Digital Forensic Analysis for Source Video Identification: a Survey

Younes Akbari*, Somaya Al-maadeed*, Omar Elharrouss*
Fouad Khelifi**, Ashref Lawgaly**, and Ahmed Bouridane***

*Department of Computer Science and Engineering , Qatar University, Doha, Qatar,
**Department of Computer and Information Sciences, Northumbria University, UK,
***Cybersecurity and Data Analytics Research Center,University of Sharjah, Sharjah, UAE.

Abstract

In recent years, many digital devices have been equipped with a video camera that allows videos to be recorded in good quality, free of charge and without restrictions. Concurrently, the widespread use of digital videos via web-based multimedia systems and mobile smartphone applications such as YouTube, Facebook, Twitter and WhatsApp is becoming increasing important. However, security challenges have emerged and are spreading worldwide. These issues may lead to serious problems, particularly in situations where video is a key part of decision-making in crimes, including movie piracy and child pornography. Thus, to increase the trustworthiness of using digital video in daily life, copyright protection and video authentication must be used. Although source camera identification based on digital images has attracted many researchers' attention, less research has been performed on the forensic analysis of videos due to certain challenges, such as compression, stabilization, scaling, and cropping, as well as differences between frame types that can occur when a video is stored in digital devices. Thus, there are insufficient large standard digital video databases and updated databases with new devices based on new technologies. The goal of this paper is to offer an inclusive overview of what has been done over the last decade in the field of source video identification by examining existing techniques, such as photo response nonuniformity (PRNU) and machine learning approaches, and describing some popular video databases.

Keywords:
Survey, Source camera identification, Video, PRNU, Machine learning methods.

1. Introduction

Digital video was commercially introduced in Sony's D-1 format in 1986. The recording of videos as large amounts of data with information that can be used as a type of evidence in court is now developing rapidly and widely. In addition to its use in social media, digital video data is subject to checking for authenticity, integrity, and identification. Integrity ensures that video evidence has not been altered prior to seizure, while authenticity is the ability to determine whether the evidence is credible. An important aspect of digital video recording is identifying the digital device that was used to record a video Huang et al. (2015). Many forensic methods have been developed based on digital images Lukas et al. (2006); Chen et al. (2008); Lawgaly and Khelifi (2016); Lawgaly et al. (2014); Kang et al. (2011); Ahmed et al. (2019), while less research has investigated the forensic analysis of videos. According to recent research, the identification algorithms that help identify the image sources are not as effective at identifying videos Hosler et al. (2019b). Video identification algorithms are used to identify and distinguish sensor types based on video produced by digital cameras. In recent years, forensic specialists have been particularly interested in this topic. In general, there are two primary ways to identify images and videos, examining images or videos to extract a unique fingerprint of the camera and using metadata associated with the images or videos (the DNA of a video). López et al. (2020) demonstrated that the internal elements and metadata of video can be used to identify a video's source. Because metadata (DNA) can be removed from an image or video, extracting a unique fingerprint based on an image and video is reliable. Two concepts are typically considered when identifying cameras: source camera model identification (SCMI) and individual source camera identification (ISCI). ISCI distinguishes cameras from both the different and same camera models, while SCMI is a subset of ISCI that distinguishes a particular camera model from others but cannot distinguish between camera devices from the same camera model. More research has been done on SCMI than on ISCI. In a database, if there is one device for each model, the ISCI is generally identical to the SCMI. When the two scenarios can be different, the number of devices is more than one device for each model; therefore, SCI and SCMI are two different problems that should be addressed differently. For example, if there are two iPhone 8 Plus devices in a database, two different classes and one class are considered for ISCI and SCMI, respectively.

Figure 1 shows the common techniques that have been used in source camera identification based on digital images, which can be divided into several classes: digital camera identification based on PRNU estimation Lukas et al. (2006); Mieremet (2019); Lawgaly and Khelifi (2016), statistical methods Chapman et al. (2015), dark signal identification Virmontois et al. (2010), sensor dust Dirik et al. (2008), optical defects Kordecki et al. (2015), and machine learning, such as deep models Yang et al. (2019) involving convolutional neural networks. However, source camera identification methods on videos have focused more on PRNU and machine learning methods.

This study presents an inclusive overview of what has been done over the last decade in the field of source video identification. This review provides insights based on 93 articles as well as additional articles and supporting publications. Figure 2 shows statistics about the literature surveyed from 1970 to the present (March 2022). From 2015 until today (March 2022), the figure shows an increased interest in the topic among academics. There are 56 journals, 27 conferences, 2 preprints, and 8 books/dissertations/patents.

A series of representative studies identifying source cameras were presented. A brief structure of methods in the field was examined by Lefèbvre et al. (2009), and some methods focusing on images were reviewed. Recently, Bernacki (2020) reviewed methods that were designed to identify source cameras on images, which included the categories shown in Figure 1. Milani et al. (2012); Kot and Cao (2013); Pandey et al. (2016); Pasquini et al. (2021) provided a brief overview of the state of digital image forensics. These studies focused on methods for detecting forgery and tampering. They also investigated a few methods for source camera identification using video. To our knowledge, there is no comprehensive paper that presents methods in conjunction with videos. In-depth analysis is critical when identifying source cameras with a focus on video and to expand future research pathways. Therefore, it is essential to conduct a comprehensive review and summary of existing research in order to continue to advance this field in the future, particularly for researchers who wish to enter the field.

In addition to reviewing the literature, this study provides a description of available video databases that can be used for camera identification methods.

The following cases are considered to prepare the survey:

- A literature review of papers with titles, abstracts, keywords, or experimental results that used video databases or explicitly referenced videos throughout the paper was conducted. Keywords for the search included "source camera identification on video, machine learning methods, PRNU, stabilization videos and social media videos"; however, these words were not the only ones used. We then selected only papers that referred to videos and those that created databases. Topics related to forensics, such as fake video detection, are not part of this survey.

- To be considered for inclusion in this review, papers must meet the following criteria: peer-reviewed English journals, peer-reviewed conference proceedings, or recent manuscripts from open-source archives. In addition, we have adhered to studies conducted by authors who are renowned and experts in the field. All distinguished and expert authors in the area of interest were explored in Google Scholar with key words related to topics of interest, such as media forensics, multimedia forensics, and multimedia security.

- We have also presented databases on this topic because a database is an essential part of any application in the field.

- All contributions to the sections (text, tables, graphs and plots) are sorted by category and year of publication.

The remainder of this paper is organized as follows. Section 2 presents basic definitions in this field of research that are relevant to the following sections. Section 3 considers methods that can be used to identify source camera models in two categories: PRNU and machine learning. Section 4 describes video databases that can be used to identify source cameras on videos. The final section discusses open research questions that

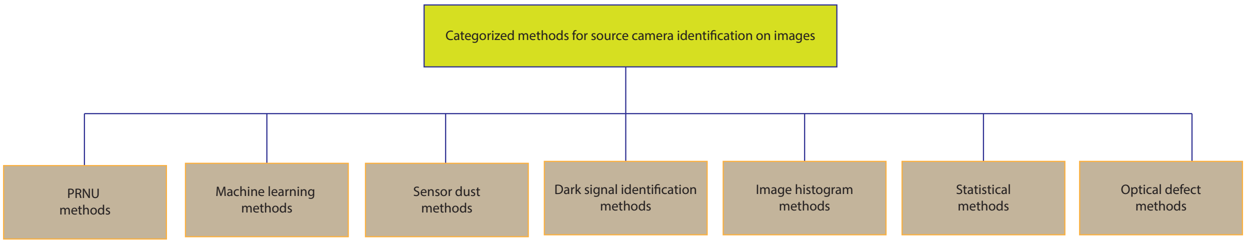

Figure 1: Categories for source camera identification on the image based on Bernacki (2020)

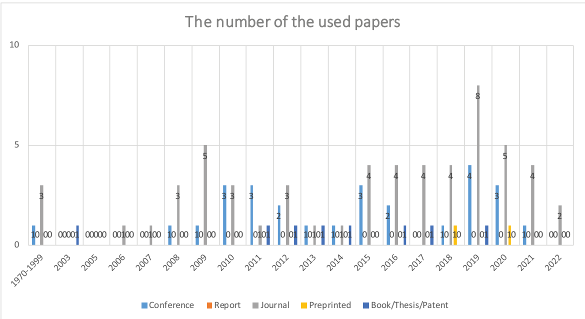

Figure 2: Number of papers used in this survey by publication year (from 1970 to August 2021).

can be addressed in the future and provides a conclusion.

2. Basic definitions

Video creating process: Understanding the basics of how digital cameras produce video is crucial and is shown schematically in Figure 3. As a first step, the light of the scene is captured by the lens. A key processing step in creating a video is to reduce the size of the full-frame sensor output to reduce the amount of data that must be processed. Both acquisition and colour processing could be performed as follows: by downsampling colour-interpolated image data and subsampling pixel readout data in acquisition. A common method of compensating blur caused by camera shake is electronic image stabilization, which modern cameras use in postprocessing. In addition, the images can also be scaled and cropped during this postprocessing, which further reduces their size. To make the storing and transferring of the postprocessed images as efficient as possible, the sequence is encoded into a standard video format.

