Contents lists available at ScienceDirect

Forensic Science International: Digital Investigation

journal homepage: www.elsevier.com/locate/fsidi

DFRWS EU 2025 - Selected Papers from the 12th Annual Digital Forensics Research Conference Europe

Video capturing device identification through block-based PRNU matching

Jian Li [a,b,*], Fei Wang [a,b], Bin Ma [a,b,**], Chunpeng Wang [a,b], Xiaoming Wu [a,b]

[a] Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China

[b] Shandong Provincial Key Laboratory of Industrial Network and Information System Security, Shandong Fundamental Research Center for Computer Science, Jinan, China

ARTICLE INFO

Keywords:
Forensics
Smartphone
Video
Stabilization
PRNU

ABSTRACT

This paper addresses the performance of a PRNU-based (photo response non-uniformity) scheme to identify the capturing device of a video. A common concern is PRNU in each frame being misaligned due to the video stabilization process compensating for unintended camera movements. We first derive the expectation of a similarity measure between two PRNUs: a reference and a test. The statistical analysis of the similarity measure helps us to understand the effect of homogeneous or heterogeneous misalignment of PRNU on the performance of identification for video capturing devices. We notice that dividing a test PRNU into several blocks and then matching each block with a part of the reference PRNU can decrease the negative effect of video stabilization. Hence a block-based matching algorithm for identifying video capturing devices is designed to improve the identification efficiency, especially when only a limited number of test video frames is available. Extensive experimental results prove that the proposed block-based matching algorithm can outperform the prior arts under the same test conditions.

1. Introduction

1.1. Motivation and problem formulation

Capturing device identification (CDI) has been serving as a valuable tool for addressing cybersecurity concerns, such as multi-factor authentication (Ba et al., 2018; Liu et al., 2023) and copyright protection (Qian et al., 2023). Video has been becoming a prevalent format of information sharing and entertainment on social media platforms, which makes video CDI of vital importance to researchers and policymakers. A promising technique for video CDI is to extract from a video distinguishable and unique traces left by its capturing device. The trace we analyze in this study is PRNU (photo-response non-uniformity) of the sensor of a capturing device (Fridrich, 2009). In the literature, the good properties of PRNU such as high dimensionality, robustness, and stability have been widely acknowledged (Altinisik et al., 2020; Al-Ani and Khelifi, 2017; Goljan et al., 2016; Mohanty et al., 2021; Kang et al., 2014; Zhang et al., 2023).

Given a test video and a capturing device, CDI can be formally formulated as a binary hypothesis testing problem, i.e.,

H_0: The device does not capture the test device.
H_1: The device captured the test video.

The testing result can be decided by checking the similarity between two PRNUs, namely,

$$\rho(P_R, P_T)\underset{H_0}{\overset{H_1}{\gtrless}}\theta, \tag{1}$$

where θ is a predefined threshold, and $\rho(\cdot,\cdot)$ represents a similarity measure which will be detailed in Section 3.1, P_R and P_T, called reference PRNU and test PRNU in this manuscript, represent the PRNU's estimated from the capturing device and the test video, respectively.

1.2. Limitation of prior art

We still face problems with video CDI if the given test video is

[*] Corresponding author. Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China.

[**] Corresponding author. Key Laboratory of Computing Power Network and Information Security, Ministry of Education, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China.

E-mail addresses: Ljian_20@163.com (J. Li), sddxmb@126.com (B. Ma).

https://doi.org/10.1016/j.fsidi.2025.301873

Available online 24 March 2025

captured by a modern smartphone, though steady progress has been made (Mandelli et al., 2020; Iuliani et al., 2019; Altinisik et al., 2020; Taspinar et al., 2020; Altinisik and Sencar, 2021). The video stabilization procedure in a smartphone to align the shaking video content makes the PRNU in each frame misaligned. As a result, we cannot congregate enough source-matched signals to remove compression and some other influences on the PRNU extracted. A possible solution is performing re-alignment for the PRNU sequence with the help of PRNU matching (Mandelli et al., 2020; Altinisik and Sencar, 2021). It requires plenty of still images available to extract a precise reference PRNU, with which we can take registration for the PRNU within a video frame. Because we usually cannot know how each video frame is transformed, the re-alignment process involves a brute-force search of geometrical transformation parameters. Considering that the transformation of a frame for stabilization could be heterogeneous, this search process is time-consuming and possibly enlarges the false alarm rate of CDI. Another solution is based on an experimental observation that we can obtain an accurate result of video CDI as long as enough video frames are available and efficiently used for extracting PRNU (Taspinar et al., 2020). However, it is still unknown how many frames are needed for a convincing CDI result.

1.3. Proposed approach

We first divide a test PRNU into blocks and then match each block with an essential part of the reference PRNU extracted from images, to attenuate the effect of video stabilization. The proposed method is based on our analysis of the varied adverse effects of video stabilization on different strategies for PRNU extraction. In addition to global transformation, the analysis concerns the local one, the parameters of which are uneasy to estimate via the registration method. We make an interesting observation from the analysis that the adverse effect of the local transformation on one strategy is similar to the global one's. The accuracy of CDI of a stabilized test video is mainly related to the number of effective matches. And the proposed block-based method can provide us with more probes.

1.4. Advantages over prior art and summary of experimental results

Our proposed block-based method integrates the advantages of the two previous solutions. In the first solution, we need to estimate a minimum of four parameters regarding the transformations of scaling (1), rotation (1), and translation (2), respectively. While our proposed block-based method only estimates the parameters associated with scaling and transition. We ignore rotation because matching with small blocks can help us decrease the rotation effect while noticeably improving the computational efficiency of CDI. The experiment is performed on a public video database. Test results show that the proposed strategy can make full use of each test video frame. The time complexity is also lower than the prior art that estimates all three kinds of geometrical transformation parameters.

Paper Organization: The rest of this paper is organized as follows. Section 2 shows the related works. In Section 3 we first introduce the basic techniques for extraction of PRNU from video and then introduce the PRNU's misalignment owing to video stabilization. Section 4 observes the effect of video stabilization on state-of-the-art strategy for video CDI, based on which a new block-based one is proposed in Section 5. We verify our analysis results in Section 6 with comprehensive experiments, followed by the conclusion in Section 7.

2. Related works

Taspinar et al. (2016) proposed to realign the frames which were assumed undergoing affine transformation caused by video stabilization. However, this method was only tested in a lab setting. Recently Mandelli et al. (2020) proposed a method to choose the frames in a video sequence that were aligned with each other. The experimental results in the published papers mentioned above showed that extracting PRNU from the video is uneasy to achieve good performance in comparison with extracting from the image. In (Iuliani et al., 2019) Iuliani et al. proposed a hybrid approach that mainly tried to use images instead of video clips to extract a reference PRNU. The motivation is that images and video clips should share the same PRNU as long as they both are captured by the same camera sensor. Furthermore (Iuliani et al., 2019), addressed the problem of how to match two PRNUs with different dimensions and the aforementioned misalignment due to video stabilization. Test results showed that realigning the PRNUs via brute-force search usually was memory- and computation-consuming work. Hence, Bellavia et al. (2019) proposed another wary for efficient PRNU alignment based on analyzing scene content. Besides, Mandelli et al. (2020) employed the particle swarm optimization method to search for the best transformation from image to video PRNU. And Altinisik et al. (Altinisik and Sencar, 2021) found the accuracy of estimation can be further improved by extending the search of the parameters of geometric transformation from 2-D to 3-D.

3. Revisiting PRNU extraction in case of video stabilisation

In this section, we first introduce the PRNU extraction method to make our manuscript self-contained, and then describe the technique of video stabilization equipped with a smartphone camera. Finally, we analyze the PRNU extracted from a stabilized video.

