A fast source camera identification and verification method based on PRNU analysis for use in video forensic investigations

Wen-Chao Yang[1] • Jiajun Jiang[2] • Chung-Hao Chen[2]

Abstract

Due to the rapid development of digital and cloud technologies, everyone can easily shoot and spread digital videos via email or social media. However, it is difficult for law enforcement to trace the origin of those digital videos, while some videos or images containing illegal information such as personal privacy, obscene pornography, and national security-related content. Recently, a significant breakthrough is achieved by using Photo-Response Non-Uniformity (PRNU) noise to characterize the camera sensor. However, PRNU analysis is often carried out on a frame-by-frame basis. As a result, the processing time is unbearable when treating a large set of videos and devices. In this paper, we propose a novel video forensic method considering both cameras rolling and I-frame of videos to improve the processing time and accuracy. Experimental results demonstrate that our proposed method is at a minimum of 15 times on average faster than the most wildly used method, PRNU analysis, and reduce the false positive rate as compared to existing methods used in the field of the forensic examination.

1 Introduction

With the widespread use of smartphones and the rapid development of corresponding software, the acquisition and processing of a video stream is no longer a difficult task. The rise of convenient social media and live broadcast platforms have made communication using videos

✉ Wen-Chao Yang
una135@mail.cpu.edu.tw

1 Department of Forensic Science, Central Police University, Taoyuan City, Taiwan, Republic of China

2 Electrical and Computer Engineering Department, Old Dominion University, Norfolk, VA, USA

easier. Consequently, some videos or images containing illegal information such as personal privacy, obscene pornography, and national security-related content have been released to the public. To prohibit these behaviors and incriminate the criminal behavior of these suspects who created illegal videos or images, it is important to identify that these videos or images were taken from devices owned by these suspects.

The task of the forensic analyst in source camera identification (SCI) is to identify the device used to shoot a desired photo or video, and in source camera verification (SCV) is to verify whether a camera has been acquired an image or video evidence by a given threshold. A large part of the research works in source camera identification and verification (SCI/V) was mainly concentrated on still images. Most of the existing methods are to extract the sensor fingerprint, which is also called the reference sensor pattern noise (SPN) or Photo-Response Non-Uniformity (PRNU), of an image [20]. In May 2009, this technique (SPN/PRNU) was used in child rape convictions in Scotland [25].

Kurosawa et al. [14] were the first investigators to study the SCI/V technique on videos. They performed SCI experiments with 9 camcorders in 4 types by using dark current on CCD chip and obtained 8 fixed pattern noise from 9 camcorders. Chen et al. [2] extended the SPN/PRNU technique [20] from image to video and showed that PRNU could be used to identify the source camcorder effectively. They extracted PRNUs from video clips using the Maximum Likelihood Estimator (MLE) and proposed the peak-to-correlation energy (PCE) as the measurement to detect the presence of PRNU. They performed SCI experiments with 25 camcorders in 3 types and got good results. McClokey [21] analyzes the influence of video content on the performance of Chen et al. [2] and proposed a confidence weight PRNU based on image gradient magnitudes to improve PRNU estimation.

Existing video coding standards such as MPEG series, H.264, or newer version, use three types of video frames, meaning that each group of pictures (GOP) is comprised of intra-coded picture (I-frame), predictive coded picture (P-frame), and bi-predictive coded picture (B-frames). I-frame is the least compressible and does not require other video frames to decode. P-frames can use data from previous frames to decompress and are more compressible than I-frames. B-frames can use both previous and forward frames for data reference to get the highest amount of data compression. Chuang et al. [4] analyzed the video compression effect on SCI based on Chen et al. [2] and McClokey [21] and described that I-frames, which are extracted from videos, are more reliable than P-frames, for PRNU analysis. Recently, Taspinar et al. [26] proposed a PRNU-based method for out-of-camera stabilized videos, such as rotation and cropping processing. They performed experiments with 100 videos from 13 smart-phone cameras of five brands. First, at most 50 I-frames are extracted as a set from each video clips. Then, to determine if a given video is stabilized using the set of the given video. Finally, compute the fingerprint by aligning the misaligned frames using an inverse affine transformation and then calculate normalized cross-correlation (NCC) values [2] between non-stabilized videos and the fingerprints from the no stabilized/stabilized videos. Otherwise, using the standard PRNU method to calculate PCE values between non-stabilized videos and the fingerprints from the no stabilized/stabilized videos. Their experimental results showed wrongly classified a stabilized video as non-stabilized in 12% cases and obtained a 91.29% average true positive rate (TPR) for SCI/V without considering the wrongly classified cases. Iuliant et al. [11] extended Taspinar et al.'s work [26] and proposed a hybrid approach that uses both videos and still images to extract video fingerprints in video source identification. They used the PRNUs extracted from still images on stabilized videos and the PRNUs extracted from videos on non-stabilized videos. The TPR in their experimental results is consistent with the results in Taspinar et al.'s work [26].

Despite that many approaches had been proposed to deal with videos, those approaches don't differentiate still images and videos to our knowledge. In other words, PRNU analysis is often carried out on a frame-by-frame basis for video, resulting in high time consumption [1]. Nevertheless, a video clip usually has a much lower spatial resolution and varying sensor sensitivity, often called the International Organization for Standardization standard values (ISO), on different frames as compared to still images. Thus, there is an urgent need to develop a faster and accurate SCI/V technique on videos. This work is targeted to address this urgent need.

The main contribution of this paper is: to our knowledge, we are the first research team to discover the effect of the camera rolling on PRNU analysis and, in the meantime, consider both cameras rolling and I-frame of video to increase processing speed and also reduce the false positive rate as compared to existing methods used in the field of the forensic examination. Besides, we provide a novel structure to analyze various SCI/V methods, which is in high demand in the SCI/V field. This paper is organized as follows: In Section 2, we provide a brief overview of the SCI/V methods based on PRNU analysis. Then, we demonstrate the effects of camera rolling[1] and I-frame of videos on PRNU in Section 3. In the following section, we introduce our proposed method considering the effects of both cameras rolling and I-frame on SCI/V, and in Section 5, we present experimental results and some discussions. This paper is concluded in the final section.

2 Related work

Throughout the past decade, many SCI/V methods which are PRNU based have been proposed. An overview and summary are described in this section. Using the SPN/PRNU extracted from images to track the source camera is one of the most widely used methods. According to the different steps in the photo production process, in general, existing approaches are grouped into three phases to improve SCI methods, including artifacts extraction phase, enhancement phase, and similarity measurement phase. The first phase, artifacts extraction phase, is to extract the rough device artifacts or called the rough PRNU. Then, techniques proposed in the enhancement phase are to estimate, enhance, or refine the rough PRNU. Finally, the sophisticated detection or similarity measurement is used to identify the source device from the enhanced PRNU. Some important works are summarized in Table 1 and will be described in the following.

Lukas et al. [19, 20] proposed the first method for source camera identification using reference SPN. Their methods de-noise each image multiple times using the wavelet-based Mihcak's filter [22] to obtain noise residuals and then use the weighted average to extract the final required SPN. Our previous work [27] showed using the Wiener filter as a de-noising algorithm can obtain better reference SPNs as compared to wavelet-based Mihcak's filter. A new de-noising algorithm, block-matching and 3D filtering (BM3D) algorithm, was subsequently proposed by Dabov et al. [6]. In the original image, for each target block, Dabov' method finds the most similar blocks nearby and stack them into a 3D array. Then, the third dimension of the array is processed through the discrete cosine transform (DCT). Finally, the inverse DCT is performed to derive aggregated values after thresholding and filtering. The advantage of this algorithm is that the identification accuracy reaches an optimal level,

---

[1] Rolling is defined as the rotation around the optical axis of the camera.