Charge-coupled device (CCD): CCD sensors are an important technology that are used in digital imaging. In the CCD, light is converted into electrical signals. A CCD array (Figure 4) generates two-dimensional image signals by arranging thousands or millions of CCD elements at regular intervals in rows and columns. The light-sensing element is known as a pixel. The output signal of CCD lines is always affected by noise. There are two types of noise: random noise and noise with a fixed pattern Boyle and Smith (1970).

Complementary metal oxide semiconductor (CMOS): In digital camera technology, such as cell phone cameras and web cameras, the CMOS sensor is similar to the CCD sensor. Metal oxide semiconductors (MOSs) are the primary component of both types of sensors, and both function similarly, although there are some differences in how the pixels are scanned and how the readout of the charges is performed.

The output information from CCD sensors must be processed in an additional chip, which increases the cost of manufacturing. A CMOS sensor has its own active pixels, and because the sensor can perform digitizing, it offers higher speeds, a smaller size, and a lower cost. Despite similar characteristics, CCD arrays have the advantage of capturing light simultaneously, providing a more consistent output. Because the readout process is typically performed as progressive scanning, which is free of blooming, CCD sensors offer a markedly larger dynamic range and better noise suppression than CMOS sensors. CMOS sensors are more sensitive to light, which makes them more effective in low-light conditions. CCD sensors used to be superior to CMOS sensors, but this gap is nearly closed today. As the strength of CCD technology has reached its limit, CMOSs are gradually being improved and developed so that most smartphones use CMOS sensors instead of CCD sensors. CMOS/CCD sensors store data in a digital format, which is then sent to a processor for processing. The image processor eliminates noise and other anomalies once it receives the digital signal. The signal is also subjected to colour interpolation, gamma correction, and colour correction.

PRNU: PRNU, which is understood to be the unique fingerprint of the camera, is often referred to as resid-

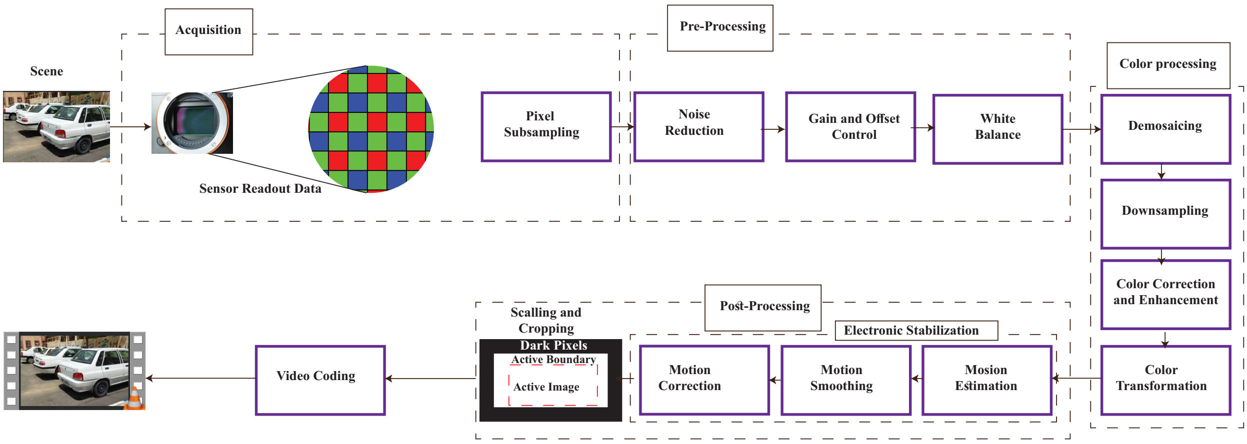

Figure 3: The steps in the image processing of a camera when generating a video based on Corcoran et al. (2014).

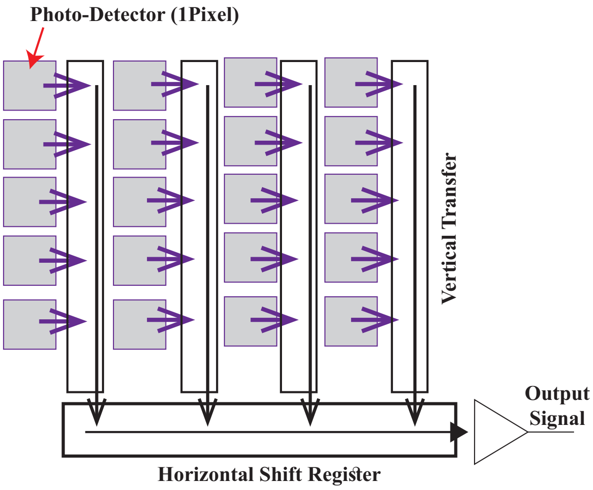

Figure 4: Overview of a CCD array (based on Kurosawa et al. (1999)).

ual noise or sensor pattern noise (SPN). Fixed pattern noise (FPN) occurs when the CCD or CMOS sensor processes the input (light) and converts it into a digital signal. FPN typically refers to two dark signal nonuniformity (DSNU) and PRNU parameters. DSNU is the offset from the average across the imaging array at a particular setting (temperature, integration time) but no external illumination. PRNU characterizes imperfections are caused by the manufacturing process due to a lack of homogeneity in the silicon area in the imaging sensor Lawgaly and Khelifi (2016). Although current imaging devices are becoming more sophisticated, making PRNU noise less effective, current research is investigating whether it can still be used as a powerful tool to identify the camera source if some improvements are considered Iuliani et al. (2021); Altinisik and Sencar (2020). The noise due to sensor imperfections is a weak signal of the same dimensions as the output image indicated in this study by $K \in R^{W \times V}$, where $W \times V$ is the dimension of the sensor. Even though sensors can be different from one device to another, the final digital image output can be expressed as Lawgaly and Khelifi (2016); Lawgaly et al. (2014):

$$I = I^0 + I^0 K + \Theta \tag{1}$$

where I^0 refers to the original input multimedia file, $I^0 K$ represents the PRNU term and Θ is a random noise factor.

To estimate PRNU based on video, first, the video V is decomposed into frames I_i, where $i = 1, ..., N$ with N being the total number of frames in the video. The reference pattern noise can be approximated by averaging I_i. During this process, each frame should only contain pattern noise, not a scene. The denoising filter F removes all scene content per frame I_i, thus leaving only the noise residual n_i, which is used to determine the pattern of noise.

$$n_i = I_i - F(I_i) \tag{2}$$

Increasing the number of frames I_i in V results in smoother spatial pattern noise P_V in the video:

$$P_V = \overline{n} \tag{3}$$

Typically, $N > 50$ is recommended Lukas et al. (2006).

Denoising filter As mentioned in the PRNU definition, a denoising filter F is used to extract pattern noise. Wavelet-based filters can be considered to have better performance than approaches such as Wiener and median filters. Areas around edges are typically misinterpreted by the latter two. For details on how this denoising filter works, see Lukas et al. (2006).

Normalized cross correlation (NCC): NCC measures the similarity between two PRNUs, which are reference and query patterns. The reference pattern is obtained by averaging the PRNUs that are extracted from frames of a video or multiple videos of a device. To evaluate the video identification system, the PRNU can be extracted from the query video, as shown in (3). The computed correlation ρ_C between the given pattern given (query) P_V and the reference pattern P_C is used to check whether video V was recorded with camera C:

$$\rho_C(V) = corr(P_V, P_C) = \frac{(P_V - \overline{P}_V).(P_C - \overline{P}_C)}{\left\|P_V - \overline{P}_V)\right\| \left\|(P_C - \overline{P}_C)\right\|} \quad (4)$$

where the mean is shown by the bar above a symbol. Using the right denoising filter can improve correlations. Due to the noise sensitivity of the NCC, these methods are not accurate when used with highly compressed and low-resolution videos.

Peak-to-correlation energy (PCE): Based on NCC, PCE, a resolution-independent similarity metric, is calculated as follows:

$$PCE(\rho) = \frac{\rho_{peak}^2}{\frac{1}{mn-|S|} \sum_{P_V, P_C \notin S} \rho(P_V, P_C)^2} \quad (5)$$

where ρ_{peak}, S, and $|S|$ are the maximum value of the NCC matrix, a small region surrounding ρ_{peak}, and the cardinality of S, respectively. ρ_{peak} can be replaced by $\rho(1,1)$ if the fingerprint and the noise estimates have the same matrix resolution. When $PCE(\rho)$ exceeds a threshold t, the query image with noise estimate P_V is considered to be taken with the same camera P_C.

There is an additional complication that arises from the observation that PCE values for PRNU matching in videos are lower than those for images. Downsizing operations and compression of video have caused this decline in PCE values. As shown in Figure 5, even though the correlation is low overall, PCE can be classified by such a peach into positive matching versus negative matching. Also, Goljan et al. (2009) have shown that peak-to-correlation energy (PCE) is an additional attribute that provides an additional level of robustness to normalized correlation.

