Multimedia Tools and Applications (2023) 82:33499–33539
https://doi.org/10.1007/s11042-023-14870-8

A comprehensive survey on state-of-the-art video forgery detection techniques

Sk Mohiuddin[1] · Samir Malakar[1] · Munish Kumar[2] · Ram Sarkar[3]

Received: 11 October 2021 / Revised: 6 October 2022 / Accepted: 6 February 2023 /
Published online: 4 March 2023

Abstract

Video plays a key role in carrying authenticity, especially in the surveillance system, medical field, court evidence, journalism, and social media among others. However, nowadays the trust in videos is decreasing day by day due to the forgery of the videos made by easily accessible video editing tools. Hence, a thrust for finding a robust solution to the problem of video forgery detection arises. As a result, researchers around the world are indulging themselves to come up with various methods for the said problem. In this article, we have comprehensively discussed many such initiatives made by researchers across the globe, keeping the focus on recent trends. In addition to this, we have also covered a wide range of forgery detection techniques that follow either an active or a passive approach, while the state-of-the-art surveys made so far on this research topic include only a few specific cases. In this article, we have described some recent technologies that are used in video forging, made a summary of the performances (provided categorically) of all the techniques discussed here, and briefed the available datasets. Finally, we have concluded this survey by clearly mentioning some future directions of the video forgery detection research based on a thorough review of existing techniques.

✉ Munish Kumar
munishcse@gmail.com

Sk Mohiuddin
myselfmohiuddin@gmail.com

Samir Malakar
malakarsamir@gmail.com

Ram Sarkar
ramjucse@gmail.com

[1] Department of Computer Science, Asutosh College, Kolkata 700026, India

[2] Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda 151001, India

[3] Department of Computer Science and Engineering, Jadavpur University, Kolkata 700032, India

Keywords Video forgery detection · Passive approach · Active approach · Copy-move video forgery · Deepfake · Surveillance video · Survey

1 Introduction

A video is simply a sequence of images/pictures, which are also known as frames, arranged in a manner so that when encoded by codec software, it creates an illusion to the human eyes. In recent days, digital video scene/video footage plays a significant role as one of the pieces of evidence in various important sectors that include forensics, surveillance, social network, court premises, and news sectors among others. However, with the simple accessibility of the video editing tools, it turns out to be very simpler to change and/or modify the actual video content, even by an unprofessional user, thereby increasing the number of forged videos at an alarming rate. Using any freely available multimedia content editing software such as Photoshop, Adobe Premiere, Light works, and Cinelerra, one can easily extract all the frames from it and then the actual video can be modified by altering the sequences or by erasing some frames and adding objects to the frames. As a result, we find many doctored videos in our day-to-day life.

Videos are used in various domains like the film industry, product advertisement, news channels, surveillance, video-based lectures, law, enforcement, and many more. Generally, it has been observed that common users trust the contents of a video and hardly think about whether the information, conveyed through a video, is genuine or not. But many times, with this blind belief in every video we come across many problems in our social, personal, and other areas. Verdoliva [146] has presented an analysis of the methods for visual media integrity verification, that is, the detection of manipulated images and videos. In this article, the author has presented highlights of the current forensic tools, the most relevant issues, the upcoming challenges, and suggests future directions. Bonettini et al. [20] have tackled the problem of face manipulation detection in video sequences targeting modern facial manipulation techniques. They have explored the ensembling of different trained Convolutional Neural Network (CNN) models and obtained from a base network (i.e., EfficientNetB4) using two different concepts: (i) attention layers; (ii) Siamese training. They have depicted that combining these networks leads to promising face manipulation detection results on publicly available datasets with more than 119,000 videos. Shelke and Kasana [131] have presented a comprehensive survey on video forgery detection using passive techniques. The primary goal of this survey was to study and analyze the existing passive video forgery detection techniques. Initially, the preliminary information required for understanding video forgery detection is presented, and later, a brief survey of existing passive video forgery detection techniques based on the features, forgery identified, datasets used, and performance parameters detail along with their limitations are reviewed in this article. Mizher et al. [106] have presented a review of video falsifying techniques and video forgery detection techniques. In this review article, several types of techniques are studied and classified and they have presented challenges to existing forgery detection techniques. Finally, they have concluded with recommended suggestions like advanced forgeries such as object motion interpolation forgeries, and dynamic texture inpainting to increase security. Lin et al. [97] have presented a review of recent advancements in passive digital image security forensics. Some of the areas where there is a need of applying a forged video detection technique are as follows:

- ***Social impression:*** Human beings are more social than other species on this planet. All of us have some social value to the other people in society. However, this social value can be hampered just because of a video, which is often made based on lies. For example, the "Synthesizing Obama" program, published in 2017, modified the video footage of former President Barack Obama to depict him mouthing the words contained in a separate audio track [144].

- ***Surveillance system:*** Videos are collected from different places like office rooms, railway stations, and airport premises for monitoring the daily activities therein. One can alter the sequence of the frameset to conceal some event or object from these captured videos. Sometimes this information may be tampered with by eliminating, adding, or altering some frames, and consequently, when the genuine need arises, these generally trusted videos fail to show the truth.
- ***Court evidence:*** In the courtroom, video footage is treated as one of the primary shreds of evidence for a defendant. It helps to take the right decision about any crime suspect(s) by the judge. Failure in differentiating genuine and forged videos may lead to an erroneous decision and as a result, the trust in the law and order might be weakened.
- ***Social network:*** The wide use or exchange of videos on social networks such as WhatsApp, YouTube, Facebook, and news channels has a great impact on our daily lives. A forged video could put a person on the path to suicide when someone posts a fake video on social media by replacing the face with the victim's face in some unpleasant video object as it leads to harassment and humiliation.
- ***Politics:*** In recent times, some politicians have become very much involved in the game to degrade the image of other politicians belonging to other parties through social media. Many videos are commonly found on social media in which the face of some political leaders is replaced in videos of some other incident to infamy them. Any political organization can establish its great impression or degrade the impression of its rivals by spreading videos edited in the said way.
- ***Investigation sector:*** Many investigations largely relied on video footage. Therefore, making and spreading counterfeit video footage on the investigating matter may divert the investigator from the actual line of investigation.
- ***Journalism:*** Due to the widespread use of television and the internet, people follow the news daily. Journalists look for important as well as interesting local or global activities that are happening at every moment to be aware of the people. Even, sometimes it is found that information in the form of a video collected from social media is shown in the news. Any fake video, available through social media, might have a huge ill impact on society.
- ***Religious faith:*** Many times, a fake image or video drives the minds of people toward the religious faith irrespective of their basic senses. An incident like this happened a while back in Russia (2015) when a photo was posted on social media with a *swastika* symbol of a newborn baby. In the early days, all the videos were treated as genuine though they were from different sources [119]. Due to the advanced technologies, the previous faith no longer holds.
- ***Business Sector:*** A high-definition (HD) video usually means that it has a good visual quality. The better the visual quality is, the more popular the video becomes for sharing on the internet. To make it more attractive and to earn more revenue from advertisement, forgers usually prefer to re-encode lower definition videos into HDs directly without any improvement in the content using the parameters of HD videos [162].

1.1 Categorization of video tampering techniques

The process of alternations to produce forged videos is called video tampering, which can be broadly classified into three major categories [159]: (i) Spatial tampering, (ii) Temporal tampering [8, 30], and (iii) Spatio-temporal tampering [84, 96]. Spatial tampering can be done either at pixel level and/or block level at some/all frames. The operations involved are adding, eliminating, or altering the entire or part of an object in a frame in a video. Such objects may be foreground or background. In temporal tampering, the operation is involved at frame level as extra frames are added to or removed from altering one or more objects. Besides, frame sequence can be changed which in turn alters the timing sequence of some events. In the case of spatio-temporal tampering, it is the mixed operation of the said two tampering processes. Inter-frame and intra-frame operations are involved in this tampering method.

1.2 Popular methods used to create forged video

In the past, many algorithms have been proposed in the literature to identify forged videos as well as frames, where forgery took place in digital videos. However, forged video identification is a challenging task while forging is done with some artificial intelligence (AI) based technologies. Three popular methods for making a forged video are described in brief in the following sub-sections.

1.2.1 Chroma key

The principle behind the Chroma key technology is that it uses constant background colour, which is the opposite colour of skin tone [10, 99, 155]. In this manner, a reasonable differentiation between the two (i.e., object and background) is added and thereby making object detection easier. Hence, the selection of the background could be done without much worrying about the frontend object. The completely unchanged colour selection is then replaced with another frame as the background. It is often used in the film industry to replace a scene's background by using a blue or green screen as the initial background and placing the actor in the foreground. This is the most used technology to replace a coloured background with a different setting.

1.2.2 Deepfake

It refers to manipulated videos or other digital representations produced by sophisticated AI techniques that yield fabricated images and sounds that appear to be genuine. Deepfake technology relies on the auto-encoder based deep learning methods [52, 85, 105], where an encoder reduces an image to a lower-dimensional latent space and a decoder reconstructs the image from the latent representation. This means the target's detailed information will be superimposed on the underlying facial and body features of the original video, represented in the latent space.

1.2.3 Frame level operations

Another kind of forgery takes place, in which frames are added from different or from the same video, or frames are deleted to alter the sequence of the actual video, thereby hiding the evidence of a crime. In this approach, a falsifier mainly concentrates on surveillance video because a surveillance video has a fixed background. To hide something, it needs to eliminate those frames that may contain crime footage and then merge them with fresh frames.

1.3 Video forgery detection approaches

Based on the above discussion, it is understood that the impact of fake videos on society is alarming. This inspires the researchers to propose various efficient video forgery detection methods. As a result, many techniques are found in the literature that distinguishes fake videos from genuine ones [34, 81, 143, 152, 161]. The authors classify the different video forgery detection techniques based on methods applied and embedded information during creation in the following subsections.

1.3.1 Categorization based on the methods applied

Generally, two main approaches are found in the literature to detect forged videos namely, *learning based* and *learning free*. These approaches are briefly described in the following subsections. In this context, it is worth mentioning that the first operation is to extract the framesets of the video files irrespective of these said approaches.

Learning based approaches The method which follows a learning based approach [53, 61, 72] needs several video files with appropriate labeling for training the underlying system. In this category of methods, desired features are extracted first from the video frames, and then training is done using some soft computing-based approach. For the easy reference of the readers, we have made a summary of some learning-based methods with some related information in Table 1. As such a system requires a huge amount of video files to train the system, so this can be considered a limitation of the methods that follow this approach. Even, the datasets used for experiments might

Table 1 Description of some learning-based video forgery detection techniques

Method	Forgery type	Features	Classifier	Dataset	Performance
Gupta et al. [53]	Detection of frame repetition and dropping	Motion estimation based	Support vector machine (SVM)	In-house	Not mentioned
Kancherla and Mukkamala [72]	Spatio-temporal forgery	Motion based	SVM	NTHU Forensics project video [110], Video Motion Interpolation for Special Effect [148], Video Inpainting Under Camera Motion [147]	Overall accuracy= 0.87, Precession= 0.89, Recall=0.87
Huang et al. [61]	Inter-frame forgery	Triangular Polarity	SVM	Recognition of human actions [117]	Precision= 0.9576 Recall =0.9826
Yao et al. [158]	Object based forgery in video	Patch based	CNN	SYSU-OBJFORG [29]	Precision =0.9731 Recall =0.9105

face the issue of class imbalance [129, 130]. It also requires prior knowledge regarding forgery types, which is not always possible to decide. It can detect only that type of forgeries using which the underlying system is trained. However, their performance is better than its counterpart i.e., learning free methods, and this may be the prime motivation behind using this approach. A basic model is shown in Fig. 1a which is mostly followed by the researchers in a learning-based approach.

