Linear Filter Kernel Estimation Based on Digital Camera Sensor Noise

Chang Liu and Matthias Kirchner
Department of Electrical and Computer Engineering
SUNY Binghamton, Binghamton, NY/USA

Abstract

We study linear filter kernel estimation from processed digital images under the assumption that the image's source camera is known. By leveraging easy-to-obtain camera-specific sensor noise fingerprints as a proxy, we have identified the linear cross-correlation between a pre-computed camera fingerprint estimate and a noise residual extracted from the filtered query image as a viable domain to perform filter estimation. The result is a simple yet accurate filter kernel estimation technique that is relatively independent of image content and that does not rely on hand-crafted parameter settings. Experimental results obtained from both uncompressed and JPEG compressed images suggests performances on par with highly developed iterative constrained minimization techniques.

Introduction

Inferring the processing history of digital images is a core problem of digital image forensics [1, 2]. A rich body of literature reflects the plethora of different operations an image may undergo after being captured by an acquisition device, focussing on aspects of manipulation detection, localization, and parameter estimation. Purely technical solutions prevail, leaving the question as to what operations are to be considered legitimate or malicious to human interpretation [3]. Recent advances in data-driven techniques advocate the idea of general-purpose image forensics, where mostly very high-dimensional feature spaces are chosen to detect, identify, or localize a variety of operations in a unified framework [4–8]. The generality of these techniques comes at the cost of a relatively coarse granularity with respect to intra-class variations within the same type of processing (e. g., different filter sizes or strengths). Even with practicable solutions to overcome issues of scalability in multi-classification [4, 5, 8], it seems infeasible to assume that a single model will ever be informative enough to discern between a large number of parameter settings across multiple processing types. Targeted techniques with the ability to *estimate* parameters of specific processing operations, such as previous JPEG quantization tables [9, 10], the shape of non-linear intensity mappings [11], or parameters of affine transformations [12–14] will thus remain to play an important role in the realm of image forensics.

Along these lines, our focus here is on estimating the coefficients of an unknown linear filter kernel a query image may have been subjected to. In contrast to the analysis of non-linear filtering, where a large number of works have focussed on the specifics of median filtering [15, 16, among many others], the interest in linear filtering has appeared relatively scattered over the forensics community [17–20]. At the same time, however, blind deconvolution and point spread function (PSF) estimation are widely established

fields in image processing of course [21–23], with applications including image restoration and enhancement, or the removal of motion blur. The major challenge in this stream of research is that estimating a filter kernel is an ill-posed problem in the absence of the original unfiltered image. The key is to incorporate informative prior models about the original image into the estimation process. In classical blind deconvolution, such prior models will typically pertain to a generic digital image, effectively acting as a proxy of the captured scene. For forensic purposes, where the focus is clearly on processing applied after image capturing, this problem can be approached in a much more narrow sense by taking knowledge about image acquisition into account, e. g., through reasonable assumptions that the original image underwent a demosaicing procedure [17], or that the original image was stored in JPEG format [19].

In the remainder of this paper, we demonstrate how knowledge of the acquisition camera's sensor noise fingerprint [24] can facilitate efficient and effective solutions to the linear filter kernel estimation problem. Similar to digital camera identification in the presence of geometric distortion [25–27], the premise here is that the camera fingerprint, as part of the image, will undergo the same processing the image undergoes. Under the assumption of linear filtering, the effects of this processing can be measured directly in the linear cross-correlation between the camera's "clean" fingerprint and the fingerprint estimate obtained from the processed image. This way, the (assumedly) known camera fingerprint imposes a strong prior model of the unfiltered signal in the estimation process. Before we detail our approach below, the following sections give a brief overview of the general filter kernel estimation problem and some background on digital camera sensor noise forensics. We then continue with a description of a simple filter kern estimation technique and present experimental results.

Model and Notation

We adopt a generic linear shift-invariant (LSI) model

$$y[i, j] = \sum_{u,v} h[u, v] \cdot x[i - u, j - v] + n[i, j] \qquad (1)$$

in which the observed image $y[i, j]$ results from the linear two-dimensional convolution of the "clean" image $x[i, j]$ with a filter kernel $h[u, v]$, plus some measurement noise $n[i, j]$. The goal of filter kernel estimation is to determine the coefficients $h[u, v]$ of the unknown filter from the image $y[i, j]$, where it is often instrumental to assume a (without loss of generality) square kernel support of $S \times S$, $S = 2S_0 + 1$, i. e., $-S_0 \le u, v \le S_0$, and $\sum_{u,v} h[u, v] = 1$. It

is convenient to write Equation (1) as a system of linear equations,

$$\mathbf{y} = \mathbf{X} \cdot \mathbf{h} + \mathbf{n}, \tag{2}$$

where $\mathbf{y}$, $\mathbf{h}$, and $\mathbf{n}$ are the column-major vector representations of the two-dimensional signals at play. Matrix $\mathbf{X}$ stacks the vectorized local $S \times S$ support neighborhoods in a row-by-row manner:

$$\mathbf{X} = \begin{bmatrix} \vdots & & \vdots & & \vdots \\ x[i-S_0, j-S_0] & \cdots & x[i,j] & \cdots & x[i+S_0, j+S_0] \\ \vdots & & \vdots & & \vdots \end{bmatrix}.$$

With slight abuse of notation, let us also write $\mathbf{y} = \mathbf{x} * \mathbf{h}$ to denote convolution. In a similar vein, denote the linear cross-correlation between two signals as $\boldsymbol{\rho} = \mathbf{x} \otimes \mathbf{y}$, with

$$\rho[r,s] = \sum_{i,j} x[i,j] \cdot y[r+i, s+j]. \tag{3}$$

We will typically put emphasis on the cross-correlation computed over a small range of lags, $-L_0 \leq r, s \leq L_0$. In this context, recall that defining $\tilde{x}[i,j] = x[-i,-j]$ helps to express cross-correlation in terms of convolution: $\mathbf{x} \otimes \mathbf{y} = \tilde{\mathbf{x}} * \mathbf{y}$. Finally, note that we write $\mathbf{x} \odot \mathbf{y}$ for the element-wise multiplication of two vectors of equal length and $\mathbf{X} \circ \mathbf{Y}$ for the Kronecker product of two matrices, and let vector $\mathbf{1}^{(M)}$ denote an M-element column vector with all ones.

