17: if  $\text{PCE}_{\text{final}} \triangleq \max \left\{ \text{PCE}(T_{\hat{a}}(\hat{\mathbf{K}}^{(1)}), \hat{\mathbf{K}}^{(2)}), \text{PCE}(\hat{\mathbf{K}}^{(1)}, T_{\hat{a}}^{-1}(\hat{\mathbf{K}}^{(2)})) \right\} > \tau_3$  then
18:   Decide  $H_1$ , output  $\text{PCE}_{\text{final}}$ 
19: else
20:   Decide  $H_0$ 
21: end if
22: FUNCTION  $\hat{a} = \text{GOLDEN\_SEARCH}([r, s], N_{\text{gold}})$ 
23: Run  $N_{\text{gold}}$  iterations of golden section search on  $[r, s]$  with PCE as the objective function, return  $\hat{a}$ 
24: end
```

5.1 Implementation issues

To lower the complexity of the grid search as well as the golden section search, we carry out these searches with both the image and the fingerprint downsampled by a factor of three. This works better than center-cropping the signals since the distortion is the least pronounced in the center. We used simple subsampling of rows and columns starting at the image center to avoid creating an artificial shift by subsampling. Specifically, given an image/fingerprint with $2N_r \times 2N_c$ pixels, the center of the image is at $(N_r + 1/2, N_c + 1/2)$, and we downsample by skipping rows no. $N_r + 3k$ and $N_r + 1 + 3k$ and columns no. $N_c + 3k$ and $N_c + 1 + 3k$ for $k = \dots, -2, -1, 0, 1, 2, \dots$

In the computation of $\text{PCE}_{\text{final}}$ on line 17 of Algorithm 1, however, we compute both PCEs at the full resolution to avoid loss of PCE due to the downsampling.

The radial mapping T_a (6) needs to be realized to work between two square lattices of pixels, \mathcal{L}' and \mathcal{L} . It turns out that it is easier to work with the inverse T_a^{-1} . Let us assume that we have a signal sampled on the square lattice \mathcal{L}' and we wish to determine the grayscale intensity after applying the inverse transformation to \mathcal{L}' . Since we need to map to another square lattice, \mathcal{L} , let $(x, y) \in \mathcal{L}$. We transform it using (6), obtaining thus real-valued coordinates (x', y') . The intensity value at (x, y) is determined as a weighted average of intensities of the following four lattice points from \mathcal{L}' : $(\lfloor x' \rfloor, \lfloor y' \rfloor)$, $(\lfloor x' \rfloor, \lceil y' \rceil)$, $(\lceil x' \rceil, \lfloor y' \rfloor)$, and $(\lceil x' \rceil, \lceil y' \rceil)$, where the two types of brackets denote the operations of rounding down and up. The weights are determined by square-distances between the lattice points and (x', y') . We use the approximation (12) when implementing the forward transform.

5.2 False alarms and threshold setting

The large scale test⁵ provides useful practical data to estimate the probability of obtaining a false match in the proposed algorithm. Let $P(\tau_3) = \text{Prob}(\text{PCE}_{\text{final}} > \tau_3 | H_0)$. Since under H_0 the PCE values obtained in the search

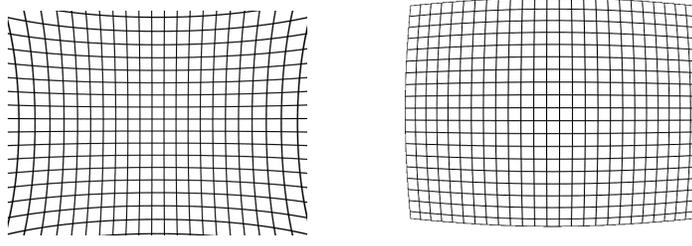


Figure 4. Example of the most severe pincushion (left) and barrel (right) distortion correction encountered in our tests with Canon Powershot SX230 HS.

