




# Source camera identification via low dimensional PRNU features

Yihua Zhao<sup>1</sup> · Ning Zheng<sup>1,2</sup> · Tong Qiao<sup>2,3</sup>  · Ming Xu<sup>2</sup>

Received: 13 June 2018 / Revised: 3 September 2018 / Accepted: 23 October 2018 /

Published online: 31 October 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

Identifying the source of digital images is the key task in the community of image forensics. Sensor pattern noise dominantly serves as an intrinsic fingerprint or feature for dealing with the problem of source camera identification. However, how to decrease the dimensionality of the pattern noise while guaranteeing the detection power remains a hot topic. The goal of this paper is to investigate the problem of source camera identification for natural images in JPEG format. By considering the image texture, we propose to design a new classifier with adopting a weight function, leading to the remarkable reduction of the feature dimensionality. In the extensive experiments, it is verified that our proposed algorithm performs comparably with the prior art. Besides, the robustness of the proposed classifier is also evaluated when the query images are attacked by post-processing techniques such as JPEG compression, noise adding, noise removing and image cropping.

**Keywords** Image origin identification · Sensor pattern noise · Photo-response non-uniformity (PRNU) · Weight function

## 1 Introduction

In the past two decades, the study of multimedia security has received remarkable attention (see [17, 18, 23, 31] for instance). Typically, digital image forensics has played an important role in this field, which can be arbitrarily divided into the following two categories: origin identification and tampering detection. As Fig. 1 illustrates, process history recovery refers to as detection of non-malicious operation, e.g., lossy compression, filtering, resizing (see [25] for instance). Anomaly detection is applied to test whether the query image suffers a malicious attack, such as copy-move or splicing (see details in [1, 20]).

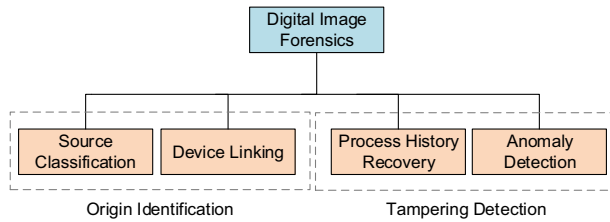
---

✉ Tong Qiao  
tong.qiao@hdu.edu.cn

<sup>1</sup> School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, China

<sup>2</sup> School of Cyberspace, Hangzhou Dianzi University, Hangzhou, China

<sup>3</sup> Zhengzhou Science and Technology Institute, Zhengzhou, China



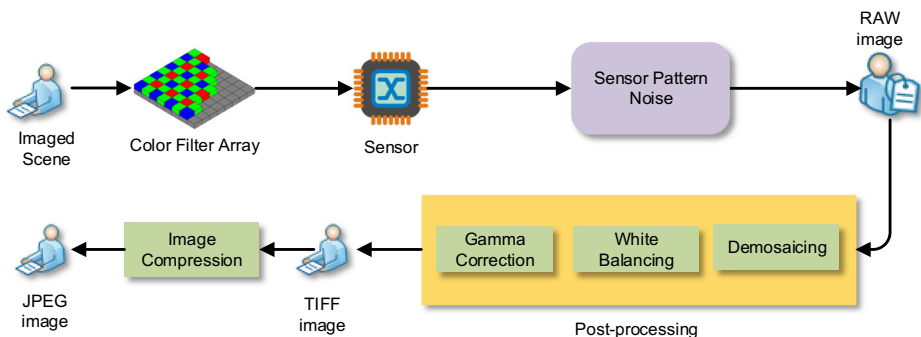
**Fig. 1** Classification of digital image forensics

With the rise of digital photography, an increased number of digital images have become associated with evidentiary pools for criminal and civil proceeding. Criminal activities like child pornography, fraud and terrorism are proliferating by misusing such digital contents. Thus, image origin identification is an important branch of digital image forensics, serving as a forensic tool to authenticate a digital image posted by an unknown account possibly used for illegal purpose. However, current camera fingerprints, used for identification, with the high dimensional features indeed become the obstacle of designing an efficient and reliable detector. Therefore, let us study an effective classifier via low dimensional features for identifying image origin.

## 1.1 State of the art

First of all, an intuitive idea is to use the metadata as the feature (e.g., EXchangeable Image File, JPEG headers) of the digital image, which contains information of the digital image, such as camera type and imaging conditions. However, due to the popularity of low-cost photo-editing software (e.g., Adobe Photoshop and IrfanView), it is easy to manipulate and remove metadata of digital image. Besides, a plenty of images shared on the social networks (e.g., Facebook, Twitter or WeChat) probably lose their header files. Therefore, metadata are usually considered as unreliable fingerprint for forensic purpose.

The main idea behind the method of image origin identification is that each stage of the acquisition process leaves a unique fingerprint within the image content itself due to intrinsic imperfections in the procedure of image acquisition. The image acquisition process is illustrated in Fig. 2 (see details in [19, 24, 26]). First of all, the photographed scene passes through the camera lens and a color filter array (CFA), to reach into the sensor. In particular, the CFA is defined as a mosaic, which only allows one color channel (red, or green, or blue) of each pixel to be recorded; and the remaining two color channels are filled (or



**Fig. 2** Illustration of a typical imaging pipeline within a digital camera

demosai cked) with a typical color interpolation algorithm. Afterwards, the acquired signal is further processed using white balancing adjustment and gamma correction, which effectively strengthens the visualization of images and generates a natural image (TIFF image for instance). Generally, a digital image can be saved as the compression format, such as JPEG compression (see [21]). In the image acquisition process, some traces are left in the image by different hardwares or post-processing algorithms, which can be used as unique features to link the image to its source device (see [1, 16]). The methods of image origin identification are mainly aimed at solving the following problems:

- Using the intrinsic fingerprint for image origin identification with reliability and accuracy.
- Extracting the intrinsic fingerprint from a single image (or several images) that may contain scene details or nuisance noise.
- Determining whether the query image is acquired from a typical device.

Currently, in order to link an image to its source device, many approaches have been proposed. The Photo-Response Non-Uniformity (PRNU) of both CCD and CMOS sensor for camera (see [10, 13]), caused by slight variations among pixels in their sensitivity to light, nearly dominates the research of source camera identification. In [16], the PRNU noise as the primary part (sensor intrinsic characteristics) of the Sensor Pattern Noise (SPN) was first proposed to deal with the problem of image origin identification. Authors of [16] estimated the PRNU by averaging the residual noise of multiple images from the same camera using a denoising filter, such as wavelet denoising filter. Thereafter, the pioneer work was improved in [1]. A unified framework for both image origin identification and image tampering detection could be established by using PRNU. Besides, the method proposed in [1] used less number of samples (images) to estimate the PRNU, in comparison with the number reported in [16]. Furthermore, the approaches presented in [1, 16] was validated under a large-scale image database in [7]. However, the dimensionality of the SPN (or PRNU) can be up to millions w.r.t. the same dimension as the image size. The large dimensionality of the feature might generate more information describing the camera sensor while more nuisance noise (generated when acquiring a digital image) probably interferes the accuracy of the sensor pattern, which is definitely a trade-off problem. Thus, it is proposed to design an effective and reliable classifier via a low dimensional features, that is the main problem we address in this context.

Recently, a novel SPN [3] unlike PRNU was proposed to describe a RAW image for source camera identification (see [27]), where Poisson-Gaussian-distributed noise, characterizing the features of each camera model, being used to design the binary classifier under the framework of hypothesis testing theory. Thereafter, the detector of [22] was proposed to address the problem of camera device classification (classifying different instances with the same camera model). Besides, the novel SPN-based classifiers can be extended into the design of classifying images based on JPEG format images [24, 28]. Generally, given that series of statistical detectors [22, 24, 27, 28] can be successfully established, the upper bound of the detection at the prescribed false alarm probability can be theoretically described as well as their theoretical and empirical performance. However, the feature is easily interfered by the content of tested images, leading to the inaccuracy of camera model description.

Meanwhile, many supervised detectors have been designed. Authors of [2] described a new feature generation method for identifying image origin, which addresses the open set classification problem by using a supervised learning method (support vector machine, SVM), as well as adjusting the decision boundary of the classifier to improve classification

accuracy. Additionally, authors of [29] presented a device identification method via image texture features, extracted from well-selected color channel model. Similarly, the SVM classifier was trained to identify source camera. In recent study, under the theoretical framework of deep learning, authors of [30] proposed a robust multi-classifier based on an improved architecture of Convolution Neural Network (CNN), which is mainly used in camera model identification. However, it could not perform very well with identifying source camera instance. In the practical detection, several problems such as the robustness to training and test set mismatch remain open. Furthermore, it is nearly impossible that forensic analysts can collect a large scale of data (such as labeled images) for training. In addition, the supervised detector need a large scale of fingerprints, referring to the high dimensional features, possibly leading to the inefficiency of classification.