Table 1 Related SCI methods based-on PRNU analysis

Phases	Method	Authors	Purpose
Artifacts extraction (PRNU Extraction)	Basic	Lukas et al. [19, 20]	The basic method for SCI using PRNU
	Wiener filter	Yang et al. [27]	Using Wiener filter as a denoising algorithm to extract better reference SPN
	Color-Decoupled PRNU (CD-PRNU)	Li and Li [17]	To prevent the interpolation noise from propagating into the physical components
	Block-matching and 3D filtering (BM3D)	Dabov et al. [6]	Grouping 2D image patches with similar structures into 3D arrays and collectively filtering the grouped image blocks.
	Predictor based on Context Adaptive Interpolation (PCAI)	Kang et al. [12]	To suppress the effect of scenes and edges.
Enhancement (Estimation)	Zero-Mean (ZM) Wiener Filter (WF)	Chen et al. [3]	To further remove the artifacts produced by camera processing operations
	Enhance components of PRNUs	Li [16]	The stronger component of PRNU is associated with strong scene details.
	Maximum Likelihood Estimation (MLE)	Chen et al. [3]	To estimate reference PRNU from several residual images.
	Phase SPN	Kang et al. [12]	To remove the periodic noise and other non-white noise contaminations in the reference PRNU.
	Spectrum equalization algorithm (SEA)	Lin and Li [18]	To detect and eliminate periodic peaks in the spectrum
Similarity Measurement	Peak-to-Correlation Energy (PCE)	Chen et al. [2]	To attenuate the influence of periodic noise contaminations.
	Circular Cross-correlation Norm (CCN)	Kang et al. [13]	To further decrease the false-positive rate by adding negative.
	Signed PCE (SPCE)	Goljan et al. [10]	Same purpose as CCN

especially for small- and medium-size images. However, the computational complexity is extremely high. Chen et al. [3] extracted PRNU by using the Maximum-likelihood Estimator (MLE) and further eliminating artifacts caused during the photographing process by combining the pre-processing steps of zero-mean (ZM) and Wiener filter (WF). A method namely peak-to-correlation energy (PCE) was proposed for measuring the similarity between the reference PRNU and extracted noise residual of query image by Chen et al. [2] to decrease the false-positive rate of source camera identification. The core idea of PCE is to reduce the similarities between noise residuals of images taken by different cameras and PRNU of a particular camera.

Similarly, in [9, 10], Goljan et al. proposed a method by adding a sign function to PCE, called signed PCE (SPCE), to obtain the positive and negative values, which reduces the false positive rate by half compared to PCE. Li [16] proposed an algorithm to eliminate the contamination of scene details by adding a factor that is inversely proportional to the size of the PRNU component. Cortiana et al. [5] compared the noise reduction effects of different denoising filters and summarized their impact on the performance of identifying source cameras. Li and Li [17] proposed a couple-decoupled PRNU (CD-PRNU) extraction method to reduce the interpolation noise caused by the color filter array (CFA). They divided the

original image into different sub-images based on each color channel to extract PRNU and then combined PRNUs of each sub-image to obtain the final CD-PRNU. Kang et al. [12] designed a PRNU predictor based on eight-neighbor context adaptive interpolation (PCAI8) algorithm to reduce the contamination caused by the scene details, such as edges and context. In [13], they extracted the reference PRNU only by using the phase from the noise residual after discrete Fourier transform (DFT). At the same time, they also improved the PCE algorithm by calculating the square root of PCE, which is named the correlation over the circular cross-correlation norm (CCN). Since CCN computes negative and positive values of the correlation, it has as similar effects as SPCE.

Based on Chen et al.'s [3], Lin and Li [18] proposed an improved algorithm, spectrum equalization algorithm (SEA), to remove periodic artifacts caused by detecting and eliminating periodic peaks in the spectrum. The SEA algorithm improves slightly accuracy rate of PRNU analysis [3] in low-resolution images but requires more processing time. Lawgaly and Khelifi [15] proposed a new source camera identification and verification system based on three processes. The first process is to improve the original Locally Adaptive DCT (LADCT) [23] filter by using varying block thresholds and operations in both horizontal and vertical directions to extract PRNU features. The second process is to use a new weighted averaging algorithm to obtain higher quality PRNU. The last one is to extract PRNU from the whole color plane. Their method slightly reduces false rates (both false positive and false negative rates) of PRNU analysis [3] in low-resolution images but spends twice processing time.

The current demand for multimedia forensics is in a great need to produce a faster and more accurate method to address the problem of source camera identification and verification for videos. The main contribution in this paper is that our proposed method considers both camera rolling and I-frame of videos, thus producing a significant increase in speed and reduce the false positive rate as compared to existing methods used in the field of the forensic examination. Moreover, the proposed method is modularized and flexible where each step of our proposed method could adapt to different approaches.

3 Effect of cameras rolling and I-frame on PRNU analysis

This section provides the analyses of the effect of cameras rolling and the I-frame of videos on PRNU analysis.

3.1 Mobile phone rolling

Due to the gyroscopes and new sensor techniques, a mobile phone can automatically rotate the video 180 degrees while recording videos with rolling 180 degrees. It causes difficulty to distinguish the rolling degrees from the video content, which is a serious problem of PRNU analysis in forensic investigation and will be illustrated below. We use an Apple iPhone 7 plus to illustrate the rolling phenomenon on PRNU analysis. Rolling 0 degrees is set as the plug port (USB, lightning, etc.) being on the right side of a mobile phone when someone uses a mobile phone to shoot images or videos. The analysis is focused on whether videos with different rolling degrees, 0, 90, 180, and 270 degrees, impact the PRNU analysis or not. Two videos are shot for each rolling degrees and then each of them is selected to be in either one of training or testing groups. The details are shown in Table 2.

Table 2 Rolling test videos (|.| denotes the cardinality of a set)

Rolling degrees	Training group				Testing group			
	Video Name	Duration Time (Sec)	|frames|	|I-frames|	Video Name	Duration Time (Sec)	|frames|	|I-frames|
0	Tr01	14.67	440	15	Te01	05.57	167	6
90	Tr02	15.57	467	16	Te02	05.80	174	6
180	Tr03	15.57	467	16	Te03	05.67	170	6
270	Tr04	18.07	542	19	Te04	05.43	163	6

The analysis is divided into two processes, reference PRNU extraction, and PRNU matching processes. The steps of the reference PRNU extraction process are described as follows:

(1) Frame extraction: each video in the training group is extracted to frames, I_i, where I_i denotes the ith frame in the video.
(2) PRNU extraction: according to [3], the model of the sensor output I_i is defined as

$$I_i = I_i^o + I_i^o K + \Theta \tag{1}$$

where I_i^o denotes the sensor output without noises, $I_i^o K$ denotes the PRNU of the sensor, and Θ denotes a complex of independent random noises. PRNU factor, K, is obtained from the following equations:

$$K = \left(I_i - I_i^o\right)/I_i^o + \Theta/I_i^o \tag{2}$$

I_i^o can be estimated by $\hat{I}_i^o$,

$$I_i^o \approx \hat{I}_i^o = f(I_i), \tag{3}$$

where $f(.)$ denotes Mıhçak et al.'s [22] wavelet-based de-noising function. From Eqs. (2) and (3), K can be estimated by $\widehat{K}$

$$K \approx \widehat{K} = (I_i - f(I_i))/f(I_i) + \Theta/f(I_i). \tag{4}$$

3   Enhancement: the zero-mean, ZM, and the Wiener filter (WF) using in the Fourier domain operations [3] are used to enhance the estimated PRNU factor, denoted as $\widehat{K}$.
4   Rotation normalization: each estimated PRNU factor with a resolution of m rows by n columns is rotated 90 degrees if m is larger than n, that is

$$\widehat{K} = R(K, 90), \text{if } m > n \tag{5}$$

where $R(A, B)$ is a rotation operation that rotates A by B degrees.