are independent realizations, we obtain the following bound on the probability of false alarm

$$P_{\text{FA}}(\tau_3) = 1 - (1 - P(\tau_3))^{N_{\text{eval}}} \approx N_{\text{eval}}P(\tau_3), \quad (13)$$

where $N_{\text{eval}} = N_{\text{grid}} + N_{\text{gold}}$ is the maximal total number of PCE evaluations. Since the interval width in the golden section search decreases by the factor of $\varphi = \frac{\sqrt{5}+1}{2}$ per evaluation,

$$\frac{A/2^{k^*-1}}{\varphi^{N_{\text{gold}}}} = \frac{1}{4d}. \quad (14)$$

For example, for an image with $d = 2500$ and $A = 0.3$, $N_{\text{gold}} \approx 12 \dots 8$ for $k^* = 5 \dots 7$.

The threshold τ_3 thus controls the false alarm probability. For $\tau_3 = 60$, it was experimentally determined using a large scale test that $P(\tau_3) \approx 10^{-5}$. The experiment could not be used to estimate $P(\tau_3)$ for larger values of τ_3 due to lack of points in the right tail of the distribution of PCE under H_0 . Since for $k_{\text{max}} = 7$, $N_{\text{eval}} \leq 138$, to keep P_{FA} below 10^{-6} , the threshold needs to be increased. In this paper, we used $\tau_3 = 75$ to compensate for the increase in P_{FA} due to the search. Unfortunately, since no experimental data is available to estimate $P(\tau_3)$ for such a high value of the threshold, we hypothesize that the false alarm probability is sufficiently low given the character of the right tail in the distribution of H_0 (see Ref. 5).

Finally, in our practical implementation, we set $\tau_1 = 15$ and $\tau_2 = \tau_3$. The thresholds τ_1 and τ_2 affect the missed detection probability, which is why we set τ_1 to a rather low and τ_2 to a high value.

6. EXPERIMENTS

In order to validate the methodology from Section 5 we added two Canon SX230 HS cameras with 14×optical zoom to our experimental setup. In the camera menu, we selected the image resolution to match the Panasonic’s 3000×4000. We intentionally chose two cameras of the same model to investigate the effect of leftover non-unique artifacts on the probability of false alarm. Below, we summarize the image sets taken with each camera.

- Panasonic Lumix DMC-ZS7 (further denoted as Panasonic), focal length range 4.1 – 49.2 mm, 225 test images, 34 fingerprints each estimated from 10 images of overcast sky with focal lengths $f = 4.1, 4.4, 4.7, 5, 5.3, 5.7, 6.1, 6.5, 6.9, 7.3, 7.9, 8.5, 9.1, 10.3, 11, 12.8, 13.7, 14.7, 15.8, 17, 18.2, 19.5, 21, 22.5, 24.1, 25.9, 27.8, 29.9, 32.1, 37, 39.7, 42.6, 45.7, 49.2$. With increased zoom, the sky shots became increasingly dark, which lowered the quality of the estimated fingerprints.
- Canon SX230 HS (Canon-I), focal length range 5 – 70 mm, 189 test images, 41 fingerprints, from 10 images (taken in the burst mode) of overcast sky with focal lengths $f = 5, 5.58, 5.68, 6, 6.7, 7.09, 7.52, 7.98, 8.13, 9.75, 10.8, 11.73, 12.22, 12.48, 14.12, 15.01, 16.29, 17.66, 18.02, 20.72, 21.14, 22.42, 22.86, 23.31, 24.24, 25.2, 25.69, 26.7, 28.29, 29.97, 31.15, 36.42, 37.91, 39.49, 42.04, 42.94, 44.86, 53.03, 54.47, 57.62, 70$.

- Canon SX230 HS (Canon-II), focal length range 5–70 mm, 50 fingerprints, 275 test images, 41 fingerprints of cloudy sky with focal lengths $f = 5, 5.58, 5.78, 6.33, 6.58, 6.96, 7.52, 7.67, 7.82, 8.64, 8.99, 9.18, 9.36, 9.95, 11.49, 11.73, 11.97, 12.22, 12.48, 12.74, 13, 13.27, 13.83, 14.12, 14.41, 15.01, 16.29, 16.62, 16.96, 17.66, 18.39, 20.32, 21.56, 27.22, 28.29, 29.97, 30.56, 31.15, 32.38, 33.66, 36.42, 37.16, 37.91, 39.49, 42.04, 44.86, 48.04, 55.99, 57.62, 70$.