Everywhere in this article, we use a capital letter in Italics to represent a matrix of image, frame, or PRNU and use a capital letter in Blackboard font like $\mathbb{F}$ to represent a set of images, frames, or blocks. $|\mathbb{F}|$ represents the number of elements in $\mathbb{F}$. The PRNU extracted from an image or frame set $\mathbb{F}$ is denoted by $K(\mathbb{F})$. The probability of a random variable X is denoted by $P(X)$. Unless specified otherwise, the product of two matrices is element-wise, namely $Z[i,j] = X[i,j]Y[i,j]$ if $Z = XY$.

3.1. Video PRNU extraction and matching

We usually start PRNU extraction by removing the irrelevant signal from an image or video frame F, namely

$$W = F - \mathrm{f}(F), \tag{2}$$

where $\mathrm{f}(\bullet)$ is a filter designed to generate an ideal noise-free frame from the input. When only one image or frame is available for probing into video CDI, the residue W is taken as an estimated PRNU, even though it is usually inaccurate as no real filter $\mathrm{f}(\bullet)$ can provide us with a noise-free frame. On the other hand, if a large number of frames are available, we have more than one residual extracted from each frame to further improve the accuracy of extraction. Chen et al. proposed a video PRNU extraction method (Chen et al., 2007) based on an output model of a camera sensor. Specifically, they resolved one frame F into three parts,

$$F = F^{(0)} + F^{(0)} \cdot K + \Xi, \tag{3}$$

where $F^{(0)}$ represented an ideal frame depending exclusively on the input light, K was a PRNU's multiplicative factor together with $F^{(0)}$ producing PRNU, and Ξ represented all of the noises left, like distortion due to image compression and modeling error. Then $\widehat{K}$ was extracted from a sequence of frames via the following equation derived from max-likelihood estimation, namely

$$\widehat{K} = \frac{\sum_{k=1}^{m} W^k F^k}{\sum_{k=1}^{m} (F^k)^2}. \tag{4}$$

It is noted that most state-of-the-art methods employ (4) in estimating the reference PRNU for a given capturing device.

Refer to (1), PRNU-based CDI checks the similarity between a reference and the given test. One similarity measure used widely in the literature is normalized cross-correlation (NCC for short), i.e.,

$$\varrho(P_R, P_T) = \max_{m,n} \langle \frac{P_R - \overline{P_R}}{\|P_R - \overline{P_R}\|}, \frac{P_T[m,n] - \overline{P_T}}{\|P_T[m,n] - \overline{P_T}\|} \rangle, \tag{5}$$

where m, n defines the scope of sliding search to compensate for the translation between two PRNUs, viz P_R and P_T.

3.2. Video stabilization

There is an implied condition to use (4) to obtain an estimate of PRNU, i.e., all the images and frames involved should be aligned with each other. Nowadays most smartphones have an in-camera video stabilization system (VSS for short) to account for or compensate for the undesired movements of the camera during video capturing. The VSS obtains information on camera movement from motion sensors equipped with a smartphone. Each video frame is reversely transformed according to the sensor-recorded information. Because the shakiness of the camera varies with time, each frame undergoes different geometric transformations.

Furthermore, we notice that a frame may not be warped uniformly during video stabilization. Instead, sometimes a video frame is divided into pieces each of which makes an individual transformation. For instance, a technique called rolling shutter correction is usually integrated into VSS to reduce the so-called rolling shutter effect. This effect specifically occurs in complementary-metal-oxide-semiconductor (CMOS) sensor that is widely adopted by the current smartphone cameras (Saffih and Hornsey, 2007). Because a CMOS sensor sequentially outputs recorded pixel charge data row by row, there is a time lag between the top and the bottom of the pixels in one frame. As a consequence, the undesired movement of the camera will lead to the lines being uncoordinated with each other when great movement occurs. Because global transformation cannot reduce the rolling shutter effect, VSS tends to carry out heterogeneous transformations to a frame.

It is easy to see that VSS can introduce misalignment of PRNU in a sequence of frames. In particular, different parts of a PRNU noise may undergo diverse geometric transformations. While the traditional video PRNU extraction method requires that all the frames are aligned perfectly with each other, so as to isolate a reliable estimate of PRNU. In the next subsection, we will theoretically analyze the effect of video stabilization on PRNU estimates.

3.3. PRNU extracted from video transformed globally or locally for stabilization

Let $\mathbb{F} = \{F^1, F^2, \ldots, F^n\}$, where $F^i \in \mathbb{R}^{w \times h}$, represents a set of n video frames captured by the same camera. The frames in $\mathbb{F}$ may be transformed globally or locally for stabilization. The former processes every frame isometrically. The latter divides a frame into several blocks each of which is transformed with a different transformation matrix. In this light, we universally represent a frame F^i by the sum of a sequence of matrices, i.e., $F^i = \sum_{j=1}^{m} S_j^i$. The matrix $S_j^i \in \mathbb{R}^{w \times h}$, is obtained by keeping the j^{th} strip of F^i unchanged while setting all of the others to 0. Furthermore we have m sets each of which is $\mathbb{S}_j = \{S_j^1, S_j^2, \ldots, S_j^n\}, j \in [1, m]$, representing n strips with the same location in each frame of $\mathbb{F}$. The relationship between frames and strips is illustrated in Fig. 1.

For the clarity of analysis, in what follows we first focus our analysis on global transformation, i.e., the cardinality of $\mathbb{S}$ is equal to 1. Then we will further consider local transformation.

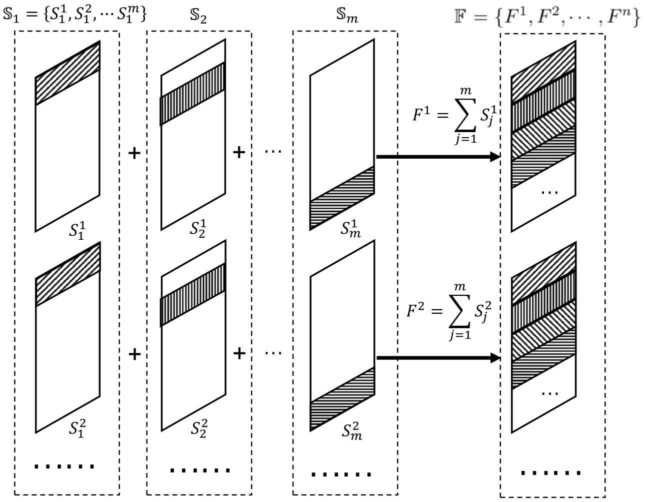
$\mathbb{S}_1 = \{S_1^1, S_1^2, \cdots S_1^m\}$
$\mathbb{S}_2$
$\mathbb{S}_m$
$\mathbb{F} = \{F^1, F^2, \cdots, F^n\}$
+
+
···
$F^1 = \sum_{j=1}^{m} S_j^1$
···
S_1^1
S_2^1
S_m^1
+
+
···
$F^2 = \sum_{j=1}^{m} S_j^2$
···
S_1^2
S_2^2
S_m^2
......
......
......
......

Fig. 1. Relationship between $\mathbb{F}$, $\mathbb{S}$, S and F.