5   RPRNU obtaining: Each reference PRNU (RPRNU) factor of a training video with N frames is obtained by averaging the PRNU factors from frames in the same video. Because Θ is a complex of independent random noises, the term $\sum_{i=1}^{N} [\Theta/(I_i - f(I_i))] / N$ in Eq. (4) can be erased.

$$
\begin{aligned}
RPRNU &= \sum_{i=1}^{N} \widehat{K}/N \\
&= \sum_{i=1}^{N} [(I_i - f(I_i))/f(I_i)]/N + \sum_{i=1}^{N} [\Theta/f(I_i)]/N \\
&= \sum_{i=1}^{N} \Big[(I_i - f(I_i))/f(I_i) \Big].
\end{aligned}
\tag{6}
$$

The steps of the PRNU matching process are similar to the reference PRNU extraction process. First, each video in the testing group is processed as the first four steps in the reference PRNU extraction process. Finally, the SPCE value between each PRNU factor of the testing frame and each RPRNU factor using the following Eqs. [16].

$$
SPCE(RPRNU, K) = \frac{\rho^2(RPRNU, K; 0, 0) \times sign(\rho(RPRNU, K; 0, 0))}{\frac{1}{mn - |N_{\max}|} \sum_{(m_s, n_s) \notin N_{\max}} \rho^2(RPRNU, K; m_s, n_s)},
\tag{7}
$$

$$
\rho(U, V; m_s, n_s) = \frac{\sum_{x,y} \left(U_{x,y} - \overline{U} \right) \left(V_{x+m_s, y+n_s} - \overline{V} \right)}{\sqrt{\sum_{x,y} \left(U_{x,y} - \overline{U} \right)^2} \sqrt{\sum_{x,y} \left(V_{x+m_s, y+n_s} - \overline{V} \right)^2}},
\tag{8}
$$

where $\rho(U, V; m_s, n_s)$ denotes a cross-correlation function standing for the range of indices, x, y, m_s, n_s, is $1 \le x$, $m_s \le m$, $1 \le y$, $n_s \le n$. $\overline{U}$ and $\overline{V}$ respectively denotes the sample mean of $U_{x, y}$ and $V_{x, y}$. $sign(.)$ stands for a signum function. N_{max} is a small neighborhood around the origin.

The results are shown in Fig. 1. In Fig.1, the x-axis is the frame order of a testing video and the y-axis is the SPCE value obtained by Eqs. (7) and (8). The testing video which has the same rolling degree corresponding to the training video obtains the highest SPCE value. In other words, the higher the SPCE value is, the more similar between testing and training videos. As the description in Section 1, the SPCE threshold is needed for SCV. We refer to the SPCE analysis in Goljan et al.'s [8]. In [8], they identify the distribution of the SPCE value between two PRNU factors, extracted from two different images shot by the same device, is approximately equal to a chi-square distribution with one degree of freedom. In other words, the SPCE threshold should be larger than 41.82 to verify two different images shot by the same device when the device population is about 10 billion, Pr {SPCE >41.82} $\approx 10^{-10}$. The statistical analysis is shown in Table 3 where the SPCE value in Mean or Max field is shown in bold when it is larger than 41.82.
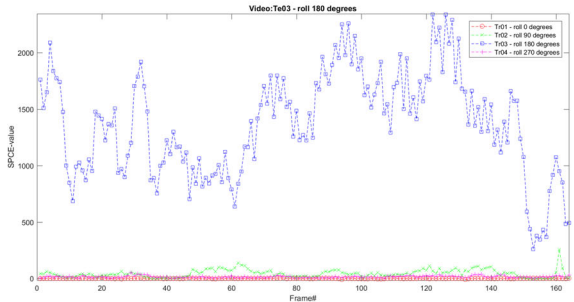
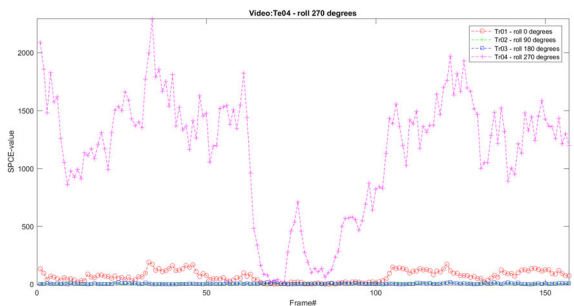**Fig. 1** The rolling analysis experimental results
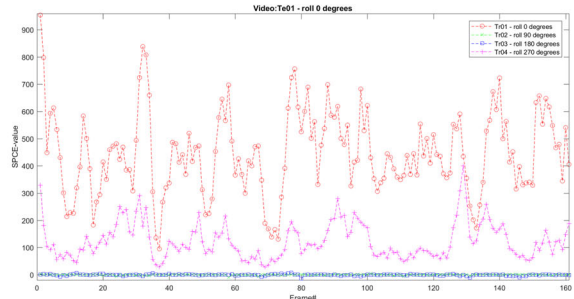
(a) Testing video: Te01 (roll 0 degrees)
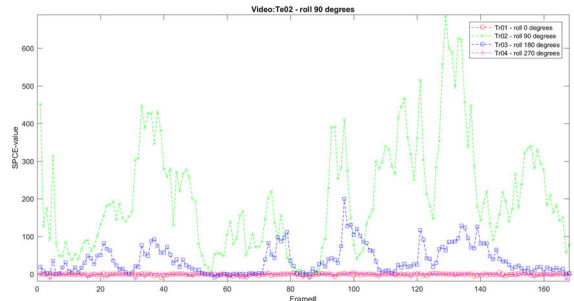
(b) Testing video: Te02 (roll 90 degrees)

(c) Testing video: Te03 (roll 180 degrees)

(d) Testing video: Te04 (roll 270 degrees)

According to the "Max" field in Table 3. The testing videos with rolling 0 or 270 degrees and the training videos with rolling 0 or 270 degrees are verified as the same device. The

Table 3 The details of the rolling experimental results (Min, Max, Mean, and Std respectively denote the minimum, the maximum, the sample mean, and standard deviation of the SPCE values)

Reference PRNU	Testing video: Te01 (roll 0 degrees)				Testing video: Te02 (roll 90 degrees)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Tr01	95.09	**953.92**	**445.24**	153.75	−8.54	3.96	−0.94	1.96
Tr02	−10.09	3.89	−0.44	1.55	−0.02	**683.81**	**204.54**	148.59
Tr03	−14.73	7.7	−0.83	2.8	−8.98	**199.56**	36.83	37.42
Tr04	27.69	**399.41**	**123.75**	66.45	−13.76	5.77	−1.69	2.87
Reference PRNU	Testing video: Te03 (roll 180 degrees)				Testing video: Te04 (roll 270 degrees)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Tr01	−11.06	7.34	−0.28	2.19	−0.05	**190.02**	**68.25**	47.63
Tr02	−0.69	**256.24**	44.28	34.91	−1.57	16	3.63	3.66
Tr03	261.27	**2339.93**	**1369.69**	460.03	−0.93	19.06	4.04	3.99
Tr04	1.24	50.49	14.21	8.55	14.93	**2289.67**	**1144.74**	530.29

testing videos with rolling 90 or 180 degrees and the training videos with rolling 90 or 180 degrees are verified as the same device. That means the rotation normalization step can make the RPRNU factor extracted from the video with rolling 90 degrees as rolling 180 degrees and the RPRNU factor extracted from the video with rolling 270 degrees as rolling 0 degrees. However, the experimental results also show that the testing video with rolling 180 degrees and the training video with rolling 0 degrees cannot be verified as the same device even both videos shot by the same device. In addition, the testing videos with rolling 270 degrees and the training videos with rolling 90 degrees cannot be verified as the same device. That means there still exists a rolling effect on PRNU analysis. The videos with rolling 180 degrees may cause a false-negative result in PRNU analysis. To erase the rolling effect on PRNU analysis, we add Eq. (9) in the SPCE computing step in PRNU matching process,

$$MSPCE(SPN, K) = \max(SPCE(SPN, K), SPCE(SPN, R(K, 180))) \qquad (9)$$

where max(,) stands for a function that returns maximum among the given numbers.