Both camera models offer a large zoom range sampled in discrete steps. The zoom in the Canons is hard to control if one needs to adjust it by the smallest step. This explains why we did not prepare fingerprints at all possible (more than 50) zoom settings.

6.1 Open set test with three cameras

For each fingerprint, we randomly chose 20 test images from each of the three cameras and ran the detection algorithm 1. The total of $3 \times (34 + 41 + 50) \times 20 = 7500$ image/fingerprint pairs entered the tests. We ran the entire experiment as an “open set problem” in the sense that we do not a priori assume that one of the three cameras took a given test image and thus may decide in some cases that the image does not match any of the three cameras. One technical note: based on preliminary tests, we set $A = 0.22$ to cover the parameter range for all the three cameras.

The main diagonals in Table 1 are the true detection rates (in percents) observed in the experiment, off diagonal values represent the false alarm rates. The effect of the resynchronization is apparent by comparing the left and right tables. While the detection rates for the original camera ID method without resynchronization³ fell below 50% (left), when supplied with the proposed resynchronization (right), all tested images from both Canons and about 92% of images from Panasonic were correctly identified.

The lower detection rate (91.62%) for Panasonic can be partly explained by the lower quality of fingerprints due to dark sky images taken at long focal lengths. Repeating this experiment with fingerprints estimated from 20 images gave us a slightly higher detection rate of 93.74%. Five missed detections occurred for images taken with the maximum zoom at $f = 49.2$; four of them were indoor snapshots. Even though taking pictures with a maximum zoom indoors is not common, as part of our future research we plan to investigate whether the camera image stabilization or other processing associated with extreme zoom may have affected the pixel synchronization between the sensor and the camera image output.

The detection rates obviously depend on the prior distribution of focal lengths for images and fingerprints. Most missed detections occurred when either the fingerprint or the image were taken at focal length $f < 10$ as this is the range where the distortion is the most severe and also the range where most photographs are taken under normal circumstances. In the bottom portion of Table 1, we report the detection rate only for images and fingerprints taken at $f < 10$. In this case, the original identification method without resynchronization performs even worse, while identification with the resynchronization is almost unaffected.

The most severe cases of distortion that we had to correct for in our test both occurred with the Canon Powershot and are shown in Figure 4. The largest barrel distortion had T_a^{-1} with $a = 0.19765$ (occured for a fingerprint/image pair at focal length 5 mm and 12.222 mm) and the largest pincushion had T_a^{-1} with $a = -0.17413$ (fingerprint taken at focal length 33.663 mm and image at 5 mm).

The resampling needed to resynchronize the test image and the fingerprint weakens the fingerprint, which translates to a lower value of PCE and thus an increased missed detection rate. This loss can be quite substantial when both the image and the fingerprint were taken at two very different focal lengths. This is apparent from Figure 5 showing the PCE between three selected test images and camera fingerprints estimated at different zooms.

6.2 Analysis of false alarms

In our open test above, we observed two false alarms (FA rate of 0.002) when testing images from Canon-I, both at $f = 5$, against a Canon-II fingerprint estimated also at $f = 5$ (PCE values were 207 and 83). We also saw two false alarms (FA rate 0.0024) when testing images from Canon-II at $f = 7.667$ and $f = 8.298$ against a Canon-I fingerprint estimated at $f = 5.583$ and $f = 8.134$ (PCEs 88 and 76). A false-alarm rate this high contradicts the

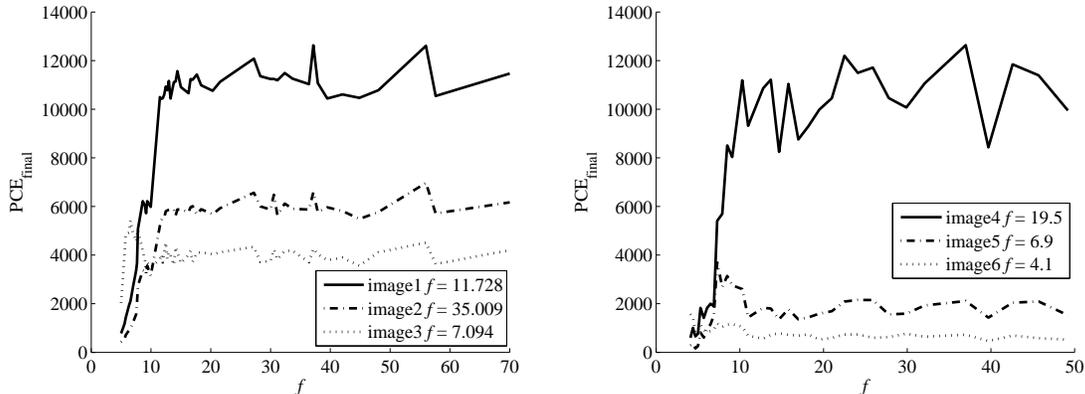


Figure 5. PCE after resynchronization as a function of fingerprint focal length f for three images from Canon-II (left) and Panasonic (right). The varying levels of the PCE are also influenced by the image content.