Table 1
Transformation parameters estimated from ten consequent frames of a video. The transformation includes 3-D rotation, translation, and scale. We use quaternion $a + b\vec{i} + c\vec{j} + d\vec{k}$ to represent the 3-D rotation. And we use $\mathbf{x_0}$ and $\mathbf{y_0}$ to represent the translation between each frame PRNU and a reference PRNU. The reference PRNU is extracted from 50 flat images. We do not give the scaling parameter here for simplicity, considering that all the frames share the same value of 0.531.

a	b	c	d	$\mathbf{x_0}$	$\mathbf{y_0}$
0.544	−3.27319E-05	−0.006228296	0.01051058	521	256
0.544	−3.27319E-05	−0.006228296	0.01051058	521	256
0.543	−8.08286E-05	−0.009787058	0.01651693	520	255
0.543	−8.08286E-05	−0.009787058	0.01651693	520	255
0.543	−8.08286E-05	−0.009787058	0.01651693	520	255
0.543	−8.08286E-05	−0.009787058	0.01651693	520	255
0.543	−8.08286E-05	−0.009787058	0.01651693	520	256
0.543	−8.08286E-05	−0.009787058	0.01651693	520	256
0.543	−8.08286E-05	−0.009787058	0.01651693	520	256
0.543	−8.08286E-05	−0.009787058	0.01651693	520	256

The transformations that occurred in some video frames are broadly similar. Referring to Table 1, we estimate the parameters associated with frame transformation owing to stabilization via matching each frame PRNU with a reference PRNU. If the correlation coefficient ρ is larger than a predefined threshold ϵ, the parameters associated with the transformation taking place in this frame are recorded. This observation inspires us that the frame set $\mathbb{F}$ can be divided into many subsets, namely

$$\mathbb{F} = \mathbb{F}_0 \cup \mathbb{F}_1 \cup \mathbb{F}_2 \cup \cdots, \tag{6}$$

where a subset $\mathbb{F}_i = \{F_i^1, F_i^2, \ldots, F_i^t\}$ contains the frames that are associated with a transformation $\mathscr{T}_i$. In particular, $\mathbb{F}_0$ represents a set in which frames do not undergo any significant transformations. In this light, the PRNU associated with a stabilized video is a mixture of transformed matrices, mathematically stated with the following theorem.

Theorem 1. *The expectation of PRNU directly extracted from a stabilized video is a linear combination of PRNU each of which is a geometrical transformation from* $K(\mathbb{F}_0)$.

Proof. According to (4), the PRNU extracted from $\mathbb{F}$ is given below,

$$\begin{aligned} K(\mathbb{F}) &= \frac{\sum_{k=1}^{|\mathbb{F}|} W^k F^k}{\sum_{k=1}^{|\mathbb{F}|} (F^k)^2} \\ &= \frac{\sum_{k=2}^{|\mathbb{F}_0|} W_0^k F_0^k + \sum_{k=1}^{|\mathbb{F}_1|} W_1^k F_1^k + \cdots}{\sum_{k=1}^{|\mathbb{F}|} (F^k)^2} \\ &= \frac{\sum_{k=1}^{|\mathbb{F}_0|} (F_0^k)^2}{\sum_{k=1}^{|\mathbb{F}|} (F^k)^2} \cdot \frac{\sum_{k=1}^{|\mathbb{F}_0|} W_0^k F_0^k}{\sum_{k=1}^{|\mathbb{F}_0|} (F_0^k)^2} + \frac{\sum_{k=1}^{|\mathbb{F}_1|} (F_1^k)^2}{\sum_{k=1}^{|\mathbb{F}|} (F^k)^2} \cdot \frac{\sum_{k=1}^{|\mathbb{F}_1|} W_1^k F_1^k}{\sum_{k=1}^{|\mathbb{F}_1|} (F_1^k)^2} + \cdots \end{aligned} \tag{7}$$

It is easy to see that $\frac{\sum_{k=1}^{|\mathbb{F}_i|} (F_i^k)^2}{\sum_{k=1}^{|\mathbb{F}|} (F^k)^2}$ is an unbiased statistic of the probability of geometrical transformation performed on each frame. In this light, the expectation of $K(\mathbb{F})$ is

$$E(K(\mathbb{F})) = \sum_i P(F \in \mathbb{F}_i)K(\mathbb{F}_i) = \sum_i P(\mathscr{T}_i)K(\mathbb{F}_i), \tag{8}$$

where $K(\mathbb{F}_i)$ is the PRNU extracted from the frame set $\mathbb{F}_i$. Given the definition of $\mathbb{F}_i$, $K(\mathbb{F}_i)$ is a geometrical transformation from $K(\mathbb{F}_0)$.

Equation (8) gives the result that does not consider the effect of local transformation. We next consider the cardinality of $\mathbb{S}$ larger than one, in which case each frame is divided into a sequence of strips in the same manner. As a consequence, the video PRNU $\widehat{K}$ is segmented into a sequence of sub-PRNU noises,

$$K(\mathbb{F}) = K\left(\sum_{i=1}^{m} \mathbb{S}_i\right) = \sum_{i=1}^{m} K(\mathbb{S}_i), \tag{9}$$

where m is the number of strips in a frame or image, and $K(\mathbb{S}_i)$ represents a PRNU extracted from the i^{th} strip in each frame. We note again that $\mathbb{S}_i$ and $\mathbb{F}$ are equal in size. The elements in $\mathbb{S}_i$ are also misaligned with each other due to video stabilization. So $K(\mathbb{S}_i)$ can also be expressed in a linear combination manner as shown in (8), i.e.,

$$\mathbf{E}[K(\mathbb{S}_i)] = \sum_j P(\mathscr{T}_j)K_{j,i}, \tag{10}$$

where $K_{j,i}$ is a PRNU corresponding to the strips in $\mathbb{S}_i$ that undergo the same geometric transformation, i.e., $K(\mathbb{S}_{j,i})$. It is reasonable to assume that the probability distribution of the transformation to a strip is independent of its position. Hence we have $P(\mathbb{S}_{j,i}) = P(\mathbb{S}_j)$. Take (10) into (9), we obtain the PRNU regarding local geometric transformation as follows,

$$K(\mathbb{F}) = \sum_{i=1}^{m} \sum_j P(\mathscr{T}_j)K_{j,i} \tag{11}$$

In summary, the PRNU extracted from a sequence of frames $K(\mathbb{F})$ is a linear combination depending on the probability distribution of geometrical transformations performed on the frames. Hence, we cannot guarantee a reliable result of video CDI by matching the reference and the test PRNU unless they both are extracted from video sequences long enough to incorporate a sufficient quantity of identical geometrical transformations. Given the difficulty of estimating the distribution of geometrical transformations, accurately predicting the requisite number of frames for obtaining a dependable PRNU is a challenging task.

4. Effect of video stabilization

The reduction of NCC caused by video stabilization is related to the PRNU extraction strategy applied for video CDI. Hence in what follows we first summarize the possible strategies, and then further discuss the strategy most favorable for stabilized video in terms of the expectation of NCC. According to the central limit theorem, NCC under the H_0 hypothesis can be seen as a normally distributed random variable with zero mean. Hence, unless specified otherwise, in what follows we only discuss the NCC under the H_1 hypothesis.

4.1. CDI strategies based on PRNU extraction from video

The first plausible scenario is that only a set of video frames $\mathbb{F}$ are available to extract a reference PRNU $K(\mathbb{F})$. The following two strategies can be employed.

- $V1$: We estimate a test PRNU $K(\mathbb{T})$ from the frames $\mathbb{T}$ of the test video. And then calculate NCC for video CDI by $\rho(K(\mathbb{F}), K(\mathbb{T}))$.
- $V2$: A residue W is extracted from each frame of the test video. The maximum of all the correlation coefficients of all frames in $\mathbb{T}$ is used as a similarity measure, i.e., $\max_{F\in\mathbb{T}}\rho(K(\mathbb{F})\cdot F, W)$.

The following two strategies, $I1$ and $I2$, are applicable to CDI of an input test video when an image set $\mathbb{I}$ is available to extract a reference PRNU $K(\mathbb{I})$. The test PRNUs associated with $I1$ and $I2$ are extracted from the test video similarly with $V1$ and $V2$, respectively.

- $I1$: $\mathbb{T}$ denotes a frame set picked from the test video for extracting a test PRNU $K(\mathbb{T})$. CDI is determined by observing the similarity between K_T and K_R which can be calculated by correlation coefficient $\rho(K(\mathbb{I}), K(\mathbb{T}))$[1].
- $I2$: A sequence of test residue is extracted from each frame of a given test video (with all frames denoted by $\mathbb{T}$). And then match the residue sequence with the reference PRNU one by one, and the largest NCC is taken as the similarity measure, mathematically, $\max_{F\in\mathbb{T}}\rho(K(\mathbb{I})\cdot F, W)$.