Learning free approaches This category of methods [8, 27, 40, 48, 67, 95] has used various techniques like sensor pattern noise (SPN) estimation, the motion of objects, and the contrast of pixels of each frame to check the authenticity of the videos. It is more advantageous than the previous approach as it does not require any labeled video data. Even, the methods using this approach do not require information related to the forgery type. Despite its benefit, this approach mostly suffers from poor performance than its counterpart [53, 61, 72]. A general framework of video forgery detection using the learning free approach is depicted in Fig. 1b while some techniques along with associated information related to learning free methods are provided in Table 2.

1.3.2 Forged video detection approaches based on embedded information during creation

A device captures the scene frame by frame in front of its lens to take a video. Thus, the digital video is a sequence of frames. These frames can now easily be manipulated, synthesized, and tampered with in different ways without leaving any visible clues. Alteration of frame sequences or tampering with the frames can change the basic property of the original video. Therefore, during video forgery detection one needs to identify whether the actual property (i.e., device information) remains as it is or has been modified. Based on this fact, we can categorize forgery detection techniques into two major approaches namely, active [27, 113] and passive [7, 12, 59, 142]. We describe these categories in the following sub-sections and a comparative study is provided in Table 3.

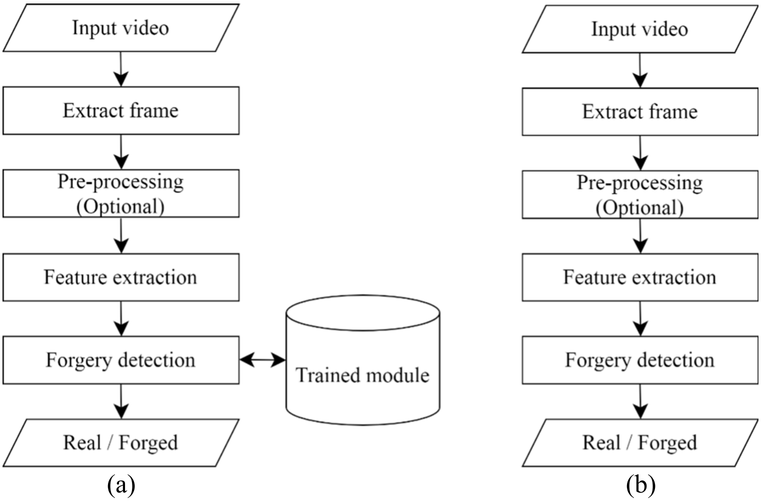

Fig. 1 General framework for video forgery detection methods: **a** learning based, and **b** learning free

Table 2 Tabular description of some learning free based video forgery detection techniques

Method	Forgery type	Features	Dataset	Performance
Elrowayati et al. [43]	Double compression	Prediction Block Feature (PBF)	YUV dataset [164]	Accuracy in [0.79, 0.90]
Su et al. [140]	Copy-move	Exponential Fourier Moment (EFM)	Fast Compressive Tracking [165]	Accuracy = 0.93
Feng et al. [48]	Frame insertion/ deletion	Motion adaptive	YUV Dataset [164]	Accuracy = 0.90
Aghamaleki and Behrad [3]	Frame insertion/ deletion	DCT coefficients	YUV data [164]	Precision in [0.81, 0.975] Recall in [0.773, 0.977]

Active approach The active approach embeds footprint data like watermarks or digital signatures in the video, either at the time of recording or later with the help of some specialized software, to enable verification of the originality, authenticity, and integrity of its contents afterward. The problem with this approach is that sometimes pre-embedding of watermarks or signatures degrades the quality of the video. Besides, all camera manufacturers do not support such pre-embedding. Due to these facts, less attention is paid by the researchers to design methods that follow an active approach. The idea behind such strategies is to check if the source data is intact or not in a questioned video. We have described some of the techniques that follow an active approach in Section 2.

Passive approach For the videos which do not contain any footprint like a digital watermark or digital signature, a passive approach is used to decide whether these are forged or genuine. This approach relies on their intrinsic characteristic to identify their genuineness, and therefore it is called the passive or blind approach. The basic strategy that is mostly followed in this approach is continuity (in terms of bitrates, quality, etc.) among the frames in a video. A passive approach is considered an advanced and the most widely utilized approach in the video forensics domain due to its significant advantages over the active approach. Some of the techniques that follow the passive approach have been described in Section 3.

1.4 Aim of the study

In the previous section, we briefly describe the technologies to make forged videos, and we witness a considerable number of cases that hampers human life due to forged videos. Because of this, forgery detection in digital videos has received much attention from researchers around

Table 3 Comparative study of active and passive approaches used for video forgery detection

Attribute	Active approach	Passive approach
Footprints	To detect video forgery, one requires prior knowledge of information like a watermark or digital signature.	It does not require such prior knowledge to detect forged videos.
Complex Hardware	To generate such types of videos it requires extra hardware to embed information in each frame while capturing the video.	No need for complex hardware to generate such types of videos, and because of this, they are prone to forgery.
Method complexity	Methods for detecting forgery are less complex as they need to check the footprints frame by frame.	More complex to detect forged videos as the videos have fewer or almost no clues left by a forger.

the world. In this context, we would like to mention that we have come across some review works [70, 74, 80, 127] in the literature on this topic. In these works, the authors have emphasized describing some special types of video forgery e.g., intra-frame (or image based) copy-move forgery detection [74], inter-frame based copy-move forgery detection [80], and both types of copy-move forgery detection [70] and the corresponding techniques for detection. To be more specific, these works have mainly focused on image based forgery detection but paid less attention to the video forgery detection algorithms. Moreover, these works have not provided any insightful overview of the techniques that deal with forgery in the H.265/ HEVC videos [32, 57, 59], a relatively new video compression technology.

Considering the above facts, in this paper, we have included various techniques for detecting video forgeries that are created under several scenarios like copy-move, Deepfake, Chroma key, and HEVC standards. Additionally, we have provided a comparative analysis (refer to Table 4) to show clearly how this work is different from the other surveys made on this topic [70, 74, 80, 127]. It is to be mentioned that in this study, our focus is on the recently proposed methods. The distribution of the cited works is shown in Fig. 2.

Recent methods mostly follow a passive approach to identifying whether a video is fake or not [44, 45, 79]. We have found a very smaller number of researchers that are following the active approaches because due to the shortage of real-life data, these approaches are not preferable. Hence, our primary focus of this work is to review the methods that follow the passive approach, yet we have included some of the notable methods that employ the active approach. Fake video detection categories that are considered in the current survey are shown in Fig. 3.

The rest of the paper is organized as follows. Section 2 provides detection methodologies under the active approach. The passive approaches are described in Section 3. Different dataset descriptions are available in Section 4. Future research direction and conclusion are presented in Section 5 and Section 6 respectively. A detailed organization of this survey work is shown in Table 5.

2 Different video forgery detection methods under an active approach

In this section, we have described a forged video detection strategy based on digital signatures and watermarks. However, most of the active forensic approaches are devoted to still image/ frame based analysis. A few of these are discussed here.

A wide variety of video authentication techniques have been developed for the detection of video tampering where pre-embedded information like a watermark or digital signature is attached during the creation of the videos. Fadl et al. [46] have proposed a novel inter-frame forgeries detection system using a 2D convolution neural network of spatio-temporal information and fusion of deep features [36, 38]; Gaussian RBF multi-class support vector machine is used for the classification process. The experimental results show that the efficiency of the proposed system for detecting all inter-frame forgeries is satisfactory even when the forged videos have undergone additional post-processing operations. Park et al. [113] have proposed an algorithm that can detect malicious frames in MPEG-2 video by using a hash function. It first partitions video frames into 8 × 8 discrete cosine transformation (DCT) blocks and then embeds the watermark information into the least significant bit (LSB) based on that DCT. Experimental results show low precision compared to Chen et al. [27], where the reversible semi-fragile method has been proposed to identify video authenticity by embedding two

Table 4 Comparative analysis of our work with some of the recent video forgery survey papers in terms of different aspects covered by these works

Methods	Works covered till the year	Forgery detection using								
		Active approach	Passive approach							
		Digital watermark and digital signature-based forgery detection	Upscale crop forgery detection	Copy-move forgery detection	Double compression forgery detection	Feature based frame insertion/ deletion/ shuffling forgery detection	Motion estimation of object(s) based forgery detection	Source camera artifacts-based forgery detection	Deepfake based video forgery detection	Chroma key-based video forgery detection
Present survey	2022	✓	✓	✓	✓	✓	✓	✓	✓	✓
Kaur and Jindal [74]	2020			✓		✓				
Johnston and Elyan [70]	2018			✓		✓	✓			
Kingra et al. [80]	2016	✓	✓		✓	✓	✓			

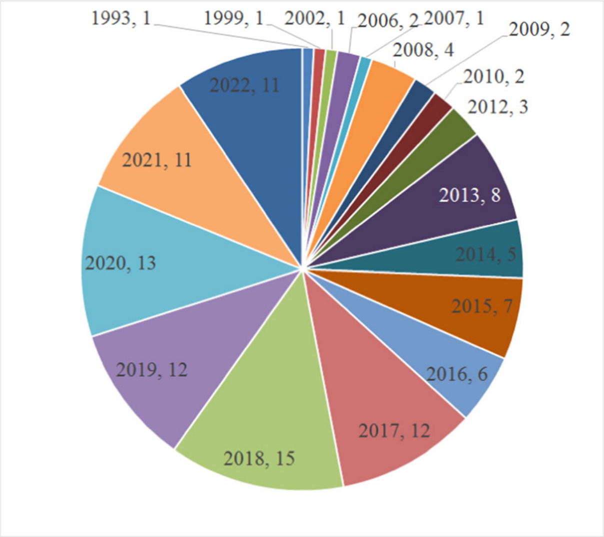

Fig. 2 Shows the year-wise distribution of different methods considered for preparing this survey

footprints into I-frame while encoding it into a MPEG-4 format. One footprint using a hash function aims to authenticate the contents, and embeds the frame number for an alteration location between frames, and the other one is based on discrete cosine (DC) coefficients which are used for the detection of tampered location within the frame.