error analysis of Section 5.2. It is caused by unremoved non-zero means of rows and columns common to both Canons. (The row and column means form the so-called “linear pattern”.¹) Even though we suppress NUAs when estimating the fingerprints by zero-meaning the rows and columns as described in Section 2, when the images from which we estimate the fingerprints were corrected for lens distortion, the sensor rows and columns become “bent” in the corrected image, which prevented us from removing their means. More accurately, we remove the means only in the central area of the fingerprint. The size of this area decreases with increased lens distortion and the remnants of the linear pattern cause unwanted correlation that can become large enough to cross the detection threshold when both the image and the fingerprint were estimated at short focal lengths where the distortion is the largest.

| $P_D(\%)$ | Panasonic | Canon-I | Canon-II |
|-----------|-----------|---------|----------|
| Panasonic | 36.03 | 0 | 0 |
| Canon-I | 0 | 43.9 | 0.40 |
| Canon-II | 0 | 0.37 | 46.5 |

| $P_D(\%)$ | Panasonic | Canon-I | Canon-II |
|-----------|-----------|---------|----------|
| Panasonic | 91.62 | 0 | 0 |
| Canon-I | 0 | 100 | 0.20 |
| Canon-II | 0 | 0.24 | 100 |

| $P_D(\%)$ | Panasonic | Canon-I | Canon-II |
|-----------|-----------|---------|----------|
| Panasonic | 12.42 | 0 | 0 |
| Canon-I | 0 | 13.74 | 0.62 |
| Canon-II | 0 | 0.62 | 14.99 |

Table 1. Experimental detection rates P_D . Left: without resynchronization, Right: with resynchronization, Top: all image/fingerprints, Bottom: restricted images/fingerprints with $f < 10$.

The removal of the linear pattern from distorted fingerprints is a challenging task but one that must be tackled to avoid an increased false alarm rate between cameras of the same model. The cleanest solution, that would not require modification of Algorithm 1, is to remove the linear pattern during the fingerprint estimation. One possibility that appears rather promising is to subject the estimated fingerprint $\hat{\mathbf{K}}$ to distortion T_a^{-1} while covering a range of a and measuring the energy of the linear pattern

$$E(a) = \sum_{i=1}^M \mu_i^2(a) + \sum_{j=1}^N \gamma_j^2(a), \tag{15}$$

where $\mu_i(a)$ and $\gamma_j(a)$ denote the sample mean of the i th row and j th column of $T_a^{-1}(\hat{\mathbf{K}})$. When T_a^{-1} becomes near exact inverse of the distortion correction applied in the camera, we observe a peak in $E(a)$ (see Figure 6) as in this case we reestablish pixel-to-pixel correspondence with the sensor. Having determined a , we can now remove the linear pattern by zero meaning the rows and columns of $T_a^{-1}(\hat{\mathbf{K}})$ and transform the fingerprint back using T_a , at which point we could apply Algorithm 1 as before. We refrain from further analysis of this issue in

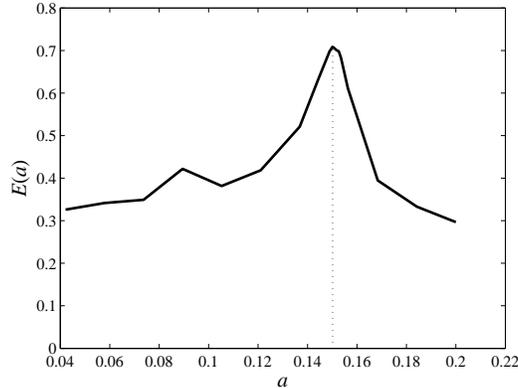


Figure 6. Linear pattern energy $E(a)$ exhibits a peak when T_a^{-1} is the inverse of the distortion correction applied in the camera. In this case, T_a is the barrel distortion.