Table 2 summarizes the four strategies given above. Strategy $I2$ is more recommended (Iuliani et al., 2019; Mandelli et al., 2020) because of its good performance on CDI of stabilized video.

4.2. Expectation of NCC with strategy I2

Define $\mathscr{G}$ the NCC between two specific PRNUs. One is extracted from a set of images $\mathbb{I}$. The other is extracted from a set of non-transformed frames $\mathbb{F}_0$. If $\mathbb{I}$ and $\mathbb{F}_0$ are captured by the same camera and have the same size, we can assume that $\mathscr{G}$ is a random variable with an expectation depending only on the camera. I.e.,

$$E(\mathscr{G}) = E(\varrho(K(\mathbb{I}), K(\mathbb{F}_0))) = E(\varrho(W, K(\mathbb{I})\cdot F)|F \in \mathbb{F}_0). \tag{12}$$

In this light, we have a theorem asserting the detection accuracy of video CDI with $I2$.

Theorem 2. *Let $K(\mathbb{I})$ be the reference PRNU extracted from a set of images, and $\mathbb{F}$ be a set of test frames possibly transformed for video stabilization. The expectation of the NCC associated with strategy* I2 *satisfies the following inequality,*

$$E(\max_{F\in\mathbb{F}}\varrho(W, K(\mathbb{I})\cdot F)) \leqslant \min(n\cdot P(\mathscr{T}_0)E(\mathscr{G}), E(\mathscr{G})), \tag{13}$$

where n is number of the frames in $\mathbb{F}$, and $P(\mathscr{T}_0)$ represents the probability of a frame in $\mathbb{F}$ being not transformed. **Proof.** See Appendix A for the proof.

Theorem 2 shows that the local and global transformations have a similar effect on strategy $I2$. This is not intuitive because the local transformation usually makes the misalignment of PRNU more complex in comparison with the global one. Besides, we note that the performance of strategy $I2$ relates to the number of video frames. The upper

Table 2
The similarity measures of the four possible strategies for PRNU extraction, viz V1, V2, I1, and I2.

Strategy	Reference	Test	Similarity Measure
V1	$\mathbb{F}$	$\mathbb{T}$	$\rho(K(\mathbb{F}), K(\mathbb{T}))$
V2	$\mathbb{F}$	$F \in \mathbb{T}$	$\max_{F\in\mathbb{T}}\rho(K(\mathbb{F})\cdot F, W)$
I1	$\mathbb{I}$	$\mathbb{T}$	$\rho(K(\mathbb{I}), K(\mathbb{T}))$
I2	$\mathbb{I}$	$F \in \mathbb{T}$	$\max_{F\in\mathbb{T}}\rho(K(\mathbb{I})\cdot F, W)$

[1] Here for simplicity of theoretical analysis, we assume K_R and K_T share the same dimension to avoid the analysis of transformation effect. But in our experiment, this kind of geometric transformation is involved.

bound of expectation of NCC associated with $I2$ can be close to the ground truth $E(\mathscr{G})$ with the increment of frame number. In other words, the effect of video stabilization can be well attenuated for strategy $I2$ if we have a large number of frames, which is well consistent with the statements given in (Taspinar et al., 2020). However, we cannot ignore that the increment of frame number also leads to a worsening of the false alarm rate. This issue will be addressed in the next section.

5. Block-based matching algorithm

5.1. Algorithm

The performance of strategy $I2$ can be improved further by a registration process (Iuliani et al., 2019; Mandelli et al., 2020; Altinisik and Sencar, 2021). PRNU registration is likely to provide us with a variety of transformed test PRNU to match with a reference. However, an accurate registration result is difficult to achieve, especially for a random signal like PRNU. The aforementioned prior arts tend to use 5–10 keyframes from a test video as long as an acceptable CDI result is obtained. Although most favorable for CDI, keyframes take a rather small proportion of the total frames in a test video. It may need one or more minutes of video to obtain the 5–10 keyframes, depending on the setup of video encoding. Furthermore, Fridrich has proved that the false alarm rate of detection will increase almost linearly with the number of matching tests (Fridrich, 2009). Namely, PRNU registration has to limit its searching space of parameters for good performance.

We can obtain hundreds of frames from a few seconds of video. Every frame is with different geometrical transformations. It is possible to attenuate the effect of video stabilization to make full use of all these frames. Specifically, we divide the test PRNU into a number of equal-sized blocks each of which is matched with a part of the reference PRNU. The proposed block-based PRNU matching algorithm is given in Algorithm 1.

Algorithm 1. Block-based PRNU Matching Algorithm for Video CDI

```
Algorithm 1: Block-based PRNU Matching Algorithm for Video CDI
Input: (1) Threshold $\tau$
       (2) Reference PRNU $K(\mathbb{I})$ extracted from an image set
       (3) Frames $\mathbb{F}$ from test video with unknown stabilization
Output: NCC result
Initialize N:=0
while $\mathbb{F}$ is not empty do
    Take a frame $F$ from $\mathbb{F}$ to extract a test PRNU $K(F)$
    for $s \in [a, b]$ do
        Resize $K(F)$ with scaling parameter $s$
        Segment the scaled $K(F)$ into $n$ blocks
        Match each block with a region in $K(\mathbb{I})$ as illustrated in Fig. 2
    end
    if the largest NCC is larger than $\tau$ then
        return 1
    end
end
```

The inputs of our proposed algorithm are reference PRNU extracted from images $\mathbb{I}$ and test video frames $\mathbb{F}$. After obtaining a test PRNU from each frame, we segment it into equal-sized blocks. Then each block is matched with a part of the reference PRNU to calculate NCC. CDI result is determined by comparing the threshold τ with the maximum NCC among all the divided blocks.

The proposed block-based PRNU matching algorithm is illustrated in Fig. 2.We experimentally find that a test PRNU extracted from a video with normal size, say 1920x1280, can be equally divided into 6 blocks. In other words, each block is with the same width and 1/6 height as the test PRNU. Besides, given a block of test PRNU, we should limit its matching to a partial reference PRNU to further improve the efficiency and performance of the proposed algorithm. Specifically, the heights of the upper and the lower regions that are excluded from matching in the reference PRNU are, respectively, equal to the heights of the upper and lower excluded regions in the test PRNU. As a result, no matter what translation occurs we can obtain an overlap between two matching regions. Refer to Fig. 2, each of the six blocks and its associated matching region in a reference is presented. Finally, we note that the proposed blocking method can overcome the influence of scaling because the video is usually downsized from the image.

5.2. Analysis

Video CDI can benefit from our proposed block-based matching algorithm in the following two aspects. First, we can exclude rotation from matching for registration. Referring to Fig. 3, the blocks located near the axis of rotation are much less modified than those located far from the axis. We note that video stabilization usually leads to a rather slight rotation of a frame. In this light, we can ignore the rotation of the blocks located near the rotation axis. Moreover, because the proposed algorithm takes the maximum of NCC calculated from all the blocks as the final matching result, it is possible to ignore rotation for all the blocks.

Second, dividing a PRNU into blocks can enlarge the number of probes. Because we can see PRNU as white noise, its statistics remain unchanged after blocking. Hence, given a PRNU without any geometrical transformation, the expectation of NCC associated with our proposed block-based PRNU matching (block NCC for short) is equal to that associated with traditional frame-based PRNU matching (frame NCC for short), i.e., $\mathscr{G}$. Furthermore, the block from a PRNU with geometrical transformation can be seen as a small-sized PRNU. Its expectation of NCC satisfies the inequality given by (13). In this light, we can moderately increase the accuracy of video CDI as the number of blocks is larger than the number of frames.