3 Different video forgery detection methods under a passive approach

As mentioned earlier, in a passive approach, no source of camera information is used for video forgery detection. This category of methods largely depends on the spatial and temporal features of the video frames. Consequently, most of the works have tried to find inter-frame similarities for solving the said problem. We have grouped the methods that follow the passive approach into two major categories that are (i) intra-frame video forgery and (ii) inter-frame

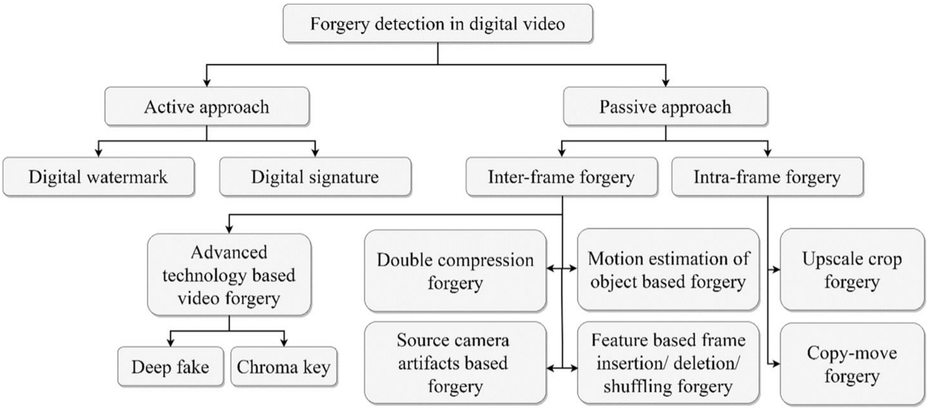
Forgery detection in digital video
Active approach
Passive approach
Digital watermark
Digital signature
Inter-frame forgery
Intra-frame forgery
Advanced technology based video forgery
Double compression forgery
Motion estimation of object based forgery
Upscale crop forgery
Deep fake
Chroma key
Source camera artifacts based forgery
Feature based frame insertion/ deletion/ shuffling forgery
Copy-move forgery

Fig. 3 Classification of video forgery detection techniques that are covered in this survey

Table 5 The overall organization of this article

Topics	Work references
1. **Introduction**	
1.1 Categorization of video tampering techniques	
1.2 Popular methods used to create forged video	
1.2.1 Chroma key	
1.2.2 Deepfake	
1.2.3 Frame level operations	
1.3 Video forgery detection approaches	
1.3.1 Categorization based on the methods applied	
1.3.1.1 Learning based approaches	
1.3.1.2 Learning free approaches	
1.3.2 Forged video detection approaches based on embedded information during the creation	
1.3.2.1 Active approach	
1.3.2.2 Passive approach	
1.4 Aim of the study	
2. **Different video forgery detection methods under the active approach**	
3. **Different video forgery detection methods under a passive approach**	
3.1 Intra-frame forgery detection techniques	
3.1.1 Upscale crop forgery detection techniques	[62, 63, 135]
3.1.2 Intra-frame copy-move forgery detection techniques	[4, 5, 7, 8, 19, 33, 34, 112, 139, 140],
3.2 Inter-frame forgery detection techniques	
3.2.1 Double compressed video forgery detection techniques	
3.2.1.1 Double compression detection techniques in H.265/HEVC format videos	[43, 59, 69, 91, 92, 95, 157, 161]
3.2.1.2 Double compression detection techniques in H.264/MPEG format videos	[25, 68, 142, 143, 150, 151, 156]
3.2.2 Frameset level forged video detection techniques	
3.2.2.1 Forgery detection techniques involving motion estimation of the object(s)	[3, 23, 29, 48, 53, 71–73, 75, 83, 98, 102, 123, 124, 134, 141, 152, 153, 158, 168]
3.2.2.2 Forgery detection techniques utilizing source camera artifacts	[26, 31, 40, 47, 82, 90, 101]
3.2.2.3 Forgery detection techniques employing different features	[2, 11, 12, 44, 45, 57, 60, 67, 76, 79, 81, 87, 88, 100, 107, 128, 132, 137, 166, 167]
3.3 Deepfake forgery detection	[1, 49, 52, 93, 108, 109]
3.4 Chroma key-based forgery detection	[35, 99, 138]
4. **Dataset description**	
4.1 Computational Vision and Active Perception Laboratory (CVAP) Database	[117]
4.2 SULFA dataset	[116]
4.3 REVerse engineering of audio-VIsual coNtent Data video (REWIND) project database	[118]
4.4 GRIP dataset	[51]
4.5 Nimble challenge 2017 dataset	[21]
4.6 Faceforensics++	[120]
4.7 Celeb-DF (V2)	[94]
4.8 Deepfake Detection Challenge (DFDC)	[41]
5. **Future research directions**	
5.1 Need for a realistic solution	
5.2 Improvement of the existing methods	
5.3 Deciding on the video forgery type	

Table 5 (continued)

Topics	Work references
5.4 Dealing with the audio component in digital videos	
5.5 Creation of standardized open access datasets	
6. **Conclusion**	

video forgery. It is to be noted that these categories are further divided into more specific sub-categories while methods of these categories are described. Apart from this, we have described methods that have been introduced to handle forged video generated using Deepfake and Chroma key-based technologies. It is to be noted that these detection strategies follow passive forgery detection protocol and are very popular nowadays.

Apart from the above mentioned categories, we have found a work proposed by He et al. [54] where a CNN model is used to detect relocated I-frames in double compressed H.264 video i.e., this technique handles intra-frame copy-move forgery in double compressed videos. To extract high-frequency components, this method designs a model that is initialized with a pre-processing layer, followed by the CNN model. The three consecutive frame samples are the input to the proposed system that aims to utilize temporal information. The output score of each corresponding input sample is compared to the pre-defined threshold to get the final frame-wise detection result. In this work, they have used a dataset containing YUV video files and the average detection accuracy obtained is 96.73%.

3.1 Intra-frame forgery detection techniques

In the upcoming sub-sections, we have discussed various intra-frame forgery detection methods under different categories. This kind of forgery takes place when editing is done within some frames of a video. Object suppressing, adding, or copy-move inside the same frame, some frame rescaling, etc. are the operations involved while intra-frame forgery is performed. We have categorized the methods as (i) Upscale crop forgery detection techniques and (ii) Intra-frame copy-move forgery detection techniques.

3.1.1 Upscale crop forgery detection techniques

The upscale crop is used to eliminate the evidence of criminal activities or to hide some unnecessary objects that are present at the boundary of a video. For example, a video contains 1–100 frames of dimension 600 × 800 pixels, and the frame sequences 40–70 contain some criminal activities that capture the boundary region of the frames. To hide such details, frame sequences 40–70 can be enlarged beyond size 600 × 800 and then cropped back using the center-crop technique to the original size. Finally, the modified frames are placed at the same position in the video file so that background remains intact. Upscale forgery can be done by erasing or cropping the frames from the first video and then enlarging the affected frames to maintain consistent resolution across the whole video. Figure 4 demonstrates the upscale forgery generation process with an example.

For detecting upscale crops in videos, Hyun et al. [62] have proposed a model which is based on SPN information extracted from the source camera. They have used minimum

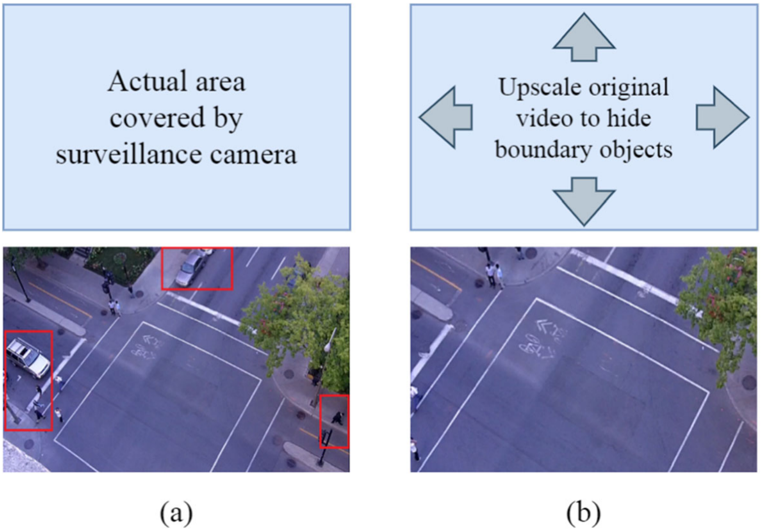
Actual area covered by surveillance camera
Upscale original video to hide boundary objects
(a)
(b)

Fig. 4 Upscale video forging process (top-row): (**a**) frame taken from original video, (**b**) frame taken from forged video made by upscaling the original video to hide boundary regions (red marked)

average correlation energy Mellin radial harmonic (MACE-MRH) correlation filter [78] to unveil the traces of upscaling in videos. This technique can detect partially manipulated regions by excluding the high-frequency components of the investigated video. Singh and Aggarwal [135] have proposed a strategy for the automatic identification of upscale crop fraud video using pixel correlation, noise variations, and inconsistencies between consecutive video frames. They have also used SPN information and frame manipulation detector (FMD) to identify upscale cropping. Movement residuals (P-frames/B-frames in GOP corresponding to I-frames) of the objective video are used to build FMD. Hyun et al. [63] have suggested an approach for the detection of such forgery in surveillance videos with the help of both RGB and infrared videos. They have used the MACE filter to detect the source device and then extracted cropped regions for upscale crop forgery or localize partially manipulated regions. However, the accuracy decreases when the scaling factors do not remain constant.

3.1.2 Intra-frame copy-move forgery detection techniques

Copy-move is one of the popularly used tampering attacks, with variants for the duplicate object(s). A different methodology has been applied for such type of video forensics purposes. A few of these are discussed in this section and a summary of these methods is depicted in Table 6. Aparicio-Díaz et al. [8] have developed a copy-move forgery detection method that is applicable for the forged videos affected by both sub-region and full-frame duplication. This work has first formed a correlation matrix and then used it to detect copy-move attacks. To obtain a set of correlation vectors, they have used block level correlation to form a block level correlation matrix (BCM). They have used the REWIND project video dataset [118] for estimating the performance of their work. This approach may not work well when the threshold and other parameters (e.g., size of tamper region and length of continuous frames) are changed. Another copy-move forgery detection technique proposed by Al-Sanjary et al. [4] has used the optical flow inconsistencies of the objects to detect object cloning. The proposed

Table 6 Comparative analysis of different intra-frame copy-move forgery detection methods

Method	Feature	Dataset	Performance
Aparicio-Díaz et al. [8]	Block correlation based	REWIND [118]	Not mentioned
Al-Sanjary et al. [4]	Optical flow based	SULFA [116]	Accuracy = 0.952
Antony and Devassy [7]	SIFT	Data used in the work [18]	Precision = 0.955 F score = 0.976
Pandey et al. [112]	SIFT	SULFA [116]	Accuracy = 0.996
Su et al. [140]	EFM	Fast Compressive Tracking [165]	Accuracy = 0.93
Bidokhti and Ghaemmaghami [19]	Optical flow based	REWIND [118]	Not mentioned
Su and Li [139]	MI-SIFT	In-house	Accuracy = 0.926

method is based on three steps. The first one is the extraction of all moving objects in the video, followed by the link of each moving object in two successive frames as a displacement path. Afterward, using the dynamic time warping (DTW) based matching algorithm, it detects the clone object by comparing the displacement paths between each other. This method performs well but if the cloned objects are resized before pasting to their new location, then the detection technique may fail.

Antony and Devassy [7] have proposed an image and video copy-move forgery detection method in which the object tracking approach is used in the case of video while adaptive over-segmentation is used for the image. At the beginning of video forgery detection, it divides the input copy-move video into frames and later, converts them to grayscale images. This method first extracts the scale invariant feature transforms (SIFT) features from the frames and then a brute-force matching technique is employed to find similarities. The choice of a brute-force detection mechanism is the main drawback of the method. As a result, Pandey et al. [112] have proposed an intra-frame level copy-move forgery detection method in the spatial and temporal domain of video where the authors have used SIFT features with a dynamic thresholding technique to detect the forged objects in the spatial domain. However, they have used cross-correlation noise residue among the frames to perform object copy-move forgery detection in the temporal domain.