The modified SPCE (MSPCE) value of each frame would be the maximum value of the original SPCE value and the SPCE value obtained between its PRNU factor with rolling 180 degrees and each RPRNU factor. Table 4 shows the experimental result. The result shows that the training video and the testing videos with four kinds of rolling degrees could be verified as

Table 4 The details of the updating rolling experimental results (Min, Max, Mean, and Std respectively denotes the minimum, the maximum, the sample mean, and standard deviation of the MSPCE values)

Reference PRNU	Testing video: Te01 (roll 0 degrees)				Testing video: Te02 (roll 90 degrees)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Tr01	95.09	**953.92**	**445.24**	153.75	−3.77	**150.30**	30.53	28.1
Tr02	1.08	**199.58**	**59.64**	42.23	0.15	**683.81**	**204.54**	148.58
Tr03	8.18	**207.84**	**52.16**	31.27	−3.89	**199.56**	37.20	37.07
Tr04	27.69	**399.41**	**123.75**	66.45	−8.57	**62.83**	6.98	11.86
Reference PRNU	Testing video: Te03 (roll 180 degrees)				Testing video: Te04 (roll 270 degrees)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Tr01	0.00	**75.13**	25.15	16.8	0.2	**190.02**	**68.26**	47.61
Tr02	0.36	**256.24**	**45.14**	33.98	0.5	**115.52**	**45.35**	28.46
Tr03	261.27	**2339.93**	**1369.69**	460.03	0.21	**301.63**	**113.86**	69.76
Tr04	4.26	**431.97**	**119.07**	77.16	14.93	**2289.67**	**1144.74**	530.29

the same device when they are shot by the same device. Therefore, the rolling effect on PRNU analysis can be eliminated.

3.2 PRNU analysis of each frame in a compression video

As mentioned previously, video usually is recorded in a lossy compression format. The famous video coding standards use three types of video frames, meaning that each group of pictures (GOP) is comprised of I-, P-, and B-frames. I-frame, intra-coded picture, uses the intra-frame coding technique that is compressing independently of all other frames. P- and B-frames, predictive coded picture and bi-predictive coded picture, respectively, use the inter-frame coding technique that presents redundancies in other frames to increase compression. That means I-frames usually have more details than P- and B-frames. Chuang et al. [4] also considered the I-frames are more reliable than P-frames for PRNU estimation from their experimental results. Moreover, according to Fig. 1, we discover the SPCE values that are significant to the corresponding PRNU factors are not smooth and may exist frequent peaks. Through the observation of frames, the SPCE peaks are obtained from I-frames. Therefore, the bold assumption that I-frames are vital than other frames in PRNU analysis is proposed. Here, I-frames instead of all frames of videos (all videos in two groups) are used to analyze the effect of I-frames in PRNU analysis. I-frames of each video are extracted by the third-party software, ffmpeg, and do the same experiment for mobile phone rolling. The experimental results are shown in Table 5. According to Tables 2 and 5, we find that using I-frames in PRNU analysis can speed up about 24 times on average faster (|frames|/|I-frames|) and maintain similar accuracy. The more evidence will be shown in Experimental Results and Discussions section. The lower SPCE may be caused by using fewer I-frames in Eq. (6) while we could not remove the term $\sum_{i=1}^{N} \left[\Theta/(I_i - f(I_i)) \right] / N$ in Eq. (4) in the PRNU factors.

4 The proposed method

As mentioned in Section 3, the existence of the rolling effect has an impact on PRNU analysis and the I-frames have more reliable outcomes than other frames for PRNU estimation. Fig. 2

Table 5 The result of the I-frames experimental results (Min, Max, Mean, and Std respectively denotes the minimum, the maximum, the sample mean, and standard deviation of the MSPCE values)

| Reference PRNU	Testing video: Te01 (roll 0 degrees)				Testing video: Te02 (roll 90 degrees)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Tr01	96.7	**262.11**	**167.21**	58.07	0.84	31.98	12.27	11.05
Tr02	7.87	**62.18**	32.29	24.29	8.41	**109.38**	**49.25**	33.74
Tr03	5.89	**49.31**	20.78	17.96	11.3	**64.57**	24.97	20.05
Tr04	9.75	**82.42**	41.64	26.67	0.34	13.07	7.25	4.18
Reference PRNU	Testing video: Te03 (roll 180 degrees)				Testing video: Te04 (roll 270 degrees)			
	Min	Max	Mean	Std	Min	Max	Mean	Std
Tr01	3.63	21.68	14.45	7.45	1.83	**76.6**	**45.11**	27.34
Tr02	2.23	31.29	13.74	10.51	2.45	11.5	5.84	3.6
Tr03	7.25	**191.83**	**120.46**	78.49	7.22	29.87	19.45	8.98
Tr04	3.63	**69.81**	29.04	23.08	14.48	**186.69**	**141.01**	64.51
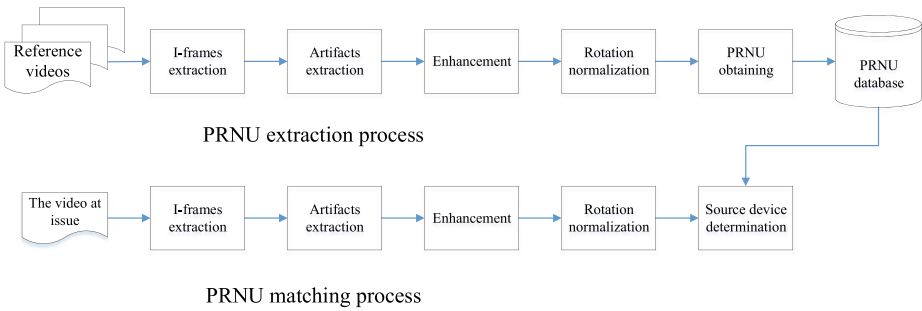
Fig. 2 The block diagram of the proposed method

illustrates the flow chart of our proposed SCI method. First, I-frames are extracted from each video. Then, the adaptive de-noising filter is applied to extract the rough device artifacts and reduce the effect of shot scenes on residues. Next, the preprocessing and refining PRNU algorithm is applied to remove the non-unique artifacts. In order to reduce the effect of rolling, the obtained PRNU frames are rotated to the adaptive direction. For reference videos, the refined noise residues are averaged and then stored in the database. For testing videos, the modified SPCE statistics and a theoretical threshold are used to determine the source camera. Each step of the proposed method will be discussed in more detail in the following subsections.

- PRNU extraction process: before analyzing the video evidence, the reference source must be built. The main steps are summarized below.

Step 1. I-frames extraction: in order to decrease the computational cost and maintain the accuracy of PRNU analysis, the reliable frames, I-frames, of the videos are extracted in the proposed method. Here, we use the third-party software, ffmpeg, to extract the I-frames of videos and save in bitmap (BMP) format.