Video Codec: Video files are compressed with codecs, which must always balance quality and size (better quality vs. larger file size). Video files can be compressed to reduce their size, which can reduce bandwidth usage and increase streaming speed. For encoding high-definition video, AVC is the standard codec used by several online video services, including YouTube and Vimeo. The MPEG-4 and H.264 standards were implemented by the library 'libx264' in

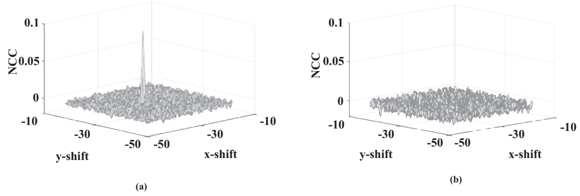

Figure 5: Normalized cross-correlation matrices when matching is positive (a) versus negative (b). (based on Ferrara and Beslay (2020)).

FFmpeg. The constant rate factor (CRF) setting is the easiest setting to change the applied compression to a certain quality for the output video. With CRF set to 0, you can achieve loss less compression. Although the compressed video with CRFs 18 and 20 is still of high quality, it is smaller than the original.

One of the most widely used video compression standards is H.264/AVC, which is managed by the JVT (Joint Video Team) and is currently used in nearly all smartphones and social media. Detailed descriptions of video coding are beyond the scope of this paper; Van Houten and Geradts (2009) and Telecom et al. (2003) provide a detailed description of the H.264/AVC video coding standard and technical approaches. The H.264/AVC encoder shares several basic steps, such as block processing, prediction, transformation, quantization, and entropy coding, as shown in Figure 6.

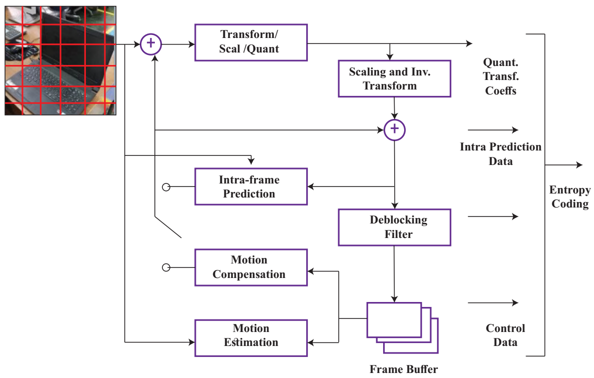

Figure 6: Primary steps of a H.264/AVC encoder (based on Van Houten and Geradts (2009)).

FFmpeg [1]: FFmpeg is one of the most popular multimedia frameworks that can decode, encode, transcode, mux, demux, stream, filter and play anything that humans and machines have created. From ancient to modern formats, the software supports different frameworks:

[1]https://www.ffmpeg.org/

Linux, Mac OS X, Microsoft Windows, the BSDs, Solaris, etc. FFmpeg is used to re-encode original flatfield videos to have the same resolution as natural videos if the resolution changes. A software package included in ffmpeg (ffprobe) is used to extract video frames (type, frame rate, bit rate, resolution, etc.) and save them in an uncompressed format. Additionally, to extract I, B, and P frames and motion vectors, these tools can be used.

Group of Pictures (GOP): GOP consists of I-frames, P-frames and B-frames as intra-coded picture, predictive-coded picture, and bi-predictive coded picture, respectively in coding standards such as MPEG series and H.264. Although modern codecs are based on instantaneous decoder refresh (IDR) frames, all methods investigated in this review used I-frames to identify the source camera. I-frames are the least compressible and do not require other video frames for decoding. P-frames can be decompressed using data from previous frames and are more compressible than I-frames. For B-frames, previous and forward viewed frames can also be used as data references to achieve the highest compression. An I-frame is typically more detailed than a P- or B-frame. GOP size can generally be divided into fixed vs. unfixed and is the number of the B- and P-frames between two consecutive I-frames. The three types of frames are divided into blocks of pixels (macroblocks) that are represented by motion vectors in the H.264 standard, which can play an important role in the compression step and estimating fingerprint patterns. The first step to compress any macroblock is to search in the current frame, or in previous or future frames, for a macroblock that is similar to the macroblock we want to compress. Then, the best location of the macroblock in the frame is found, and the location information is sent to the decoder. The samples of the motion vectors predicted in a P-frame are shown in Figure 7.

Stabilization: Video stabilization uses digital processing to minimize the effects of vibration and camera shake on video quality, which are common in handheld devices. Video stabilization compensates for involuntary user movements using geometric transformations (e.g., translations, similarities, homographs, etc.) to capture video frames Grundmann et al. (2016). The result is a misalignment of individual pixels across frames, which complicates PRNU fingerprint estimation. PRNU estimation is much more difficult for videos than for images because videos are typically highly compressed and are often subject to video stabilization.

Optical stabilization and digital stabilization are methods that can stabilize images or videos. In the former, a hardware-based stabilization mechanism is used, and camera movement is not fully transferred to the video. In the latter, image processing methods are applied by moving and warping to align frames. It is possible to combine both stabilization approaches in one camera to achieve more effective results Altinisik and Sencar (2020).

Video stabilization has less effect on the first frame of the video, which is not the case with the newly created iPhone Grundmann et al. (2016). Thus, some cameras seem to activate stabilization when set to video, even before recording begins Altinisik and Sencar (2020). During the digital stabilization process, video frames are cropped, warped, and inpainted Elharrouss et al. (2020) to eliminate unwanted camera movement. To identify a source camera, these transformations must be blindly inverted.

The effect of video stabilization shown in Figure 8, where the visual content is aligned between two frames. The example shows that the line describing the edge of a wall changes its orientation when it appears in successive frames of an unstabilized video, but such a line is properly aligned after video stabilization, producing more stable visual content.

Deep learning methods: A special class of machine learning techniques called deep learning (or deep structured learning) is based on artificial neural networks with a deep learning structure that can be supervised, semi-supervised or unsupervised. Convolutional neural networks (CNNs) form the basis of most deep learning models, although they may also contain propositional formulae or latent variables organized in layers, such as the nodes in deep belief networks or deep Boltzmann machines LeCun et al. (2015). Currently, convolutional layers are primarily used to capture scene content instead of camera detection features such as noise patterns. In addition, deep learning methods such as CNN and Siamese networks Cozzolino and Verdoliva (2018) can be used for the aim.

Challenges and metrics: When a video is produced, challenges such as compression, stabilization, scaling, cropping, and differences between frame types can be considered. A new challenge is related to new databases that include devices with new technologies to test the effectiveness of existing methods. While reported results based on ISCI and SCMI scenarios on the new databases based on existing methods show that an improvement is essential, ISCI compared to SCMI needs more improvement. Thus, developing and introducing new databases, improving existing methods and introducing new methods to address the ISCI and SCMI scenarios are essential. Additionally, moving videos (videos captured while the device was moving) are more challenging compared with still videos (videos captured

(a) (b)

Figure 7: Motion vector calculated by FFmpeg.

by fixed devices).

To evaluate the performance of source camera identification techniques, the common measures used by the different authors are mentioned in this section. In addition to NCC and PCE values, accuracy, true positive rate (TPR) and false positive rate (FPR) can be considered. Additionally, receiver operating characteristic (ROC) curves can be used to obtain the area under the curve (AUC).

3. Methods

Source camera identification is reviewed on videos in two categories: PRNU and machine learning methods.

3.1. PRNU methods

The methods are classified into general, stabilization and resizing, social media videos, and network streamed videos. The methods presented based on PRNU are summarized in Table 1. There are three primary steps in the type of methods: frame extraction, PRNU estimation for both query and reference videos, and matching query with reference video.

3.1.1. General

In this subsection, we focus on methods that improve PRNU estimation to identify source cameras in videos. In many methods, the improvement is performed by introducing denoising filters.

McCloskey (2008) extended and improved methods proposed by Chen et al. (2007). They focused on videos with edge and texture content, and their method was based on a map function that gives each pixel a weight. Using a Gaussian kernel, the mapping function emulates the spreading of errors caused by wavelet denoising at multiple scales. Canon and Kodak cameras with indoor scenes were the primary target in the study to obtain videos with high-frequency content (e.g., edges).

Chuang et al. (2011) and Goljan et al. (2016) studied the impact of compression on estimating PRNU in video frames. While Chuang et al. (2011) considered reordering and weighting the frame, Goljan et al. (2016) applied an adjustment step to make a correct decision to choose the threshold. They tested different scenarios to find the best condition in selecting the threshold using with frames and real images in the M-JPEG and MPEG-4 formats. Based on Monte Carlo simulations, theoretical analysis and experiments show a marked increase in the variances of normalized correlation and PCE when JPEG is compressed. Therefore, it is necessary to adjust the decision threshold to reduce the probability of a false alarm. The authors calculated this adjustment factor experimentally by comparing different JPEG CRFs with 1-Mpixel still images. Their experiments showed that the quality of the compression and the increase in

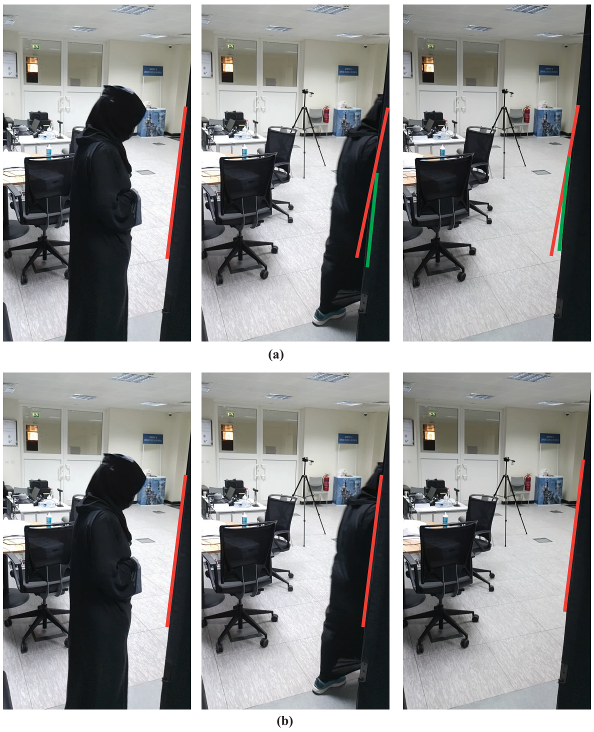

Figure 8: Stabilization effects on subsequent frames in digital video. (a) The frames are not stabilized, which means that the red and green lines that highlight the content do not align with the content due to hand movements. (b) Stabilizing the frames ensures alignment (lines) of the content based on Ferrara and Beslay (2020).

positive bias both affect the normalized correlation.