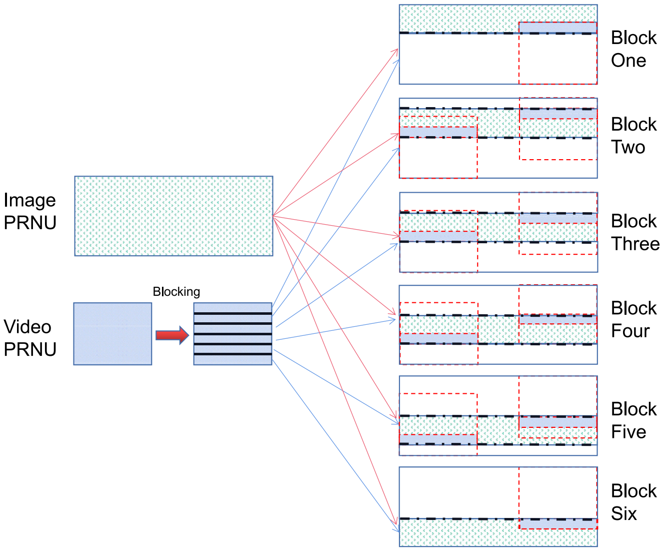

Fig. 2. The method for blocking a test PRNU to match with a part of a reference PRNU.

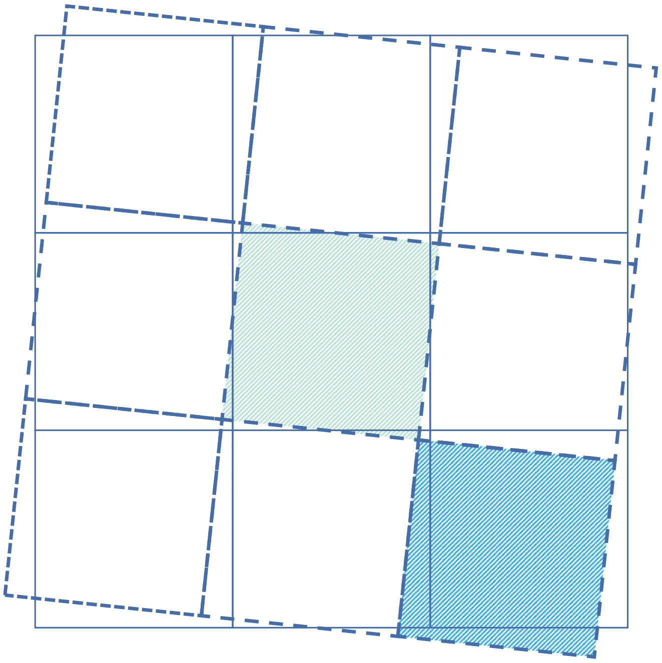

Fig. 3. Illustration of frame ration. The blocks on the edge of the frame are modified more greatly than those on the frame center.

6. Experimental results

6.1. Setup

The information about the capturing devices employed in our test is given in Table 3. We employ six Apple devices from the VISION dataset (Shullani et al., 2017) to verify our analysis and the proposed algorithm. These devices all automatically turn on the function of video stabilization. The videos collected in VISION dataset have plenty of content recording various outdoor and indoor scenarios. Besides, each scenario is captured with three different modes: still mode where the capturing device is stably held; move mode where the capturing device is held by a walking user; panrot mode where the user deliberately performs a video record combining a pan and a rotation on the devices.

6.2. Performance of comparison test

To show the performance of our proposed block-based algorithm, we first compare it with a traditional algorithm (Iuliani et al., 2019) based on frame matching under the same experimental setting. NCC is resistant to translation between two PRNUs. Hence we only need to consider how to compensate for the scale or rotate transformation. A PRNU extracted from the test video frame is scaled up first to fit with the reference PRNU. To speed up the brute-force searching process, we limit the search for the scale parameter in the range estimated by (Shullani et al., 2017) which has been given in Table 3. Besides, for the traditional frame-based algorithm, we need to estimate an exact rotation parameter within a search range of −2:0.2:2 in degree. For our proposed block-based algorithm, we only consider the estimation of scale parameters for each block.

Table 3
Detailed information of devices used in our experiments.

Model	ID in (Shullani et al., 2017)	Scaling (Iuliani et al., 2019)	#Videos	#Images
iPhone4S	D02	[0.748, 0.753]	13	50
iPhone5c	D05	[0.681, 0.691]	19	50
iPhone6	D06	[0.696, 0.713]	17	50
iPhone5c	D14	[0.681, 0.691]	19	50
iPhone6	D15	[0.696, 0.713]	18	50
iPad Mini	D20	[0.806, 0.821]	16	50

The performance of the video CDI test is related to the capturing mode and the number of available frames. Fig. 4 presents the ROC (receiver operating characteristic) curves associated with the block-based and the frame-based algorithms employed in identifying the video captured with move mode. Besides, to show the performance of the two algorithms being given a limited number of frames, our test only considers the second to the tenth frames, a total of nine frames. The first frame is not involved because it usually does not undergo geometrical transformation. Then we randomly choose different numbers of frames from each test video to perform CDI. It is easy to see from Fig. 4 that our proposed block-based algorithm outperforms the traditional frame-based algorithm. We know that the videos captured with move mode are strongly influenced by video stabilization. The test results show that our proposed block-based video CDI algorithm can solve this problem efficiently.

Besides, Fig. 5 shows the performance of CDI on the videos captured with still mode. This kind of video is seldom influenced by video stabilization. The experimental results indicate that our proposed block-based algorithm has achieved detection performance similar to that of the comparison algorithm. And by comparing the results of Figs. 4 and 5, it can be observed that the proposed algorithm is more capable of resisting the influence of video stabilization. Fig. 6 shows the performance of CDI on panrot videos. While this type of video is less common in reality, it can be used to assess the robustness of the video CDI algorithm against video stabilization. We can see that both of the two algorithms cannot achieve good performance when only a small number of frames is available. Nonetheless, the detection performance becomes acceptable when all of the nine frames are used for CDI. Actually, in our test using nine frames, the proposed algorithm demonstrates the ability to accurately associate each video with its respective capturing device, with no occurrence of false alarms.

Then we record the time cost associated with the two algorithms for video CDI to observe the speed improvement. The results in Table 4 show the average running time to test one frame of the videos given in Table 3. We test the two algorithms on the same computer with Core-i7 CPU and 32G RAM. The experimental results show that the proposed block-based algorithm is much faster than the traditional one. This is because estimating the parameters of rotation is rather time-consuming and challenging to obtain accurate results.

Finally, we note that the efficiency of the scheme for video CDI can be further improved by replacing the brute-force algorithm employed in the process of PRNU matching with particle swarm optimization (PSO) algorithm (Mandelli et al., 2020). Fig. 7 shows that the proposed block-based strategy can be used to further improve the performance of the PSO-based scheme. Here we only present the results associated with the case that only one frame is taken from the test video for simplicity. The AUC (area under the curve) values associated with the three algorithms for comparison are 0.91 (red line), 0.85 (dotted yellow line), and 0.72 (dotted blue line), respectively. Furthermore, the test results show that it can achieve comparatively good performance to limit the search range of rotation from $\pm 2^\circ$ to $\pm 1^\circ$. This observation is consistent with the analysis that extending the search range can improve both the accuracy rate and the false alarm rate. The proposed block-based algorithm of video CDI can achieve a better trade-off between the two rates than the prior arts. From Fig. 7, it can be observed that the green line eventually above the red line and the blue dashed line, indicating that the denser search parameter space can improve the final accuracy. However, this improvement is very limited.