1   Artifacts extraction: By considering the computational cost, accuracy, and whether the method has been used in court, the Mıhçak et al.'s [22] wavelet-based de-noising filter is selected in the proposed method. The artifacts are extracted as Eqs. (3) and (4).

2   Enhancement: the artifacts extracted by step 2 may include non-unique noises. ZM and WF operations [3] are used to refine artifacts.

3   Rotation normalization: in order to decrease the rolling effect, Eq. (5) is used to obtain refined artifacts with a consistent rotation.

4   PRNU obtaining: the averaging operation, Eq. (6), is used to remove some independent and identically distributed noises to obtain the PRNU factors of each corresponding device. Then, the PRNU factors are stored in the PRNU database as the reference source.

- PRNU matching process: the video evidence is analyzed through the following steps.

Step 1–4   Operations are the same as those in steps 1–4 of the PRNU extraction process.

1   Source device determination: The modified SPCE statistics and the suggestive threshold value of 41.82 as described in Section 3.1 is used to determine the source device of the

targeting video. The detailed operations of MSPCE statistics are shown in Eqs. (7), (8), and (9).

5 Experimental results and discussions

The public vision dataset [24] is selected to verify the effectiveness of our proposed method. Vision dataset has 1914 videos in the native format (Facebook, YouTube, and WhatsApp) from 35 portable devices. Two videos for each portable device are randomly chosen from native format videos and then each of them is selected to be either the training or testing group. Some images of both groups are shown in Appendix. Our algorithms are developed in Matlab R2017 and the experimental platform is Microsoft Windows 7, Intel i7-7600U 2.80 GHz with 16GB Ram.

5.1 Evaluation statistics

In the following experiment, the overall false-positive rate (FPR) and the overall false-negative rate (FNR) of maximum MSPCE and mean MSPCE statistics are computed to assess the performance. $\widehat{K}_i^j$ is the estimated PRNU factor of the ith frame/I-frame shot by a camcorder j. I_i^j is the ith frame/I-frame of a testing video with N_j frames/I-frames shot by a camcorder j whose PRNU factor is K^j. Given C different testing camcorders, $I_{i^p}^j$ denotes the frame/I-frame with a maximum of MSPCE value between a testing video shot by the camcorder j and a testing camcorder p and is defined as

$$i^p = \underset{i}{\arg\max} \, MSPCE\left(K^p, \widehat{K}_i^j\right) \coloneqq \left\{i | \forall l : MSPCE\left(K^p, \widehat{K}_i^j\right) \leq MSPCE\left(K^p, \widehat{K}_i^j\right)\right\} \quad (10)$$

The FPR of a testing video shot by the camcorder j based on the maximum statistic of MSPCE values, $FPR_{\max}(j)$, is defined as

$$FPR_{\max}(j) = 100 \times \sum_{p=1}^{C} D1(p, j)/C \quad (11)$$

$$D1(p, j) = \begin{cases} 1, & \text{if } p \neq j \text{ and } MSPCE\left(K^p, \widehat{K}_{i^p}^j\right) > TH \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where TH denotes a threshold value, 41.82, in the experiment. The FPR of a testing video shot by a camcorder j based on the mean statistic of MSPCE values, $FPR_{mean}(j)$, is defined as

$$FPR_{\mean}(j) = \sum_{p=1}^{C} D2(p, j)/C, \quad (13)$$

$$D2(p, j) = \begin{cases} 1, & \text{if } p \neq j \text{ and } \overline{MSPCE}(p, j) > TH \\ 0, & \text{otherwise} \end{cases}, \quad (14)$$

$$\overline{MSPCE}(p,j) = \sum_{i=1}^{N_j} MSPCE\left(\mathrm{K}^p, \widehat{\mathrm{K}}_i^j\right)/N_j. \tag{15}$$

The FNR of a testing video shot by the camcorder j based on the maximum statistic of MSPCE values, $FNR_{\max}(j)$, is defined as

$$FNR_{\max}(j) = 100 \times \sum_{p=1}^{C} D3(p,j)/C, \tag{16}$$

$$D3(p,j) = \begin{cases} 1, & \text{if } p = j \text{ and } MSPCE\left(\mathrm{K}^p, \widehat{\mathrm{K}}_{\bar{p}}^j\right) \leq TH \\ 0, & \text{otherwise} \end{cases}. \tag{17}$$

The FNR of a testing video shot by a camcorder j based on the mean of MSPCE values, $FNR_{mean}(j)$, is defined as

$$FNR_{\mean}(j) = \sum_{p=1}^{C} D4(p,j)/C \tag{18}$$

$$D4(p,j) = \begin{cases} 1, & \text{if } p = j \text{ and } \overline{MSPCE}(p,j) \leq TH \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

The overall FPR and FNR as $FPR_{\max} = \sum_{j=1}^{C} FPR_{\max}(j)/C$, $FPR_{mean} = \sum_{j=1}^{C} FPR_{mean}(j)/C$, $FNR_{\max} = \sum_{j=1}^{C} FNR_{\max}(j)/C$, and $FNR_{mean} = \sum_{j=1}^{C} FNR_{mean}(j)/C$, respectively, are used to evaluate the proposed method. Using the SCI/V in forensic investigations, the false-positive error, well known as the type I error, means the video in litigation is verified to an irrelative device by mistake. That means it might cause an innocent person to be arrested. In contrast, the false-negative error, well known as type II error, means the video in litigation is not verified to its source device. This means the video evidence might be useless. However, the suspect still can be identified according to other evidence. Therefore, false-positive errors should receive more attention than false-negative errors in forensic examinations [7].

5.2 Analysis of components of the proposed method

As discussed in section 2, the proposed method is modularized and flexible. In this experiment, three forms of our proposed methods by adapting different techniques in different steps are

Table 6 The different techniques applied to the three methods

Steps ^{Methods}	Our1	Our2	Our3
Artifacts extraction	Wavelet-based de-noising filter [22]	Wavelet-based de-noising filter [22]	BM3D [6]
Enhancement	ZM + WF [3]	SEA [18]	ZM + WF [3]
Source device determination	MSPCE and *TH* is 41.82	MSPCE and *TH* is 41.82	MSPCE and *TH* is 41.82

Table 7 The average processing time of the three methods of a 1920 × 1080 color image

	Our1	Our2	Our3
Average processing time	1.78 s	76.58 s	14.64 s

Table 8 The experimental result of the three methods

Error rate (%)	Our1	Our2	Our3
FPR_{max}	0	0.1796	0.0023
FPR_{mean}	0	0	0
FNR_{max}	0.0140	0.0163	0.0140
FNR_{mean}	0.0187	0.0280	0.0210

analyzed. Table 6 shows how we apply different techniques to those three forms in our proposed method, Our1, Our2, and Our3. The averaged processing time shown in Table 7 is conducted by processing a 1920 × 1080 color image 100 times and then calculates the average processing time. The videos in two groups, training and testing, each of which has 35 native format videos from Vision database. They are used to evaluate each form of our proposed method. The experimental result is shown in Table 8.

The experimental result in Table 7 shows the speed analysis result. Our proposed method (Our1) is the fastest and more than 8 times on processing time faster than the second one (Our2). The main reason is the BM3D de-noising algorithm [6] that processes in two dimensions cost too much time than the Wavelet-based de-noising filter that processes in one dimension. The experimental result in Table 8 shows the accuracy analysis result with zero FPR and very low FNR. According to [7], the false-positive errors in the forensic examination should receive more attention than false-negative errors in forensic examinations. It shows that the proposed method is very suitable for use in forensic investigation.