A PRNU method based on a minimum average correlation energy (MACE) filter Mahalanobis et al. (1987) was used by Hyun et al. (2012). The filter was used to reduce the impact of noise on NCC. After extracting PRNU from reference videos, this filter is applied but did not affect the test (query) videos. This method was performed on 7 camcorders, and results showed that this filter can increase accuracy up to a maximum of 10%.

Villalba et al. (2016) followed the structure of a classification approach for identifying source cameras based on PRNU. They extracted features from PRNU estimated from frames to classify videos. The features are based on decomposing wavelet subbands. Support vector machine (SVM) was used to classify the feature space.

The idea behind the method presented by Al-Athamneh et al. (2016) was simple, but results showed that they achieved promising results. They extracted PRNU only from the green channel of each frame instead of all three RGB channels because the green channel is the noisiest channel among the three RGB channels. Once the frames were extracted, they were resized to 512×512 pixels. Wavelet denoising was applied to obtain the residual signal for each frame. The method was applied to 256 videos captured by 6 devices. Resizing was thus shown to effectively improve the results of source camera identification.

In parts of the thesis of Wales (2019), a case study was considered to evaluate existing methods to identify source cameras. The study further can be considered for fresh researchers in the field.

Yang et al. (2021) estimated PRNU from extracted I-frame camera video rolls, which is defined as rotation around the camera axis. When capturing videos with 180-degree rolling, the video can be rolled back 180 degrees with a mobile phone. After obtaining the PRNU for reference video, an enhancement step based on the Wiener filter (WF) and the zero-mean (ZM) was used in the Fourier domain as described in Chen et al. (2008). Then, a rotation normalization step is applied to the enhanced frames. This method was evaluated on the VISION database Shullani et al. (2017). Results showed some improvements compared with a few methods.

López et al. (2021) aimed to improve PRNU using an enhancement step and clustering approach. In the first step, PRNU was estimated from a macroblock within the frame. Then, to enhance it, the method introduced in Li (2010) was considered. In this step, the high frequency components of the scene are represented in such a way that their magnitude surpasses that of the noise pattern due to the details of the scene. Therefore, it is important to remove fragments from the high-frequency scene to improve the fingerprint of sensor noise. To achieve this, on PRNU, smaller weighting factors were applied to the strong components of the signal in the wavelet transform domain. Finally, the unsupervised agglomerative clustering used in Caldelli et al. (2010) was applied to classify videos of the VISION database Shullani et al. (2017).

Ferrara et al. (2022) explored which type of frame is suitable for source camera identification while considering compression and stabilization. They showed that I-frames obtain better results when stabilization occurs and that the most significant PRNU information is provided by the first I-frame. Among the P-frames, the most reliable PRNU information is provided within the P-frames of the first GOP. This finding is explored on the VISION database.

3.1.2. Stabilization and resizing

A major challenge in PRNU-based camera identification from a video is image stabilization during capture and/or post-processing. Essentially, digital stabi-

Table 1: Methods presented based on PRNU.

Method	Category	Year	Content
McCloskey (2008)	General	2008	dealing with videos with edge and texture content
Chuang et al. (2011)	General	2011	estimating PRNU in video frames on compression status
Hyun et al. (2012)	General	2012	reduce impact of noises on NCC
Goljan et al. (2016)	General	2016	estimating PRNU in video frames on compression status
Villalba et al. (2016)	General	2016	extracting features from PRNU pattern and using SVM classifier
Al-Athamneh et al. (2016)	General	2016	extracted PRNU only from Green channel of each frame instead of all three channel RGB
Wales (2019)	General	2019	case study to evaluate existing methods to identify source camera
Yang et al. (2021)	General	2021	estimated PRNU from I-frame extracted cameras rolling
López et al. (2021)	General	2021	improve PRNU by using a enhancement step and clustering approach
Ferrara et al. (2022)	General	2022	exploring which kind of frame is useful in the field
Höglund et al. (2011)	Stabilization	2011	a simple transformation on PRNU patterns
Taspinar et al. (2016)	Stabilization	2016	to correct for both the shift and rotation applied, inverse affine transformations are used
Iuliani et al. (2019)	Stabilization	2019	searching the proper amount of scaling, shifting, and cropping that should be applied to each frame
Mandelli et al. (2019)	Stabilization	2019	searching for transform parameters using PSO
Altinisik and Sencar (2020)	Stabilization	2020	assumes a larger degree of freedom in the search for stabilization transformations that consider spatially variant natures
Ferrara and Beslay (2020)	Stabilization	2020	shifting and inverse transformation on flat I-frames creating a robust reference instead of eliminating stabilization on queries frames
Mandelli et al. (2020)	Stabilization	2020	a fast search of inverse transform (Fourier-Mellin Transform)
Taspinar et al. (2020)	Resizing	2020	considering different aspect ratios (resizing and cropping) of images and videos
Van Houten and Geradts (2009)	Social media	2009	investigating the usage of PRNU for source attribution of YouTube videos
Scheelen and van der Lelie (2012)	Social media	2012	exploring impact of video Codec on YouTube videos
Brouwers and Mousa (2017)	Social media	2017	exploring impact of PRNU videos on YouTube
Amerini et al. (2017)	Social media	2017	exploring impact of PRNU videos on Facebook and Twitter
Meij and Geradts (2018)	Social media	2018	identifying source when videos transmitted by WhatsApp
Kouokam and Dirik (2019)	Social media	2019	Estimated PRNU videos transmitted in Youtube
Pande et al. (2013)	Streamed videos in network	2013	a hardware architecture for source identification in networked cameras
Chen et al. (2014)	Streamed videos in network	2015	identify source camera in a wireless stream with blocking and blurring
Kaur and Randhawa	Streamed videos in network	2020	deal with blocking and blurring issues by adding Gaussian mixture models (GMM) using k-means clustering

lization involves three primary steps: estimating motion, smoothing, and aligning frames based on the corrected motion analysis. Using a parametric model or by considering the geometric relationship between successive frames, feature trajectories are calculated by tracking key points across frames and estimating motion Xu et al. (2012); Grundmann et al. (2011); Thivent et al. (2017). When stabilizing a camera, whether the motion can be reproduced in a two-dimensional (2D) or three-dimensional (3D) representation is important. Modern methods consider 3D motion models to overcome the limitations of 2D modelling. Because reconstructing 3D with depth information, is not trivial, these methods simplify the structure of 3D and rely heavily on feature tracking accuracy Liu et al. (2009); Wang et al. (2018). The primary goal of video stabilization is to align successive frames by geometric registration. To compensate for any type of perspective distortion, Euclidean transformations (scaling, rotation, and translation individually or in combination) can be applied to a spatially varying warping transform, depending on the complexity of the camera motion throughout the capture. Because these transformations are applied to each frame, there is sufficient variance in camera motion to easily remove pixel-to-pixel correspondences between frames. Frame-level PRNU pattern alignment or averaging is ineffective at estimating a reference PRNU pattern. Finally, in stabilized video, determining and inverting the frame-level transformations is necessary for source attribution.

In source camera identification on images, search methods for the geometric transform parameters are applied karakucuk et al. (2015) Goljan and Fridrich (2008).

The overall results of the algorithms are shown in Figure 9. The focus of algorithms is to present a transformation that can deal with alignment issues in the cases. In the methods that extract reference patterns, the transformation can be images captured by the same camera for capturing videos and flat frames and first frames from videos by averaging them.

While Höglund et al. (2011) compensated for stabilization by a simple inverted transformation method in extracting the noise pattern, Taspinar et al. (2016), stabilization in each video was detected in the first step by matching the PRNU of the beginning and end frames.

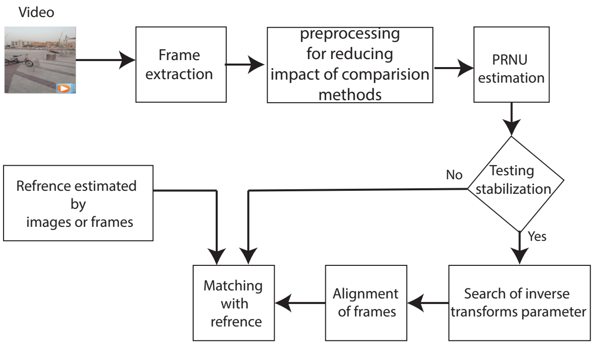

Figure 9: Overall stabilization methods

If the frames match, the video is not stabilized. After extracting I-frames, one of the I-frames is set as a reference. The inverse affine transformation to correct rotation and shift is applied to I-frames aligned to the reference. After aligning the I-frames, the PRNU pattern derived from an unstabilized video is compared to the pattern derived from the aligned I-frames.