7. Conclusion

This paper proposed an efficient block-based algorithm for video CDI via matching PRNUs. We mainly make the following two-fold

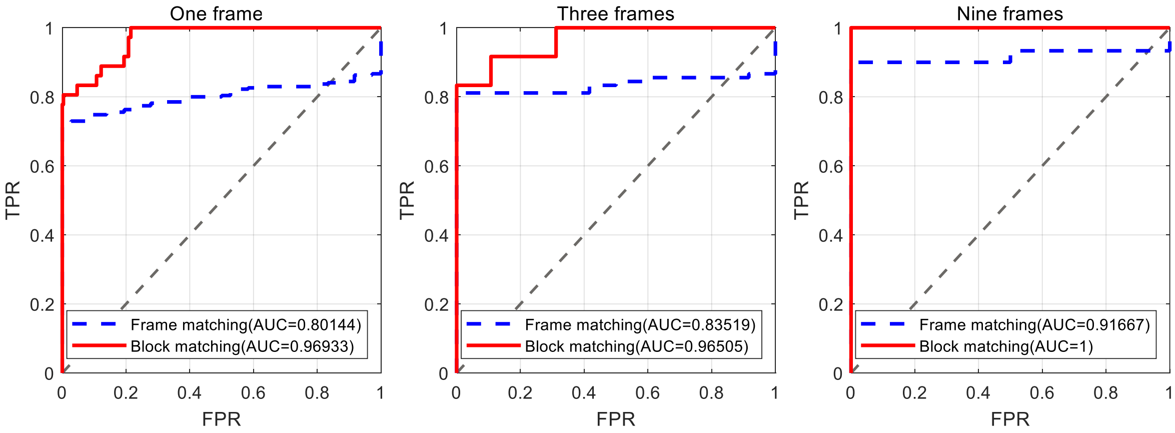

Fig. 4. ROC curves showing the performance of a test on the videos captured with move mode. The frame-based matching algorithm considers rotation, while the proposed block-based matching algorithm does not involve rotation. The test is performed by selecting from each video one frame, three frames, and nine frames.

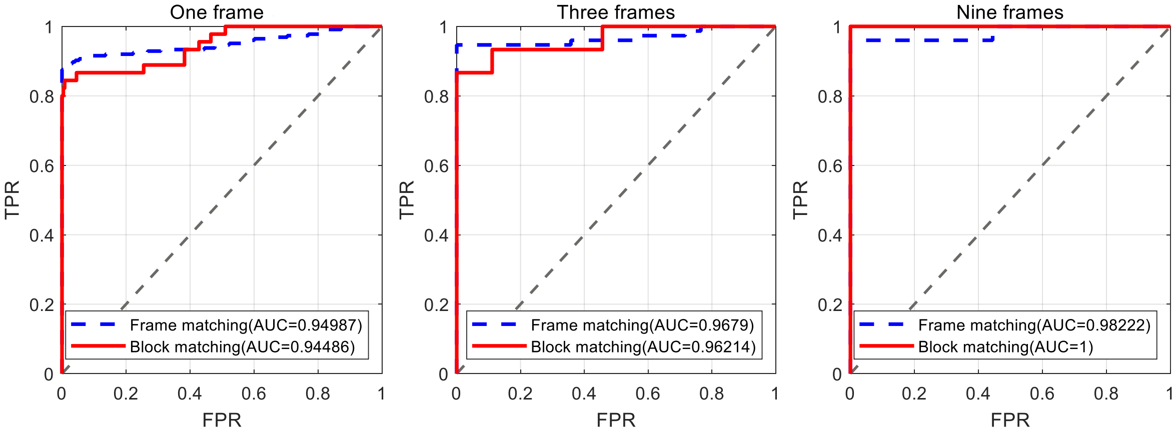

Fig. 5. ROC curves showing the performance of a test on the videos captured with still mode. In this mode, the video stabilization feature operates infrequently in comparison with move and panrot modes. The traditional frame-based algorithm achieves much better detection results than that in the move mode, while the proposed one remains stable. The test is performed by selecting from each video one frame, three frames, and nine frames.

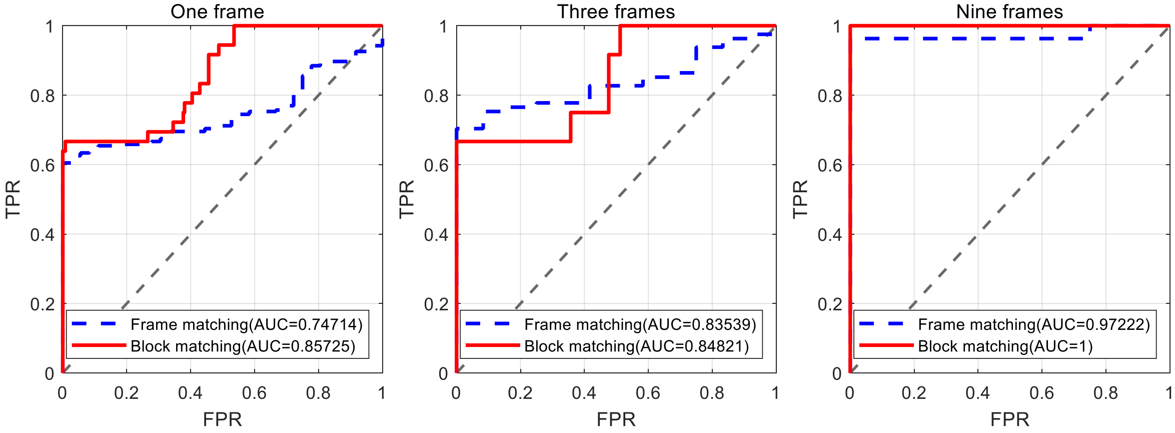

Fig. 6. ROC curves showing the performance of a test on the videos captured with panrot mode. In this capturing mode, the video stabilization feature requires significant processing of the video to stabilize the content. The test is performed by selecting from each video one frame, three frames, and nine frames.

contributions.

- We analyze the effect of video stabilization on CDI when the $I2$ strategy is employed for PRNU extraction. The results show that there is a linear relationship between the number of test video frames and the expectation of NCC.
- Based on the analysis results, we propose a block-based matching algorithm for video CDI. Because the proposed algorithm effectively enlarges the number of matches between test PRNU and reference PRNU, the performance of the video CDI is remarkably improved when a test video has a limited number of frames.

Table 4
Running time (in seconds) of video CDI with our proposed block-based algorithm and the traditional frame-based one. We show a running time per frame averaged over all test videos associated with each device.

Model	Frame-based	Block-based
iPhone4S (D02)	831.8	220.9
iPhone5c (D05)	3482.2	396.8
iPhone6 (D06)	5348.5	678.1
iPhone5c (D14)	1211.1	102.2
iPhone6 (D15)	1777.4	220.8
iPad Mini(D20)	1943.2	422.3
Average	**2432.3**	**340.2**

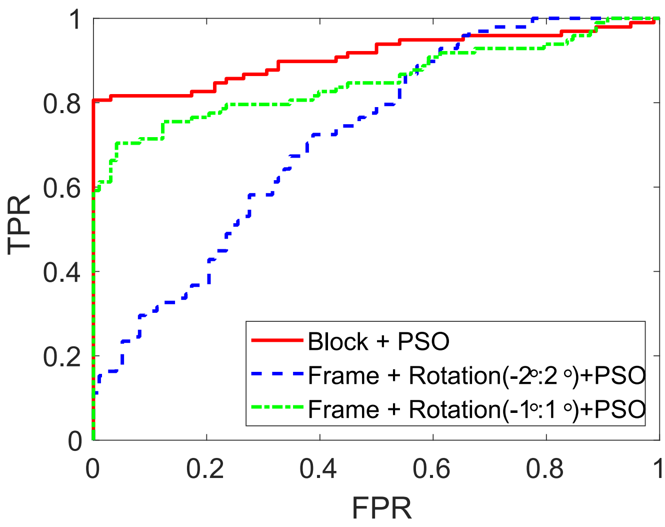

Fig. 7. ROC performance of a video CDI scheme using PSO and the proposed block-based matching algorithm (red line), and the frame-based matching algorithm. The latter algorithm additionally considers rotation with two different ranges (dotted yellow line −1°: 1° and dotted blue line −2°: 2°). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Finally, we note that extracting reference PRNU from still images may become unreliable, though it is the best choice currently. Modern smartphones like to be equipped with multiple rear cameras to obtain pro-level photography experience. And a photo is taken by combining the output images recorded by different cameras. As a result, the PRNU extracted from these images becomes a combination of PRNU blocks. We will address this issue in our future work based on the proposed block-based matching algorithm.