Linear Kernel Estimation

Without prior knowledge or assumptions about the clean image, solving Equation (2) for $\mathbf{h}$ is an ill-posed problem. The literature on blind deconvolution, for which kernel estimation is an integral step, is thus rich of proposals for prior image models and kernel estimation algorithms [21, 23]. While a detailed review is beyond the scope of this paper, it is worth pointing out that many state-of-the-art techniques approach the problem in a regularization framework, minimizing expressions of the general form

$$\|\mathbf{x} * \mathbf{h} - \mathbf{y}\|^2 + \phi(\mathbf{x}, \mathbf{h}) \to \min. \tag{4}$$

The regularization term $\phi(\mathbf{x}, \mathbf{h})$ acts as (a set of) constraint(s) in the minimization, favoring natural images and filter smoothness. The minimization may be conducted over the domain of filter kernels, clean images, or both jointly. Due to the lack of tractable parametric image models, however, many techniques work in the image gradient domain. The popular algorithm by Krishnan et al. [28] for instance employs sparsity constraints of the form $\phi(\mathbf{x}, \mathbf{h}) = \|\mathbf{x}\|_1 / \|\mathbf{x}\|_2 + \alpha \|\mathbf{h}\|_1$, alternating between a minimization over $\mathbf{x}$ and $\mathbf{h}$ iteratively.

More specific constraints may be employed in the context of image forensics, where post-capture linear filtering is of particular concern. The idea here is that clean camera images exhibit certain signal characteristics pertaining to the acquisition process and that image processing will affect these characteristics. Most prominently, Swaminathan et al. [17] leverage that demosaicing from a color filter array (CFA) introduces peculiar periodic inter-pixel dependencies in images taken with a digital camera. Incorporating knowledge or assumptions about the layout of the CFA, the authors attempt to reinstate those pixel structures while estimating the filter kernel. In a related work, Conotter et al. [19] studied the effect of linear filtering on the distribution of quantized JPEG DCT coefficients, yet with a focus on the detection of different linear filter kernels.

Correlation and Convolution

While all approaches above rely on (the restoration) of general statistical signal characteristics (e. g., sparsity, or a certain form of periodic pixel interdependencies), knowledge of the actual clean signal would clearly open more direct avenues to filter kernel estimation. Consider for instance the most simple conceivable situation, where a known i.i.d. noise signal is fed into an LSI system with unknown kernel $\mathbf{h}$. It is well-known that in a situation like this, the filter kernel can be obtained directly from the cross-correlation between the clean input and the filtered output,

$$\boldsymbol{\rho} = \mathbf{x} \otimes (\mathbf{x} * \mathbf{h} + \mathbf{n}) = \mathbf{h} * (\mathbf{x} \otimes \mathbf{x}) + \mathbf{x} \otimes \mathbf{n} \approx \alpha \mathbf{h}, \tag{5}$$

as the auto-correlation $\mathbf{x} \otimes \mathbf{x}$ of a white noise signal will converge to an impulse, and $\mathbf{x}$ and $\mathbf{n}$ are uncorrelated.

Figure 1 illustrates this circumstance without loss of generality for a one-dimensional i.i.d. Gaussian signal $\mathbf{x}$, $x[i] \sim \mathcal{N}(0,1)$, subjected to Gaussian blurs $\mathbf{h}$,

$$h[u] \propto \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-u^2}{2\sigma^2}\right), \qquad -S_0 \leq u \leq S_0, \tag{6}$$

with varying kernel support $S = 2S_0 + 1$ and filter strength σ. The graphs illustrate the cross-correlation up to lag ± 5 between the clean signal and the filtered signal, in the presence of additive white noise $n[i] \sim \mathcal{N}(0,1)$. The Gaussian density function, from which the filter coefficients are sampled according to Equation (6), is depicted for reference. Observe how the measured cross-correlation in this simulation perfectly reveals the filter the original signal went through.

Although filter kernel estimation from actual digital images clearly poses a more challenging situation, we will argue below that knowledge of an image's source camera can establish a situation very similar to the seemingly unrealistic and simplistic simulation above. Specifically, we will exploit that every image taken with a specific camera can be assumed to contain a camera-specific sensor noise fingerprint [24]. As this quasi-known noise signal will undergo the same processing the image is subjected to, filter kernel estimation can be carried out based on the cross-correlation between the fingerprint estimated from clean images and the noise estimate obtained from the filtered image.

Sensor Noise Forensics

Camera sensor noise fingerprints inevitably emerge in images captured with a digital camera due to minute differences and manufacturing imperfections of individual sensor elements [24]. Even under absolutely homogenous exposure each sensor element will react slightly different, in its own unique way. Denoting the noise-free sensor output $\mathbf{x}^{(0)}$, this photo-response non-uniformity (PRNU) can be modeled as a spatially varying noise pattern of multiplicative nature, $\mathbf{x} = (\mathbf{1} + \mathbf{k}) \odot \mathbf{x}^{(0)} + \boldsymbol{\theta}$. The PRNU factor $\mathbf{k}$ is unique to the specific source camera. It will be present in every image taken with the same camera, yet it will vary from camera to camera (sensor) [29]. Additive modeling noise $\boldsymbol{\theta}$ comprises dark current and a variety of temporally varying noise sources.

Estimating a digital camera's PRNU fingerprint requires access to a sufficiently large number of clean camera outputs, $\mathbf{x}_1, \ldots, \mathbf{x}_N$, each of which is fed into a suitable denoising filter $d(\cdot)$ to obtain noise residuals $\mathbf{w}_{\mathbf{x}_i} = \mathbf{x}_i - d(\mathbf{x}_i)$. Adopting the simple
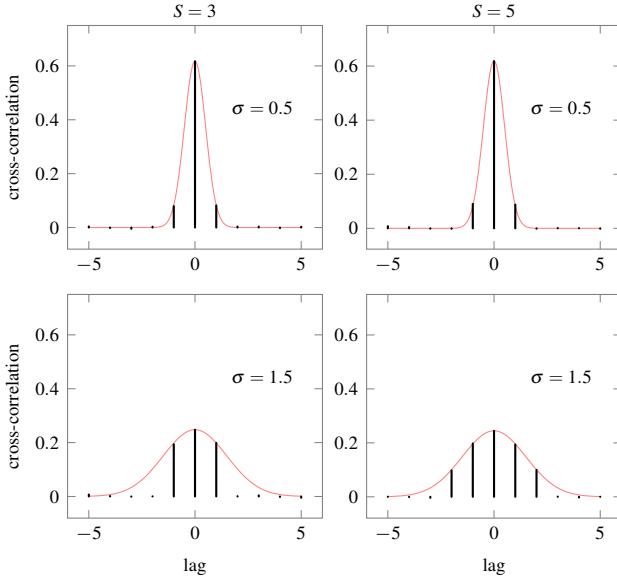
Figure 1. *Cross-correlation between a 100,000 element i.i.d. sequence* $\mathbf{x}$, $x[i] \sim \mathcal{N}(0,1)$, *and a filtered noisy version* $\mathbf{x} * \mathbf{h} + \mathbf{n}$, $n[i] \sim \mathcal{N}(0,1)$, *for Gaussian blur kernels* $\mathbf{h}$ *with different supports* S *and standard deviations* σ. *Solid lines indicate the envelope of the corresponding (scaled) filter kernel.*

multiplicative model

$$\mathbf{w}_{\mathbf{x}_i} = \mathbf{k} \odot \mathbf{x}_i + \boldsymbol{\Theta} \tag{7}$$

with white Gaussian noise $\boldsymbol{\Theta}$, the maximum likelihood PRNU fingerprint estimator is [24]

$$\hat{\mathbf{k}} = \left(\sum_{i=1}^{N} \mathbf{w}_{\mathbf{x}_i} \odot \mathbf{x}_i \right) \odot \left(\sum_{i=1}^{N} \mathbf{x}_i^2 \right)^{-1}. \tag{8}$$

All operations in the equation above are element-wise.