For detecting region duplication using exponential Fourier moments (EFMs) in video frames, a fast forgery detection algorithm is proposed by Su et al. [140]. This work first extracts EFMs based features from each block in the current frame and then employs a new block matching algorithm to detect region duplication in video with mirroring. This block matching technique finds multiple matching pairs with high computational efficiency. To eliminate falsely matched pairs, a post-verification scheme (PVS) is also utilized. At last, to track duplicate regions in the subsequent frames, an adaptive parameter-based fast compression tracking (AFCT) algorithm is proposed. Bidokhti and Ghaemmaghami [19] have proposed another MPEG video based copy-move forgery detection technique. This work has extracted the optical flow coefficient from each video frame. Forgeries are located when an abnormal value in the optical flow coefficient of the suspicious object is detected. In another work, Al-Sanjary et al. [5] have proposed another optical flow inspired method to detect forged regions created by copy-move. First, the method detects the important features (speed and direction of the motion object) of a frame that can be followed in the subsequent frames, and then the optical flow method is applied to determine the displacement of pixels that matches these features among two successive frames.

D’Amiano et al. [33] have proposed a method to detect copy-move video forgery based on a 3D patch matching algorithm [15] and later on, they have proposed another alternative dense field based method [34] to detect both additive and occlusive copy-move forgery. In contrast to the previous one [33], they have introduced a new criterion in the post-processing to tell apart copy-move from false matches [34]. In addition to this, in the work [34], the authors have improved their previous forged detector [33] using multi-scale processing and parallel implementation. In this new concept [34], features are obtained by computing densely on spatio-temporal grids, which allows detecting not only additive but also occlusive forgeries. To this end, they have used the nearest neighbor field (NNF) technique to differentiate the area of the forgery. In addition to these, Su and Li [139] have designed a two-stage method to detect intra-frame region duplication based on video forgery in digital videos. In the first stage of this work, the authors use an improved version of the previously introduced mirror and inversion invariant generalization for SIFT (MI-SIFT) algorithm [103] to extract rotation and scale invariant independent feature of the current frame and then computes the similarities of all feature points in it. In the second stage, it detects tampering areas in the current frame by measuring the ratio of the distance of the closest neighbor to that of the second closest neighbor and compares it with a threshold to determine optimal matching points. At last, it studies the tampered areas obtained from the previous step to locate the tampered areas in subsequent frames.

3.2 Inter-frame forgery detection techniques

Inter-frame forgery is the most common type of video forgery in that mostly frame insertion, deletion, shuffling, and duplication operations are performed. Besides, double compression detection is another variation, where the video is encoded/compressed for a second time. In this section, we have described different types of detection methods those fall under inter-frame forgery detection approaches.

3.2.1 Double compressed video forgery detection techniques

Earlier, researchers treated this detection approach as a traditional one that this type of method is not consider in the forensics study. However, it has been considered a video forgery assuming that there might be some editing in the original video that leads to twice compression. Researchers try to find out whether a video is encoded for the second time or not irrespective of the types like frame insertion, duplication, etc. The characteristics that differentiate between doubly compressed video and actual video, which is compressed only once during preparation, are their intrinsic properties such as frame predictive error, and a group of pictures (GOP) structures. Researchers have invented some special traces that are left when making double compressed videos. In this section, we have discussed different double compression detection methods proposed by different researchers. We have classified all these methods into two sub-categories based on the type of video format used therein namely, high efficiency video coding (H.265/HEVC) and moving picture experts’ group (H.264/MPEG) videos. The main features of H.265 are that it requires much less bandwidth as compared to H.264 codecs and it delivers higher quality video at the same bitrate as H.264. For instance, H.264 video requires 32 Mbps internet speed to broadcast 4 K video while HEVC video can easily be transmitted with just 15 Mbps of internet speed. The compression ratio of H.265 is almost double that of H.264.

Double compression detection techniques in H.265/HEVC format videos H.265/HEVC videos are a relatively the recent form of compressed videos. If a forged video is prepared using this compression mechanism, then it is critical to identify the forgery, because it is hard to identify the structure of the prediction unit (PU) and transforming unit (TU) from where quantization parameters (QPs) and DCs are extracted. In this section, we have described some of the existing methods that follow forgery detection under double compression using HEVC coding, and a summary of these methods is provided in Table 7. Different QPs for double compressed HEVC video are considered in the video tampering detection process by Li et al. [92]. In this work, the size of TU, a fundamental unit of H.265/HEVC technique, and the effects of QPs on the distribution of DCT coefficients are analyzed. Using quantized DCT coefficients and the TU size, a feature vector is designed which can describe statistical differences between single and doubly compressed videos. Finally, it identifies whether a given HEVC video is double compressed or not using the SVM classifier. Another TU based method has been proposed by Yu et al. [161] to detect double compression of HEVC videos. At the outset, they have calculated the histogram of each TU partition type in the first I/P frame of all GOP. Next, a 10-dimensional TU-based feature vector is extracted from the frameset and fed into an SVM classifier to classify single and double compressed videos. The method, proposed by Jiang et al. [69], analyzes the properties of re-encoded frames in double compressed HEVC videos based on PU, another fundamental unit of H.265/HEVC encoding technique. The authors have mainly concentrated on relocated I-frames that are used as evidence to expose double compression. For this, they have utilized the concept of abnormal statistics of PU. A 6-dimensional feature vector is extracted from each GOP collectively rather than the isolated frame. To construct PU sequences for each GOP unit, they have computed the ratio of I-frame PUs and S-frame PUs. Later, a multi-layer perceptron (MLP) is used to classify GOPs, where a classifier fusion strategy is applied.

Huang et al. [59] have proposed a method for the detection of double compression in HEVC videos based on the distribution of quantized DCT coefficients. It has used a 136-dimensional feature vector to classify a video as either double compressed video or single compressed using the SVM classifier. A similar mechanism is used by Li et al. [91], where the authors have used a feature dimension reduction strategy. In this work, the authors have first calculated four 5 × 5 co-occurrence matrices obtained from DCT coefficients along with four directions (horizontal, vertical, main diagonal, and minor diagonal) and then four 4 × 4 co-occurrence matrices are constructed like PUs. Finally, these two feature sets are combined and

Table 7 Comparative study of different double compression detection methods in HEVC format video files

Method	Feature	Feature dimension	Classifier	Dataset	Performance
Yu et al. [161]	TU-based	10D	SVM	DERF's collection [37], YUV dataset [164]	Accuracy ∈[0.84, 0.97]
Jiang et al. [69]	PU-based	6D	MLP	YUV dataset [164]	Accuracy = 0.943
Huang et al. [59]	CTU-based	34D	SVM	YUV dataset [164]	Accuracy = 0.936
Li et al. [91]	PU-based	164D	SVM	YUV dataset [164]	Accuracy = 96.6%
Liang et al. [95]	PU-based	25D	SVM	YUV dataset [164]	Not mentioned
Xu et al. [157]	SN-PUPM	Every single frame with 3 features	SVM	YUV dataset [164]	Accuracy = 0.944
Elrowayati et al. [43]	PBF	31D	–	YUV dataset [164]	Accuracy between [0.78, 0.90]

fed into SVM to detect re-compressed videos. To reduce the feature dimension, only the co-occurrence matrices of DCT coefficients and PU types in the horizontal direction are passed to identify the forgery. In a different method from others, Liang et al. [95] have proposed an effective PU partition based algorithm to detect fake bitrates in HEVC video forensics. According to the authors, PU based partition would change during compression. As a result, a feature vector is designed by the histogram of PU partition types in the first P-frame in each GOP and fed into the SVM classifier. However, the limitation of this approach is that though their system can detect frame deletion, and frame copy-paste tampering under double compression video but fails to check whether the video is double compressed or not under HEVC encoding.

Xu et al. [157] have proposed a sequence numbering process of PUs in the prediction mode (SN-PUPM) for double compression detection on HEVC standard videos. First, they have extracted a few PUs with three kinds of prediction modes, *n*amely, INTRA, INTER, and SKIP, from each frame in each video sequence. Later, SN-PUPM features are calculated and noises are removed in it by taking the difference of absolute values from three adjacent frames. Finally, they have passed these features to the SVM classifier to identify recompression. On contrary, the algorithm proposed by Elrowayati et al. [43] can distinguish single and double-compression in HEVC videos even if the same QPs are used in both compression (i.e., compression during video making and reconstruction). The authors analyzed the changes in the quantized residual discrete sine transform (DST) coefficients and the values of intra-prediction modes. Using this approach, they can differentiate single and double compression in HEVC sequence streams.

Double compression detection techniques in H.264/MPEG format videos MPEG double compression of a video describes whether a video is under double JPEG compression or not. Under double JPEG compression, a relatively larger motion error is encountered between two consecutive frames while moving frames from one GOP to another. Such errors can be estimated by distinguishing the occasional spikes in the discrete Fourier transform (DFT) of the P-frame prediction error sequence. We have described some methods that fall under this category and a summary of these methods is depicted in Table 8. We can find such attempts to detect forged video in the works reported in [150, 151]. Wang and Farid [150] have shown larger motion errors due to double compression by taking double MPEG compression while they have proved it using double compression in their work [151]. In both works, the distribution of DCT coefficients of each macro-block in I-frames is calculated and compared with the actual scenario to measure the amount of contrast deficiency. They have identified a video as forged if a slight variation is found using the Euclidean distance measure. However,

Table 8 Comparative analysis of different double compression detection methods on MPEG format video files

Method	Feature	Dataset	Performance (Accuracy)
Wang and Farid [151]	Not mentioned	In-house	0.994
Subramanyam and Emmanuel [142]	Pixel estimation based	TRACE [149]	0.98
Chen and Shi [25]	Statistical 36D feature vector for each GOP	YUV [164]	0.958
Sun et al. [143]	Statistical 12D feature vector for each GOP based	YUV [164]	0.958
Xu et al. [156]	Statistical GOP based 12D feature	In-house	0.870

such an approach fails when many frames are erased or inserted to maintain the GOP sequence.

In another research work, Subramanyam and Emmanuel [142] have proposed an algorithm that uses principles of estimation theory [77] to determine the double quantization effect that occurs due to double compression in a tampered video. They have considered that video is captured using a static camera and identified double compressed frame(s) by the estimating error between true and estimated quantization effects. One of the limitations of this algorithm is that it is not suitable for detecting a single B or P frame that is affected by double compression.

In another work, Chen and Shi [25] have proposed a method that detects double MPEG-2 compression in videos having constant bitrate (CBR) and variable bitrate (VBR). For this, the authors have considered the distribution of nonzero MPEG/MPEG-2 quantized altering components (AC) coefficients of original video which follows the first digit law (also called Benford's Law [17]) that states that for naturally occurring collection of numbers, the leading significant digit likely follows logarithmic law. However, the same is not true if the video is doubly compressed. Therefore, when there is a violation in the first digit law, they have marked them as double compressed. The GOP is proposed as the detection unit to obtain 36 features to make the detection more reliable. In their experimental work, the primary encoder generates the doubly compressed video, or in other words, the primary and secondary coding processes adopt the same encoder. However, if the secondary encoder is different from the primary one, the first digit distribution of nonzero double quantized AC coefficients may obey the generalized Bendford's Law again, and the detection performance of this algorithm will decrease. Sun et al. [143] have modified the features of the work proposed by Chen and Shi [25] to detect whether the bitrate of the double compression is bigger than that of the single compression or not. However, in their experiments, both the single and the double MPEG-2 compression processes are implemented with the same MPEG-2 encoder. This is a major drawback of the work and eventually, detection performance will decrease when a different MPEG-2 encoder is utilized to realize the double compression.