In general, the study topic of reducing the dimensionality of camera fingerprints (or fingerprint digest / reduction) always remains hot. To our knowledge, the Approximate Rank Matching Search (ARMS) proposed in [8] was one of the early fast search algorithms to deal with the problem of fingerprint reduction. For the full-size fingerprint of each image, the authors define a subset of  $k (\ll n)$  largest values as the digest fingerprint, where  $n$  is the size of each image. However, the ARMS was sensitive to noise interference. Additionally, the authors of [8] only gave the detection results under the strong assumption that both query and reference fingerprints were good quality, but not contaminated by nuisance noise. In fact, the quality of query fingerprints cannot be guaranteed in the practical detection due to different image sources and device-dependent properties.

Following the prior study [8], more methods are proposed to reduce the computational complexity by compressing the fingerprint. In [6, 12], the authors introduced a novel fingerprint reduction method, which preserved the largest elements and the corresponding locations. In [6], the authors proposed a method that only used the outlier values in the fingerprints. Therefore, the approach obtained a much better trade-off result between computational complexity and detection reliability. Unlike [6], the authors of [12] adopted a lookup table that was built on the separate-chaining hash table to speed up the search process. The proposed Search Priority Array method indeed improved the fast search algorithm while enhancing the robustness. In recent study (see [15] for instance), a framework was proposed for denoising and compressing full-sized fingerprint, where more discriminant features were extracted by using principal component analysis and linear discriminant analysis. However, the supervised mechanism adopted in [15] limits its efficiency of classification. Supposing that a new reference pattern extracted from digital images is obtained, the classifier needs to be re-trained for up-dated classification.

## 1.2 Contributions of the paper

Due to the effectiveness of the widely-used PRNU feature, it is proposed to design our PRNU-based classifier for image origin identification in this context. Different textures of the image leads to different weights of the extracted feature matrix, which is the primary motivation of our proposed classifier with low dimensional PRNU features. It should be noted that we adopt a weight function to extract “clean” PRNU<sup>1</sup> with high discriminant power. The main contributions of this paper are summarised as follows:

- The weight matrix of the image is acquired by using a designed weight function, and then the dimensionality reduced fingerprints with more discriminative features are

<sup>1</sup>“clean” PRNU refers to the PRNU obtained after feature dimensionality reduction in Section 3.

extracted, leading to reducing computational complexity and improving classification efficiency.

- Based on the “clean” PRNU, we propose to design an effective classifier dealing with the problem of source camera identification, which can be applied to two practical forensic scenarios.
- Extensive results verify that our proposed classifier can not only detect images captured by different camera models, but also different instances from the same model.
- When suffering attacks such as JPEG compression, noise adding and noise removing, our proposed classifier still performs relevantly.

### 1.3 Organization of the paper

The rest of the paper is organized as follows. Section 2 mainly describes the PRNU noise serving for source camera identification. In Section 3, we propose a weight function used for designing a novel method of reducing feature dimensionality, and then our classifier is given by considering two practical scenarios. The experimental results are presented, and the effectiveness of the proposed method is verified in Section 4. Finally, Section 5 concludes this paper.

## 2 Description of PRNU noise pattern

In virtue of [1] based on PRNU model [9], a natural image acquired by a digital camera can be formulated as follows<sup>2</sup>:

$$\mathbf{Z} = \mathbf{Z}^0 + \mathbf{Z}^0 \mathbf{P}_f + \Xi, \quad (1)$$

in which  $\mathbf{Z}$  denotes a camera output image; and meanwhile a “true scene” image that would be captured in the absence of any imperfections, denoted as  $\mathbf{Z}^0$ ; and all elements in the multiplicative factor (or PRNU factor, also defined as camera reference pattern in this context)  $\mathbf{P}_f$  are typically close to 0. In particular, the PRNU noise, referring to as  $\mathbf{Z}^0 \mathbf{P}_f$ , is unique for each camera sensor and it is generated due to the imperfections of the device production. In addition,  $\Xi$ , referring to as nuisance noise, includes all other components, such as shot noise, read-out noise, dark current, and quantization noise (see [10, 13]). Note that in the stage of PRNU noise extraction, we have to exclude the interference from  $\Xi$ .

Usually, residual noise is extracted from an image via denoising filtering operation, and the PRNU is obtained by averaging multiple  $\mathbf{R}_n$  obtained from images generated by the same digital camera. The specific calculation is as follows:

$$\mathbf{R}_n = \mathbf{Z}_n - F(\mathbf{Z}_n), \quad n = \{1, \dots, N_t\}, \quad (2)$$

where  $\mathbf{Z}_n$  denotes  $n$ -th image acquired by a camera source  $\mathcal{S}$ .  $N_t$  represents the total number of the entire dataset. The wavelet denoising filter  $F(\cdot)$  is used to filter out the scene details of the image  $\mathbf{Z}_n$ . Then residual noise  $\mathbf{R}_n$  of each acquired image can be obtained. Subsequently, the PRNU factor can be extracted by averaging on a specified number (for instance,  $N_t = 50$  in our experiments) of images from the same camera:

$$\widehat{\mathbf{P}}_f = \frac{\sum_{n=1}^{N_t} \mathbf{R}_n \mathbf{Z}_n}{\sum_{n=1}^{N_t} (\mathbf{Z}_n)^2}. \quad (3)$$

Finally, we obtain the camera reference pattern  $\widehat{\mathbf{P}}_f$  to identify image origin.

<sup>2</sup>All matrix operations are element-wise.

In general, it is proposed to use the similarity between both of noises, which are respectively generated by the query image and the reference camera, that is the residual noise  $\mathbf{R}$  of the query image and reference noise pattern  $\widehat{\mathbf{ZP}}_f$ . Let us adopt the maximum of the normalized cross correlation  $corr$  (see [14]) as the classification metric, which can be formulated by:

$$corr(\widehat{\mathbf{ZP}}_f, \mathbf{R}) = \frac{(\widehat{\mathbf{ZP}}_f - \overline{\widehat{\mathbf{ZP}}_f}) \odot (\mathbf{R} - \overline{\mathbf{R}})}{\|\widehat{\mathbf{ZP}}_f - \overline{\widehat{\mathbf{ZP}}_f}\| \cdot \|\mathbf{R} - \overline{\mathbf{R}}\|}, \quad (4)$$

where  $\odot$  represents the dot multiplication;  $\|\cdot\|$  is norm. Note that  $\widehat{\mathbf{P}}_f$  and  $\overline{\mathbf{R}}$  can be obtained by respectively averaging all the pixel-wise values of matrices  $\widehat{\mathbf{P}}_f$  and  $\mathbf{R}$ .

Furthermore, as a more stable and accurate statistic, Peak to Correlation Energy (PCE) ratio can be used as an alternative to directly normalize correlation (see details in [5, 7]). Let us denote the coordinates of the peak value, referring to as the maximum value calculated by using (4), as  $l_{peak}$  among all candidate coordinate  $l$ . The PCE ratio can be expressed as:

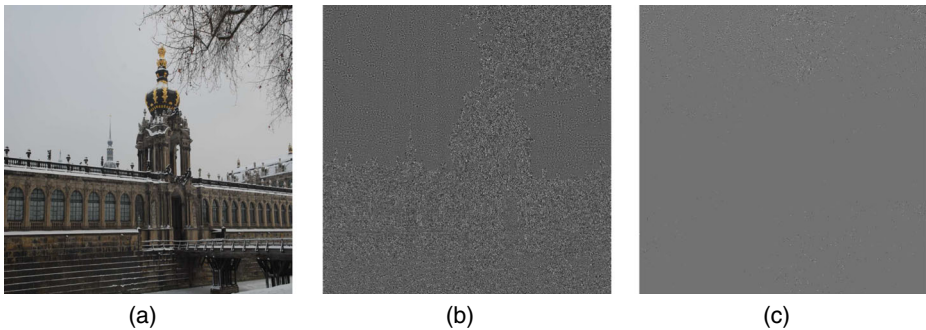
$$\mathcal{R}_{PCE} = \frac{corr(l_{peak}; \widehat{\mathbf{ZP}}_f, \mathbf{R})^2}{\frac{1}{mn - \|\mathcal{N}\|} \sum_{l, l \notin \mathcal{N}} corr(l; \widehat{\mathbf{ZP}}_f, \mathbf{R})^2} \quad (5)$$

where  $m \times n$  denotes the dimension of the query image, and  $\mathcal{N}$  is a small neighborhood region around the position of the peak value.