Iuliani et al. (2019) identified the source camera by assuming that the reference PRNU pattern could have been determined by images or from an unstabilized video. When matching, the videos and images have different resolutions. The PRNU patterns for 5-10 I-frames are aligned with the reference PRNU pattern by determining the proper amount of scaling, shifting, and cropping for each frame. An aligned PRNU pattern is created by combining frames that yield a matching.

The PRNU pattern was estimated by Mandelli et al. (2019) by taking weakly stabilized video into account. This approach generates an alignment reference based on a set of frames. Pairwise matching is performed between PRNU estimates obtained from each frame and other frames to detect stabilization. A reference PRNU pattern is then generated by combining the largest group of frames that yield a sufficient match, and the remaining frames are aligned to this pattern. When the reference PRNU pattern is already known at another resolution, as in Iuliani et al. (2019), then the reference pattern is used to compare all other PRNU patterns using particle swarm optimization (PSO) to search the transformation parameters Sammut and Webb (2011). To speed up the search for weakly stabilized videos, they observed that rotation can be ignored. An estimate of the PRNU pattern is first based on flat and still content that has been weakly stabilized. To verify the stabilized video, five I-frames are extracted and compared with this reference PRNU pattern. An evaluation of results obtained using the VISION database demonstrates the effectiveness of the method.

Altinisik and Sencar (2020) extends Altinisik and Sencar, who presented a method for verifying source cameras for a video that considers how stabilization transformations vary spatially and assumes a higher degree of freedom for their search. Transformations are identified at the subframe level, a variety of constraints are included to verify their correctness, and computational flexibility is provided in the search for the right transformation. Transformations also take a holistic approach to neutralize the disruptive effects of video coding and downsizing steps for more reliable mapping. Seven steps were involved in this method. To reduce artefacts after a filtering process in the decoder (i.e., the loop filter), the bitstream is decoded into video frames. In the next step, each extracted frame is then processed to extract the PRNU pattern. Prior to analysis, the videos are also evaluated for the level of stabilization to eliminate unstabilized and weakly stabilized videos based on Iuliani et al. (2019); Mandelli et al. (2019). Then, smaller blocks are cropped from each PRNU pattern to address the spatially variable nature of the stabilization transformations; the best results were achieved with 500×500 blocks. The PRNU transformation parameters are identified by a search for each block to avoid false inversions. A weighting procedure considers the compressed levels of inverse transformed blocks before they are combined. The final evaluation of the match is performed by comparing the estimated PRNU pattern with the reference PRNU pattern.

Ferrara and Beslay Ferrara and Beslay (2020) focused on creating a robust reference instead of eliminating stabilization on query frames as in the previous methods mentioned above. They applied the primary steps of cropping, shifting and inverse transformation to flat I-frames. They also propose an optimal strategy for comparing PRNUs extracted from motion stabilization videos.

Mandelli et al. (2020) suggested a search method for scaling and rotation parameters in the frequency domain to quickly find inverse transformations. The Fourier-Mellin transform was used to estimate transformations of scale, rotation, and shift between two images when they are in closed form Reddy and Chatterji (1996). Experiments performed on the VISION database showed that the method is much faster than existing methodologies.

Using a combination of cropping and scaling on the visual object can help downsize images or frames when the sensor resolution exceeds the desired resolution. To scale down still images, cameras use bicubic scaling or Lanczos scaling (or their derivatives). Pixel binning or line skipping techniques are also used to reduce camera processing costs. These techniques are typically used in video capture. Using only the central pixels of a sensor and discarding the surrounding pixels is one approach. There is a serious disadvantage of cropping for still images, the change in the field of view, which is narrowed when the cropped region is large. Thus, cropping is often combined with resizing. Taspinar et al. (2020) extends Taspinar et al. (2021), who considered different aspect ratios (resizing and cropping) of images and videos to identify the source camera. They also introduce a database for resizing and cropping issues.

3.1.3. Social media videos

Van Houten and Geradts (2009) investigated the use of PRNU for YouTube video source identification. Videos were recorded and encoded using a range of webcams and codecs. YouTube was used for uploading and downloading the videos. To determine a source device's PRNU, the downloaded videos were analysed. Despite positive results, the findings of this study are outdated because YouTube users now use handheld devices (video cameras) rather than fixed-lens cameras.

Scheelen and van der Lelie (2012) examined videos that are re-encoded with the Advanced Video Codec, as found on YouTube. Their results showed that YouTube applied a similar compression level as CRFs between 27 and 30. These videos were matched to the original cameras based on the noise pattern extracted from them.

As with previous tests, Brouwers and Mousa (2017) tested the PRNU correlation procedure with YouTube videos to see if it still performed effectively. They also tested how effective PRNU pattern extraction is using cell phone cameras and YouTube videos. This correlation can be demonstrated by testing various mobile cameras, but it depends on the type of camera used.

Amerini et al. (2017) evaluated methods Chen et al. (2007); Goljan et al. (2016) to build a series of experiments in social media. They built the reference pattern based on original, Facebook and Twitter videos. Also, they explored the PRNU on I-frames and showed when various videos are first uploaded and then downloaded from Facebook or Twitter can still be considered to identify the source camera.

Meij and Geradts (2018) extracted PRNU for source camera identification using software developed by the Netherlands Forensic Institute [2]. The PRNU can be extracted using four different filters: 4th order extraction filter, wavelet (Daubechies), wavelet (Coiflet) and the 2nd order (FSTV) extraction filter. This study uses a second-order (FSTV) extraction filter due to its superior performance Brouwers and Mousa (2017) compared to the other methods; the authors studied videos transmitted by YouTube. They estimated PRNU based on videos transmitted by WhatsApp 2.17.79 (Android) and 2.17.20 (IOS), and also showed that video transmissions by WhatsApp can also be identified by their source.

In Kouokam and Dirik (2019), a block-based method was used to simulate all the processes of the H.264/AVC video compression standard. In the simulation, the effects of PRNU noise on the operations applied to an encoded frame block were considered and investigated. This method was tested on videos transmitted on YouTube that are part of the VISION database Shullani et al. (2017). They showed that based on their method, their results contradict previous literature, which states that using all frames (I, B, and P) is more accurate than using only I frames, even when YouTube recompresses the videos under study. This result means that PRNU noise in P and B frames is still valid despite compression to improve source camera identification. The primary advantage of this method is that it estimates PRNU fingerprints based only on frame blocks that have the correct PRNU components.

3.1.4. Streamed videos in network

As a result of packet loss during network transmission, video streaming is typically blocked and blurred. Source camera identification methods can fail when degradation occurs with blocking and blurring. Therefore, the methods should be improved to consider the problems in the network stream.

Pande et al. (2013) introduced a hardware architecture for identifying source cameras in networked cameras. They designed the hardware to estimate PRNU by combining an orthogonal inverse wavelet transform with minimum mean square error based estimation using frames extracted from videos. Parallelism, pipelining, and hardware reuse techniques were used to maximize hardware utilization and accelerate throughput. Device prototyping was implemented on a Xilinx Virtex-6 FPGA hardware with a clock frequency of 167 MBs, processing 30 frames with a size of 640×480 in 0.17 s.

Chen et al. (2014) extends Chen et al. (2013), who identified a source camera in a wireless stream. They compared the existing PRNU-based method Lukas et al. (2006) with their own method to show impact blocking and blurring in the results. They added two steps when detecting wirelessly streamed videos with blocking and blurring that are based on wavelet to Lukas et al.

[2] https://www.forensischinstituut.nl/

(2006). Two aspects of video blocking affect sensor pattern noise extraction. First, the details and the noise in the patterns are lost within the blocks (8 or 16 pixel blocks). Second, the edges of the blocking provide a signal that survives extraction and averaging. In this study, only blur caused by packet loss was considered, although video blur can be caused by many other factors, including high compression ratios, limited lens resolutions, and fast motion. Blurring is caused by the loss of high frequency information rather than the loss of all data in a block, but this information loss is still block-based. One block may appear blurrier than another.

Kaur and Randhawa investigate blocking and blurring issues by adding Gaussian mixture models (GMMs) using k-means clustering to initialize clusters for GMM to estimate the PRNU step. The experiment involved 20 tests, each performed on eight videos from eight wireless cameras by randomly selecting five frames from each video.

3.2. Machine learning methods

The methods presented based on machine learning are summarized in Table 2. Two studies López et al. (2021) and Villalba et al. (2016) mentioned in general categories can also be considered machine learning methods.