Acknowledgements

This work was supported by Jinan “20 Universities” Project of Jinan Research Leader Studio (2020GXRC056); Jinan “New 20 Universities” Project of Introducing Innovation Team (202228016); National Natural Science Foundation of China (62272255, 62302248, 62302249); National Key Research and Development Program of China (2021YFC3340600, 2021YFC3340602); Taishan Scholar Program of Shandong (tsqn202306251); Shandong Provincial Natural Science Foundation (ZR2020MF054, ZR2023QF018, ZR2023QF032, ZR2022LZH011); Ability Improvement Project of Science and Technology SMES in Shandong Province (2022TSGC2485, 2023TSGC0217); Youth Innovation Team of Colleges and Universities in Shandong Province (2022KJ124); The “Chun-hui Plan” Cooperative Scientific Research Project of Ministry of Education (HZKY20220482); Achievement Transformation of Science, Education, and Production Integration Pilot Project (2023CGZH-05); Integration pilot Project of Science Education Industry under Grant (2023PX006, 2023PY060, 2023PX071).

Appendix A

A Expectation of NCC Regarding Strategy I2

We will separately discuss the expectation of NCC associated with the strategy *I*2 designed to CDI of a test video being stabilized in two different ways, namely, global transformation and local transformation. First, if the frames of the test video are all globally transformed, the expectation of NCC is

$$\max_{F\in\mathbb{T}} E(\varrho(W, K(\mathbb{I})\cdot F), \tag{14}$$

where W is a residue extracted from frame F via (2). Based on the assumption that PRNU is white Gaussian, the largest NCC under H_1 hypothesis (refer to (1)) is $\mathscr{G}$, obtained from the frames aligned with $\mathbb{I}$, i.e. $F\in\mathbb{T}_0$. On the other hand, if there is no frame aligned with $\mathbb{I}$, NCC is close to zero. So the result of (14) depends on the probability of $\mathbb{T}$ containing a frame without transformation, namely $P(\mathbb{T}_0)$. This probability is related to, n, the number of frames in $\mathbb{T}$ and can be modeled with a binomial distribution, mathematically,

$$\begin{aligned}\max_{F\in\mathbb{T}} E(\varrho(W,K(\mathbb{I})\cdot F) &= (1-(1-P(\mathbb{T}_0))^n)\cdot E(\mathscr{G}) \\ &\leqslant \min(n\cdot P(\mathbb{T}_0)E(\mathscr{G}), E(\mathscr{G})).\end{aligned} \tag{15}$$

Next, we consider the case of strategy *I*2 facing local transformation where a test PRNU is seldom aligned with a reference globally. Nevertheless, there are some strips of the test not undergoing any geometric transformation during the process of video stabilization. And hence these strips can be matched with the reference. Given a frame F in test frame set $\mathbb{T}$, we have a NCC associated with F as follows,

$$\varrho(K(\mathbb{I})\cdot F, W) \tag{16}$$

Furthermore, split the frame F into strips, (16) changes to

$$\varrho\left(K(\mathbb{I})\cdot\sum_{j=1}^{m}S^{j},\sum_{j=1}^{m}W^{j}\right), \tag{17}$$

where S^j represents the j^{th} strip of the frame F, and W^j represents a residue extracted from S^j. A strip without any transformations is associated with the maximum NCC, no matter where the strip is located in the frame, i.e.,

$$E\left(\varrho(K(\mathbb{I})\cdot S^{j}, W^{j}|S^{j}\in\mathbb{S}_{0}\right) = E\left(\varrho\left(K(\mathbb{I})\cdot S^{i}, W^{i}|S^{i}\in\mathbb{S}_{0}\right).\tag{18}$$

Adding together all the non-transformed strips from various positions we have a NCC as follows,

$$\begin{aligned}\sum_{j=1}^{m}E\left(\varrho(K(\mathbb{I})\cdot S^{j}, W^{j})|S^{j}\in\mathbb{S}_{0}\right) &= E(\varrho(K(\mathbb{I})\cdot F, W)|F\in\mathbb{F}_{0})\\ &= E(\mathscr{G}),\end{aligned}\tag{19}$$

Combining (18) and (19) we have

$$\varrho\left(K(\mathbb{I})\cdot S^{j}, W^{j}\right) = \begin{cases}\dfrac{\mathscr{G}}{m}, & S^{j}\in\mathbb{S}_{0}\\ 0, & S^{j}\notin\mathbb{S}_{0}\end{cases}\tag{20}$$

where m is the number of strips in each frame. It is easy to see that the number of non-transformed strips in a frame is closely related to the expectation of NCC.

Given a test frame $F\in\mathbb{T}$, we denote the set of non-transformed strips in F by

$$\mathbb{S}_{0}^{F} = \left\{S^{i}|S^{i}\in\mathbb{S}_{0}, F=\sum_{i}S^{i}, F\in\mathbb{F}\right\}.\tag{21}$$

Such that the expectation of NCC associated with $\mathbb{T}$ is

$$\begin{aligned}&E(\max\{\varrho(K(\mathbb{I})\cdot F, W)|F\in\mathbb{T}\}) = P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_{0}^{F}|\}=1)\cdot E(\varrho_{1})\\ &+P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_{0}^{F}|\}=2)\cdot E(\varrho_{2}) + \cdots P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_{0}^{F}|\}=m)\cdot E(\varrho_{m})\\ &=\sum_{i=1}^{m}P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_{0}^{F}|\}=i)\cdot E(\varrho_{i}),\end{aligned}\tag{22}$$

where $P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_0^F|\}=i)$ represents the probability that in one frame of $\mathbb{F}$ there are i non-transformed strips at most, and ϱ_i represents the NCC associated with one frame containing i non-transformed strips. Refer to (20) which gives us the NCC associated with one non-transformed strip in a frame, we have

$$\varrho_{i} = i\frac{\mathscr{G}}{m}.\tag{23}$$

So the challenge of calculating the expectation in (22) is to obtain the probability $P(\max_{F\in\mathbb{F}}\{|\mathbb{S}_0^F|\}=i)$. Denote P_i the probability that a frame has i non-transformed strips. Such that P_i should follow a binomial distribution, i.e.,

$$P_{i} = C_{m}^{i}\cdot P^{i}(1-P)^{m-i},\tag{24}$$

where P is the probability of a strip being not transformed, i.e., $P(\mathbb{S}_0)$. The simplest case is $i=1$, i.e. in any of n frames at most one strip is non-transformed. So the possible cases include there is only one frame with a non-transformed strip, and all the other $n-1$ frames are all with transformed strips, or there are two or more frames individually with a non-transformed strip, and the strips in the rest frames are all transformed, mathematically,

$$\begin{aligned}P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_{0}^{F}|\}=1) &= C_{n}^{1}\cdot P_{1}P_{0}^{n-1} + C_{n}^{2}\cdot P_{1}^{2}P_{0}^{n-2} + \cdots C_{n}^{n}\cdot P_{1}^{n}P_{0}^{0}\\ &= (P_{0}+P_{1})^{n} - C_{n}^{0}\cdot P_{1}^{0}P_{0}^{n}\\ &= (P_{0}+P_{1})^{n} - P_{0}^{n}.\end{aligned}\tag{25}$$