### 3 Our proposed methodology

Generally, in the stage of classification, the designed classifier with the high dimensional PRNU unavoidably causes a high computational cost. Moreover, the PRNU tends to contain more or less redundancy and interfering components, leading to kind of mismatch of the final detection. Therefore, in this practical context, it is proposed to remove those nuisance noises while reducing the fingerprint dimensionality for our establishment of the source camera classifier. Here, we adopt a simple model to represent the original PRNU noise  $\mathbf{X}$  (see Fig. 3b), which can be described as:

$$\mathbf{X} = \mathbf{X}^0 + \xi, \quad (6)$$



**Fig. 3** **a** A digital still image captured by Nikon D200. **b** The original PRNU noise extracted from (a) using the wavelet de-noising filter. **c** The “clean” PRNU noise based on the proposed feature dimensionality reduction method with  $k = 5\%$ . For better visualization, (b) and (c) represent contrast-adjusted and normalized results

where  $\mathbf{X}^0$  denotes the “clean” PRNU noise;  $\xi$  is an additive mixture of unwanted interferences, which may include scene details and the artifacts introduced during color interpolation, JPEG compression and other image processing operations (see [1]). Scene details is specific to the scene when shooting; the artifacts can be shared among cameras of the same camera model. Hence, they are non-unique, less discriminant and redundant. We need remove those nuisance noises as much as possible.

The intuitive idea behind nuisance noise removing is to adopt low-dimensional PRNU features, only containing  $\mathbf{X}^0$ , not involving  $\xi$ . Therefore, we propose a feature dimensionality reduction algorithm using a weight function, in which the non-deterministic (less important) pixel-wise reference pattern is excluded while more important one is retained. That drives our designed classifier of source camera identification more faster, and meanwhile maintaining a comparable performance to the prior art.

### 3.1 Design of weight function

The core idea behind the weight function is that the high texture region of an image contains nuisance noise, leading to the contamination of the PRNU noise. Thus, we design a weight function of selecting all the useful smooth region of the image. Inspired by the algorithm proposed in [11], we investigate the texture of the image relying on scalable cost of the whole image caused by each pixel-wise flipping.

Given an image  $\mathbf{Z} = \{z_{ij}\}$  acquired by the reference camera  $\mathcal{S}$ , we can estimate each weight value  $\omega$  of the image by using the weight function  $D(\mathbf{Z}, \mathbf{Z}^{flip})$ . In our assumption, the cost caused by each pixel flipping can map to the weight of the corresponding position.  $\mathbf{Z}^{flip}$  represents an image  $\mathbf{Z}$  by flipping pixel-wise value in spatial domain. Straightforwardly, the weight function can be expressed by:

$$D(\mathbf{Z}, \mathbf{Z}^{flip}) \triangleq \sum_{t=1}^3 \frac{|W_{(\mathbf{Z})}^t - W_{(\mathbf{Z}^{flip})}^t|}{\sigma + |W_{(\mathbf{Z})}^t|}, \quad (7)$$

where  $W_{(\mathbf{Z})}^t$  and  $W_{(\mathbf{Z}^{flip})}^t$ ,  $t \in \{1, 2, 3\}$ , respectively denote wavelet coefficients in the  $t$ -th subband of the first decomposition level. In addition,  $\sigma > 0$  is a constant stabilizing the numerical calculations.

Immediately, let us calculate the cost of the  $\mathbf{Z}_{ij}$ , changing  $\mathbf{Z}_{ij}$  to  $\mathbf{Z}_{ij}^{flip}$  while leaving all other pixel-wise elements of  $\mathbf{Z}$  unchanged, formulated by:

$$\omega_{ij}(\mathbf{Z}, \mathbf{Z}_{ij}^{flip}) \triangleq D(\mathbf{Z}, \mathbf{Z}_{\sim ij} \mathbf{Z}_{ij}^{flip}), \quad (8)$$

where  $\mathbf{Z}_{\sim ij} \mathbf{Z}_{ij}^{flip}$ ,  $i \in \{1, \dots, I\}$ ,  $j \in \{1, \dots, J\}$ , denote the image  $\mathbf{Z}$  with only its  $ij$ -th element flipping.

Finally, we obtain the weight value  $\omega$  over the whole image  $\mathbf{Z}$ . The bigger  $\omega$  is, the greater the cost of the pixel caused by flipping is, meaning that clustering pixels with small  $\omega$  consist a flat region (relatively smooth); high texture region can be represented using the pixels with large  $\omega$ . Thus, the weight function can guide us to design the algorithm of dimensionality reduction.

### 3.2 Feature dimensionality reduction

Based on the weight function  $\omega$  proposed in Section 3.1, we can further extract more representative “clean” PRNU with low dimensionality. For instance, given the original PRNU

noise  $\mathbf{X}$ , we prescribe its largest elements  $K \ll I \times J$ , and set the remaining elements in  $\mathbf{X}$  to zero. That operation preserves the region with relatively flat field which can effectively describe the fingerprint of the source camera. Also, we decrease the computation cost. For each matrix  $\mathbf{X}$  extracted from an image, we need merely select  $K$  elements, instead of considering all elements  $I \times J$ . The expression is formulated by:

$$\omega_{ij}^d = \begin{cases} \omega_{ij} & \text{if } \omega_{ij} \text{ is one of the largest } d \text{ elements,} \\ 0 & \text{if otherwise.} \end{cases} \quad (9)$$

where  $\Omega^d = \{\omega_{ij}^d\}$ ,  $i \in \{1, \dots, I\}$ ,  $j \in \{1, \dots, J\}$ . Then we need estimate a weight matrix  $\Omega^k$  from images acquired by the reference camera. Let us select the first  $K$  representative elements over the weight matrix  $\Omega^k$ , and label the corresponding locations. Then we obtain the “clean” PRNU  $\mathbf{X}^0$  (see Fig. 3c), written as:

$$\mathbf{X}^0 = \mathbf{X}(\text{index}(\Omega^k)) = \mathbf{X} \left( \text{index} \left( \bigcap_{n=1}^{N_t} \Omega_n^d \right) \right), \quad (10)$$

where  $\text{index}(\cdot)$  represents the locations of  $K$  elements in the matrix; the  $N_t$  represents the number of the images used to extract the reference pattern of camera  $\mathcal{S}$ . It should be noted that the ratio  $k$  controls the number of the used elements, which can be formulated as:

$$k = \frac{K}{I \times J}. \quad (11)$$

where “clean” PRNU  $\mathbf{X}^0$  degrades to original PRNU  $\mathbf{X}$  when  $k = 100\%$ .

Note that when we extract PRNU features, original images captured by a camera need to be available. Because the original image retains the most comprehensive information of camera fingerprint. For instance, once the image is post-processed, the information of fingerprint will be possibly removed or contaminated.

### 3.3 Establishment of our proposed classifier

Let us denote  $\mathbf{Z}$  as a query image.  $\mathcal{S}_0$  and  $\mathcal{S}_1$  represent two camera sources to be identified. In this scenario, forensic analysts make a choice between the two following hypotheses:

$$\begin{cases} \mathcal{H}_0 : \mathbf{Z} & \text{acquired by the source } \mathcal{S}_0 \\ \mathcal{H}_1 : \mathbf{Z} & \text{acquired by the source } \mathcal{S}_1 \text{ different from } \mathcal{S}_0. \end{cases}$$

Note that two different camera sources  $\mathcal{S}_0$  and  $\mathcal{S}_1$  might refer to as models / instances. In this content, it is proposed to establish two types of practical detectors for source camera identification. Supposing that the fingerprints extracted from  $\mathcal{S}_0$  and  $\mathcal{S}_1$  are both known, then our proposed classification addresses the problem of identifying whether the query image  $\mathbf{Z}$  is captured by camera source  $\mathcal{S}_0$  or by camera source  $\mathcal{S}_1$ . Let us define it as close-set classification. In another scenario, where only the fingerprints from  $\mathcal{S}_0$  are known (without knowing the other possible source  $\mathcal{S}_1$ ), the proposed classification detects whether the query image  $\mathbf{Z}$  has been captured by camera source  $\mathcal{S}_0$ . Let us define it as open-set classification.

To determine the results of the classification, it is necessary to set a threshold  $\tau$  in advance, and compare it with the obtained  $\mathcal{R}_{PCE}$  to identify the source camera. Next, we give two types of classifiers to deal with our proposed two scenarios.



The first scenario (close set) can be described as identifying source camera ( $\mathcal{S}_0$  or  $\mathcal{S}_1$ ) of the query image  $\mathbf{Z}$ :

$$\begin{cases} \mathbf{Z} \in \mathcal{S}_0 & \text{if } \mathcal{R}_{PCE}(\mathbf{Z}, \mathbf{X}_{\mathcal{S}_0}^0) \geq \tau \\ \mathbf{Z} \in \mathcal{S}_1 & \text{if } \mathcal{R}_{PCE}(\mathbf{Z}, \mathbf{X}_{\mathcal{S}_0}^0) < \tau \end{cases} \quad (12)$$

It means that when  $\mathbf{Z} \in \mathcal{S}_0$ ,  $\mathcal{H}_0$  is true and  $\mathcal{H}_0$  is accepted; when  $\mathbf{Z} \in \mathcal{S}_1$ ,  $\mathcal{H}_1$  is true and  $\mathcal{H}_1$  is accepted.