Su et al. (2010) identified source cameras based on the features extracted from bit stream, quantification factor and motion vectors and classified them into camera classes by an SVM classifier. They extracted the motion vectors of each macroblock in P-frames. For bit stream features, the number of bits, P-frames, and B-frames of a GOP, and the average and variance of the relative difference between adjacent P-frames and B-frames. For quantization factor features, the maximum number of consecutive macroblocks with the same quantization parameter in a frame of type I, P and B with the order left to right and top to bottom; the average and variance of the number of consecutive macroblocks with the same quantization parameter in a frame of type I, P and B with the above order; the maximum difference value of the quantization parameters between adjacent macroblocks in a type I, P and B frame; and the average difference value of the quantization parameters between adjacent macroblocks in a type I, P and B frame were extracted. For motion vector features, a search window was considered such that the maximum horizontal and vertical dimensions were estimated.

Yahaya et al. (2012) introduced a feature extraction method based on conditional probability (CP) for source camera identification. The features reported promising results in steganalysis applications Wahab et al. (2009), where conditional probabilities refers to the probability of B when A has already occurred. The JPEG DCT coefficient array was extracted from each frame, and then, the CP features were obtained from the coefficients. The features were extracted in blocks with size 8 in each frame. The features were classified by an SVM classifier to identify the source camera. The method was only tested on videos captured by four devices.

In Kirchner and Johnson (2019), a CNN based on sensor pattern noise (SPN) called SPN-CNN was presented. They implemented the CNN based on the idea that CNN has the ability to extract signals characterized by noise from a set of images Zhang et al. (2017). Therefore, they trained a network to identify a noise pattern. The method was tested on the VISION database Shullani et al. (2017), and experiments showed that the experiments obtained better results than the Wavelet denoiser. Additionally, they showed that when I-frames were considered to feed into CNN, the results improved.

Timmerman et al. (2020) and Hosler et al. (2019a) proposed a deep learning method (MISLnet CNN architecture) for source camera identification using frames to train the network. They extended a version of a constrained convolutional layer introduced in Bayar and Stamm (2018). A majority vote was used to make decisions at the video level using frames fed into the network. The constrained convolutional layer was added as the first layer that used three kernels with size 5. This layer is constructed such that there are relationships between adjacent pixels that are independent of the content of the scene. The methods were tested on the VISION database Shullani et al. (2017). The experiments showed that the layer can improve the results compared with deep learning architectures without the layer. The difference between the two methods is the size of the images and the type of color modes. Timmerman et al. (2020) and Hosler et al. (2019a) used RGB and gray scale modes, respectively. The patches used in the former are 480, while the latter are patched with 256.

Mayer et al. (2020) used a CNN introduced in Bayar and Stamm (2018), similar to the two previous studies, to extract features and a similarity network to verify the source camera. The similarity network maps two input deep feature vectors to a 2D similarity vector. To do this, they followed the design of the similarity network developed by Mayer and Stamm (2019). To obtain a decision at the video level, a fusion approach based on the mean of the inactivated output layer from the similarity network was presented. This method was tested on the SOCRatES dataset Galdi et al. (2019). The experiments showed that the method improved traditional methods, such as Goljan et al. (2009).

Table 2: The methods presented based on machine learning.

Method	Machine learning approach	Year	Content
Su et al. (2010)	SVM classifier	2010	extracting features: bit stream, quantization factor and motion vectors
Yahaya et al. (2012)	SVM classifier	2012	feature extraction method based on Conditional Probability (CP)
Villalba et al. (2016)	SVM classifier	2016	features based on wavelet transform
Kirchner and Johnson (2019)	CNN architecture	2019	extracting well-characterized noise signals from a given frame
Hosler et al. (2019a)	MISLnet CNN architecture	2020	adding a constrained convolutional layer on gray scale mode
Timmerman et al. (2020)	MISLnet CNN architecture	2020	adding a constrained convolutional layer on RGB mode
Mayer et al. (2020)	MISLnet CNN architecture	2012	the similarity network maps two input deep feature vectors to a 2D similarity vector

The structure of the CNN for the three studies is shown in Figure 10. As shown in the figure, a constrained convolutional layer is added to a simple CNN.

4. Video databases

Although there are several databases that can be used to identify source cameras for images Gloe and Böhme (2010); Shaya et al. (2018), there are few databases for videos. The databases presented based on videos are summarized in Table 3. For a database, it is important to have the following options: number of videos and camera, resolution, codec, GOP, suitable for ISCM or ISCI.

One of the primary reasons for the lower exploration of videos compared to images is that there are few standard digital video databases to develop such methods Hosler et al. (2019b). We thus explore these databases in this section.

CAMCOM2010 Houten et al. is a contest designed to identify source YouTube videos. Two participants submitted results despite a satisfactory number of participants at first. However, the database is not available publicly.

The University of Surrey's website provides access to the SULFA database Qadir et al. (2012) [3] that contains original videos and forged videos. The original videos are suitable for source camera identification purposes. Approximately 150 videos were collected from three sources. The method presented in Rosenfeld and Sencar (2009) was tested in the study. This dataset was also extended by D'Amiano et al. (2015).

The VISION database [4] was introduced in Shullani et al. (2017), which is the most popular database in the field. In total, 35 portable devices from 11 major brands contributed 34,427 images and 1914 videos, all in native format and social format: Facebook, YouTube and WhatsApp are included. There are videos captured in indoor, outdoor, and flat scenarios. Videos of flat surfaces such as walls and sky are included in the flat scenario. Videos depicting offices or shops are included in the indoor scenario, while videos depicting gardens are included in the outdoor scenario. Three recording modes were used for each scenario. In still mode, the user stands still while the video is recorded. While capturing the video, the user walks. The panrot mode combines a pan with a rotation to achieve a recording. YouTube and WhatsApp social media platforms were used to exchange videos belonging to each scenario. In the study, they evaluated the database by the method presented in Chen et al. (2008).

The video-ACID database [5] was presented in Hosler et al. (2019b) to source camera identification that is accessible publicly. Over 12,000 videos were collected from 46 physical cameras representing 36 different camera models in the video-ACID database. All of these videos were shot manually to represent a range of lighting conditions, content, and motion. This database is suitable for both ISCM or ISCI scenarios, which can be used to evaluate the deep learning method presented in Hosler et al. (2019a).

Tian et al. (2019) presented a Daxing smartphone identification database [6], which include both images and videos from extensive smartphones of different brands, models and devices. The data from 90 smartphones, representing 22 models and 5 brands, includes 43400 images and 1400 videos. In the case of the iPhone 6S

[3]http://sulfa.cs.surrey.ac.uk/

[4]https://lesc.dinfo.unifi.it/VISION/

[5]misl.ece.drexel.edu/video-acid

[6]https://github.com/xyhcn/Daxing

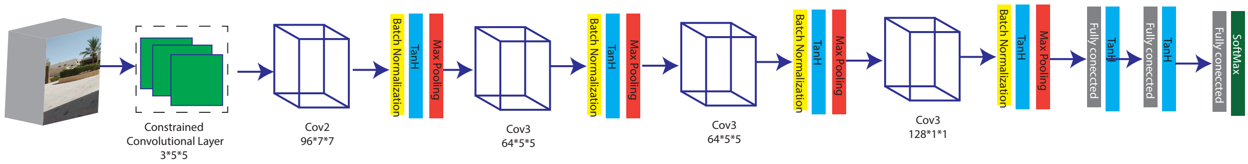

Figure 10: Architecture of ConstrainedNet. (Based on Bayar and Stamm (2018))

Table 3: The databases presented based on videos.

Method	Availability	Year	Description
SULFA Qadir et al. (2012)	Public	2012	150 videos collected from three camera sources
VISION Shullani et al. (2017)	Public	2017	34,427 images and 1914 videos, 35 portable devices of 11 major brands
Video-ACID Hosler et al. (2019b)	Public	2019	12 000 videos from 46 physical devices representing 36 unique camera models
Daxing Tian et al. (2019)	Public	2019	43400 images and 1400 videos captured by 90 smartphones of 22 models belonging to 5 brands
SOCRatES Galdi et al. (2019)	Public	2019	9700 images and 1000 videos captured with 103 different smartphones
QUFVD Akbari et al. (2022)	Public	2022	6000 videos captured with 20 different smartphones

(Plus), 23 different smartphone models are available. Scenes selected normally include a sky, grass, rocks, trees, stairs, a vertical printer, a lobby wall, and a white wall in a classroom, among others. The videos were shot vertically in each scene. Each scene contains at least three videos. In addition, all videos were recorded over 10 seconds. The database was evaluated by the method presented in Goljan et al. (2009).

SOCRatES database [7] Galdi et al. (2019) was captured by smartphones. Approximately 9700 images and 1000 videos were taken by 103 different smartphones from 15 different brands. Lukas et al. (2006) and Li (2010) were assessed on the database.

QUFVD database [8] Akbari et al. (2022) included 6000 videos from 20 modern smartphones representing five brands, each brand has two models, and each model has two identical smartphone devices. This database was evaluated by the deep learning method presented in Bayar and Stamm (2018). This database is suitable for deep learning methods, and the results from this database show that new databases with devices based on new technologies need more improvement in both ISCI and SCMI scenarios.