In the context of digital image forensics, camera sensor noise fingerprints have been successfully employed in camera identification [29], camera-based blind image clustering [30], and image manipulation detection [31, 32], amongst many others. A common building block for all those applications is the evaluation of the cross-correlation between a camera-specific PRNU term and the noise residual obtained from a query image $\mathbf{x}$, $\boldsymbol{\rho}_{\mathbf{x}} = (\hat{\mathbf{k}} \odot \mathbf{x}) \otimes \mathbf{w}_{\mathbf{x}}$. A match is declared when a sufficiently large correlation-based similarity score is observed, e. g., in terms of the maximum normalized cross-correlation, or the peak-to-correlation energy.

While PRNU-based forensic techniques based on such correlation measures have been reported to be fairly robust against various types of image processing, we argue here that any type of processing will inevitably have an impact on the image's inherent sensor noise pattern, and ultimately also on the resulting cross-correlation between the processed image's noise residual and the camera fingerprint estimate from unprocessed images. Inspired by the previous section on the interplay of convolution and correlation, the following section inspects more closely how linear image filtering translates into the PRNU cross-correlation domain, and thus how knowledge of the clean image's source camera may inform linear filter kernel estimation.

PRNU-based Linear Filter Kernel Estimation

We base the estimation of the kernel $\mathbf{h}$ from a query image $\mathbf{y} = \mathbf{x} * \mathbf{h} + \mathbf{n}$ on the following assumptions:

1. the source camera of the clean image $\mathbf{x}$ is known;
2. the denoising filter to compute noise residuals for all PRNU-related analyses is linear, i. e.,

$$\mathbf{w}_{\mathbf{x}} = \mathbf{x} - d(\mathbf{x}) = \mathbf{x} - \mathbf{x} * \mathbf{d}; \quad \text{and} \tag{9}$$

3. C clean (unprocessed) images $\mathbf{x}_1, \dots, \mathbf{x}_C$ from the same camera are available to compute clean cross-correlations

$$\boldsymbol{\rho}_{\mathbf{x}_c} = \hat{\mathbf{k}} \otimes \mathbf{w}_{\mathbf{x}_c} = \hat{\mathbf{k}} \otimes (\mathbf{x}_c - \mathbf{x}_c * \mathbf{d}). \tag{10}$$

The first assumption is reasonable considering that moderate post-processing affects source camera attribution only mildly [33, 34]. For a better understanding of the other two assumptions, let us consider the cross-correlation

$$\boldsymbol{\rho}_{\mathbf{y}} = \hat{\mathbf{k}} \otimes \mathbf{w}_{\mathbf{y}} = \hat{\mathbf{k}} \otimes (\mathbf{y} - \mathbf{y} * \mathbf{d}) \tag{11}$$

between the camera's PRNU fingerprint estimate and the query image's noise residual, for which we may also write

$$\boldsymbol{\rho}_{\mathbf{y}} = \hat{\mathbf{k}} \otimes ((\mathbf{x} * \mathbf{h} + \mathbf{n}) - (\mathbf{x} * \mathbf{h} + \mathbf{n}) * \mathbf{d}) \tag{12}$$

$$= \hat{\mathbf{k}} \otimes ((\mathbf{x} - \mathbf{x} * \mathbf{d}) * \mathbf{h} + (\mathbf{n} - \mathbf{n} * \mathbf{d})) \tag{13}$$

$$= (\hat{\mathbf{k}} \otimes \mathbf{w}_{\mathbf{x}}) * \mathbf{h} + \hat{\mathbf{k}} \otimes \mathbf{w}_{\mathbf{n}} \tag{14}$$

$$= \boldsymbol{\rho}_{\mathbf{x}} * \mathbf{h} + \boldsymbol{\nu}. \tag{15}$$

Similar to Equation (5), above lines state that the PRNU cross-correlation obtained from a filtered image $\mathbf{y} = \mathbf{x} * \mathbf{h} + \mathbf{n}$ is the filtered PRNU cross-correlation obtained from the corresponding clean image, plus an additive noise term $\boldsymbol{\nu}$. Note that this result only emerges because of the linearity of the denoising filter $d(\cdot)$.

Figures 2 and 3 illustrate the relation between $\boldsymbol{\rho}_{\mathbf{x}}$ and $\boldsymbol{\rho}_{\mathbf{y}}$ by depicting the center 11×11 portions ($L_0 = 5$) of those quantities, averaged over a number of clean images and filtered images (5×5 Gaussian blurring with $\sigma \in \{0.5, 2.5\}$).[1] Besides clearly promoting the idea that stronger blurring in the image domain translates to stronger blurring in the PRNU cross-correlation domain, the graphs also suggest an important difference to the simple simulation in Figure 1: we cannot assume $\boldsymbol{\rho}_{\mathbf{x}}$ to have impulse-like qualities. Although the PRNU noise pattern itself may very well exhibit i.i.d. characteristics, demosaicing and other post-capture operations will introduce inevitable inter-pixel dependencies.

Assuming for a moment that the clean cross-correlation $\boldsymbol{\rho}_{\mathbf{x}}$ for the given query image was known, we may obtain the standard least squares kernel estimate from minimizing $\|\boldsymbol{\rho}_{\mathbf{x}} * \mathbf{h} - \boldsymbol{\rho}_{\mathbf{y}}\|^2$,

$$\hat{\mathbf{h}} = (\mathbf{P}'_{\mathbf{x}} \mathbf{P}_{\mathbf{x}})^{-1} \mathbf{P}'_{\mathbf{x}} \boldsymbol{\rho}_{\mathbf{y}}. \tag{16}$$

Matrix $\mathbf{P}_{\mathbf{x}}$ is the equivalent of matrix $\mathbf{X}$ in Equation (2), here filled with elements of $\boldsymbol{\rho}_{\mathbf{x}}$ for an assumed $S \times S$ filter support, i. e., the m-th row of $\mathbf{P}_{\mathbf{x}}$ contains—vectorized in column-major order—the S^2 clean cross-correlation values from the $S \times S$ portion of $\boldsymbol{\rho}_{\mathbf{x}}$ corresponding to the m-th element in $\boldsymbol{\rho}_{\mathbf{y}}$.