In another work, Xu et al. [156] have proposed a more robust algorithm to detect double MPEG-2 compression in CBR videos under a more general condition. The output bitrate can be set in a large range, regardless of whether the primary compression and the second compression use the same MPEG-2 encoder or not. The variation process of the DCT coefficient is analyzed in depth during double MPEG-2 compression and convex pattern features are exploited to build a complete detection scheme with the SVM classifier. Their detection algorithm is heuristically designed based on the qualitative impact caused by MPEG-2 compression on the distribution of reconstructed DCT coefficients. It identifies qualitative impact using some statistical analysis on DCT coefficients. This work also demonstrates the differences in the distribution of quantized DCT coefficients between single compression and double compression videos. Another algorithm, which used quantized DCT coefficients during the frame decoding process, has been proposed by Jiang et al. [68]. Markov statistics are proved to be distinguishable for single compression and double compression in JPEG images [28]. The authors have used this concept in videos and extracted statistical Markov model based 162-dimensional features for adaptation to detect double compression.

3.2.2 Frameset level forged video detection techniques

Different forms of frameset level manipulations like copy and paste, deletion, insertion, and shuffling could be made in digital videos. Based on the type of manipulation, the video forgery techniques could be categorized as (i) copy-move, (ii) insertion (iii) deletion and (iv) shuffling. In the case of copy-move, several frames are copied and pasted somewhere in the same or different video to overlay some activity. Besides, this can also be done by inserting frames from other videos or by deleting the frameset. To change the actual sequence of some events in a video, frame shuffling is used. Figure 5 illustrates how these video forgeries are made. Several methods have been developed by researchers to detect such types of video forgery using different properties of video. We have categorized these methods as (i) techniques involving motion estimation of the object(s), (ii) techniques utilizing source camera artifacts, and (iii) techniques consuming different features. All these categories of methods are described hereafter in this subsection.

Forgery detection techniques involving motion estimation of the object(s) Optical flow is the pattern of apparent motion of objects between consecutive frames caused by the movement of the object or camera or both. It is a 2D vector field where each vector is a displacement vector showing the movement of points from the first frame to the second. The motion of an object and the corresponding brightness are important features in video forensics. Based on

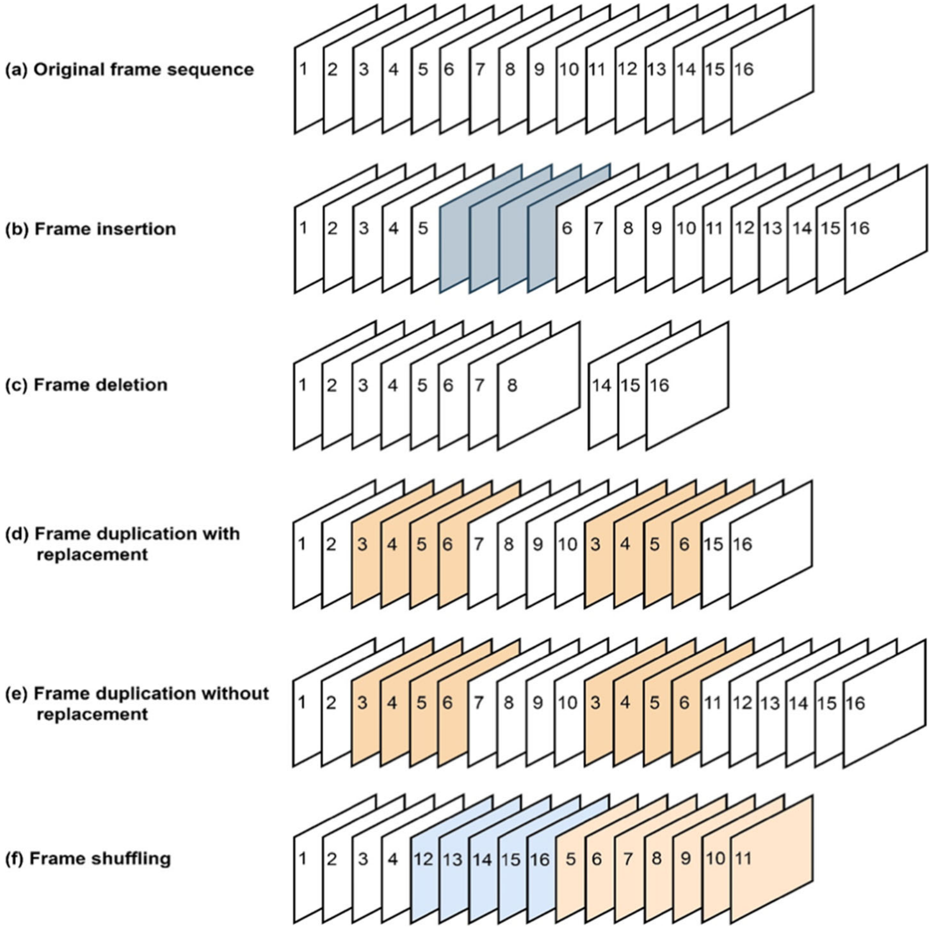

Fig. 5 General processes for different inter-frame video forgeries

these characteristics, many techniques are developed by the researchers. Aghamaleki and Behrad [3] have suggested a method for the detection of forgery in videos using spatial and time domain analysis of quantization effect on I-frame whereas, residual errors in the case of P-frames. First, they have detected single or double compressed video with the help of the SVM classifier using the first significant digit law of DCT coefficients in I-frames. Next, they have utilized time domain analysis of quantization effect on residual errors of P- frames to identify malicious inter-frame forgery comprising frame insertion or deletion.

Gupta et al. [53] have proposed a two-step verification method for frame deletion video forgery detection. First, it computes the time difference between every consecutive frame of a video and removes the low-level noise from them. They have used SVM to classify the video as tampered or non-tampered. However, this technique cannot handle forgery caused by frame alteration, insertion, etc. In another work, Zheng et al. [168] have introduced a block-wise brightness variance descriptor (BBVD) based model that can detect video tampering as well as the location of the frame at which forgery is made. The core idea of their work is that the correlation of adjacent frames in the original video is always very high, which means the corresponding variation is relatively rather low. Therefore, the variance of brightness will be changed largely if some frames are inserted or deleted to tamper with the original content of the original videos.

Liu et al. [98] have proposed an algorithm that can detect frame deletion in the video using the periodicity of the P-frame prediction error sequence. Kang et al. [73] have designed an improved version of the frame deletion detection method using the concept of the work [98] and added the magnitude of the fingerprint in the P-frame prediction error. They have analyzed and proved that the fingerprint of frame deletion still can occur with the help of the anti-forensic method of video frame deletion. They have further estimated the true prediction error and compared it with the prediction error stored in the videos using a counter anti-forensics approach. For this reason, they have claimed that their work is capable of not only detecting video frame deletion but also can identify frame insertion forgery.

Singh and Aggarwal [134] have developed a DCT coefficient analysis based hybrid system for the detection of frame deletion in re-compressed videos. To do this, they have first detected the presence of decompressed frames in MPEG-4 and H.264/AVC videos and then labeled them as 'possibly forged' because of the presence of re-compression artifacts that do not ensure forgery. As a result, after that, they have used the optical flow component in the direction of the brightness gradient only to be sure of whether the frame deletion happened or not. Another optical flow based technique is proposed by Wang et al. [152]. It facilitates an authentication technique for digital videos by identifying the inter-frame forgery process. First, they have calculated the optical flow variation sequence and decided the presence of anomalies to locate discontinuity points. By exploiting characteristics of the optical flow sequence and discontinuity points, they have analyzed the affected areas (i.e., frame deletion, insertion, and duplication) in a surveillance video.

Wu et al. [153] have checked the consistency of the velocity field in the surveillance video to detect inter-frame forgery. The proposed method can be used to identify different forgery types like deletion or duplication of consecutive frames. In addition to this, it can distinguish the tampered videos from the genuine ones and locate the manipulated positions in forged videos. To perform the said tasks, the authors have first used block based cross-correlation coefficients to obtain the velocity field sequence and then calculated the corresponding relative factors from this velocity field sequence. Finally, they have determined the forgery type and manipulated locations with a generalized extreme studentized deviate (ESD) algorithm [64].

Chao et al. [23] have proposed an inter-frame video forgery detection method based on optical flow consistency. For detection of inserted frames, a window based rough detection method followed by a binary search scheme is employed. However, frame-to-frame optical flow and double adaptive threshold are used for identifying frame deletion.

Chen and Tan [29] have proposed another type of object based forgery detection mechanism which can automatically identify videos, encoded with advanced frameworks and their forged segment location. Frame manipulation detector using motion residual of the target video frame sequence is used to detect the traces in the video frames. Kancherla and Mukkamala [72] have also focused on detecting frame based tampering by detecting motion in videos. In their work, first, a base frame is captured by applying window based averaging on successive frames, and then motion information is extracted by subtracting the base frame from the actual frame of the video. Finally, they have applied Markov models with an SVM classifier to this motion residue.

Based on the texture and micro-patterns of two consecutive frames, a method is proposed by Saddique et al. [123] for detecting forged video segments and marking the forged frames. Employing the discriminative robust local binary pattern (LBP) [14], dubbed DRLBP, and chrominance value of consecutive frame difference (CCD) a new descriptor is designed to model the inconsistency embedded among the frames at the inter-frame level due to forgery. It detects consecutive frames as either forged or genuine using the SVM classifier. Object-based forgery detection in the advanced video using deep learning based approach is proposed by Yao et al. [158]. To extract high dimensional features automatically from input image patches, it uses a pre-trained CNN model. Before being fed into the CNN model, the video frames are passed through some pre-processing steps. They have also used an asymmetric data augmentation strategy to get a similar number of positive and negative image patches for model training.

Saddique et al. [124] have proposed another method that focuses on tampering detection in the spatial domain of videos. The method is based on a deep CNN model, which is composed of three layers. Motion residual (MR) is the first layer, which takes a video clip as input and computes the motion residual for each of the R, G, and B channels. Later, it first concatenates these motion residuals and then passed them to the next CNN layers. CNN layers involve convolutional, pooling, and fully connected layers to compute the hierarchical representation of the video clip. Finally, parasitic layers work as a classifier and predict whether the video clip is authentic or tampered with using discriminant features extracted from the CNN layers and motion residual layer. Similarly, a two-step CNN based method is proposed by Kohli et al. [83] that determines forgery in HD videos. In the beginning, the CNN model decides whether the video is affected by double compression or not, and if affected then it detects forged frames from double compressed videos using temporal features. Finally, it localizes the forged region using spatial features. To detect an object-based forgery in HD videos, it uses motion residual.

A slightly different CNN based model is proposed by Kaur and Jindal [75]. This work detects inter-frame tampering in the videos as well as localizes the externally added object in the video if any. In this work, first, a 12-layered CNN model takes frames as input and then each layer of the model extracts some significant features based on temporal and spatial correlation values of the consecutive frames. Once the forged frames are classified using correlation values, the forged frames are fed into another 90-layered CNN model for testing the localization of the counterfeit region. In the second part of this method, frames are segmented into meaningful fragments and the network has trained with colour intensity values. If any abnormality is present in the copied region, then they are considered as the regions with the counterfeit. The proposed work has computed its simulation results on the REWIND [118] and GRIP [51] video datasets.

Based on the statistical properties of interfering frames, such as relocated I-frames, Feng et al. [48] have designed a motion adaptive forensic method for detecting frame deletion in digital video. Features are obtained using frame motion residuals to detect frame deletion points. Later, it identifies frame deletion points using the fluctuation feature (FF) which is further enhanced by an intra-prediction elimination procedure. To this end, it eliminates the minor interference of sudden lighting change, vibration in focusing, and frame jitter using a post-processing procedure.