[7]http://socrates.eurecom.fr/

[8]https://www.dropbox.com/sh/nb543na9qq0wlaz/AAAc5N8ecjawk2KlVF8kfkrya?dl=0

5. Discussion and conclusion

As reported in recent studies such as Akbari et al. (2022); Tian et al. (2019), due to the development of devices with new technologies in imaging, existing methods must be improved or new methods must be developed. Concurrently, developing new databases with devices based on new technologies is important. In practice, identifying source video devices in a large dataset is more challenging than model classification. Although both scenarios need improvement, as shown in the evaluation of new databases, there is more room for improvements on ISCI systems. This fact may lead to more studies being performed recently that focus more on ISCI. Among new databases, only a few databases are suitable for the application of deep learning methods; thus, a combination of existing and new databases may be a solution. In addition, the daily development of deep learning methods in the field can be considered, and deep learning should focus on independence from the content of the video scene (i.e., separating content from noise). Another way to improve results may be to combine the methods and take advantage of both PRNU and machine learning approaches (e.g., using loss functions that can be based on PRNU). In addition, providing pretrained models based on old and new databases for use in new deep learning methods can be useful to improve results.

It is important to present studies that address some questions and open research issues. What is the difference between old and new databases? Whether PRNU is extracted by devices with new compression and stabilization technologies is different from older devices. A comparison showing the impact of compression, resolution, and codec for the new databases compared to existing databases may be useful. The introduction of new databases, particularly video smartphone databases, is essential due to rapid growth. Although deep learning methods are explored in more depth in new databases, a comprehensive comparison between the PRNU and deep learning methods is important for future research. Although there are methods that consider video stabilization, these methods must be improved. Existing methods can also be tested on new social media such as Telegram and TikTok. Another aspect that has been less studied in the literature is the identification of source cameras based on stream networks.

The well-studied topic of digital camera identification on video was examined in this review, and an overview of the definitions and challenges in this field of study was provided. PRNU and machine learning are the two research categories into which we place the methods. A description of video databases that can be used to evaluate camera identification methods is also included. Challenges such as ISCI, SCMI, stabilization, and compression were discussed. As discussed in this review, deep learning methods must be improved to achieve better results in both ISCI and SCMI scenarios.

Acknowledgment

This publication was made possible by NPRP grant # NPRP12S-0312-190332 from Qatar National Research Fund (a member of Qatar Foundation). The statement made herein are solely the responsibility of the authors.

Ahmed, F., Khelifi, F., Lawgaly, A., Bouridane, A., 2019. Comparative analysis of a deep convolutional neural network for source camera identification, in: 2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3), IEEE. pp. 1–6.

Akbari, Y., Al-Maadeed, S., Almaadeed, N., Al-ali, A., Khelifi, F., Lawgaly, A., et al., 2022. A new forensic video database for source smartphone identification: Description and analysis. IEEE Access .

Al-Athamneh, M., Kurugollu, F., Crookes, D., Farid, M., 2016. Digital video source identification based on green-channel photo response non-uniformity (g-prnu), in: Proceedings of the 9th International Conference on Utility and Cloud Computing.

Altinisik, E., Sencar, H.T., . Source camera attribution from strongly stabilized videos .

Altinisik, E., Sencar, H.T., 2020. Source camera verification for strongly stabilized videos. IEEE Transactions on Information Forensics and Security 16, 643–657.

Amerini, I., Caldelli, R., Del Mastio, A., Di Fuccia, A., Molinari, C., Rizzo, A.P., 2017. Dealing with video source identification in social networks. signal processing: Image communication 57, 1–7.

Bayar, B., Stamm, M.C., 2018. Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. IEEE Transactions on Information Forensics and Security 13, 2691–2706.

Bernacki, J., 2020. A survey on digital camera identification methods. Forensic Science International: Digital Investigation 34, 300983.

Boyle, W.S., Smith, G.E., 1970. Charge coupled semiconductor devices. Bell System Technical Journal 49, 587–593.

Brouwers, M., Mousa, R., 2017. Automatic comparison of photo response non uniformity (prnu) on youtube. System and Network Engineering”, styczeń .

Caldelli, R., Amerini, I., Picchioni, F., Innocenti, M., 2010. Fast image clustering of unknown source images, in: 2010 IEEE International Workshop on Information Forensics and Security, IEEE. pp. 1–5.

Chapman, G.H., Thomas, R., Thomas, R., Koren, Z., Koren, I., 2015. Enhanced correction methods for high density hot pixel defects in digital imagers, in: Image Sensors and Imaging Systems 2015, International Society for Optics and Photonics. p. 94030T.

Chen, M., Fridrich, J., Goljan, M., Lukáš, J., 2007. Source digital camcorder identification using sensor photo response nonuniformity, in: Security, steganography, and watermarking of multimedia contents IX, International Society for Optics and Photonics. p. 65051G.

Chen, M., Fridrich, J., Goljan, M., Lukás, J., 2008. Determining image origin and integrity using sensor noise. IEEE Transactions on information forensics and security 3, 74–90.

Chen, S., Pande, A., Zeng, K., Mohapatra, P., 2013. Video source identification in lossy wireless networks, in: 2013 Proceedings IEEE INFOCOM, IEEE. pp. 215–219.

Chen, S., Pande, A., Zeng, K., Mohapatra, P., 2014. Live video forensics: Source identification in lossy wireless networks. IEEE Transactions on Information Forensics and Security 10, 28–39.

Chuang, W.H., Su, H., Wu, M., 2011. Exploring compression effects for improved source camera identification using strongly compressed video, in: 2011 18th IEEE International Conference on Image Processing, IEEE. pp. 1953–1956.

Corcoran, P., Bigioi, P., Chen, J., Cranton, W., Fihn, M., 2014. Consumer imaging i—processing pipeline focus and exposure, in: Handbook of Visual Display Technology. Springer, pp. 1–25.

Cozzolino, D., Verdoliva, L., 2018. Noiseprint: A cnn-based camera model fingerprint. arXiv preprint arXiv:1808.08396 .

D'Amiano, L., Cozzolino, D., Poggi, G., Verdoliva, L., 2015. Video forgery detection and localization based on 3d patchmatch, in: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), IEEE. pp. 1–6.

Dirik, A.E., Sencar, H.T., Memon, N., 2008. Digital single lens reflex camera identification from traces of sensor dust. IEEE Transactions on Information Forensics and Security 3, 539–552.

Elharrouss, O., Almaadeed, N., Al-Maadeed, S., Akbari, Y., 2020. Image inpainting: A review. Neural Processing Letters 51, 2007–2028.

Ferrara, P., Beslay, L., 2020. Robust video source recognition in presence of motion stabilization, in: 2020 8th International Workshop on Biometrics and Forensics (IWBF), IEEE. pp. 1–6.

Ferrara, P., Iuliani, M., Piva, A., 2022. Prnu-based video source attribution: Which frames are you using? Journal of Imaging 8, 57.

Galdi, C., Hartung, F., Dugelay, J.L., 2019. Socrates: A database of realistic data for source camera recognition on smartphones., in: ICPRAM, pp. 648–655.

Gloe, T., Böhme, R., 2010. The'dresden image database'for benchmarking digital image forensics, in: Proceedings of the 2010 ACM Symposium on Applied Computing, pp. 1584–1590.

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

Goljan, M., Fridrich, J., 2008. Camera identification from cropped and scaled images, in: Security, Forensics, Steganography, and Watermarking of Multimedia Contents X, International Society for Optics and Photonics. p. 68190E.

Goljan, M., Fridrich, J., Filler, T., 2009. Large scale test of sensor fingerprint camera identification, in: Media forensics and security, International Society for Optics and Photonics. p. 72540I.

Grundmann, M., Kwatra, V., Essa, I., 2011. Auto-directed video stabilization with robust l1 optimal camera paths, in: CVPR 2011, IEEE. pp. 225–232.

Grundmann, M., Kwatra, V., Essa, I., 2016. Cascaded camera motion estimation, rolling shutter detection, and camera shake detection for video stabilization. US Patent 9,374,532.

Höglund, T., Brolund, P., Norell, K., 2011. Identifying camcorders using noise patterns from video clips recorded with image stabilisation, in: 2011 7th International Symposium on Image and Signal Processing and Analysis (ISPA), IEEE. pp. 668–671.

Hosler, B., Mayer, O., Bayar, B., Zhao, X., Chen, C., Shackleford, J.A., Stamm, M.C., 2019a. A video camera model identification system using deep learning and fusion, in: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 8271–8275.

Hosler, B.C., Zhao, X., Mayer, O., Chen, C., Shackleford, J.A., Stamm, M.C., 2019b. The video authentication and camera identification database: A new database for video forensics. IEEE Access 7, 76937–76948.

Houten, W., Geradts, Z., Franke, K., Veenman, C., . Verification of video source camera competition (camcom 2010), in: Recognizing Patterns in Signals, Speech, Images and Videos.

Huang, Y., Zhang, J., Huang, H., 2015. Camera model identification with unknown models. IEEE Transactions on Information Forensics and Security 10, 2692–2704.

Hyun, D.K., Choi, C.H., Lee, H.K., 2012. Camcorder identification for heavily compressed low resolution videos, in: Computer Science and Convergence. Springer, pp. 695–701.

Iuliani, M., Fontani, M., Piva, A., 2021. A leak in prnu based source identification—questioning fingerprint uniqueness. IEEE Access 9, 52455–52463.

Iuliani, M., Fontani, M., Shullani, D., Piva, A., 2019. Hybrid reference-based video source identification. Sensors 19, 649.