---

[1]We refer to the following section for details about our datasets and experimental setup.
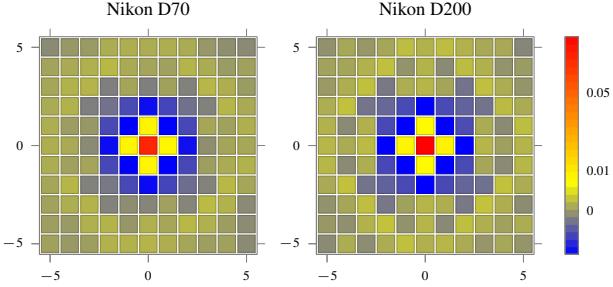
Figure 2. *Average cross-correlation ($L_0 = 5$) between camera fingerprint estimate from a Nikon D70 and a Nikon D200 camera and noise residuals from clean images of the same camera (correlation displayed with non-linear color mapping $y = x^{2/3}$ for better visibility).*

Figure 3. *Average cross-correlation ($L_0 = 5$) between Nikon D200 camera fingerprint estimate and noise residuals from blurred images of the same camera (5×5 Gaussian blur with standard deviation σ; non-linear color mapping $y = x^{2/3}$ for better visibility).*

In reality, the situation is more involved as we cannot assume direct access to $\boldsymbol{\rho}_{\mathbf{x}}$. Different from working in the image domain directly, however, it is more reasonable to assume here that the clean signal under consideration—$\boldsymbol{\rho}_{\mathbf{x}}$ instead of $\mathbf{x}$—should be relatively independent of the image content. Specifically, we can expect that the distinct characteristics of $\boldsymbol{\rho}_{\mathbf{x}}$ will depend to a large degree on $\hat{\mathbf{k}}$ and its characteristics due to camera-internal processing. A simple and straight-forward remedy is to work with an estimate of $\boldsymbol{\rho}_{\mathbf{x}}$, obtained from a number of clean images,

$$\hat{\boldsymbol{\rho}}_{\mathbf{x}} = \frac{1}{C} \sum_{c=1}^{C} \left(\hat{\mathbf{k}} \otimes \mathbf{w}_{\mathbf{x}_c} \right), \tag{17}$$

with mean values computed element-wise (see also Figure 2). The filter kernel is then estimated based on Equation (16), with matrix $\mathbf{P}$ replaced by the estimate $\hat{\mathbf{P}}$.

Since an ordinary least squares kernel estimate is not guaranteed to sum to unity, we make use of the fact that the convolution $\hat{\boldsymbol{\rho}}_{\mathbf{x}} * \mathbf{h}$ can also be written as

$$\sum_{|u|+|v| \neq 0} h[u,v] \cdot \hat{\rho}_{\mathbf{x}}[i-u, j-v] + \left(1 - \sum_{|u|+|v| \neq 0} h[u,v] \right) \cdot \hat{\rho}_{\mathbf{x}}[i,j] \tag{18}$$

to define auxiliary column vector $\boldsymbol{\eta} = \mathbf{Rh}$, with $(S^2 - 1) \times S^2$ matrix $\mathbf{R}$ removing the center element ($|u| + |v| \neq 0$) from $\mathbf{h}$,

$$\mathbf{R} = \begin{bmatrix} \ddots & & & & & \\ & 1 & & & \mathbf{0} & \\ & & 1 & & & \\ & \mathbf{0} & & 0 & 1 & \\ & & & & 0 & 1 \\ & & & & & \ddots \end{bmatrix}. \tag{19}$$

We estimate $\boldsymbol{\eta}$ instead of $\mathbf{h}$, i.e., all but the center coefficient, by reformulating Equation (15) based on Equation (18) to give

$$\boldsymbol{\rho}_{\mathbf{y}} - \hat{\boldsymbol{\rho}}_{\mathbf{x}} = \tilde{\mathbf{P}} \cdot \boldsymbol{\eta} + \boldsymbol{\nu}, \tag{20}$$

with matrix $\tilde{\mathbf{P}} = \hat{\mathbf{P}} \mathbf{R}' - \hat{\boldsymbol{\rho}}_{\mathbf{x}}' \circ \mathbf{1}^{(S^2-1)}$ being obtained from $\hat{\mathbf{P}}$ by removing the center column and subtracting the m-th element of $\hat{\boldsymbol{\rho}}_{\mathbf{x}}$ from the m-th row of the matrix. Vector

$$\hat{\boldsymbol{\eta}} = \left(\tilde{\mathbf{P}}_{\mathbf{x}}' \tilde{\mathbf{P}}_{\mathbf{x}} \right)^{-1} \tilde{\mathbf{P}}_{\mathbf{x}}' \left(\boldsymbol{\rho}_{\mathbf{y}} - \hat{\boldsymbol{\rho}}_{\mathbf{x}} \right) \tag{21}$$

then holds all $S^2 - 1$ kernel coefficient estimates $\hat{h}[u,v]$ except for the center one ($|u| + |v| \neq 0$), which can be computed from,

$$\hat{h}[0,0] = 1 - \sum_i \hat{\eta}_i. \tag{22}$$

Experimental Results

In the following experimental validation of the proposed approach we consider the estimation of Gaussian blurs $\mathbf{h}$ of support $S \times S$, $S \in \{3, 5, 7\}$ and strengths $0.5 \leq \sigma \leq 5$. We worked with a total of 1,445 never-compressed "Lightroom" images from the Dresden Image Database [35] in our experiments, stemming from three different camera models and six unique digital cameras (see table). All images were cropped to their center 1000×1000 portion and converted to grayscale prior to any processing. Camera-specific sensor noise fingerprints were estimated from 20 flat-field images per camera, employing the post-processing suggested in [24]. We found a 3×3 Gaussian blur kernel with standard deviation $\sigma = 0.5$ to produce suitable noise residuals $\mathbf{w}$ for our purposes, cf. Equation (9). Cross-correlation estimates $\hat{\boldsymbol{\rho}}_{\mathbf{x}}$ were obtained from $C = 50$ randomly chosen images per camera. Those images were otherwise excluded from the kernel estimation tests.