Another kind of spatial forgery that takes place in digital video is moving object removal to hide some unwanted scenes in front of the camera. Su et al. [141] have proposed a novel forgery detection method for the detection of moving foreground pixels removal from video. Initially, the proposed method calculates the energy factor (EF) of each frame to identify forged frames. An adaptive parameter based visual background extractor (AVIBE) algorithm is then used to detect suspected regions from the forged frames determined in the first stage. After eliminating false detection by calculating the difference of EF between suspected regions in the forged frames and the corresponding regions in the authentic frames, the algorithm finally locates the tampering traces. However, the accuracies of the proposed system for the static background and complex background are 93.17% and 86.58%, respectively. Joshi and Jain [71] have proposed an algorithm to detect video forgery where they have first decomposed all frames from a video and then measured frame level prediction error between two consecutive frames. Finally, to detect and localize video tampering, they have used an optical motion score that is calculated using the method proposed by Lucas-Kanade [102]. The main drawback of the proposed system is that it needs manual intervention to decide the originality of the video in question. Besides, the accuracy of their method is low as compared to other methods. Table 9 depicts the comparative analysis of different motion estimation and object based forgery detection methods.

Forgery detection techniques utilizing source camera artifacts Recording devices keep some detectable traces, generated during the video creation and subsequent processing, as a part of the recorded videos. These detectable traces are termed as the source camera artifacts in literature. These artifacts are used by many researchers to separate genuine videos from forged ones. One such example is SPN which has been used to trace the source identification and thereby for detection of forgeries. In the early days, such artifacts were mostly utilized for source camera identification, but some authors exploited them as a means of tamper detection as well. Here, the basic idea is to determine whether all the scenes of the video were recorded using the same camera or not. Source camera identification based methods that are found in the literature for forgery detection follow this approach. Pioneer work, designed by Li [90], can identify the source camera based on SPN which it extracts from digital images. This SPN serves as fingerprints of recording devices and this characteristic has widely been used as an effective way for digital device identification. With the help of SPN, Kirchner and Johnson [82] have presented an advanced source camera identification based on sensor noise in a data-driven framework. This method relies on a deep learning approach and uses CNN that extracts SPN from a single image during test time. In the preliminary stage, they have developed this approach for image based camera identification and in the later stage, they have applied it to video frames and got satisfactory results. In another work, Corripio et al. [31] have proposed a

Table 9 Comparative analysis of different inter-frame forgery detection methods involving motion estimation of objects

Method	Dataset	Performance
Aghamaleki and Behrad [3]	YUV data [164]	Precision ∈[0.81, 0.975] Recall ∈ [0.773, 0.977]
Kang et al. [73]	YUV data [164]	Accuracy (frame deletion) ∈ [0969, 0.993]
Singh and Aggarwal [134]	SULFA [116], Change detection video database [22]	Accuracy = 0.993
Wang et al. [152]	TREC Video Retrieval Evaluation [145]	Accuracy ∈ [0.867, 0.933]
Wu et al. [153]	TREC Video Retrieval Evaluation [145]	Accuracy ∈ [0.80, 0.90]
Kohli et al. [83]	SYSU-OBJFORG dataset [29]	Accuracy ∈ [0.974, 0.989]
Saddique et al. [123]	SULFA [116] and databases used in the works [9, 18, 58]	Accuracy = 0.966
Yao et al. [158]	SYSU-OBJFORG data set [29]	Accuracy = 0.9845±0.0037

Kancherla and Mukkamala [72]	NTHU Forensics project [110],
	Video Motion Interpolation for Special Effect [148],
	Video Inpainting Under Camera Motion [147].
		Accuracy = 0.871
		Precision = 0.89
		Recall = 0.86

Kaur and Jindal [75]	REWIND [118] and GRIP [51] video dataset	Accuracy = 0.98
Feng et al. [48]	YUV Dataset [164]	Accuracy = 0.90
Saddique et al. [124]	SULFA [116] and databases used in the works [9, 18, 58]	Accuracy = 0.989

Su et al. [141]	Self-made	Accuracy = 0.9317 for static background
		Accuracy = 0.866 for complex background

machine learning based source camera identification technique where they have used wavelet based features along with SPN and SVM as the classifier.

López et al. [101] have proposed a clustering based method to identify the source of digital videos. They have identified the source of video acquisition through the characteristics of its internal elements and the metadata. Later, they have used two clustering algorithms - one is density based ordering points to identify the clustering structure [6] and another is a hierarchical clustering algorithm to make the grouping of brand and model as well as digital videos based on the source. A customized parser is used to extract all file format structures of videos. The work has been tested on videos captured by 14 mobile phones, 19 digital cameras, and 6 video editing toolboxes. There is another technique to detect forged videos by identifying the source device. The identification technique by Chen et al. [26] uses a maximum likelihood estimator (MLE) to generate estimates of photo response non-uniformity (PRNU) from video clips whereas Dirik et al. [40] have proposed a method to identify the source camera by detecting dust spot on the camera lens which creates noise in the image frames of the video. These dust spots have been used as footprints for detecting forged video. Fayyaz et al. [47] have designed an improved surveillance video forgery detection technique using SPN and the correlation of noise residues. SPN based methods have some problems like the attacker can induce the SPN by first de-noising the forged frames and then adding the SPN to each of them. However, the authors have claimed that they can identify an attack where the attacker induces SPN to the forged frames to make the frames SPN vulnerable. They can detect such forgeries using the integrated forgery detection system, which is based on the correlation between noise residues of consecutive frames to the SPN based forgery detection system.

Forgery detection techniques employing different features In the past few years, a significant number of techniques have been developed that have relied on the discriminating power of texture based features [24, 121] while performing inter-frame forgery detection. Hong et al. [57] have presented a forgery detection approach based on the deletion of frames from H.265/HEVC coded videos. They have focused on specific coding pattern changes due to the deletion of frames, which break some of the regular patterns of CU/PU/TU coding information under a double compression domain. It has identified discriminating coding patterns and made them a single feature vector by considering three features of the intra-CU area (ICUA), total residual TU energy, and the number of TUs combined for each GOP. Later they have used the MLP classifier for the detection of deleted frames.

Fadl et al. [45] have developed a robust method to detect inter-frame duplication forgery using a temporal average (TP) of each shot with high accuracy and low running time. First, they have partitioned the input video sequence into smaller clips based on edge change ratio instead of using fixed size for clips which helps in reducing the running time. It uses the TP of each clip as a discriminating frame instead of entire frames, which is invariant to frame order. After that, they have used a gray level co-occurrence matrix (GLCM) [122] to extract statistical textural features of each TP image. These features are utilized to investigate TP similarities that have achieved improved accuracy while performing frame duplication detection. In the method proposed by Fadl et al. [44], a histogram of oriented gradients (HOG) features [16, 125] and a motion energy image (MEI) are used. From the different images, HOG is used to acquire discriminative features for detecting anomalies and localize frame insertion and frame deletion forms of video forgery. To detect frame deletion and frame insertion, it relies on correlation coefficients, and abnormal points are extracted via Grabb's test. Besides, MEI is used to edge images of each shot to detect frame duplication and shuffling.

Zhao et al. [167] have broken the sequence of frames using a boundary detection method and grouped them as shots. To detect and locate tampered frames in the video shot, they have calculated Hue-Saturation and Saturation-Value colour histograms of every frame in a video shot and compared the similarity between histograms. Further, they have detected the types of forgery like frame insertion, deletion and duplication using speeded up robust features (SURF) together with the fast library for approximate nearest neighbors (FLANNs) matching for double checking. Bakas et al. [12] have presented a two-step technique to detect frame insertion, deletion, and duplication in forged video. In the first step, they have detected outlier frames [56], based on Haralick coded frame correlation, and in the second step, they have fine-tuned the same to eliminate false positive cases.

Huang et al. [60] have proposed an algorithm called multi-level subtraction (MLS), which can detect frame insertion forgery using subtraction of a pre-set threshold value. In this work, block wise intensity change ratios have been calculated for a frame first and then subtracted the change ratio features from the consecutive frames to a certain level for the recognition of forged boundary. Later, they have modified the old technique [61] and designed a robust and threshold-less framework dubbed triangular polarity feature classification (TPFC) for the detection of video frame insertion and deletion based forgeries. In this work, they have generated feature vectors from the whole video by using three levels of subtraction based on MLS. After extraction of discriminative features, it performs video forgery classification using an SVM classifier. For experimental work, they have considered the human actions database [117]. However, the dataset does not contain forged videos and therefore, they have built an in-house dataset by applying frame insertion and deletion.

Kingra et al. [81] have introduced a method for automatic detection and localization of said type of video tampering by utilizing prediction residual gradient (PRG) and optical flow gradient (OFG) based features. This method has measured the degree of variation of PRG's in two consecutive frames and if it generates false alarms, then only they have made use of OFG features to enable forgery detection in all kinds of video sequences. A hybrid technique is used by Kaur and Kaur [76] for differentiating original and forged frames where they have used discrete wavelet transform (DWT) [154] and SIFT features while the optical flow is estimated to detect forged frames. The major drawback of the technique is that the detection accuracy is too low as compared to other inter-frame forgery detection methods. Long et al. [100] have proposed a learning based method to detect frame deletion from single shot video with the help of local spatio-temporal relationships within a portion of a video. They have used a 3D CNN that can detect the exact location of frame dropping.

Shanableh [128] has proposed a machine learning based technique to detect frame deletion in videos. The proposed solution can detect forged videos regardless of the number of deleted frames if it is not a multiple of the length of a GOP. It uses 8 features from P and B frames like mean prediction residual energy, mean of quantization scale values, etc. After extracting features, it checks with a pre-trained module to detect frame deletion forgery. Compared to the previous one, a different approach is proposed by Aghamaleki and Behrad [2], which can detect frame insertion or deletion and double compression with different GOP structures and lengths. Further, it detects the abnormal properties in the P frames, i.e., quantization error in rich areas, and reduces the effect of motion on residual errors of P frames. Later, they have proposed a wavelet transforms-based algorithm to detect and localize video forgery based on the quantization error in the frequency domain and spatially constrained residual errors of P frames. They have used 22 YUV raw video sequences in QCIF and CIF formats from the video trace library for their experiment [149].

A two-step algorithm is proposed by Kharat et al. [79], in which the suspicious (duplicated) frames are identified and their features are extracted to compare with other frames of the test video. They have used SIFT for comparison. To take the decision, the Random Sample Consensus algorithm is used to locate duplicate frames. Singh and Singh [137] have proposed a relatively different approach for detecting duplicate frame(s) as well as the region(s) at which duplication is made in forged videos. Their work has considered two algorithms. The first algorithm can detect three different forms of frame duplication viz., i) consecutive frame sequences of large length at the continuous position, ii) consecutive frame sequences with varying lengths and positions, and, iii) any length consecutive frame sequences from other videos with varying positions. To do this, the authors have converted grayscale frames into the DCT matrix. Again, they have calculated the correlation coefficient of each sequence with other sequences of the entire video. Whereas, the second algorithm has performed regional duplication detection for regular and irregular regions by locating the position of the error with the thresholding process to calculate the similarity between regions of two frames or within the affected frame. Jia et al. [67] have suggested a novel approach to detect frame level copy-move forgery. Initially, they have calculated the optical flow sum consistency to find suspected tampered points. After that, they have compared optical flow correlation to match duplicated frame pairs and followed by validation checks to reduce false detections.