The second one (open set) can be described as determining whether the query image  $\mathbf{Z}$  is taken by camera  $\mathcal{S}_0$ :

$$\begin{cases} \mathbf{Z} \in \mathcal{S}_0 & \text{if } \mathcal{R}_{PCE}(\mathbf{Z}, \mathbf{X}_{\mathcal{S}_0}^0) \geq \tau \\ \mathbf{Z} \notin \mathcal{S}_0 & \text{if } \mathcal{R}_{PCE}(\mathbf{Z}, \mathbf{X}_{\mathcal{S}_0}^0) < \tau, \end{cases} \quad (13)$$

where  $\mathbf{Z}$  might come from an unknown camera. When  $\mathbf{Z} \in \mathcal{S}_0$ ,  $\mathcal{H}_0$  is true and  $\mathcal{H}_0$  is accepted; when  $\mathbf{Z} \notin \mathcal{S}_0$ ,  $\mathcal{H}_0$  is rejected. In other words,  $\mathbf{Z} \in \mathcal{S}_0$  denotes the query image  $\mathbf{Z}$  taken by camera  $\mathcal{S}_0$ ;  $\mathbf{Z} \notin \mathcal{S}_0$  means  $\mathbf{Z}$  is not taken by  $\mathcal{S}_0$ .  $\mathcal{R}_{PCE}(\mathbf{Z}, \mathbf{X}_{\mathcal{S}_0}^0)$  represents PCE value. In the following simulation, we propose evaluation metrics to demonstrate the performance of our proposed classifier.

## 4 Experiments

In this section, we conduct all the experiments on the Dresden image database (see [4]). In Section 4.1, we describe the experimental setups, which contain the process of obtaining “clean” PRNU  $\mathbf{X}^0$  and the image database used in the experiment. In Section 4.2, we experimentally evaluate the performance of our proposed classifier via feature dimensionality reduction method. In Section 4.3, we study the robustness of our proposed algorithm, with considering JPEG compression, noise adding, noise removing, and image cropping attacks. And, we use the overall evaluation metrics, including accuracy  $\mathcal{V}_A$ , recall  $\mathcal{V}_R$ , precision  $\mathcal{V}_P$  and F1-score  $\mathcal{V}_F$ .

### 4.1 Experiment setups

We use a weight function  $\omega$ , which has been described in detail in Section 3.1, and obtain the weight matrix  $\Omega^k$ . Then, we further deal with the estimated PRNU  $\mathbf{X}$  through weight matrix  $\Omega^k$ . The aim of that step is to remove the nuisance noise. Finally, we get “clean” PRNU  $\mathbf{X}^0$  with different dimensionality, which is used as the reference pattern in the design of our classifier. It is important to point out that the reference pattern for each model / instance is extracted from  $N_t = 50$  images.

Next, we select a total of 3364 JPEG images with quality factors ranging from 93 to 98. The benchmark Dresden database [4] is used, stemming from 11 models and 20 unique digital camera instances, as shown in Table 1. These images contain scene with different contents as daylight outdoor scenes or indoor with poor illumination, and will then be used for various experiments in this research. In particular, in order to exclude the interference and the inconsistency of cameras resolution, we crop the central of images to the same size in each experiment. In addition, we carry out our experiments in two scenarios, including open set and close set. The open set indicates that query images in the testing set might not be captured by the camera with reference pattern; meanwhile, the close set indicates that all the query images in the testing set can be matched to the corresponding camera with reference

**Table 1** List of digital cameras and their statistic

Camera model	#Devices	Resolution	Image format
Canon Ixus55	1	2592 × 1944	JPEG
Canon Ixus70	2	3072 × 2304	JPEG
Kodak M1063	2	3664 × 2748	JPEG
Nikon CoolPixS710	3	4352 × 3264	JPEG
Nikon D200	2	3872 × 2592	JPEG
Panasonic DMC-FZ50	1	3648 × 2736	JPEG
Pentax OptioA40	2	4000 × 3000	JPEG
Pentax OptioW60	1	3648 × 2736	JPEG
Samsung L74wide	3	3072 × 2304	JPEG
Samsung NV15	2	3648 × 2736	JPEG
Sony DSC-H50	1	3456 × 2592	JPEG

pattern. For example, in Section 4.2.3, the source model identification is carried out in the open set, and the source instance identification is carried out in the close set.

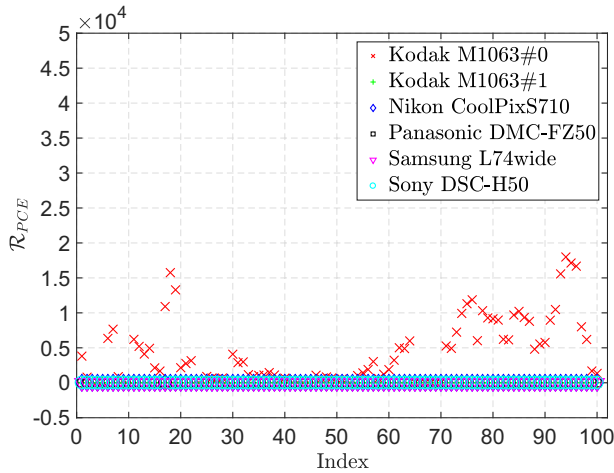
## 4.2 Evaluation of image origin identification

In this section, we use a total of 1864 images from 10 individual cameras for image origin identification based on PRNU noise. In the case of an open set, we describe the distribution of the PCE value  $\mathcal{R}_{PCE}$ . Then, we plot the scatter diagram of  $\mathcal{R}_{PCE}$  to verify that the weight matrix  $\Omega^k$  can be used for feature (or PRNU) dimensionality reduction. Finally, in the case of the open and close set, we use ROC curves to show the performance of the proposed classifier. Besides, it is proposed to compare the results of our classifier with that of the prior-art method [1].

### 4.2.1 Effectiveness of PCE from original PRNU-based classifier

First of all, the goal of this experiment is to investigate the distribution of PCE value  $\mathcal{R}_{PCE}$  in the open set scenario, which serves as similarity evaluation measurement. The reference pattern can be effectively estimated in virtue of (3) in Section 2. Let us respectively test 100 images  $\mathbf{Z}$  from 6 different camera devices. The results are reported in Fig. 4. All images are cropped to the resolution of 2592 × 2304 with JPEG format.  $\mathcal{R}_{PCE}$  (red cross label) represents a query image in the positive class ( $\mathbf{Z} \in \mathcal{S}_0$ ), meaning that the query image is captured by the reference camera  $\mathcal{S}_0$ ;  $\mathcal{R}_{PCE}$  (other labels) represents a query image in the negative class ( $\mathbf{Z} \notin \mathcal{S}_0$ ). In our experiment,  $\mathcal{S}_0$  refers to as Kodak M1063#0.

From the results shown in Fig. 4, most of the  $\mathcal{R}_{PCE}$  with images captured by Kodak M1063#0 camera are larger, while the  $\mathcal{R}_{PCE}$  with Kodak M1063#1, Nikon CoolPixS710, Panasonic DMC-FZ50, Samsung L74wide, and Sony DSC-H50 are close to zero, as we expect. In other words, the query images are correctly associated with the Kodak M1063#0 camera in Fig. 4. As a result, a good performance is obtained by testing whether the query image is captured by the reference camera. To sum up, we can conclude that the prior PRNU-based method can effectively identify the source devices of query images, no matter dealing with the problem of identifying different models or different instances. However, the feature dimensionality of the prior-art algorithm with considerable computational complexity is relevantly high.

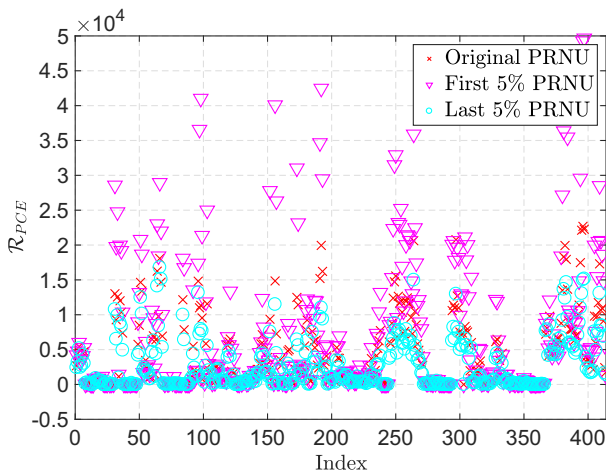


**Fig. 4** Distribution of  $\mathcal{R}_{PCE}$ , in which reference pattern from Kodak M1063#0, and residual noise from  $6 \times 100$  query images acquired by 6 camera devices

#### 4.2.2 Comparison of PCE between original and dimensionality reduced PRNU classifier

An efficient classifier plays an important role when a large scale of images are required to be identified. In order to verify the feasibility of our algorithm proposed in Section 3, let us compare our algorithm with the prior-art algorithm. All images are cropped to the resolution of  $2744 \times 2744$  (approximately 7.53 million pixel-wise) with JPEG format in the experiment. First of all, we estimate the original PRNU  $\mathbf{X}$  on the basis of (3). Then, we obtain the weight matrix  $\Omega^k$ . Finally, we use (10) to obtain the “clean” PRNU  $\mathbf{X}^0$  with low dimensionality.