For an assumed kernel support $\hat{S} \geq S$, kernel estimates $\hat{\mathbf{h}}$ are obtained from Equations (21) and (22) by fitting the adopted linear model to the zero-centered $\hat{S} \times \hat{S}$ portion of the observed cross-correlation $\boldsymbol{\rho}_{\mathbf{y}}$. We are interested in the estimation accuracy when the true kernel size is assumed to be known ($\hat{S} = S$), for instance as the result of a prior feature-based inference or classification step [6, 8]. In addition, we also consider the more generic case where such knowledge is not available. In the latter scenario, we set $\hat{S} = 7$ independent of the true kernel size. In all cases, the mean squared error (MSE) between the true and the estimated kernel,

$$\text{MSE} = \|\mathbf{h} - \hat{\mathbf{h}}\|^2 / \hat{S}^2, \tag{23}$$

Number of images from the Dresden Image Database (Lightroom, never-compressed) [35] used in the experiments.

camera model	device 0	device 1
Nikon D70	175	188
Nikon D70s	175	177
Nikon D200	360	370
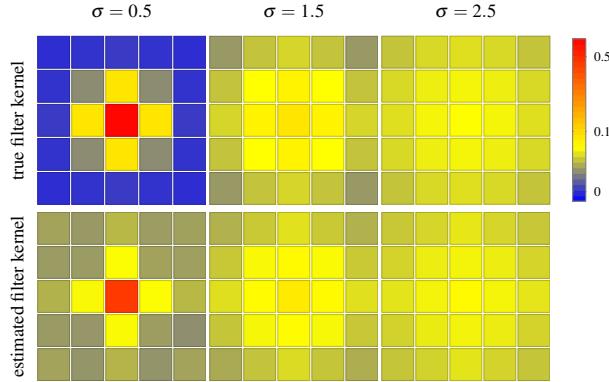
Figure 4. *Gaussian kernels (S = 5) (top) and corresponding average kernel estimates from Nikon D200 images (bottom). Filter size assumed to be known ($\hat{S} = S$); non-linear color mapping $y = \sqrt{x}$ for better visibility.*

averaged per kernel instance over all images of the same camera, serves as quantitative criterion. For comparison, we report results obtained with the state-of-the-art iterative blind deconvolution algorithm by Krishnan et al. [28], which we operate for a given filter support $\hat{S}$ with the default settings provided in the authors' reference implementation.[2]

Figures 5 through 9 present quantitative estimation results for the described experimental setup. Specifically, Figures 5 and 6 cover the estimation of 3×3 Gaussian blurs with known ($\hat{S} = 3$) and unknown ($\hat{S} = 7$) filter support, respectively, while Figures 7 and 8 report equivalent results for 5×5 filters. Figure 9 examines the estimation of filters with support $S = 7$. Each figure comprises six panels, in correspondence with the six digital cameras in our test set. The six graphs per panel depict the average MSE values of the two estimation methods under test for different filter strengths σ, as obtained after storing the blurred images in uncompressed 8-bit bitmap format (Q_*), or as JPEGs with quality factors 90 or 80 (Q_{90}, Q_{80}), respectively.

Across all tested settings, the graphs indicate that the proposed PRNU-based filter kernel estimation performs as good as or better than the computationally expensive iterative benchmark approach. The reported results suggest a particularly strong advantage over the benchmark method when small kernel sizes are concerned and when the true kernel size is assumed to be known (cf. Figures 5 and 7). Under such circumstances, the PRNU-based estimation benefits notably from knowledge of the actual filter kernel size. In general, the estimation errors are considerably lower for stronger blurs (i. e., larger kernel supports and more uniform filter kernels), as the impact of high-frequency image content becomes less predominant (with respect to characteristics of the noise residuals, and similarly also with regards to the computation of image gradients for the benchmark method). The average estimated 5×5 filter kernels from blurred Nikon D200 images in Figure 4 highlight this effect for filter strengths $\sigma \in \{0.5, 1.5, 2.5\}$, in correspondence with the respective Q_* data points in Figure 7 (e). Finally, observe that the estimation accuracy remains relatively stable with respect to JPEG post-compression, although differences exist for images from different source cameras. Overall, the results suggest that the estimation of small filter kernels of unknown size appears to be affected by lossy compression the most, cf. Figure 6.

---

[2] http://cs.nyu.edu/~dilip/research/blind-deconvolution

In summary, we conclude that our experiments demonstrate the potential of PRNU-based linear filter kernel estimation clearly. For the chosen type of Gaussian kernels, our simple linear least squares estimator achieves highly competitive estimation accuracies under the premise that the source camera is known, even in the presence of considerable JPEG compression. We see the presented results as a first step and thus leave the examination of a broader class of linear kernels to future work, in which also the impact of image size should be investigated.

Concluding Remarks

Blind linear filter kernel estimation can benefit from knowledge of or assumptions about the image capturing process. Along those lines, this paper has leveraged camera-specific PRNU sensor noise [24] as a proxy to model the impact of linear filtering on digital images. Instead of estimating the filter kernel in the pixel or gradient domain, we have identified the cross-correlation between a camera fingerprint estimate and a noise residual extracted from the filtered query image as a viable domain to perform filter estimation. The result is a simple yet effective linear least squares algorithm that gives estimation accuracies on par with highly developed iterative constrained minimization techniques.

Future generations of our proof-of-concept estimator may draw on the rich body of literature on kernel estimation and blind deconvolution [21–23] to strive for improvements along several dimensions. Specifically, it seems viable to incorporate more elaborate probabilistic models of the clean cross-correlation "signal" $\boldsymbol{\rho}_\mathbf{x}$ to account more explicitly for its (co-)variance structure. On possibility is a Bayesian framework, which would lead to regularization terms already known from state-of-the-art techniques. A content-based correlation predictor, adopting techniques along the lines of [24], may be another approach. As for the data fed into the estimator, it seems promising to explore to what degree larger portions of the computed cross-correlation might contribute to more accurate estimates. The reader should keep in mind that current estimates of an $\hat{S} \times \hat{S}$ kernel are obtained from merely $\hat{S} \times \hat{S}$ data points. Finally, it also seems worthwhile to revisit the implementation of the unity constraint in the estimation procedure.

We close this paper on a more general note, as our results suggest that the PRNU cross-correlation domain bears valuable information beyond the predominant examination of the most prominent peak for the sake of fingerprint synchronization and camera identification [24–26]. A broadened view may not only open up new avenues to the analysis and detection of fingerprint de-synchronization attacks based on seam-carving [36] or patch replacement strategies [37], but also contribute to a deepened understanding of source camera attribution in the presence of sophisticated in-camera processing [38]—thus underlining and strengthening the role of camera sensor noise as one of the most valuable image characteristics in digital image forensics further.