Kang, X., Li, Y., Qu, Z., Huang, J., 2011. Enhancing source camera identification performance with a camera reference phase sensor pattern noise. IEEE Transactions on Information Forensics and Security 7, 393–402.

karakucuk, A., Dirik, A.E., Sencar, H.T., Memon, N.D., 2015. Recent advances in counter prnu based source attribution and beyond, in: Media Watermarking, Security, and Forensics 2015, International Society for Optics and Photonics. p. 94090N.

Kaur, J., Randhawa, D.K.K., . Source identification of videos transmitted in lossy wireless networks .

Kirchner, M., Johnson, C., 2019. Spn-cnn: Boosting sensor-based source camera attribution with deep learning, in: 2019 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE. pp. 1–6.

Kordecki, A., Palus, H., Bal, A., 2015. Fast vignetting reduction method for digital still camera, in: 2015 20th International Conference on Methods and Models in Automation and Robotics (MMAR), IEEE. pp. 1145–1150.

Kot, A.C., Cao, H., 2013. Image and video source class identification, in: Digital Image Forensics. Springer, pp. 157–178.

Kouokam, E.K., Dirik, A.E., 2019. Prnu-based source device attribution for youtube videos. Digital Investigation 29, 91–100.

Kurosawa, K., Kuroki, K., Saitoh, N., 1999. Ccd fingerprint method-identification of a video camera from videotaped images, in: Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348), IEEE. pp. 537–540.

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 Transactions on Information Forensics and Security 12, 392–404.

Lawgaly, A., Khelifi, F., Bouridane, A., 2014. Weighted averaging-based sensor pattern noise estimation for source camera identification, in: 2014 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 5357–5361.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. nature 521, 436–444.

Lefèbvre, F., Chupeau, B., Massoudi, A., Diehl, E., 2009. Image and video fingerprinting: forensic applications, in: Media Forensics and Security, International Society for Optics and Photonics. p. 725405.

Li, C.T., 2010. Source camera identification using enhanced sensor pattern noise. IEEE Transactions on Information Forensics and Security 5, 280–287.

Liu, F., Gleicher, M., Jin, H., Agarwala, A., 2009. Content-preserving warps for 3d video stabilization. ACM Transactions on Graphics (ToG) 28, 1–9.

López, R.R., Luengo, E.A., Orozco, A.L.S., Villalba, L.J.G., 2020. Digital video source identification based on container's structure analysis. IEEE Access 8, 36363–36375.

López, R.R., Orozco, A.L.S., Villalba, L.J.G., 2021. Compression effects and scene details on the source camera identification of digital videos. Expert Systems with Applications 170, 114515.

Lukas, J., Fridrich, J., Goljan, M., 2006. Digital camera identification from sensor pattern noise. IEEE Transactions on Information Forensics and Security 1, 205–214.

Mahalanobis, A., Kumar, B.V., Casasent, D., 1987. Minimum average correlation energy filters. Applied Optics 26, 3633–3640.

Mandelli, S., Argenti, F., Bestagini, P., Iuliani, M., Piva, A., Tubaro, S., 2020. A modified fourier-mellin approach for source device identification on stabilized videos, in: 2020 IEEE International Conference on Image Processing (ICIP), IEEE. pp. 1266–1270.

Mandelli, S., Bestagini, P., Verdoliva, L., Tubaro, S., 2019. Facing device attribution problem for stabilized video sequences. IEEE Transactions on Information Forensics and Security 15, 14–27.

Mayer, O., Hosler, B., Stamm, M.C., 2020. Open set video camera model verification, in: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 2962–2966.

Mayer, O., Stamm, M.C., 2019. Forensic similarity for digital images. IEEE Transactions on Information Forensics and Security 15, 1331–1346.

McCloskey, S., 2008. Confidence weighting for sensor fingerprinting, in: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, IEEE. pp. 1–6.

Meij, C., Geradts, Z., 2018. Source camera identification using photo response non-uniformity on whatsapp. Digital Investigation 24, 142–154.

Mieremet, A., 2019. Camera-identification and common-source identification: The correlation values of mismatches. Forensic science international 301, 46–54.

Milani, S., Fontani, M., Bestagini, P., Barni, M., Piva, A., Tagliasacchi, M., Tubaro, S., 2012. An overview on video forensics. APSIPA Transactions on Signal and Information Processing 1.

Pande, A., Chen, S., Mohapatra, P., Zambreno, J., 2013. Hardware architecture for video authentication using sensor pattern noise.

IEEE transactions on circuits and systems for video technology 24, 157–167.
Pandey, R.C., Singh, S.K., Shukla, K.K., 2016. Passive forensics in image and video using noise features: A review. Digital Investigation 19, 1–28.
Pasquini, C., Amerini, I., Boato, G., 2021. Media forensics on social media platforms: a survey. EURASIP Journal on Information Security 2021, 1–19.
Qadir, G., Yahaya, S., Ho, A.T., 2012. Surrey university library for forensic analysis (sulfa) of video content .
Reddy, B.S., Chatterji, B.N., 1996. An fft-based technique for translation, rotation, and scale-invariant image registration. IEEE transactions on image processing 5, 1266–1271.
Rosenfeld, K., Sencar, H.T., 2009. A study of the robustness of prnu-based camera identification, in: Media Forensics and Security, International Society for Optics and Photonics. p. 72540M.
Sammut, C., Webb, G.I., 2011. Encyclopedia of machine learning. Springer Science & Business Media.
Scheelen, Y., van der Lelie, J., 2012. Camera identification on youtube. Chin. J. Forensic Sci 64, 19–39.
Shaya, O.A., Yang, P., Ni, R., Zhao, Y., Piva, A., 2018. A new dataset for source identification of high dynamic range images. Sensors 18, 3801.
Shullani, D., Fontani, M., Iuliani, M., Al Shaya, O., Piva, A., 2017. Vision: a video and image dataset for source identification. EURASIP Journal on Information Security 2017, 1–16.
Su, Y., Xu, J., Dong, B., Zhang, J., Liu, Q., 2010. A novel source mpeg-2 video identification algorithm. International Journal of Pattern Recognition and Artificial Intelligence 24, 1311–1328.
Taspinar, S., Mohanty, M., Memon, N., 2016. Source camera attribution using stabilized video, in: 2016 IEEE International Workshop on Information Forensics and Security (WIFS), IEEE. pp. 1–6.
Taspinar, S., Mohanty, M., Memon, N., 2020. Camera identification of multi-format devices. Pattern Recognition Letters 140, 288–294.
Taspinar, S., Mohanty, M., Memon, N., 2021. Effect of video pixel-binning on source attribution of mixed media, in: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 2545–2549.
Telecom, I., et al., 2003. Advanced video coding for generic audiovisual services. ITU-T Recommendation H. 264 .
Thivent, D.J., Williams, G.E., Zhou, J., Baer, R.L., Toft, R., Beysserie, S.X., 2017. Combined optical and electronic image stabilization. US Patent 9,596,411.
Tian, H., Xiao, Y., Cao, G., Zhang, Y., Xu, Z., Zhao, Y., 2019. Daxing smartphone identification dataset. IEEE Access 7, 101046–101053.
Timmerman, D., Bennabhaktula, S., Alegre, E., Azzopardi, G., 2020. Video camera identification from sensor pattern noise with a constrained convnet. arXiv preprint arXiv:2012.06277 .
Van Houten, W., Geradts, Z., 2009. Source video camera identification for multiply compressed videos originating from youtube. Digital Investigation 6, 48–60.
Villalba, L.J.G., Orozco, A.L.S., López, R.R., Castro, J.H., 2016. Identification of smartphone brand and model via forensic video analysis. Expert Systems with Applications 55, 59–69.
Virmontois, C., Goiffon, V., Magnan, P., Girard, S., Inguimbert, C., Petit, S., Rolland, G., Saint-Pé, O., 2010. Displacement damage effects due to neutron and proton irradiations on cmos image sensors manufactured in deep submicron technology. IEEE Transactions on Nuclear Science 57, 3101–3108.
Wahab, A.W., Briffa, J.A., Schaathun, H.G., Ho, A.T., 2009. Conditional probability based steganalysis for jpeg steganography, in: 2009 International Conference on Signal Processing Systems, IEEE. pp. 205–209.
Wales, G.S., 2019. Proposed Framework for Digital Video Authentication. Ph.D. thesis. University of Colorado at Denver.
Wang, Z., Zhang, L., Huang, H., 2018. High-quality real-time video stabilization using trajectory smoothing and mesh-based warping. IEEE Access 6, 25157–25166.
Xu, J., Chang, H.w., Yang, S., Wang, M., 2012. Fast feature-based video stabilization without accumulative global motion estimation. IEEE Transactions on Consumer Electronics 58, 993–999.
Yahaya, S., Ho, A.T., Wahab, A.A., 2012. Advanced video camera identification using conditional probability features, in: IET Conference on Image Processing (IPR 2012), IET. pp. 1–5.
Yang, P., Ni, R., Zhao, Y., Zhao, W., 2019. Source camera identification based on content-adaptive fusion residual networks. Pattern Recognition Letters 119, 195–204.
Yang, W.C., Jiang, J., Chen, C.H., 2021. A fast source camera identification and verification method based on PRNU analysis for use in video forensic investigations. Multimedia Tools and Applications 80, 6617–6638.
Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L., 2017. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing 26, 3142–3155.