As Fig. 5 illustrates, most of the  $\mathcal{R}_{PCE}$  (magenta downward-point triangle label) of our algorithm with  $k = 5\%$  (approximately 0.38 million pixel-wise) are significantly higher than the  $\mathcal{R}_{PCE}$  (red cross label) of the prior PRNU-based method. Meanwhile, the “clean” PRNU



**Fig. 5** Distribution of  $\mathcal{R}_{PCE}$  between original and dimensionality reduced PRNU classifier, in which reference pattern from Kodak M1063#0, and residual noise from 414 query images captured by Kodak M1063#0

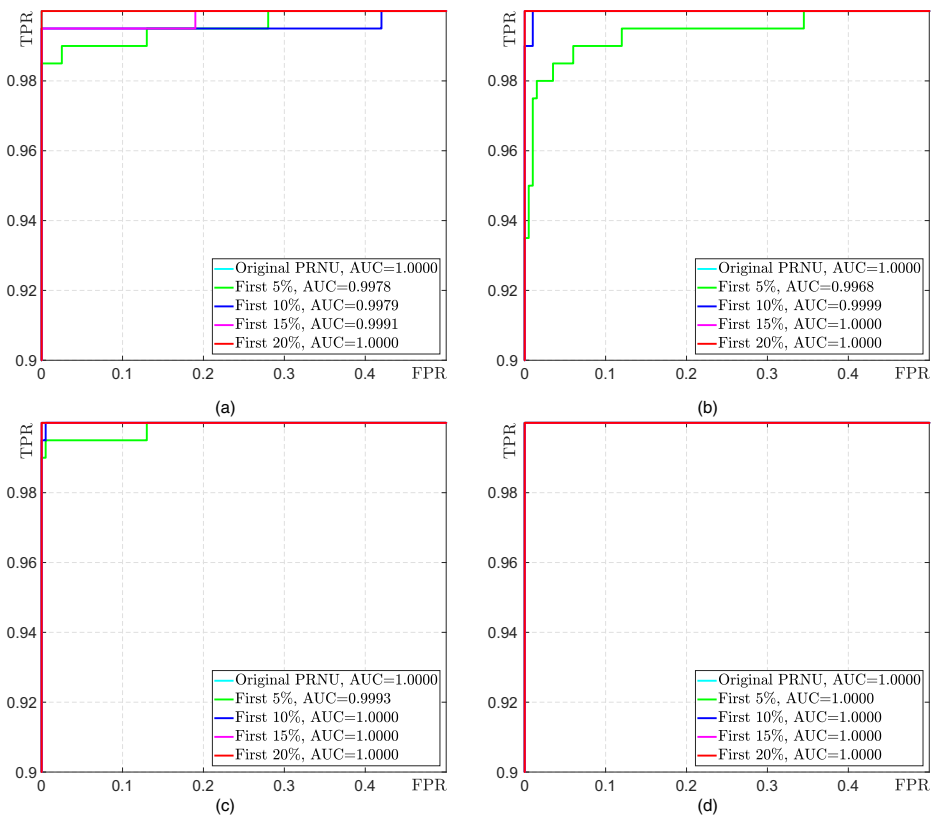
$\mathbf{X}^0$  with  $k = 5\%$  retains deterministic pixel-wise reference pattern that can be used for image origin identification. Besides, the  $\mathcal{R}_{PCE}$  (cyan circle label) of the last 5% PRNU are nearly close to zero, since the last 5% PRNU contains the non-deterministic pixel-wise reference pattern. Obviously, our proposed algorithm with low dimensional PRNU features can perform very well, comparable to that of the prior-art algorithm with original PRNU features.

### 4.2.3 Overall evaluation of our proposed classifier

In this section, we further verify our proposed classifier with low dimensional PRNU in two scenarios, including open set and close set. For instance, images from Nikon D200 and Nikon CoolPixS710 are used in the experiment. All images are cropped to the resolution of  $2592 \times 2592$  with JPEG format. The image set of each model consists of the same number of images from multiple instances.

As Fig. 6 displays, we report the ROC curves of our proposed classifier with  $k \in \{5\%, 10\%, 15\%, 20\%\}$ , represented as the first 5% PRNU (green), the first 10% (blue), the first 15% (magenta) and the first 20% (red), together with the prior PRNU-based method (cyan).

In the open set, we display the ROC curves and Area Under Curve (AUC) of source model identification. In Fig. 6a, the reference pattern comes from  $S_0 = \text{Nikon D200}$ , the query



**Fig. 6** The performance of ROC curves **a**  $S_0 = \text{Nikon D200}$ ; **b**  $S_0 = \text{Nikon CoolPixS710}$ ; **c**  $S_0 = \text{Nikon D200\#0}$ ,  $S_1 = \text{Nikon D200\#1}$ ; **d**  $S_0 = \text{Nikon D200\#1}$ ,  $S_1 = \text{Nikon D200\#0}$ . Note that in each test,  $S_0$  serves as reference camera

images from Nikon D200 and Nikon CoolPixS710. In Fig. 6b, the reference pattern comes from  $\mathcal{S}_0$ =Nikon CoolPixS710, the query images from Nikon D200 and Nikon CoolPixS710. With decreasing ratio  $k$  of “clean” PRNU  $\mathbf{X}^0$ , the performance of our classifier does not fall down sharply in Fig. 6a. The ROC curve of our algorithm with  $k = 20\%$  appears to be comparable with the ROC curve of the prior-art algorithm. Besides, the performance of our algorithm is relatively stable and shows comparably results by observing Fig. 6b.

In the close set, we show the ROC curves and AUC of source instance identification. In Fig. 6c, the reference patterns come from  $\mathcal{S}_0$  = Nikon D200#0 and  $\mathcal{S}_1$  = Nikon D200#1, the query images from Nikon D200#0. In Fig. 6d, the reference patterns come from  $\mathcal{S}_0$  = Nikon D200#1 and  $\mathcal{S}_1$  = Nikon D200#0, the query images from Nikon D200#1. Compared with the prior PRNU-based algorithm, the performance of our algorithm is slightly reduced when “clean” PRNU  $\mathbf{X}^0$  with  $k = 5\%$ ; meanwhile the performance is not decreased when “clean” PRNU  $\mathbf{X}^0$  with  $k = \{10\%, 15\%, 20\%\}$ . In Fig. 6d, the performance of our algorithm with  $k = \{5\%, 10\%, 15\%, 20\%\}$  is on par with the prior PRNU-based algorithm.

In Table 2, we give the average  $TPR$  with  $FPR = 0.005$ , which is an alternative illustration of the results from Fig. 6. In the row of “Model”, we show the average  $TPR$  of Fig. 6a, b. In the row of “Instance”, we display the average  $TPR$  of Fig. 6c, d. The results of Table 2 directly verify the effectiveness of our algorithm. Besides, the performances of source instance identification are better than that of model identification, since the images acquired by one camera instance come from a single camera while the images captured by one camera model might come from multiple instances.

Due to the same number of images acquired by each camera are used in the experiment, we can formulate the  $TPR$  and  $FPR$  as follows:

$$\begin{aligned} TPR &= \frac{D_{tp}}{N_q} \\ FPR &= \frac{D_{fp}}{N_q} \end{aligned} \quad (14)$$

where  $D_{tp}$  (true positive), the number of the query images is captured by the reference camera;  $D_{fp}$  (false positive), the number of the query images belonging to other camera is different from the reference camera.  $N_q$  denotes the number of query images from each camera.