References
[1] R. Böhme and M. Kirchner, "Media forensics," in *Information Hiding*, S. Katzenbeisser and F. Petitcolas, Eds. Artech House, 2016, ch. 9, pp. 231–259.
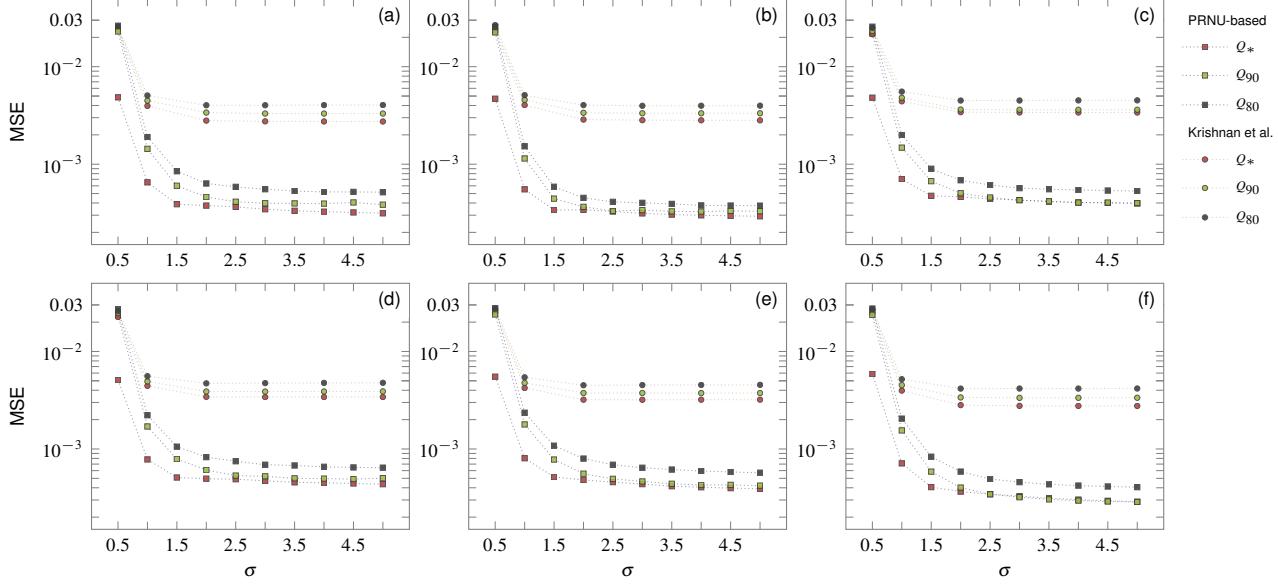
Figure 5. Kernel estimation accuracies for images blurred with a 3×3 Gaussian kernel of strength σ. Kernel size assumed to be known ($S = \hat{S}$). Nikon D70 (a) & (b), Nikon D70s (c) & (d), Nikon D200 (e) & (f). Images stored in uncompressed bitmap format (Q_*), or with JPEG qualities 90 (Q_{90}) and 80 (Q_{80}).
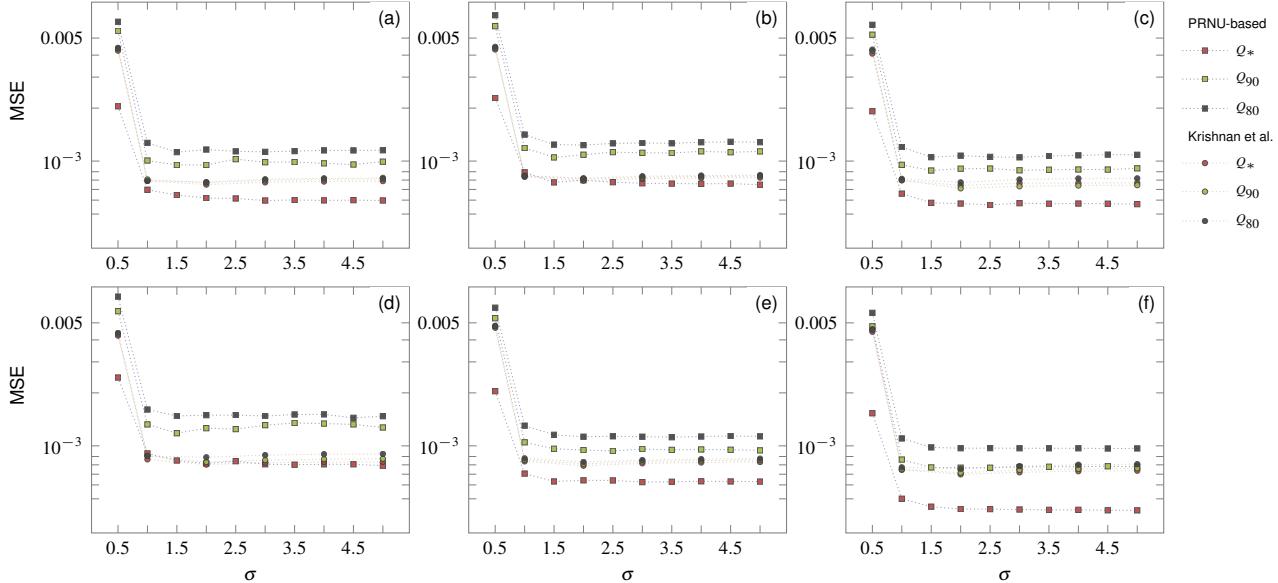

Figure 6. Kernel estimation accuracies for images blurred with a 3×3 Gaussian kernel of strength σ. Kernel size unknown ($\hat{S} = 7$). Nikon D70 (a) & (b), Nikon D70s (c) & (d), Nikon D200 (e) & (f). Images stored in uncompressed bitmap format (Q_*), or with JPEG qualities 90 (Q_{90}) and 80 (Q_{80}).

[2] H. Farid, *Photo Forensics*. MIT Press, 2016.

[3] R. Böhme and M. Kirchner, "Counter-forensics: Attacking image forensics," in *Digital Image Forensics: There is More to a Picture Than Meets the Eye*, H. T. Sencar and N. Memon, Eds. Springer, 2013, pp. 327–366.

[4] L. Verdoliva, D. Cozzolino, and G. Poggi, "A feature-based approach for image tampering detection and localization," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014.

[5] W. Fan, K. Wang, and F. Cayre, "General-purpose image forensics using patch likelihood under image statistical mod-els," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2015.

[6] H. Li, W. Luo, X. Qiu, and J. Huang, "Identification of various image operations using residual-based features," *IEEE Transactions on Circuits and Systems for Video Technology*, in press.

[7] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in *ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*, 2016, pp. 5–10.