5.3 Comparison of other works

In this section, our proposed method is assessed in comparison with Yang et al. [27] and Chen et al. [3]. According to our best knowledge, Yang et al.'s method [27] was applied to the surveillance video forensic examination in 2010 in Taiwan and Chen et al.'s method [3] passed the Daubert challenge in July 2011 [8]. Both comparison methods are the only ones that have been used in real forensic examination cases. The comparative analysis includes two different aspects, running time and source camera forensic. The videos in two groups used in the foregoing experiment are used in this comparison experiment.

Table 9 Running time comparison (sec)

	Yang et al. [27]	Chen et al. [3]	Proposed method
The processing time per frame	0.9804	1.5406	1.7789
The average of processing frames per video	2048.77	2048.77	72.54
Processing time per video	2008.61	3156.33	129.04

Table 10 The result of the comparison

| Error rate (%)	Yang et al. [27]	Chen et al. [3]	Proposed method
FPR_{max}	0.0350	0.0070	0
FPR_{mean}	0.0047	0.0047	0
FNR_{max}	0.0093	0.0093	0.0140
FNR_{mean}	0.0140	0.0163	0.0187

A   Running time

The running time of different SCI/V methods is listed in Table 9. First, we ran each scheme to process 100 frames of a video and calculate the average of running time per frame. Then, the average of processing frames of videos is calculated. Besides, only I-frames are processed in our proposed method. Finally, the processing time is obtained by the result of the average running time multiplied by the average of processing frames. In Table 9, the result shows our proposed method is approximately 15 times on average faster than our previous work [27] and 24 times on average faster than Chen et al. [3]. So in practice, the proposed method is in great need, especially for large-scale video forensic investigations.

B   Source camcorder forensic

In this subsection, the accuracy of SCI/V of the proposed method is assessed in comparison with [3, 27]. The same videos in two groups chosen from Vision Database are used in this experiment. The MSPCE threshold set to 41.82 is used for all schemes. The experimental results of FPR and FNR are shown in Table 10. According to [7], false-positive errors should receive more attention than false-negative errors in forensic examinations. Our proposed method has zero FPR and lows FNR in the experiment. That means the proposed method does not cause an innocent person to be arrested and demonstrates our proposed method is suitable for use in video forensic investigation.

6 Discussions

In this paper, we propose a modularized and flexible method that can adapt different techniques in different steps. Our proposed method is assessed in comparison with methods
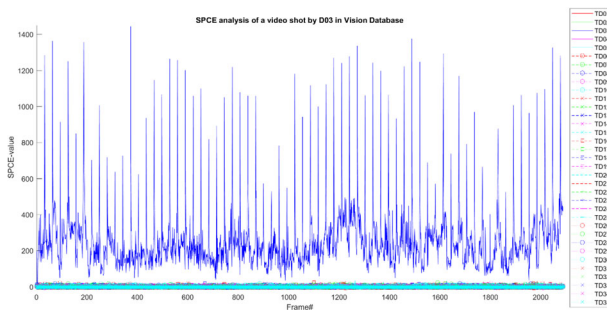
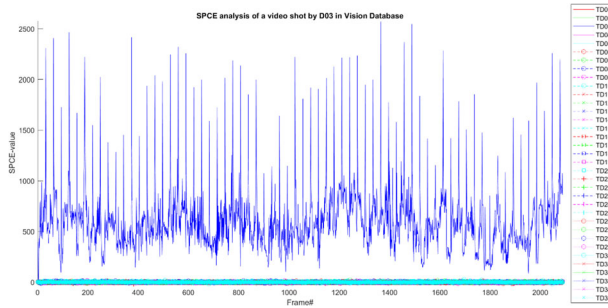**Fig. 3** (Best viewed in color) MSPCE analysis result of D03 video using [27]

Fig. 4 (Best viewed in color) MSPCE analysis result of D03 video using [3]

of Yang et al. [27] and Chen et al. [3]. Yang et al. [27] used Wiener filter to extract the device artifacts and skipped the enhancement step. Chen et al. [3] used the Wavelet-based de-noising filter to extract the device artifacts and the zero-mean and Wiener filter processes to enhance the extracted artifacts. Therefore, the running time is shorter than the work of Chen et al. [3]. Meanwhile, Chen' FPR [3] is lower than the work of Yang et al. [27] because of the enhanced PRNUs. Our proposed method can produce a significant increase in speed and reduce the false positive rate as compared to those two methods used in the video forensic examination.

Furthermore, we discover MSPCE values are significant to corresponding PRNU factors as regular peaks. Figs. 3 and 4 demonstrate those regular peaks correspond to MSPACE values in works of [3, 27], respectively. Through further analysis, these frames which cause the regular peaks are I-frames. The inter-frame predictive coding process might be the main reason that causes smaller MSPCE values. Only P- or B-frames use the inter-frame predictive coding process in video compression. That also proves I-frames used in our proposed method is more reliable than using P- or B- frames in PRNU analysis.

7 Conclusion and future work

In this paper, a new fast and efficient source camera identification and verification method based on PRNU analysis had been proposed. Based on the analysis of the device rolling and the characteristics of frames in videos, fast and efficient processing steps were proposed. The main contribution of this paper is to our knowledge, we are the first research team to discover the effect of the camera rolling on PRNU analysis and, in the meantime, consider both cameras rolling and I-frame of video to increase processing speed and also reduce the false positive rate as compared to existing methods used in the field of the forensic examination. Experimental results demonstrated our method can provide low processing time and a very low FPR SCI/V method for use in video forensics which meets the requirement of forensic examination.

Appendix

Some I-frames of the videos used for training and testing in the experiments are shown in this appendix. Figs. 5 and 6 show 2 I-frames of 8 testing videos and 8 testing videos, respectively.

It is worth noting that there exist the similar scenes and colors in videos from different portable devices in both Figs. 5 and 6.

(a) 1st I-frame of D01   (b) 36th I-frame of D01
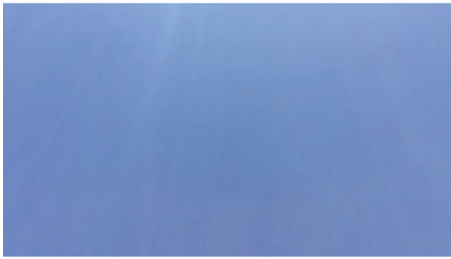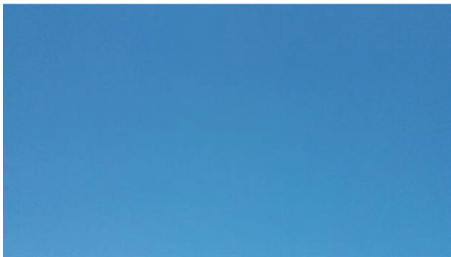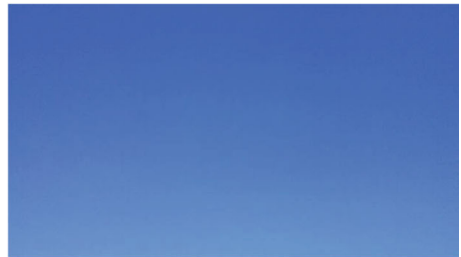
(c) 1st I-frame of D02   (d) 35th I-frame of D02