### 4.3 Robustness analysis of our proposed classifier

In this section, it is proposed to evaluate the robustness of our classifier in the close set. Here, we select four typical attacks, involving JPEG compression, noise adding, noise removing, and image cropping. In the robustness experiments, it should be noted that the attack only happens when testing query images acquired by digital cameras, are not referring to the reference patterns extracted from the images with knowing camera model or instance. In addition, it should be noted that the reference pattern come from  $\mathcal{S}_0$  and  $\mathcal{S}_1$ , when query images come from  $\mathcal{S}_0$ , we get one set of experimental results; when query images come

**Table 2** Average  $TPR$  with  $FPR = 0.005$  for image origin identification of different PRNUs

	Original	“Clean” PRNU $\mathbf{X}^0$			
		$k = 5\%$	$k = 10\%$	$k = 15\%$	$k = 20\%$
Model	1.0000	0.9675	0.9925	0.9975	1.0000
Instance	1.0000	0.9975	1.0000	1.0000	1.0000

**Table 3**  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of source model identification against JPEG compression attacks, in which reference pattern from Samsung L74wide and Samsung NV15, and query images also from Samsung L74wide and Samsung NV15

(a) Original PRNU							
Metric	Untouched	JPEG compression					
		QF 90	QF 80	QF 70	QF 60	Aver.	
$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(b) $k = 15\%$							
$\mathcal{V}_A$	1.0000	1.0000	0.9983	0.9933	0.9783	0.9925	
$\mathcal{V}_R$	1.0000	1.0000	1.0000	0.9967	0.9733	0.9925	
$\mathcal{V}_P$	1.0000	1.0000	0.9967	0.9901	0.9832	0.9925	
$\mathcal{V}_F$	1.0000	1.0000	0.9983	0.9934	0.9782	0.9925	
(c) $k = 10\%$							
$\mathcal{V}_A$	1.0000	1.0000	0.9967	0.9900	0.9750	0.9904	
$\mathcal{V}_R$	1.0000	1.0000	0.9967	0.9933	0.9767	0.9917	
$\mathcal{V}_P$	1.0000	1.0000	0.9967	0.9868	0.9734	0.9892	
$\mathcal{V}_F$	1.0000	1.0000	0.9967	0.9900	0.9750	0.9904	

from  $\mathcal{S}_1$ , we get another set of results. Tables 3, 4, 5, 6, 7, and 8 show the average of two groups of results.

Based on our analysis (see Section 4.2.3), we empirically assume that when  $k = \{10\%, 15\%\}$ , the performance of our proposed classifier via reduced dimensionality feature is close to that of the prior-art algorithm. Therefore, it is proposed to use “clean” PRNU  $\mathbf{X}^0$  with  $k = \{10\%, 15\%\}$ , and original PRNU  $\mathbf{X}$ ; three typical PRNUs with different

**Table 4**  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of source instance identification against JPEG compression attacks, in which reference pattern from Samsung L74wide#0 and Samsung L74wide#1, and query images also from Samsung L74wide#0 and Samsung L74wide#1

(a) Original PRNU							
Metric	Untouched	JPEG compression					
		QF 90	QF 80	QF 70	QF 60	Aver.	
$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(b) $k = 15\%$							
$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	0.9983	0.9996	
$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	0.9967	0.9992	
$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	0.9983	0.9996	
(c) $k = 10\%$							
$\mathcal{V}_A$	1.0000	1.0000	1.0000	0.9950	0.9967	0.9979	
$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	
$\mathcal{V}_P$	1.0000	1.0000	1.0000	0.9902	0.9967	0.9967	
$\mathcal{V}_F$	1.0000	1.0000	1.0000	0.9950	0.9967	0.9979	

**Table 5**  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of source camera identification against noise adding attack measured by the distribution Gaus(m,v)

(a) The performance of source model identification

PRNU	Metric	Untouched	Adding Gaussian distributed noise				Aver.
			Gaus(0,1)	Gaus(0,2)	Gaus(0,3)	Gaus(0,4)	
Original	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 15\%$	$\mathcal{V}_A$	0.9975	0.9975	0.9975	0.9975	0.9975	0.9975
	$\mathcal{V}_R$	0.9950	0.9950	0.9950	0.9950	0.9950	0.9950
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	0.9975	0.9975	0.9975	0.9975	0.9975	0.9975
$k = 10\%$	$\mathcal{V}_A$	0.9975	0.9975	0.9975	0.9975	0.9975	0.9975
	$\mathcal{V}_R$	0.9950	0.9950	0.9950	0.9950	0.9950	0.9950
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	0.9975	0.9975	0.9975	0.9975	0.9975	0.9975

(b) The performance of source instance identification

Original	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 15\%$	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 10\%$	$\mathcal{V}_A$	1.0000	1.0000	1.0000	0.9975	0.9950	0.9981
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	0.9950	0.9950	0.9975
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	0.9950	0.9988
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	0.9975	0.9950	0.9981

(a) Reference pattern from Canon Ixus55 and Canon Ixus70, and query images also from Canon Ixus55 and Canon Ixus70. (b) Reference pattern from Canon Ixus70#0 and Canon Ixus70#1, and query images also from Canon Ixus70#0 and Canon Ixus70#1

dimensionalities are adopted in the design of source camera classifier. The experiment results are mainly evaluated in details by using the following metrics.

4.3.1 Evaluation metrics

We evaluate the performances of the robustness of our proposed classifier using the following metrics:

- Accuracy  $\mathcal{V}_A$  is defined as the percentage of correctly classified instance among the total number of samples. It is formulated by:

$$\mathcal{V}_A = \frac{D_{tp} + D_{tn}}{D_{tp} + D_{fp} + D_{tn} + D_{fn}}, \tag{15}$$

**Table 6**  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of source camera identification against noise adding attack measured by SNR

## (a) Source model identification

PRNU	Metric	Adding Gaussian distributed noise				
		80dB	60dB	40dB	30dB	Aver.
Original	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 15\%$	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 10\%$	$\mathcal{V}_A$	1.0000	1.0000	1.0000	0.9975	0.9994
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	0.9950	0.9988
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	0.9975	0.9994

## (b) Source instance identification

Original	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 15\%$	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000
$k = 10\%$	$\mathcal{V}_A$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$	1.0000	1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$	1.0000	1.0000	1.0000	1.0000	1.0000

(a) Reference pattern from Canon Ixus55 and Canon Ixus70, and query images also from Canon Ixus55 and Canon Ixus70.  
 (b) Reference pattern from Canon Ixus70#0 and Canon Ixus70#1, and query images also from Canon Ixus70#0 and Canon Ixus70#1

where the number of true negative samples is denoted as  $D_{tn}$ , and the number of false negative samples is denoted as  $D_{fn}$ .

- Recall  $\mathcal{V}_R$  is the ratio of the number of samples  $D_{tp}$  to  $D_{fp}$  plus  $D_{tp}$ , it is given by:

$$\mathcal{V}_R = \frac{D_{tp}}{D_{tp} + D_{fn}}. \quad (16)$$

- F1-score  $\mathcal{V}_F$  considers both precision and recall, and it is calculated by:

$$\mathcal{V}_F = 2 \times \frac{\mathcal{V}_P \times \mathcal{V}_R}{\mathcal{V}_P + \mathcal{V}_R}, \quad (17)$$

where precision  $\mathcal{V}_P$  is formulated by:

$$\mathcal{V}_P = \frac{D_{tp}}{D_{tp} + D_{fp}}. \quad (18)$$



**Table 7**  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of source camera identification against noise removing attack

(a) Source model identification						
	PRNU	Metric	Wiener filtering			
			H(3,3)	H(4,4)	H(5,5)	Aver.
Original	$\mathcal{V}_A$		1.0000	1.0000	0.9900	0.9967
	$\mathcal{V}_R$		1.0000	1.0000	0.9900	0.9967
	$\mathcal{V}_P$		1.0000	1.0000	0.9900	0.9967
	$\mathcal{V}_F$		1.0000	1.0000	0.9900	0.9967
$k = 15\%$	$\mathcal{V}_A$		0.9950	0.9900	0.9550	0.9800
	$\mathcal{V}_R$		0.9950	0.9900	0.9550	0.9800
	$\mathcal{V}_P$		0.9950	0.9900	0.9550	0.9800
	$\mathcal{V}_F$		0.9950	0.9900	0.9550	0.9800
$k = 10\%$	$\mathcal{V}_A$		0.9950	0.9950	0.9050	0.9650
	$\mathcal{V}_R$		0.9950	0.9950	0.9050	0.9650
	$\mathcal{V}_P$		0.9950	0.9950	0.9050	0.9650
	$\mathcal{V}_F$		0.9950	0.9950	0.9050	0.9650
(b) Source instance identification						
Original	$\mathcal{V}_A$		1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_R$		1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_P$		1.0000	1.0000	1.0000	1.0000
	$\mathcal{V}_F$		1.0000	1.0000	1.0000	1.0000
$k = 15\%$	$\mathcal{V}_A$		1.0000	1.0000	0.9800	0.9933
	$\mathcal{V}_R$		1.0000	1.0000	0.9800	0.9933
	$\mathcal{V}_P$		1.0000	1.0000	0.9800	0.9933
	$\mathcal{V}_F$		1.0000	1.0000	0.9800	0.9933
$k = 10\%$	$\mathcal{V}_A$		1.0000	0.9950	0.9725	0.9892
	$\mathcal{V}_R$		1.0000	0.9950	0.9750	0.9900
	$\mathcal{V}_P$		1.0000	0.9950	0.9701	0.9884
	$\mathcal{V}_F$		1.0000	0.9950	0.9726	0.9892

(a) Reference pattern from Canon Ixus55 and Canon Ixus70, and query images also from Canon Ixus55 and Canon Ixus70.  
 (b) Reference pattern from Canon Ixus70#0 and Canon Ixus70#1, and query images also from Canon Ixus70#0 and Canon Ixus70#1

### 4.3.2 Robustness to JPEG compression attack

Let us test the robustness of our classifier, resisting against JPEG compression attack. For instance, images from Samsung L74wide and Samsung NV15 are used in the experiment; the concerning statistic is shown in Table 1. We test the images attacked by JPEG compression with quality factor  $QF = \{90, 80, 70, 60\}$ . In Tables 3 and 4, we illustrate the experimental results for image origin (model and instance) identification. The accuracy  $\mathcal{V}_A$ , recall  $\mathcal{V}_R$ , precision  $\mathcal{V}_P$ , and F1-score  $\mathcal{V}_F$  on the 150 query images for each device are reported.