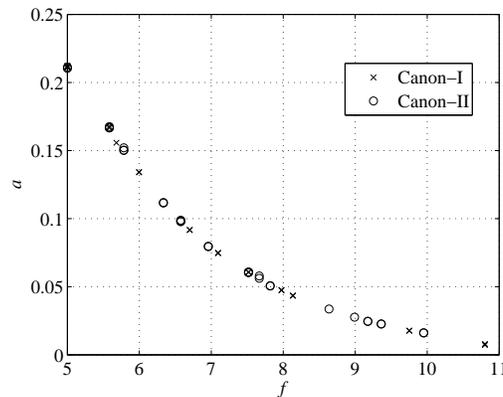


Figure 7. Parameter a for reversing the radial distortion correction in two Canon cameras.

this paper due to lack of space and postpone our work to a journal version of this paper that is currently under preparation.

6.3 Blind focal length estimation

The energy of the linear pattern (15) could potentially be used to estimate the focal length at which an image (fingerprint) was taken. This is because one can experimentally determine the relationship between the distortion parameter a and focal length f for a given camera model and a can be estimated. Figure 7 shows this relationship for both Cannon cameras. We note that $a(f) \approx 0$ for $f \gtrsim 11.5$. This also implies that if the image and the fingerprint are both taken in this focal range, no resynchronization is necessary for successful camera identification. The figure also confirms that the strongest distortion occurs for the shortest focal lengths.

7. CONCLUSIONS

A large number of modern consumer-end digital cameras correct their images on the fly for optical distortion to give the customers a wide range of optical zoom in a compact and low-cost package. Since this geometrical correction depends on the zoom, it causes a desynchronization between sensor fingerprint and an image noise residual in camera identification methods that utilize sensor fingerprint, effectively making these methods fail. The main contribution of this work is a method for estimating the distortion and reestablishing synchronization of both signals, making thus the camera ID technology work again. The algorithm can be applied to the case when an analyst has an estimate of the sensor fingerprint and one image under investigation. The task is to determine whether the image contains the estimated (and distorted) fingerprint. Our approach makes no use

of the EXIF header data as this information may be unavailable or corrupted. Neither do we assume that the camera that produced the images is available.

The identification algorithm consists of a grid search followed by a refining golden section search for the lens distortion parameter on downsampled signals and then computing the PCE detection statistics after one of the signals is geometrically transformed to synchronize with the other signal. Once resynchronized, a positive match can be obtained. We demonstrate the reliability of this approach with three cameras, two of which are of the exact same model to investigate the effects of possible non-unique artifacts caused by camera signal read-out and processing. By testing hundreds of fingerprint-image pairs with a wide range of focal lengths, we were able to achieve a detection reliability of 91% for a Panasonic camera and 99.8% for Canon, with camera fingerprints estimated from only 10 images.

We observed a higher false alarm rate between both Canon cameras than what we designed the algorithm for. Analysis showed that this was caused by remnants of a linear pattern (non-zero means of rows and columns), which is not unique to the camera and thus undesirably increases the false alarm rate. We proposed a method for removing this pattern from images corrected for lens distortion, which we plan to test in our future work.

We view this work as an initial proof of concept that is also supposed to raise the awareness of the impact of computational photography techniques on digital forensics. A lot remains to be done. Among the future topics, we plan to carry out a large scale test to better assess the extendability of the proposed measure. Together with lens distortion, camera manufacturers are likely correcting for lateral chromatic aberration as well. In this case, it will be important to resynchronize each color channel separately instead of working with grayscale versions of fingerprints and noise residuals. It is also quite intriguing to use the sensor fingerprint as a template to reverse-engineer in-camera processing and estimate the severity of the distortion in the uncorrected, raw image, and through this non-intrusively evaluate the rendering quality of optics.