[8] M. Boroumand and J. Fridrich, "Scalable processing history
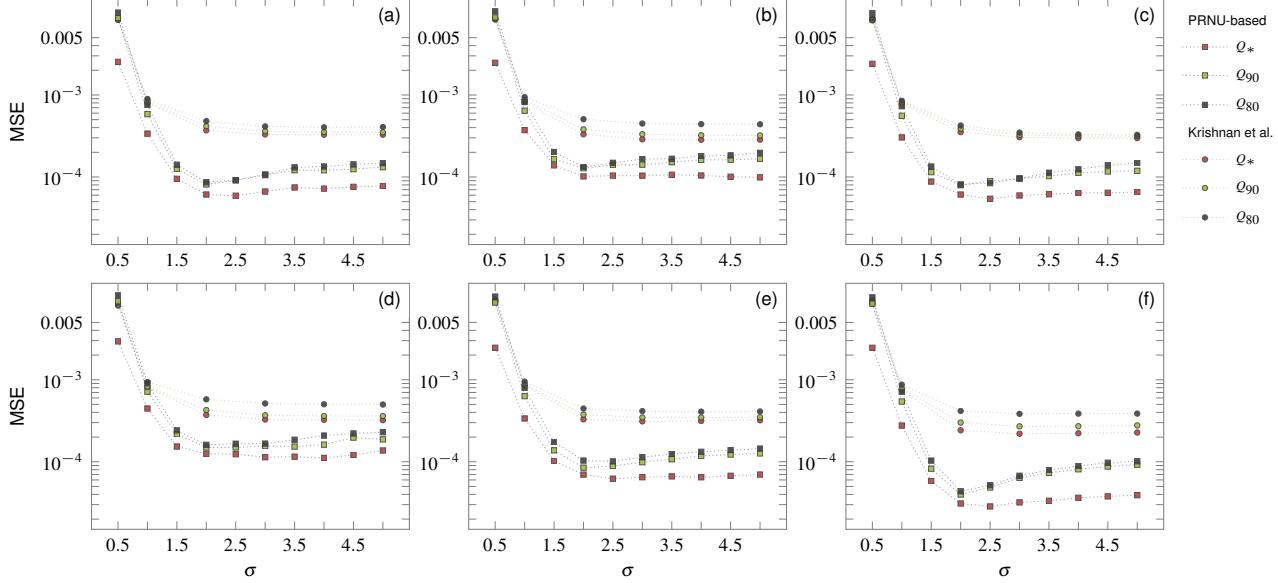
Figure 7. Kernel estimation accuracies for images blurred with a 5×5 Gaussian kernel of strength σ. Kernel size assumed to be known ($S = \hat{S}$). Nikon D70 (a) & (b), Nikon D70s (c) & (d), Nikon D200 (e) & (f). Images stored in uncompressed bitmap format (Q_*), or with JPEG qualities 90 (Q_{90}) and 80 (Q_{80}).
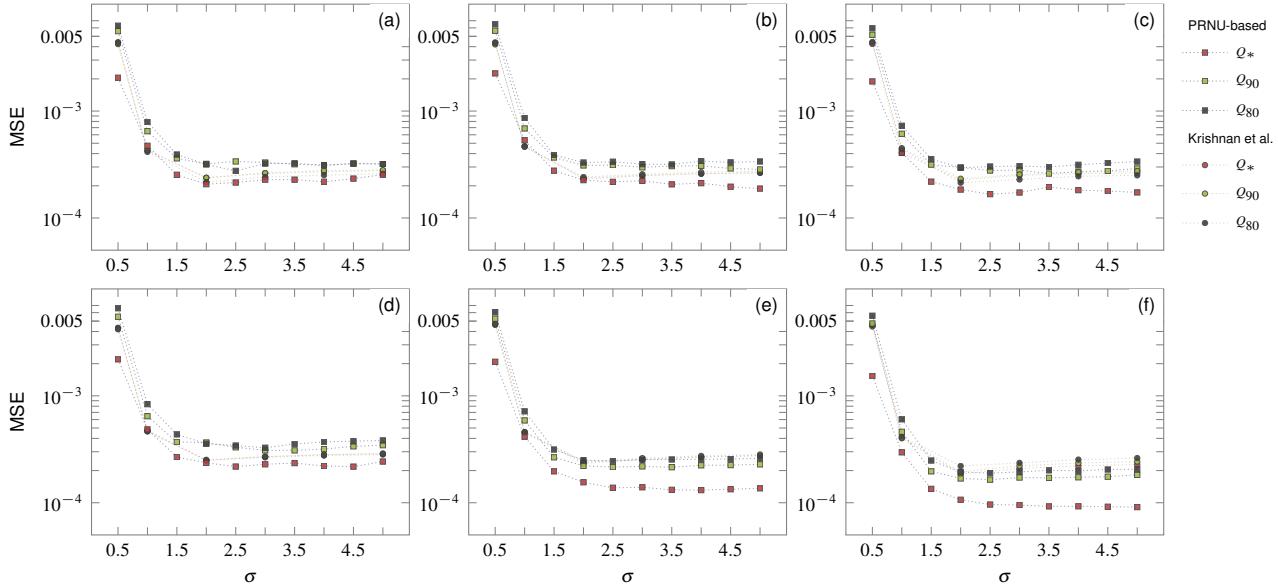
Figure 8. Kernel estimation accuracies for images blurred with a 5×5 Gaussian kernel of strength σ. Kernel size unknown ($\hat{S} = 7$). Nikon D70 (a) & (b), Nikon D70s (c) & (d), Nikon D200 (e) & (f). Images stored in uncompressed bitmap format (Q_*), or with JPEG qualities 90 (Q_{90}) and 80 (Q_{80}).

detector for JPEG images," in *IS&T Electronic Imaging: Media Watermarking, Security, and Forensics*, 2017.

[9] R. Neelamani, R. de Queiroz, Z. Fan, S. Dash, and R. G. Baraniuk, "JPEG compression history estimation for color images," *IEEE Transactions on Image Processing*, vol. 15, no. 6, pp. 1365–1378, 2006.

[10] T. Pevný and J. Fridrich, "Detection of double-compression in JPEG images for applications in steganography," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 2, pp. 247–258, 2008.

[11] M. Stamm and K. Liu, "Forensic detection of image manipu-lation using statistical intrinsic fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 492–506, 2010.

[12] S. Pfennig and M. Kirchner, "Spectral methods to determine the exact scaling factor of resampled digital images," in *International Symposium on Communications Control and Signal Processing, ISCCSP 2012*, 2012.

[13] C. Chen, J. Ni, and Z. Shen, "Effective estimation of image rotation angle using spectral method," *IEEE Signal Processing Letters*, vol. 21, no. 7, pp. 890–894, 2014.