It can be observed that with decreasing  $QF$  of query images, the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  values of our classifier are slightly reduced when  $k = 15\%$  and  $k = 10\%$ , in which column of the “Untouched” indicates the results of the query images without any attacks. In the column of “Aver.”, it can be observed that the average  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  reach to over 98.92%. Besides, the values of  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$ , for the case of adopting “clean” PRNU  $\mathbf{X}^0$  with  $k = 15\%$  and  $k = 10\%$  appear to be comparable with those obtained from original PRNU  $\mathbf{X}$ . When  $QF = 90$ , our proposed algorithm with “clean” PRNU  $\mathbf{X}^0$  performs as well as the detector with the original PRNU  $\mathbf{X}$ , meaning that all the metrics of

**Table 8**  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of source camera identification against image cropping attack

(a) The performance of source model identification

PRNU	Metric	Untouched	Image cropping		
			Crop 95%	Crop 90%	Aver.
Original	$\mathcal{V}_A$	0.9725	0.7575	0.7425	0.7500
	$\mathcal{V}_R$	0.9450	0.7600	0.7450	0.7525
	$\mathcal{V}_P$	1.0000	0.7560	0.7406	0.7483
	$\mathcal{V}_F$	0.9712	0.7580	0.7428	0.7504
$k = 15\%$	$\mathcal{V}_A$	0.9275	0.7725	0.7550	0.7638
	$\mathcal{V}_R$	0.8950	0.7750	0.7650	0.7700
	$\mathcal{V}_P$	0.9569	0.7709	0.7504	0.7607
	$\mathcal{V}_F$	0.9234	0.7729	0.7576	0.7653
$k = 10\%$	$\mathcal{V}_A$	0.9400	0.7825	0.7600	0.7713
	$\mathcal{V}_R$	0.9100	0.7750	0.7700	0.7725
	$\mathcal{V}_P$	0.9673	0.7870	0.7547	0.7709
	$\mathcal{V}_F$	0.9368	0.7809	0.7621	0.7715

(b) The performance of source instance identification

Original	$\mathcal{V}_A$	1.0000	0.7150	0.7425	0.7288
	$\mathcal{V}_R$	1.0000	0.7100	0.7100	0.7100
	$\mathcal{V}_P$	1.0000	0.7188	0.7608	0.7398
	$\mathcal{V}_F$	1.0000	0.7142	0.7344	0.7243
$k = 15\%$	$\mathcal{V}_A$	0.9950	0.7400	0.7325	0.7363
	$\mathcal{V}_R$	0.9900	0.7300	0.7250	0.7275
	$\mathcal{V}_P$	1.0000	0.7452	0.7359	0.7406
	$\mathcal{V}_F$	0.9949	0.7375	0.7304	0.7340
$k = 10\%$	$\mathcal{V}_A$	0.9925	0.7275	0.7400	0.7338
	$\mathcal{V}_R$	0.9900	0.7350	0.7250	0.7300
	$\mathcal{V}_P$	0.9949	0.7244	0.7474	0.7359
	$\mathcal{V}_F$	0.9925	0.7296	0.7360	0.7328

(a) Reference patterns from Pentax OptioA40 and Pentax OptioW60, and query images also from Pentax OptioA40 and Pentax OptioW60. (b) Reference pattern from Pentax OptioA40#0 and Pentax OptioA40#1, and query images also from Pentax OptioA40#0 and Pentax OptioA40#1

both detectors achieve 100%. Nevertheless, our proposed classifier still performs very well under JPEG compression attacks. Note that the performance of source instance identification (see Table 4) is better than that of model identification (see Table 3). Because the images acquired by one camera instance come from a single camera while the images captured by one camera model might come from multiple instances.

### 4.3.3 Robustness to noise adding attack

It is proposed to test the robustness of our classifier resisting against noise adding attack. We select images captured by Canon Ixus55 and Canon Ixus70 in the experiment. Let us test the images attacked by Additive White Gaussian Noise (AWGN), in which the model can be described by using the distribution  $\text{Gaus}(m, v)$  with expectation  $m = 0$ , variance  $v = \{1, 2, 3, 4\}$  or signal to noise ratios (SNR) = {80 dB, 60 dB, 40 dB, 30 dB}. It should be noted that the values of  $v$  and SNR are limited to the range without affecting the image visuality. In Tables 5 and 6, we illustrate the experimental results for image origin (model and instance) identification. The accuracy  $\mathcal{V}_A$ , recall  $\mathcal{V}_R$ , precision  $\mathcal{V}_P$ , and F1-score  $\mathcal{V}_F$  over the 100 query images for each device are reported.

In Table 5, with increasing the intensity of noise adding, the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  values of our classifier are nearly insusceptible. Specifically, in Table 5b, with increasing the variance  $v$ , our proposed classifier with  $k = 15\%$  remains well-performed with  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  equal to 100%; meanwhile when  $k = 10\%$ , the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  are slightly decreased. In the column of “Aver.”, it can be observed that all the average  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  arrive at the ratio not smaller than 99.50%. Compared with the prior method (presented as “Original”), the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of our classifier for image origin identification are slightly reduced, yet that is still acceptable. Note that the performances of source instance identification (see Table 5b) are better than that of model identification (see Table 5a).

In addition, the performances resisting against noise adding attack with SNR are shown in Table 6. In Table 6a, with decreasing the SNR, the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  of our algorithm with  $k = 15\%$  are not affected; meanwhile the performance of our algorithm with  $k = 10\%$  is slightly reduced with the exception of  $\mathcal{V}_R$ . And in Table 6b, our proposed classifier still performs very well suffering from noise adding attack. We can note that the performances of instance identification are preferable to the model identification. Compared with the prior method with original PRNU, our algorithm with  $k = 10\%$  and  $k = 15\%$  still perform well under noise adding attacks; and in the column of “Aver.”, all the average  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  reach to over 99.88%.

Nevertheless, our classifier still performs very well against noise adding attack. Compared with the prior method, the performances of our proposed algorithm for image origin identification are reduced, yet that is still admissible.

#### 4.3.4 Robustness to noise removing attack

Let us consider another attack, that is noise removing. It is proposed to test the robustness of our classifier against low-pass Wiener filtering, in which the filter can be formulated by  $H(s,s)$  with parameter  $s = \{3, 4, 5\}$ . The value of  $s$  is limited to the range without affecting the image visually. We still select images captured by Canon Ixus55 and Canon Ixus70 in the experiment. The accuracy  $\mathcal{V}_A$ , recall  $\mathcal{V}_R$ , precision  $\mathcal{V}_P$ , and F1-score  $\mathcal{V}_F$  on the 100 query images for each device are shown in Table 7.

It can be observed that with increasing the intensity of noise removing, the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  values of our classifier with  $k = 10\%$  and  $k = 15\%$  are slightly reduced. In the column of “Aver.”, all the average  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  reach to over 96.50%. Compared with the prior algorithm with original PRNU, the performance of our algorithm is slightly reduced. Specifically, in Table 7b, when  $s = 3$ , our proposed algorithm with  $k = 10\%$  and  $k = 15\%$  perform as well as the detector with original PRNU; meanwhile when  $s = 4$ , our algorithm with  $k = 15\%$  remains well-performed with  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  equal to 100%. Besides, we can note that the performances of instance identification (see Table 7b) are better than that of model identification (see Table 7a). Numerical results empirically verify that our proposed classifier performs robustly, resisting against noise removing attack.

#### 4.3.5 Robustness to image cropping attack

To further investigate the robustness of our proposed classifier resisting against image cropping attack, we utilize 3 unique instances from 2 camera models, involving Pentax OptioA40 and Pentax OptioW60. We detect the query images which are cropped the central portion with the ratio of 95% and 90%, respectively presented as “Crop 95%” and “Crop 90%” in Table 8. We display the accuracy  $\mathcal{V}_A$ , recall  $\mathcal{V}_R$ , precision  $\mathcal{V}_P$  and F1-score  $\mathcal{V}_F$  of image origin identification in Table 8, in which the query images composed by a total of

200 images from each device. Note that when extracting the reference pattern, we crop the central portion of untouched images.