8. ACKNOWLEDGEMENTS

The authors would like to thank Mo Chen for useful discussions.

Appendix

Let us assume that we have two estimates of fingerprints at two different focal lengths:

$$\hat{\mathbf{K}}^{(1)} = \mathbf{K}^{(1)} + \xi^{(1)} \quad (16)$$

$$\hat{\mathbf{K}}^{(2)} = \mathbf{K}^{(2)} + \xi^{(2)} \quad (17)$$

with $\mathbf{K}^{(l)}$, $l = 1, 2$, being zero-mean deterministic signals and the noise terms $\xi^{(l)} \sim N(0, \sigma_l^2)$, $\sigma_l^2 \approx \frac{1}{n} \|\xi^{(l)}\|^2$, both mutually independent, all signals of length n . Due to the lens-distortion correction, $\mathbf{K}^{(1)}$ and $\mathbf{K}^{(2)}$ will be approximately synchronized only around their center and thus $\rho(\mathbf{K}^{(1)}, \mathbf{K}^{(2)}) < 1$. The PCE can be written as

$$\text{PCE}(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)}) = \frac{\rho^2(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)})}{\frac{1}{n-|\mathcal{N}|} \sum_{\mathbf{s} \in \mathcal{I}-\mathcal{N}} \rho^2(\hat{\mathbf{K}}_{i+s_1, j+s_2}^{(1)}, \hat{\mathbf{K}}_{i, j}^{(2)})} = \frac{(n-|\mathcal{N}|) (\hat{\mathbf{K}}^{(1)} \cdot \hat{\mathbf{K}}^{(2)})^2}{\sum_{\mathbf{s} \in \mathcal{I}-\mathcal{N}} (\hat{\mathbf{K}}_{i+s_1, j+s_2}^{(1)} \cdot \hat{\mathbf{K}}_{i, j}^{(2)})^2} \quad (18)$$

because the norms in the normalized correlations in the numerator and the denominator cancel.

We will assume that both fingerprints $\mathbf{K}^{(l)}$ are fixed i.i.d. realizations of a Gaussian random variable with zero mean and variance $\nu_l^2 = \frac{1}{n} \|\mathbf{K}^{(l)}\|^2$ and independent of the noise terms. Thus, $\mathbf{K}^{(l)} + \xi^{(l)} \sim N(0, \nu_l^2 + \sigma_l^2)$ and by the Central Limit Theorem (CLT) for each $\mathbf{s} \neq 0$ $\frac{1}{\sqrt{n}} \hat{\mathbf{K}}_{i+s_1, j+s_2}^{(1)} \cdot \hat{\mathbf{K}}_{i, j}^{(2)} \sim N(0, (\sigma_1^2 + \nu_1^2)(\sigma_2^2 + \nu_2^2))$ because the variance of the product of two independent Gaussian random variables is the product of their variances. Thus, we have for the denominator:

$$E \left[\sum_{\mathbf{s} \in \mathcal{I}-\mathcal{N}} (\hat{\mathbf{K}}_{i+s_1, j+s_2}^{(1)} \cdot \hat{\mathbf{K}}_{i, j}^{(2)})^2 \right] = (n-|\mathcal{N}|) E \left[(\hat{\mathbf{K}}_{i+s_1, j+s_2}^{(1)} \cdot \hat{\mathbf{K}}_{i, j}^{(2)})^2 \right] = (n-|\mathcal{N}|) n (\sigma_1^2 + \nu_1^2) (\sigma_2^2 + \nu_2^2), \quad (19)$$

and, for the numerator:

$$E \left[\left(\hat{\mathbf{K}}^{(1)} \cdot \hat{\mathbf{K}}^{(2)} \right)^2 \right] = E \left[\hat{\mathbf{K}}^{(1)} \cdot \hat{\mathbf{K}}^{(2)} \right]^2 + \text{Var} \left[\hat{\mathbf{K}}^{(1)} \cdot \hat{\mathbf{K}}^{(2)} \right] = \left(\mathbf{K}^{(1)} \cdot \mathbf{K}^{(2)} \right)^2 + O(\sqrt{n}), \quad (20)$$

because the variance in (20) is proportional to \sqrt{n} due to the CLT. Thus, from (19) and (20) we finally have:

$$\text{PCE} \left(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)} \right) \approx \frac{(n - |\mathcal{N}|) \left(\mathbf{K}^{(1)} \cdot \mathbf{K}^{(2)} \right)^2}{(n - |\mathcal{N}|) n (\sigma_1^2 + \nu_1^2) (\sigma_2^2 + \nu_2^2)} = \frac{n^2 \nu_1^2 \nu_2^2 \left(\frac{1}{n} \frac{\mathbf{K}^{(1)} \cdot \mathbf{K}^{(2)}}{\nu_1 \nu_2} \right)^2}{n (\sigma_1^2 + \nu_1^2) (\sigma_2^2 + \nu_2^2)} \quad (21)$$

$$= n \rho^2 \left(\mathbf{K}^{(1)}, \mathbf{K}^{(2)} \right) \left(1 + \sigma_1^2 / \nu_1^2 \right)^{-1} \left(1 + \sigma_2^2 / \nu_2^2 \right)^{-1}. \quad (22)$$

The following important observations can be made from here. First, the PCE is proportional to the number of pixels, n , and the square correlation between deformed fingerprints. Second, the PCE increases with increasing SNR, $\text{SNR} \left(\hat{\mathbf{K}}^{(l)} \right) \triangleq \nu_l^2 / \sigma_l^2$, and, in the limit of perfect estimates, $\text{PCE} = n \rho^2$. Due to the CLT, when estimating the fingerprint $\hat{\mathbf{K}}^{(1)}$ from L images, the SNR increases L times w.r.t. an estimate from a single image (its noise residual \mathbf{W}). Thus, we have the following ratio between the PCEs for large L :

$$\frac{\text{PCE} \left(\mathbf{W}, \hat{\mathbf{K}}^{(2)} \right)}{\text{PCE} \left(\hat{\mathbf{K}}^{(1)}, \hat{\mathbf{K}}^{(2)} \right)} = \frac{1 + 1/(L \times \text{SNR}(\mathbf{W}))}{1 + 1/\text{SNR}(\mathbf{W})} \approx \frac{1}{1 + 1/\text{SNR}(\mathbf{W})}. \quad (23)$$

REFERENCES

1. M. Chen, J. Fridrich, and M. Goljan. Digital imaging sensor identification (further study). In E.J. Delp and P.W. Wong, editors, *Proc. SPIE, Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents IX*, volume 6505, pages 0P–0Q, January 2007.
2. 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.
3. J. Fridrich. Digital image forensic using sensor noise. *IEEE Signal Processing Magazine*, 26(2):26–37, 2009.
4. M. Goljan, M. Chen, and J. Fridrich. Identifying common source digital camera from image pairs. In *Proc. IEEE International Conference on Image Processing (ICIP)*, September 2007.
5. 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–01–0I–12, January 2009.
6. 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, *Proc. SPIE, Electronic Imaging, Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, volume 6819, pages OE1–OE13, January 2008.
7. 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.
8. W. Hugemann. Correcting lens distortions in digital photographs, 2010. http://www.imagemagick.org/Usage/lens/correcting_lens_distortions.pdf.
9. J.R. Janesick. *Scientific Charge-Coupled Devices*, volume PM83. SPIE Press Monograph, 2001.
10. H. Li and R. Hartley. A non-iterative method for correcting lens distortion from nine-point correspondences. In *Proc. OmniVision'05, ICCV-workshop*, 2005.
11. 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.
12. M.K. Mihcak, I. Kozintsev, and K. Ramchandran. Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising, March 1999.
13. J. Pers and S. Kovacic. Nonparametric, model-based radial lens distortion correction using tilted camera assumption. In *Kropatsch (Eds.), Proceedings of the Computer Vision Winter Workshop 2002*, pages 286–295, 2002.