For the detection of frame insertion and deletion, another two-step method is proposed by Zhang et al. [166]. In the first step, each frame of a video is coded by the LBP i.e., quotients of correlation coefficients, and in the second step, it calculates quotients of correlation coefficients among sequential LBP coded frames. To detect abnormal points, i.e., localization of frame insertion and deletion it has used Tchebyshev inequality twice followed by abnormal point detection using a decision threshold. Mohiuddin et al. [107] have used a structural similarity index measure (SSIM) among consecutive frames to detect frame insertion for both surveillance and non-surveillance videos. Another inter-frame forgery detection method proposed by Bakas and Naskar [11] have used 3D-CNN architecture to detect the types of inter-frame video forensics i.e., it detects whether insertion, deletion, or duplication of the frame has occurred or not. It uses two 3D-CNN networks: one is to generate spatio-temporal features while the other targets to extract the temporal information. Kumar et al. [87] have performed detection of frame insertion forgeries by measuring the correlation coefficient between extracted deep features extracted from consecutive frames in the questioned video. The model proposed by Shelke and Kasana [132] has used Polar Cosine Transform (PCT) and Neighborhood Binary Angular Pattern (NBAP) to extract features from videos and feed them to CNN based GoogleNet model for the detection of inter and intra forgeries. Frame insertion and deletion forgery is detected by Kumar and Gaur [88] and they have identified the forgery location by measuring frame correlation distance between the frames. Table 10 depicts the analysis of different frame insertion/deletion/shuffling based forgery detection techniques.

3.3 Deepfake forgery detection

As already mentioned, that Deepfake is a new video forgery technique where the face of the objects (mostly humans) is targeted and the recent trend to make such tampering in digital videos is widespread over social media. Figure 6 illustrates some real faces and their corresponding Deeepfakes faces.

Table 10 Comparative analysis of different feature-based video forgery detection techniques

Method	Feature	Database source(s) of the videos in use	Performance
Fadl et al. [45]	GLCM	SULFA [116], LASIESTA [89], IVY LAB [66]	Precision = 0.94 for frame duplication with shuffling Precision = 0.99 for frame duplication
Fadl et al. [44]	HOG, MEI	SULFA [116], LASIESTA [89], IVY LAB [66]	Precision = 0.96 and Recall = 0.94 for frame duplication

Huang et al. [61]	TPFC	Recognition of human action [117]	Recall = 0.982
			Precision = 0.957

Zhao et al. [167]	HSV, SURF	In-house	Precision = 0.98
			Recall = 1.00

Bakas et al. [12]	GLCM, Haralick	SULFA [116], TRACE [149], and YouTube videos [160]	Precision = 0.98
			Recall = 0.97

Kingra et al. [81]	PRG and OFG	DIC-Panjab University	Frame detection accuracy = 0.89 Duplicate region localization accuracy = 0.81
Kaur and Kaur [76]	DWT, SIFT	REWIND [118]	Precision = 0.61
Long et al. [100]	–	Nimble Challenge [21]	Accuracy = 0.981
Bakas and Naskar [11]	CNN	In-house	Accuracy = 0.96

Kharat and Chougule [79]	SIFT	YouTube video	Precision = 0.99
			Recall = 0.99

Jia et al. [67]	Video motion based	VTL, SULFA [116], DERF collections [37]	Precision in [0.912, 0.985] Recall in [0.747, 0.985]

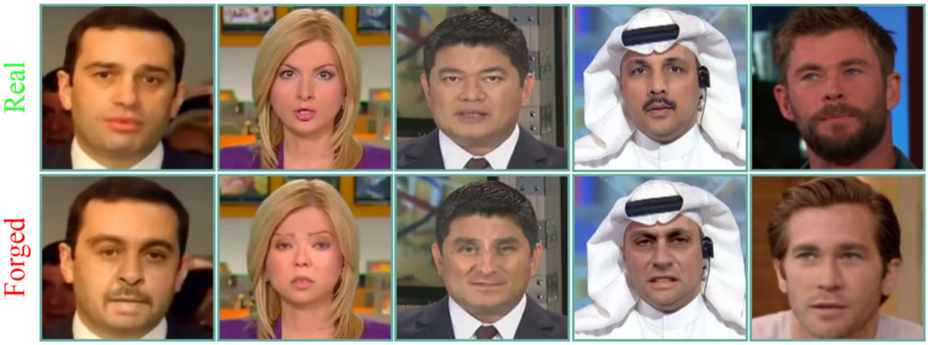

Fig. 6 Some sample real faces along with their corresponding Deepfake faces

Afchar et al. [1] have proposed a method to automatically and efficiently detect face tampering. They have designed two CNN models with a small number of layers dubbed Meso-4 and MesoInception-4 for detecting such types of forgery. Their designed networks are based on well-established CNN networks [86, 133], mostly used for image classification. In the same context, Guera and Delp [52] have used a long short-term memory (LSTM) network [55], which is a particular type of recurrent neural network (RNN). The authors have used a collection of 600 videos (fake video: real video = 1:1) to evaluate the method and they have reported the highest accuracy of 97.1%. The method proposed by Nguyen et al. [109] can detect various kinds of forgeries like replay attack detection, facial re-enactment detection, and also face-swapping detection using a Capsule network with the visual geometry group (VGG-19) network [133] as a backbone CNN model. In contrast to these works, Li et al. [93] have used eye blinking characteristics in the videos for the detection of a similar kind of forgery. The proposed method has used a deep neural network model that combines CNN and RNN, called long-term recurrent CNN (LRCN) [42], to distinguish open and closed eyelids. LRCN model first locates the face areas in each frame of the video using a face detector model and then the eye regions are classified as either closed state or open state with the help of an RNN-LSTM model [55]. In the RNN-LSTM, the features from a pre-trained VGG-16 framework [133] are used. Recently, Ganguly et al. [49] have proposed local-global attention based CNN network [50] which is empowered by using a visual transformer while Mohiuddin et al. [108] have concatenated deep features extracted from RGB, HSV, YCbCr colour space for detection of Deepfake videos.

3.4 Chroma key-based forgery detection

Chroma key-based video forgery is another recent video forgery technique like Deepfake technology. Unlike Deepfake, it normally targets background forging to remove or add different background other than the actual one to hide some information. Figure 7 shows some forgery samples based on the Chroma key method. To detect such forgery, Liu et al. [99] have designed a 3-stage foreground analysis and tracking (3FAT) algorithm that relies on blue screen compositing. In the beginning, the foreground blocks are identified by a multi-pass foreground locating method. For this, they have used a series of operations like a Gaussian mixture model (GMM), binarization, de-noising, morphological opening, and connected component labeling. In the second stage, it measures the similarity between the foreground blocks and the background using local features consisting of luminance and contrast of the frame. Finally, a fast target search algorithm is designed based on compressive tracking that

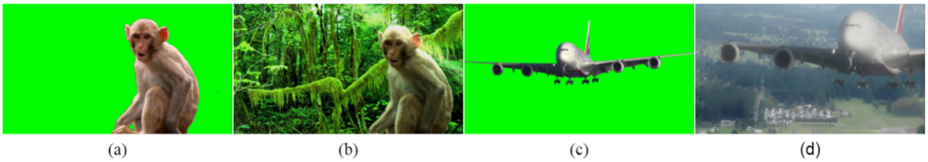

Fig. 7 Chroma key-based forged sample frames: (**b**) and (**d**) are frames taken from fake videos while the corresponding real frames are shown in (**a**) and (**c**) respectively. The fake frames are made by removing the green background and substituting it with other backgrounds

first detects the tampered block in the first tampered frame and then goes on tracking the block of subsequent frames. Singh and Singh [138] have computed the frame differences of each frame pair of the video and identified the edges of different frames. After that, it differentiates the edge pixels of each edge frame into large and small edge pixel values by applying a threshold to detect Chroma key foreground forgery. Whereas, for the same purpose, D'Avino et al. [35] have designed a splicing based model that moves with Chroma key composition. An auto-encoder based RNN model that follows an anomaly detection strategy is adopted. During anomaly detection, the method checks the feature vector that does not fit with the intrinsic model stored in the network parameter.

3.5 Dataset description

Datasets are an integral part of any research and development work. Without an appropriate dataset, testing the efficiency of any proposed model is not possible, and it limits the usage of the model in real-life scenarios. Most of the research attempts mentioned above have used some video datasets for experimenting with their work. However, in some cases, the authors preferred to work with their self-made datasets, which are not available to the research community. In this context, it is worth mentioning that many researchers have used some available and standard video datasets databases, and they have also modified the datasets for their experimental needs, but such datasets are also not available to the research community. Here we have provided a brief description of available datasets on which most of the referred work relied and Table 11 provides the downloadable links of the datasets along with the forgery types for which the datasets are used.

3.6 Computational vision and active perception laboratory (CVAP) database [117]

It contains videos of six types of human actions (i.e., boxing, handclapping, hand waving, jogging, running, and walking) performed by 25 subjects under four different recording conditions labeled as S1-S4 (S1: outdoor, S2: scaled version of S1, S3: outdoor with different clothing and S4: indoor). All videos are recorded with 25 frames per second using a static camera. The objects are in front of homogeneous background for all the recording conditions. The entire database is freely available to the research community.

3.7 SULFA dataset [116]

SULFA dataset is designed by the University of Surrey, which contains original as well as forged video files. There are overall1 150 videos collected from different 3 camera sources, which are Nikon S3000 (codec MJPEG), Canon SX220 (codec H.264), and Fujifilm

Table 11 Description of some important video datasets

Dataset	Types of forgery	#Videos	Link
CVAP [117]	Original with homogeneous background	2391	https://www.csc.kth.se/cvap/actions/
SULFA [116]	Spatial and temporal domain	150	Not directly downloadable
REWIND [118]	Copy-move	20	https://sites.google.com/site/rewindpolimi/downloads/datasets/video-copy-move-forgeries-dataset
GRIP [51]	Copy-move	154	http://www.grip.unina.it/
	Splicing	10	
	Deepfake	5000	
Panchal and Shah [111]	Frame deletion, Frame duplication, and Frame insertion	210	https://drive.google.com/drive/folders/1y_TVO6-ow2yoKGSLLvj-GAYf_-HtCLw4
Al-Sanjary et al. [65]	Copy-move, Splicing, and Swapping-frames.	52	https://www.youtube.com/channel/UCZuuu-iyZvPptbIUHT9tMrA/videos
Fadl et al. [45]	Frame duplication, Frame shuffling	53	https://drive.google.com/open?id=1CE6_X010dyQiMBpAlgoU18yWQxD1ckeU
Faceforensics++ [120]	Deepfakes, FaceSwap, Face2Face, NeuralTextures	5000	https://github.com/ondyari/FaceForensics
Celeb-DF (V2) [94]	Deepfakes	6229	https://github.com/yuezunli/celeb-deepfakeforensics
DFDC [41]	Deepfakes	138,154	https://ai.facebook.com/datasets/dfdc/

S2800HD (codec MJPEG). Each video length is approximately 10 seconds with a resolution of 320×240 and 30 frames per second. All videos preserve both spatial and temporal video characteristics.

3.8 REVerse engineering of audio-VIsual coNtent data video (REWIND) project database [118]

This dataset contains 20 videos (10 original and 10 forged) with a resolution of 320×240 pixels and a frame rate of 30 fps. Original sequences have been recorded using low-end devices, thus all are compressed at the origin (using either MJPEG or H.264 codecs). Forged sequences have been saved as an uncompressed file (RV24, 24-bit RGB). To make all the sequences to the same standard, all are converted into uncompressed YUV (4:2:0) files. It is to be noted that some of the original sequences come from the Surrey University Library for Forensic Analysis (SULFA) database [116]. An additional. MAT file (i.e., MATLAB file) is supplied, with each forged sequence, which contains the differences among the Y, U, and V components for each frame of the original and the forged sequences.