With slightly shrinking the area of cropping query images, the  $\mathcal{V}_A$ ,  $\mathcal{V}_R$ ,  $\mathcal{V}_P$  and  $\mathcal{V}_F$  values are sharply degraded. In fact, the results obtained by the prior-art method are also very sensitive to image cropping attack. Besides, it should be noted that the performance of instance identification is better than that of model identification when the query images are untouched; meanwhile when the query images suffer from cropping attack, the results of instance identification are not better than that of model identification.

Like the performance of the prior-art method, our proposed classifier is vulnerable to image cropping attack. We can not guarantee that the selected elements in the weight matrix  $\Omega^k$  of PRNU are the first  $K$ ; meanwhile it is not possible to guarantee the reliability of the “clean” PRNU  $\mathbf{X}^0$  extracted from the image that suffers from cropping attack. In addition, when dealing with the images acquired by Pentax camera, the detection accuracy is relatively low with respect to the other camera brand (e.g., Samsung, Nikon).

## 5 Conclusions and discussions

This paper has proposed a classifier for image origin identification, which is established via low dimensional PRNU features. It can distinguish images even if they are captured by different instances of the same camera model. Specifically, we select 3364 images with JPEG format taken by 20 cameras from the Dresden Image Database. Both theoretical analysis and experimental results verify that our proposed classifier can effectively identify different camera models / instances by adopting “clean” PRNU with the ratio  $k$ , remarkably lower dimensionality than that of the original PRNU. Significantly, our proposed algorithm can achieve the comparable results to that of the prior art for dealing with the problem of source camera identification.

To evaluate the robustness of our algorithm, the experimental results show that the proposed classifier performs robustly to some post-processing techniques, such as JPEG compression, noise adding attacks and noise removing attacks, yet vulnerably to image cropping attack. Besides, when detecting images acquired from Pentax or Canon camera, the overall performance of classification is worse with respect to the other brands.

In the future, we will focus on the problem of how to enhance the robustness with suffering from image cropping attack. Furthermore, it could be interesting to investigate the case study of applying image identification to the identification of users over social networks.

**Acknowledgements** This work is funded by the Cyberspace Security Major Program in National Key Research and Development Plan of China under grant No. 2016YFB0800201, the Natural Science Foundation of China under grant No. 61702150 and No. 61572165, the State Key Program of Zhejiang Province Natural Science Foundation of China under grant No. LZ15F020003, the Key Research and Development Plan Project of Zhejiang Province under grant No. 2017C01062 and No.2017C01065.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## References

1. Chen M, Fridrich J, Goljan M, Lukás J (2008) Determining image origin and integrity using sensor noise. *IEEE Trans Inf Forensics Secur* 3(1):74–90

2. Costa FDO, Silva E, Eckmann M, Scheirer WJ, Rocha A (2014) Open set source camera attribution and device linking. *Pattern Recogn Lett* 39:92–101
3. Foi A, Trimeche M, Katkovnik V, Egiazarian K (2008) Practical poissonian-gaussian noise modeling and fitting for single-image raw-data. *IEEE Trans Image Process* 17(10):1737–1754
4. Gloe T, Böhme R (2010) The dresden image database for benchmarking digital image forensics. *Journal of Digital Forensic Practice* 3(2–4):150–159
5. Goljan M (2008) Digital camera identification from images-estimating false acceptance probability. In: *International workshop on digital watermarking*, pp 454–468
6. Goljan M, Fridrich J (2013) Sensor fingerprint digests for fast camera identification from geometrically distorted images. In: *Media watermarking, security, and forensics 2013*, vol 8665, p 86650b
7. Goljan M, Fridrich J, Filler T (2009) Large scale test of sensor fingerprint camera identification. In: *Media forensics and security*, vol 7254, p 72540i
8. Goljan M, Fridrich J, Filler T (2010) Managing a large database of camera fingerprints. In: *Media forensics and security II*, vol 7541, p 754108
9. Healey GE, Kondepudy R (1994) Radiometric ccd camera calibration and noise estimation. *IEEE Trans Pattern Anal Mach Intell* 16(3):267–276
10. Holst GC (1998) Ccd arrays, cameras, and displays. SPIE Optical Engineering Press, Bellingham
11. Holub V, Fridrich J, Denemark T (2014) Universal distortion function for steganography in an arbitrary domain. *EURASIP J Inf Secur* 2014(1):1
12. Hu Y, Li CT, Lai Z (2015) Fast source camera identification using matching signs between query and reference fingerprints. *Multimedia Tools and Applications* 74(18):7405–7428
13. Janesick JR (2001) Scientific charge-coupled devices, vol 83. SPIE Press, Bellingham
14. Kay SM (1998) Fundamentals of statistical signal processing, vol. ii: Detection theory. Signal Processing. Upper Saddle River, NJ: Prentice Hall
15. Li R, Li CT, Guan Y (2018) Inference of a compact representation of sensor fingerprint for source camera identification. *Pattern Recogn* 74:556–567
16. Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. *IEEE Trans Inf Forensics Secur* 1(2):205–214
17. Luo X, Song X, Li X, Zhang W, Lu J, Yang C, Liu F (2016) Steganalysis of hugo steganography based on parameter recognition of syndrome-trellis-codes. *Multimedia Tools and Applications* 75(21):13557–13583
18. Ma Y, Luo X, Li X, Bao Z, Zhang Y (2018) Selection of rich model steganalysis features based on decision rough set  $\alpha$ -positive region reduction *IEEE Transactions on Circuits and Systems for Video Technology*
19. Nakamura J (2017) Image sensors and signal processing for digital still cameras. CRC Press, Boca Raton
20. Pandey RC, Singh SK, Shukla KK (2016) Passive forensics in image and video using noise features: a review. *Digit Investig* 19:1–28
21. Qiao T, Retraint F, Cogranne R (2013) Image authentication by statistical analysis. In: 2013 proceedings of the 21st European signal processing conference (EUSIPCO), pp 1–5
22. Qiao T, Retraint F, Cogranne R, Thai TH (2015) Source camera device identification based on raw images. In: 2015 IEEE international conference on image processing (ICIP), pp 3812–3816
23. Qiao T, Retraint F, Cogranne R, Zitzmann C (2015) Steganalysis of jsteg algorithm using hypothesis testing theory. *EURASIP J Inf Secur* 2015(1):2
24. Qiao T, Retraint F, Cogranne R, Thai TH (2017) Individual camera device identification from jpeg images. *Signal Process Image Commun* 52:74–86
25. Qiao T, Zhu A, Retraint F (2018) Exposing image resampling forgery by using linear parametric model. *Multimedia Tools and Applications* 77(2):1501–1523
26. Ramanath R, Snyder WE, Yoo Y, Drew MS (2005) Color image processing pipeline. *IEEE Signal Proc Mag* 22(1):34–43
27. Thai TH, Cogranne R, Retraint F (2014) Camera model identification based on the heteroscedastic noise model. *IEEE Trans Image Process* 23(1):250–263
28. Thai TH, Retraint F, Cogranne R (2016) Camera model identification based on the generalized noise model in natural images. *Digital Signal Process* 48:285–297
29. Xu B, Wang X, Zhou X, Xi J, Wang S (2016) Source camera identification from image texture features. *Neurocomputing* 207:131–140
30. Yao H, Qiao T, Xu M, Zheng N (2018) Robust multi-classifier for camera model identification based on convolution neural network. *IEEE Access* 6:24973–24982
31. Zhang Y, Qin C, Zhang W, Liu F, Luo X (2018) On the fault-tolerant performance for a class of robust image steganography. *Sig Process* 146:99–111



**Yihua Zhao** received the B.S. degree in Computer Science and Technology in 2016 from Heilongjiang Bayi Agricultural University, Da Qing, China, and since then she has joined the Laboratory of Internet and Network Security of Hangzhou Dianzi University, Hangzhou, China. Her current research interest is digital image forensics.



**Ning Zheng** received his M.S. degree from Zhejiang University, in 1987. He is currently a full professor of Hangzhou Dianzi University. His research interest is digital forensics.



**Tong Qiao** received the B.S. degree in Electronic and Information Engineering in 2009 from Information Engineering University, Zhengzhou, China, and the M.S. degree in Communication and Information System in 2012 from Shanghai University, Shanghai, China, and the Ph.D. degree in University of Technology of Troyes, Laboratory of Systems Modelling and Dependability, Troyes, France. Since 2016, he works as an assistant professor in School of Cyberspace from Hangzhou Dianzi University. His current research interests focus on steganalysis and digital image forensics.



**Ming Xu** received his M.S., and Ph.D. degrees from Zhejiang University, in 2000, and 2004, respectively. He is currently a full professor of Hangzhou Dianzi University. His research interest is digital forensics.