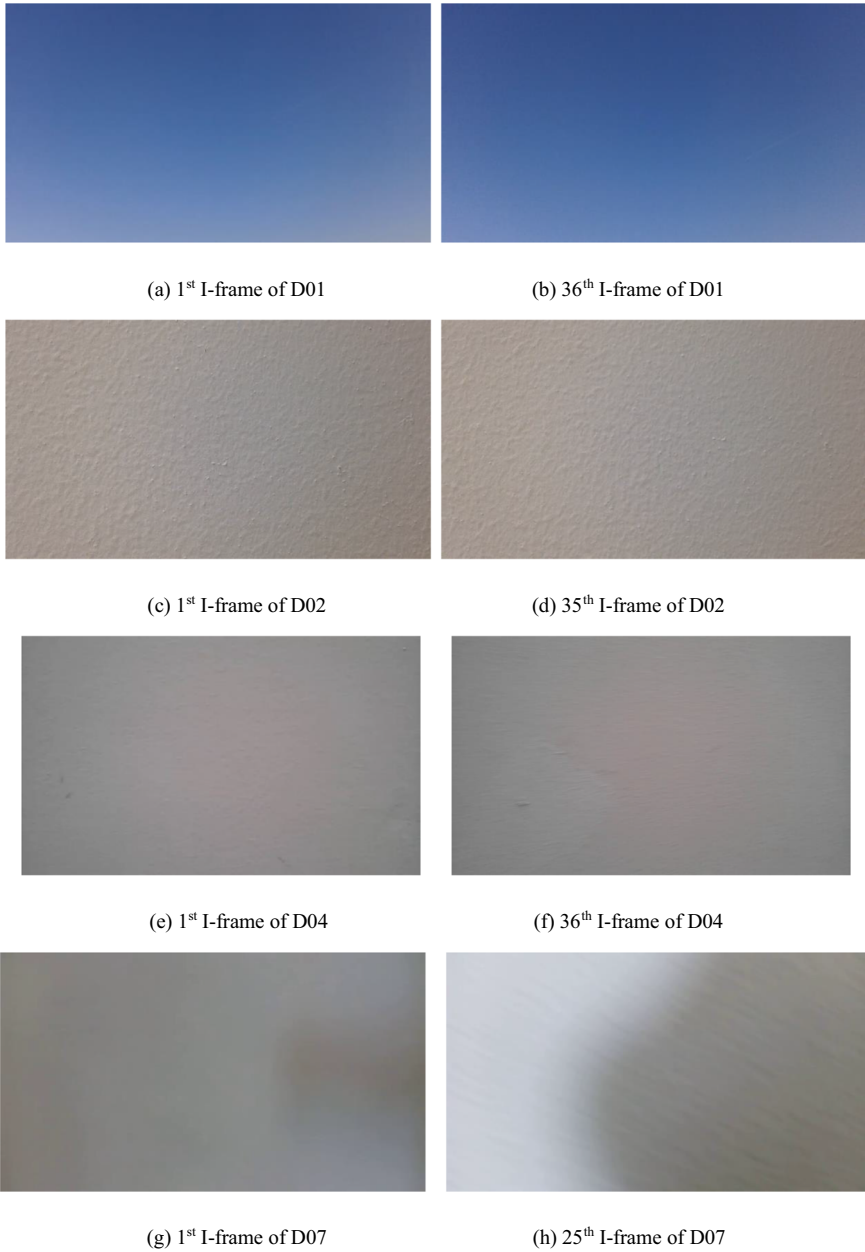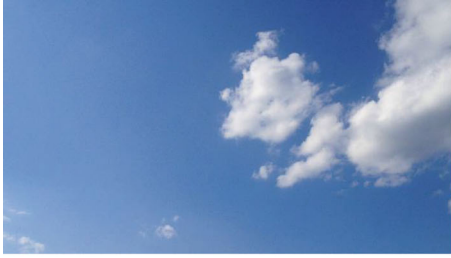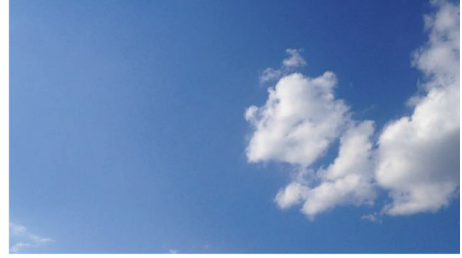(e) 1st I-frame of D04   (f) 36th I-frame of D04

(g) 1st I-frame of D07   (h) 25th I-frame of D07

Fig. 5 Some I-frames of the training videos in the experiments
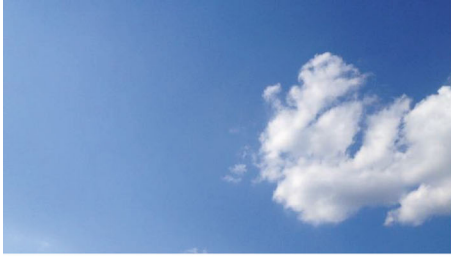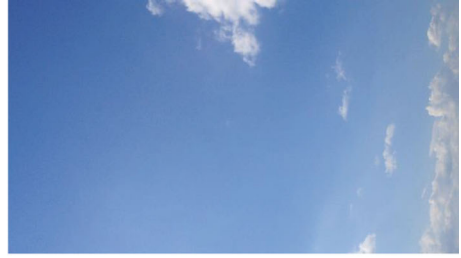
(i) 1st I-frame of D10   (j) 29th I-frame of D10
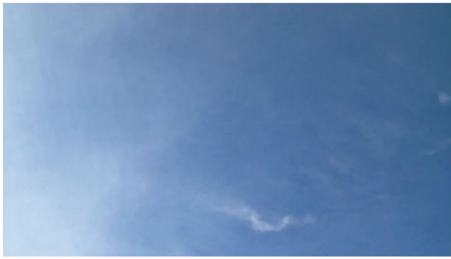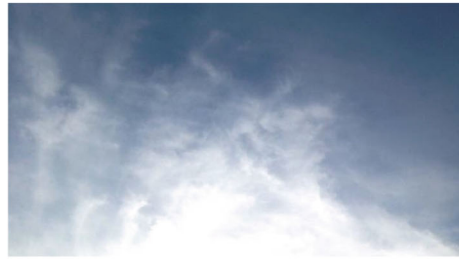
(k) 1st I-frame of D13   (l) 37th I-frame of D13

(m) 1st I-frame of D20   (n) 36th I-frame of D20

(o) 1st I-frame of D35   (p) 36th I-frame of D35

Fig. 5 (continued)

(a) 1st I-frame of D01   (b) 36th I-frame of D01

(c) 1st I-frame of D02   (d) 35th I-frame of D02

(e) 1st I-frame of D04   (f) 36th I-frame of D04

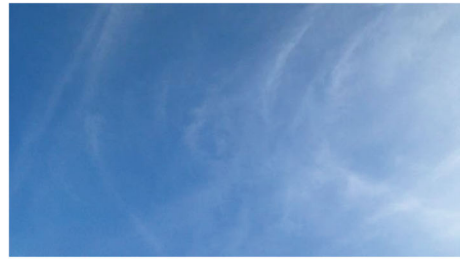(g) 1st I-frame of D07   (h) 25th I-frame of D07

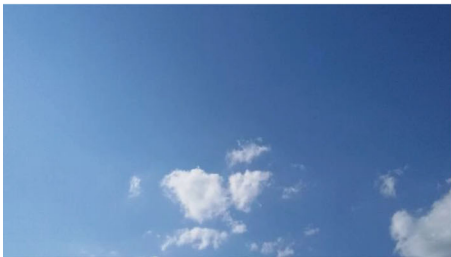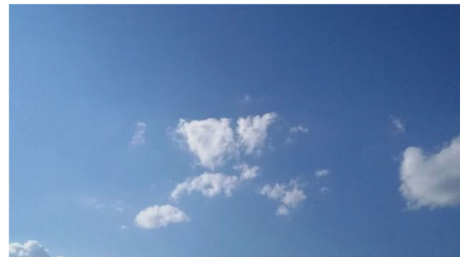**Fig. 6** Some I-frames of the testing videos in the experiments