In the same manner, we can derive the probability there are at most two non-transformed strips in one frame as follows,

$$\begin{aligned}P(\max_{F\in\mathbb{T}}\quad\{|\mathbb{S}_{0}^{F}|\}=2) &= C_{n}^{1}\cdot P_{2}^{1}(P_{0}+P_{1})^{n-1} + C_{n}^{2}\cdot P_{2}^{2}(P_{0}+P_{1})^{n-2}+\\ &\cdots C_{n}^{n}\cdot P_{2}^{n}(P_{0}+P_{1})^{0} = (P_{2}+P_{1}+P_{0})^{n} - (P_{0}+P_{1})^{n}.\end{aligned}\tag{26}$$

Equations (25) and (26) interprets the probability $P(\max|\mathbb{S}_0^F| = n|F\in\mathbb{T})$ from another point of view. I.e., in each frame at most, there are i strips non-transformed, such that the probability for this event is $P_0+P_1+\cdots P_i$. And given n frames independently transformed, the probability that we have a frame with i strips non-transformed is

$$P(\max_{F\in\mathbb{T}}\{|\mathbb{S}_0^F|\}=n) = (P_i + P_{i-1} + \cdots P_0)^n - (P_{i-1} + P_{i-2} + \cdots P_0)^n, \tag{27}$$

where minus $(P_{i-1} + P_{i-2} + \cdots P_0)^n$ indicates to remove the cases that some frames have less than i non-transformed strips.

With (27) and (23), we are ready to calculate the expectation given in (22), i.e.,

$$E(\max\{\varrho(K(\mathbb{I})\cdot F, W)|F\in\mathbb{T}\}) = U\cdot\frac{E(\mathscr{G})}{m}, \tag{28}$$

where

$$\begin{aligned} U &= (P_0 + P_1)^n - P_0^n + 2*[(P_0 + P_1 + P_2)^n - (P_0 + P_1)^n] \\ &\quad + \cdots + m*[(P_0 + P_1 + \cdots P_m)^n - (P_0 + P_1 + \cdots P_{m-1})^n] \\ &= m\cdot\left(\sum_{i=0}^{m} P_i\right)^n - \left(\sum_{i=0}^{m-1} P_i\right)^n - \left(\sum_{i=0}^{m-2} P_i\right)^n \cdots - P_0^n \end{aligned} \tag{29}$$

Considering $\sum_{i=0}^{m} P_i = 1$, we have

$$\begin{aligned} U &= m - [(1 - P_m)^n + (1 - P_m - P_{m+1})^n + \cdots] \\ &< m - [1 - nP_m + 1 - n(P_m + P_{m-1}) + 1 - n(P_m + P_{m-1} + P_{m-2})\cdots \\ &\quad + 1 - n(P_m + P_{m-1} + \cdots + P_1)] = n\cdot[mP_m + (m-1)P_{m-1} + \cdots P_1] \end{aligned} \tag{30}$$

It is easy to see that the term in the square bracket is the expectation of a random variable that indicates the number of strips non-transformed. As defined in (24), the random variable follows a binomial distribution. Hence its expectation is $m\cdot P$ and we have

$$E(\max\{\varrho(K(\mathbb{I})\cdot F, W)|F\in\mathbb{T}\}) < n\cdot m\cdot P\cdot\frac{E(\mathscr{G})}{m} = n\cdot P\cdot E(\mathscr{G}) \tag{31}$$

Considering the correlation coefficient cannot be larger than $\mathscr{G}$, we obtain the result,

$$E(\max\{\varrho(K(\mathbb{I})\cdot F, W)|F\in\mathbb{T}\})\leqslant\min(n\cdot P\cdot E(\mathscr{G}), E(\mathscr{G})). \tag{32}$$

References

Al-Ani, M., Khelifi, F., 2017. On the spn estimation in image forensics: a systematic empirical evaluation. IEEE Trans. Inf. Forensics Secur. 12, 1067–1081.

Altinisik, E., Sencar, H.T., 2021. Source camera verification for strongly stabilized videos. IEEE Trans. Inf. Forensics Secur. 16, 643–657.

Altinisik, E., Tasdemir, K., Sencar, H.T., 2020. Mitigation of h.264 and h.265 video compression for reliable prnu estimation. IEEE Trans. Inf. Forensics Secur. 15, 1557–1571.

Ba, Z., Piao, S., Fu, X., Koutsonikolas, D., Mohaisen, A., Ren, K., 2018. Abc: Enabling smartphone authentication with built-in camera. In: 25th Annual Network and Distributed System Security Symposium. NDSS.

Bellavia, F., Iuliani, M., Fanfani, M., Colombo, C., Piva, A., 2019. Prnu pattern alignment for images and videos based on scene content. In: IEEE International Conference on Image Processing. ICIP), pp. 91–95.

Chen, M., Fridrich, J., Goljan, M., Lukáš, J., 2007. Source digital camcorder identification using sensor photo response non-uniformity. In: III, E.J.D., Wong, P.W. (Eds.), Security, Steganography, and Watermarking of Multimedia Contents IX. International Society for Optics and Photonics. SPIE, pp. 517–528.

Fridrich, J., 2009. Digital image forensic using sensor noise. IEEE Signal Process. Mag. 26, 26–37.

Goljan, M., Chen, M., Comesaña, P., Fridrich, J., 2016. Effect of compression on sensor-fingerprint based camera identification. Electron. Imag. 2016, 1–10.

Iuliani, M., Fontani, M., Shullani, D., Piva, A., 2019. Hybrid reference-based video source identification. Sensors(Basel) 19 (3).

Kang, X., Chen, J., Lin, K., Peng, A., 2014. A context-adaptive spn predictor for trustworthy source camera identification. EURASIP J.Image Video Process. 2014, 1–11.

Liu, L., Fu, X., Chen, X., Wang, J., Ba, Z., Lin, F., Lu, L., Ren, K., 2023. Fits: matching camera fingerprints subject to software noise pollution. In: Proceedings of the 2023 ACM SIGSAC Conference on Computer and Communications Security, New York, NY, USA, pp. 1660–1674.

Mandelli, S., Bestagini, P., Verdoliva, L., Tubaro, S., 2020. Facing device attribution problem for stabilized video sequences. IEEE Trans. Inf. Forensics Secur. 15, 14–27.

Mohanty, M., Zhang, M., Asghar, M., Russello, G., 2021. e-prnu: Encrypted domain prnu-based camera attribution for preserving privacy. IEEE Trans. Dependable Secure Comput. 18, 426–437.

Qian, F., He, S., Huang, H., Ma, H., Zhang, X., Yang, L., 2023. Web photo source identification based on neural enhanced camera fingerprint. In: Proceedings of the ACM Web Conference, pp. 2054–2065.

Saffih, F., Hornsey, R., 2007. Reduced human perception of fpn noise of the pyramidal readout cmos image sensor. IEEE Trans. Circ. Syst. Video Technol. 17, 924–930.

Shullani, D., Fontani, M., Iuliani, M., Shaya, O.A., Piva, A., 2017. Vision: a video and image dataset for source identification. EURASIP J. Inf. Secur. 2017, 15.

Taspinar, S., Mohanty, M., Memon, N., 2016. Source camera attribution using stabilized video. In: IEEE International Workshop on Information Forensics and Security (WIFS), pp. 1–6.

Taspinar, S., Mohanty, M., Memon, N., 2020. Camera fingerprint extraction via spatial domain averaged frames. IEEE Trans. Inf. Forensics Secur. 15, 3270–3282.

Zhang, Y., Tan, Q., Qi, S., Xue, M., 2023. Prnu-based image forgery localization with deep multi-scale fusion. ACM Trans. Multimed Comput. Commun. Appl 19, 1–20.