[14] T. R. Goodall, I. Katsavounidis, Z. Li, A. Aaron, and A. C.
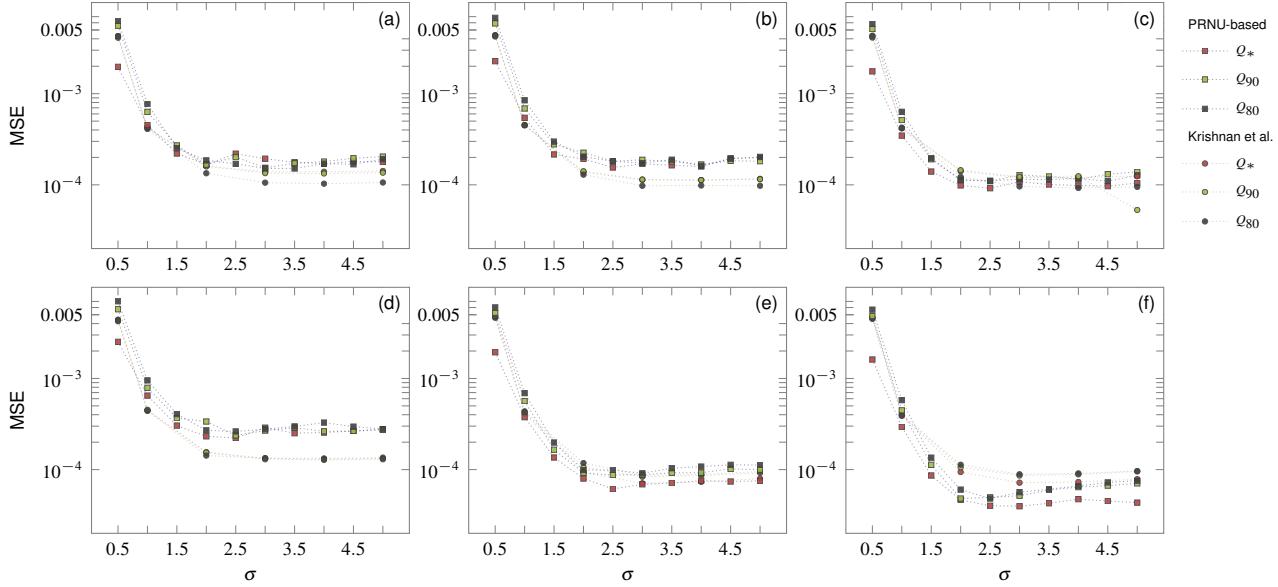
Figure 9. *Kernel estimation accuracies for images blurred with a 7 × 7 Gaussian kernel of strength σ. Kernel size assumed to be known (S = Ŝ). Nikon D70 (a) & (b), Nikon D70s (c) & (d), Nikon D200 (e) & (f). Images stored in uncompressed bitmap format (Q_*), or with JPEG qualities 90 (Q_{90}) and 80 (Q_{80}).*

Bovik, "Blind picture upscaling ratio prediction," *IEEE Signal Processing Letters*, vol. 23, no. 12, pp. 1801–1805, 2016.

[15] M. Kirchner and J. Fridrich, "On detection of median filtering in digital images," in *Media Forensics and Security II*, ser. Proceedings of SPIE, N. D. Memon, J. Dittmann, A. M. Alattar, and E. J. Delp, Eds., vol. 7541, 2010, 754110.

[16] C. Chen, J. Ni, and J. Huang, "Blind detection of median filtering in digital images: A difference domain based approach," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4699–4710, 2013.

[17] A. Swaminathan, M. Wu, and K. Liu, "Digital image forensics via intrinsic fingerprints," *IEEE Transactions on Information Forensics and Security*, vol. 3, no. 1, pp. 101–117, 2008.

[18] W.-H. Chuang, A. Swaminathan, and M. Wu, "Tampering identification using empirical frequency response," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009, pp. 1517–1520.

[19] V. Conotter, P. Comesaña, and F. Pérez-González, "Forensic detection of processing operator chains: Recovering the history of filtered JPEG images," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2257–2269, 2015.

[20] H. Ravi, A. V. Subramanyam, and S. Emmanuel, "Forensic analysis of linear and nonlinear image filtering using quantization noise," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 12, no. 3, 2016.

[21] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.

[22] M. R. Banham and A. K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Magazine*, vol. 14, no. 2, pp. 24–41, 1997.

[23] T. E. Bishop, S. D. Babacan, B. Amizic, A. K. Katsaggelos, T. Chan, and R. Molina, "Blind image deconvolution: Problem formulation and existing approaches," in *Blind image deconvolution: Theory and applications*, P. Campisi and K. Egiazarian, Eds. CRC Press, 2007, ch. 1.

[24] J. Fridrich, "Sensor defects in digital image forensics," in *Digital Image Forensics: There is More to a Picture Than Meets the Eye*, H. T. Sencar and N. Memon, Eds. Springer, 2013, pp. 179–218.

[25] M. Goljan and J. Fridrich, "Camera identification from cropped and scaled images," in *Security, Forensics, Steganography, and Watermarking of Multimedia Contents X*, ser. Proceedings of SPIE, E. J. Delp and P. W. Wong, Eds., vol. 6819, 2008, p. 68190E.

[26] ——, "Sensor-fingerprint based identification of images corrected for lens distortion," in *Media Watermarking, Security, and Forensics 2012*, ser. Proceedings of SPIE, N. Memon, A. M. Alattar, and E. J. Delp, Eds., vol. 8303, 2012, p. 83030H.

[27] T. Gloe, S. Pfennig, and M. Kirchner, "Unexpected artefacts in PRNU-based camera identification: A 'Dresden Image Database' case-study," in *ACM Multimedia and Security Workshop, MM&Sec'12*, 2012, pp. 109–114.

[28] D. Krishnan, T. Tay, and R. Fergus, "Blind deconvolution using a normalized sparsity measure," in *Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 233–240.

[29] M. Goljan, J. Fridrich, and T. Filler, "Large scale test of sensor fingerprint camera identification," in *Media Forensics and Security*, ser. Proceedings of SPIE, E. J. Delp, J. Dittmann, N. D. Memon, and P. W. Wong, Eds., vol. 7254, 2009, 72540I.

[30] F. Marra, G. Poggi, C. Sansone, and L. Verdoliva, "Correlation clustering for PRNU-based blind image source identification," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2016.

[31] G. Chierchia, G. Poggi, C. Sansone, and L. Verdoliva, "A Bayesian-MRF approach for PRNU-based image forgery

detection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 4, pp. 554–567, 2014.

[32] S. Chakraborty and M. Kirchner, "PRNU-based image manipulation localization with discriminative random fields," in *IS&T Electronic Imaging: Media Watermarking, Security, and Forensics*, 2017.

[33] J. Lukáš, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, no. 2, pp. 205–214, 2006.

[34] A. Karaküçük, A. E. Dirik, H. T. Sencar, and N. D. Memon, "Recent advances in counter PRNU based source attribution and beyond," in *Media Watermarking, Security, and Forensics*, ser. Proceedings of SPIE, A. M. Alattar, N. D. Memon, and C. D. Heitzenrater, Eds., vol. 9409, 2015, 94090N.

[35] T. Gloe and R. Böhme, "The Dresden Image Database for benchmarking digital image forensics," *Journal of Digital Forensic Practice*, vol. 3, no. 2–4, pp. 150–159, 2010.

[36] S. Bayram, H. T. Sencar, and N. D. Memon, "Seam-carving based anonymization against image & video source attribution," in *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2013, pp. 272–277.

[37] J. Entrieri and M. Kirchner, "Patch-based desynchronization of digital camera sensor fingerprints," in *IS&T Electronic Imaging: Media Watermarking, Security, and Forensics*, 2016.

[38] S. Taspinar, M. Mohanty, and N. Memon, "Source camera attribution using stabilized video," in *IEEE International Workshop on Information Forensics and Security*, 2016.