(i) 1st I-frame of D10   (j) 29th I-frame of D10
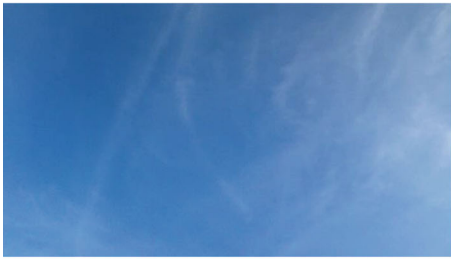
(k) 1st I-frame of D13   (l) 37th I-frame of D13
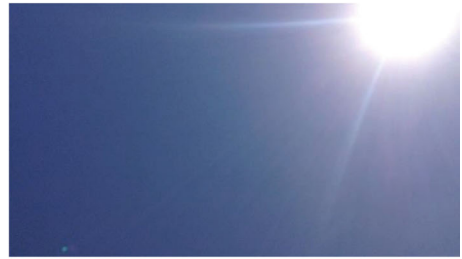
(m) 1st I-frame of D20   (n) 8th I-frame of D20

(o) 1st I-frame of D35   (p) 36th I-frame of D35

Fig. 6 (continued)

References

1. Bestagini P, Lubin L (2019) Special session 15: new challenges in video PRNU. In: IEEE Int. Conf. Image process. (ICIP) http://2019.ieeeicip.org/upload/files/201812121036101863.pdf. Accessed 10 Sept 2019
2. Chen M, Fridrich J, Goljan M, Lukás J (2007) Source digital camcorder identification using sensor photo-response non-uniformity. In: SPIE electron. Imag, Photonics West, pp 1G–1H. https://doi.org/10.1117/12.696519
3. Chen M, Fridrich J, Goljan M, Lukás J (2008) Determining image origin and integrity using sensor noise. IEEE Trans Inf Forensics Security 3(1):74–90. https://doi.org/10.1109/TIFS.2007.916285
4. Chuang W, Su H, Wu M (2011) Exploring compression effects for improved source camera identification using strongly compressed video. In: IEEE Int. Conf. Image process. (ICIP), pp.1953-1956. https://doi.org/10.1109/ICIP.2011.6115855
5. Cortiana A, Conotter V, Boato G, De Natale FGB (2011) Performance comparison of denoising filters for source camera identification. In: SPIE Media Watermarking, Security, and Forensics III 7880:788007. https://doi.org/10.1117/12.872489
6. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process 16(8):2080–2095. https://doi.org/10.1109/TIP.2007.901238
7. Du M (2017) Analysis of errors in forensic science. J Forensic Sci Med 3(3):139–143. https://doi.org/10.4103/jfsm.jfsm_8_17
8. Goljan M, Chen M, Comesaña P, Fridrich J (2016) Effect of compression on sensor-fingerprint based camera identification. In: IS&T Int. Symp. Electron. Imag., Media Watermarking, Security, and Forensics, San Francisco, CA, p 1–10. https://doi.org/10.2352/ISSN.2470-1173.2016.8.MWSF-086
9. Goljan M, Fridrich J (2012) Sensor-fingerprint based identification of images corrected for lens distortion. SPIE Media Watermarking, Security, and Forensics 8303:83030H. https://doi.org/10.1117/12.909659
10. Goljan M, Fridrich J, Filler T (2009) Large scale test of sensor fingerprint camera identification. In: SPIE Media Forensics and Security 7254:72540I. https://doi.org/10.1117/12.805701
11. Iuliani M, Fontani M, Shullani D, Piva A (2019) Hybrid reference-based video source identification. Sensors 19(3):649. https://doi.org/10.3390/s19030649
12. Kang X, Chen J, Lin K, Anjie P (2014) A context-adaptive spn predictor for trustworthy source camera identification. EURASIP J Image Video Process 2014(1):1–11. https://doi.org/10.1186/1687-5281-2014-19
13. Kang X, Li Y, Qu Z, Huang J (2012) Enhancing source camera identification performance with a camera reference phase sensor pattern noise. IEEE Trans Inf Forensics Security 7(2):393–402. https://doi.org/10.1109/TIFS.2011.2168214
14. Kurosawa K, Kuroki K, Saitoh N (1999) CCD fingerprint method-identification of a video camera from videotaped images. In: Int. Conf. Image Process. p 537–540. https://doi.org/10.1109/ICIP.1999.817172
15. Lawgaly A, Khelifi F (2016) Sensor pattern noise estimation based on improved locally adaptive DCT filtering and weighted averaging for source camera identification and verification. IEEE Trans Inf Forensics Security 12(2):392–404. https://doi.org/10.1109/TIFS.2016.2620280
16. Li C (2010) Source camera identification using enhanced sensor pattern noise. IEEE Trans Inf Forensics Security 5(2):280–287. https://doi.org/10.1109/TIFS.2010.2046268
17. Li C, Li Y (2012) Color-decoupled photo response non-uniformity for digital image forensics. IEEE Trans Circuits Syst Video Technol 22(2):260–271. https://doi.org/10.1109/TCSVT.2011.2160750
18. Lin X, Li C (2016) Preprocessing reference sensor pattern noise via spectrum equalization. IEEE Trans Inf Forensics Security 11(1):126–140. https://doi.org/10.1109/TIFS.2015.2478748
19. Lukas J, Fridrich J, Goljan M (2005) Digital 'bullet scratches' for images. In: IEEE Int. Conf. Image process. (ICIP), Genova, Italy, p III-65–III-68
20. Lukas J, Fridrich J, Goljan M (2006) Digital camera identification from sensor pattern noise. IEEE Trans Inf Forensics Security 1(2):205–214. https://doi.org/10.1109/TIFS.2006.873602
21. McCloskey S (2008) Confidence weighting for sensor fingerprinting. In: IEEE Computer Society Conf. Computer Vision and Pattern Recognition Workshops, p 23–28. https://doi.org/10.1109/CVPRW.2008.4562986
22. Mıhçak MK, Kozintsev I, Ramchandran K (1999) Spatially adaptive statistical modeling of wavelet image coefficients and its application to denoising. In: IEEE Int. Conf. Acoust., Speech, Signal Process. https://doi.org/10.1109/ICASSP.1999.757535
23. Öktem R, Egiazarian K, Lukin VV, Ponomarenko NN, Tsymbal OV (2007) Locally adaptive DCT filtering for signal-dependent noise removal. EURASIP J Adv Signal Process 2007:042472. https://doi.org/10.1155/2007/42472
24. Shullani D, Fontani M, Iuliani M, Shaya OA, Piva A (2017) VISION: a video and image dataset for source identification. EURASIP J Inf Security 2017:15. https://doi.org/10.1186/s13635-017-0067-2

25. Spy Blog - Watching Them, Watching Us (2019). Operation Algebra child rape convictions in Scotland: open WiFi tracking, digital camera image forensics. https://spyblog.org.uk/ssl/spyblog/2009/05/09/operation-algebra-child-rape-convictions-in-scotland-open-wifi-tracking-digi.html, Accessed 10 Sept 2019

26. Taspinar S, Mohanty M, Memon N (2016) Source camera attribution using stabilized video. In: IEEE Int Workshop Inf Forensics Security (WIFS). https://doi.org/10.1109/WIFS.2016.7823918

27. Yang W, Tsai L, Chen C (2017) Applying the sensor noise based camera identification technique to trace origin of digital images in forensic science. Forensic Sci J 16(1):19–42. https://doi.org/10.6593/FSJ.2017.1601.03