3.9 GRIP dataset [51]

GRIP is a research group working on image processing at the University Federico II, Naples. The main interests of this research group are image segmentation, denoising, image coding, and, multimedia forensics. The group uploads codes and datasets for different research contributions [51]. As our current intention covers video forgery only, thus we describe some video datasets provided by this group. They have made two forged video datasets namely, Copy-move and Deepfake which are freely available to the research community. The forged video dataset is categorized into splicing and copy-move. The splicing dataset contains 10 forged videos along with the original ones while the other one contains 60 original videos and 95 forged videos with associated ground truth information.

3.10 Nimble challenge 2017 dataset [21]

This dataset contains 209 investigation videos with different manipulations. Among these, there are 6 videos manipulated with “Temporal Remove” or “frame dropping”.

3.11 Faceforensics++ [120]

Faceforensics++ is a dataset with 1000 original videos that have been tampered with by four automated face manipulation methods: Deepfakes, FaceSwap, Face2Face, and NeuralTextures. The samples are gathered from 977 YouTube videos, and all videos contain a trackable mostly frontal face without occlusions.

3.12 Celeb-DF (V2) [94]

The Celeb-DF dataset has 590 original videos gathered from YouTube with subjects of different ages, ethnic groups, and genders, and 5639 corresponding synthetic videos (i.e., forged) of different celebrities generated by an improved Deepfakes manipulation method.

3.13 Deepfake detection challenge (DFDC) [41]

The Deepfake Detection Challenge (DFDC) dataset was introduced in a Kaggle contest and later made public by Facebook AI, which is one of the most comprehensive third-generation forensics datasets to date. The training, validation, public test, and private test sets consist of 119,154, 4000, 5000, and 10,000 video clips of 10 seconds in length respectively.

4 Future research directions

From the above discussion, we understand that the research topic, considered here for surveying, has the utmost importance in different fields that include surveillance, law, social media, and many more. Although a decent number of research works are found in the literature, still these mentioned efforts are yet to provide a real-life solution. In this section, we have tried to pinpoint some of the future research directions that need to be taken care of by the researchers.

4.1 Need for a realistic solution

In our day-to-day life, we find many doctored videos on social media like WhatsApp, Facebook, and Instagram that are shared for different purposes like fun, to make someone or some community popular or infamous, spreading of religious faith, advertisement, and so forth. Despite the research growth on video forensics, in the past few years, no such forensics strategy is available using which we can confirm that this is a doctored video with 100% assurance. Even on the internet, we find the same video with different resolutions while the truth is that these videos are captured with a single resolution and then converted to different resolutions using a standard video converter. Although we do not consider this type of video as doctored, such videos are also forged as source camera information gets altered. These are some examples that convey the need for hours in the domain of video forgery detection. Therefore, one obvious future research direction is to come up with some comprehensive methods that might help in identifying such doctored videos and eventually stop spreading these.

4.2 Improvement of the existing methods

In Sections 2 and 3, we have described that many researchers have tried hard to make a robust model to deal with the different kinds of forgeries in videos, still, there are certain areas where more attention is utmost required. In this subsection, we try to find out the limitations of the existing methods as well as identify the possible measures that could be followed to improve the same keeping in mind the current video forgery technologies.

In the methods which follow the active approach, video forgery detection algorithms have used either watermark information or source camera artifacts. To obtain this information, researchers have mainly relied on DCT coefficients, which are extracted from the frames using some image forgery detection based concepts. However, these works have mostly ignored the temporal relationship among the frames. Considering the recent advancements in deep learning models, the use of CNN could be beneficial to solve the problems of active video forgery detection. More specifically the use of a graph neural network (GNN) can be a viable option as it can model the input of elements and their inter-dependency.

Under the passive approach, a significant number of research attempts are found for double compressed forgery, upscale cropping, and copy-move forgery (intra−/inter-frame) detection. Most of the works have used fixed GOP structures as video encoders as they are easier to implement while widely used compression standards like H.265 and H.264 have employed adaptive GOP structures [136]. Hence, such techniques might fail for videos that are encoded using these newer video codecs. Mandelli et al. [104] tackled the problem of blind video temporal splicing detection leveraging PRNU-based source attribution. They considered videos composed of a few seconds of shots coming from various sources and they performed analysis on a recently released dataset composed of videos acquired with mobile devices. The method is validated on both non-stabilized and stabilized videos.

Apart from these, if we solely consider the intra-frame forgery detection methods, then we can see two variants of approaches (upscale crop and object-based copy-move forgery detection) are found in the literature. Methods, mentioned in the literature, can identify forged videos that are affected by upscale crop forgery that has performed well for the constant scaling factor. But, in reality, we can come across videos with variant scaling factors [63]. Therefore, improvement in such methods is required so that they can perform well for the variant scaling factor. On the other hand, in copy-move forgery detection methods tampering with a single object is considered which is not always true in real life. Hence, there is an urge to devise techniques that can identify frames that are altered by modifying multiple objects. Besides, the features used by both categories of methods that have dealt with intra-frame forgery detection are hand-engineered and in some cases, the feature dimension is too high. Therefore, we can apply some feature selection approaches, specifically wrapper [13, 126] or hybrid approach [39, 163], to get rid of redundant and noisy features. We can also rely on CNN-based feature extraction to upgrade the performance of the existing methods that deal with intra-frame forgery detection. It is also to be noted that these methods have made use of different classifiers i.e., followed the learning-based approach. Therefore, one can try some clustering-based methods to handle these two-class problems as it is very hard to collect sufficient data to train a learning-based model. An anti-forensic method adapts and re-models the forgery process intending to diminish the traces that are created by some object addition or removal from digital video frames and thereby making the detection process critical. The use of anti-forensic strategy to prepare object-based forgery and thereby detect them using counter anti-forensic strategy got much attention from the researchers for images but similar tries in the case of video are less in literature. Very few methods [73, 98] could be found in the literature that have dealt with such anti-forensically prepared forged videos. Therefore, in the future, attention can be switched to such kind of forged video detection method.

Deepfake and Chroma key are the most recent AI-based forged video-making technologies. We have already described some methods that can detect the forged videos prepared using these technologies but those are not enough to get a competent solution for such a multifaceted problem. Existing methods have mostly relied on CNN based models like Capsule network [109], LRCN [42], and LSTM [55] with VGG-16/19 CNN architecture as the backbone. Even the performances of these models are not sufficient when considering the inter-dataset evaluation strategy [50, 108]. Therefore, other standard CNN based architectures could be used in the future. Researchers may think of forming an ensemble of different deep learning-based models [114, 115].

4.3 Deciding the video forgery type

An interesting observation about the methods mentioned in this study is that mostly these methods are devised for some specific type of forgery. In other words, the authors of the methods have first considered the type of forgery and then proposed a model to detect such forgery. This is one of the pure limitations of the existing methods if we would like to apply these in real-life scenarios. Because in practical cases, it is not an easy task to predict what kind of video forgery we are dealing with. This implies that unless we know the forgery type, available methods cannot be applied. Furthermore, a doctored video can be affected by multiple tampering attacks. Therefore, we need a serious thought over this research gap so that we can have a viable solution to the video forgery detection problem irrespective of its type.

4.4 Dealing with the audio component in digital videos

A video is not all about visual components only. Sound/audio is also an important part of this. All the methods related to forgery detection, described above, have taken the decision based on visual components of the video only, while entirely ignoring the other part. But the recent doctored videos that spread over the internet are not always tampered with the visual part but also in some cases the audio part is altered. Therefore, the decision-making process (i.e., to decide whether a video in question is forged on not) should incorporate audio components in the future to have a more generic and practical solution for the video forgery detection problem.

4.5 Creation of standardized open access datasets

Many researchers, described earlier, have experimented with their methods on self-made/synthetic datasets that are prepared by introducing some form of forgeries. In the few cases, open-access datasets, described in the previous section, are used. However, these datasets are mostly prepared for experimenting with some special type of video forgery detection. The forged video dataset containing multiple tampering attacks is absent in the literature. Therefore, there is a need to have such types of datasets. Even, the datasets, mentioned earlier, are not close to the viral videos that are found every moment on social media networks. Therefore, the preparation of such forged video data is an essential requirement before designing a multifaceted video forensic method. Besides, more open-access datasets are required and this would eventually help the research community to design practical solutions, as they no more would evaluate their method on synthetic or self-made datasets.

Precisely, a standardized dataset, considering the real-life challenges regarding video forensics, should contain doctored videos having the following features:

- Same videos with different resolutions.
- Videos that are tampered with in visual components, audio components, and both.
- Inter-frame copy-move forgeries like deletion, duplication, and insertion of the frames in the same video.
- Videos with upscale cropping and multiple objects tampering in the same frame.
- Chroma key and Deepfake based forged videos.
- Enough video files to train the state-of-the-art deep learning based classifiers.

- Availability of the commonly found doctored videos on social media/internet.

5 Conclusion

In this survey article, we have discussed several issues related to video forgery and the methods proposed by various researchers over the last few years to deal with these. Video forgery detection is an important but critical research topic and it is getting more difficult day by day due to the continual changes in compression as well as codec technology. Specifically, the arrival of Deepfake and Chroma key technologies to make forged videos provides an added challenge to the researchers. Besides, detection of copy-move, deletion of frames, and addition of frames in a digital video require more advanced algorithms to identify forgery and the corresponding forged region. It is also observed that although being an active research area, a practical solution to the said problem is yet to achieve. Several techniques have been developed to detect forgery videos until now, but there is no universal algorithm/tool to identify the type of tampering in a video. Also, a very limited number of datasets is available to the researchers to work in this growing but complex research field. Therefore, in a nutshell, we can say that despite the significant advancements, there are many aspects of video forgery detection, which need more attention from the researchers, as discussed in Section 5, to have a practical solution for this problem.

Data availability Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Declarations

Research involving human participants and/or animals No human and animal participants were used.

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

1. Afchar D, Nozick V, Yamagishi J, Echizen I (2018) Mesonet: a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–7
2. Aghamaleki JA, Behrad A (2016) Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding. Signal Process Image Commun 47: 289–302. https://doi.org/10.1016/j.image.2016.07.001
3. Aghamaleki JA, Behrad A (2017) Malicious inter-frame video tampering detection in MPEG videos using time and spatial domain analysis of quantization effects. Multimed Tools Appl 76:20691–20717
4. Al-Sanjary OI, Ahmed AA, Bin JAA, et al (2018) Detection clone an object movement using an optical flow approach. In: 2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE). IEEE, pp 388–394
5. Al-Sanjary OI, Ahmed AA, Ahmad HB, et al (2018) Deleting object in video copy-move forgery detection based on optical flow concept. In: 2018 IEEE conference on systems, Process and Control (ICSPC). IEEE, pp 33–38
6. Ankerst M, Breunig MM, Kriegel H-P, Sander J (1999) OPTICS: ordering points to identify the clustering structure. ACM SIGMOD Rec 28(2):49–60. https://doi.org/10.1145/304181.304187

7. Antony N, Devassy BR (2018) Implementation of image/video copy-move forgery detection using brute-force matching. In: 2018 2nd international conference on trends in electronics and informatics (ICOEI). IEEE, pp 1085–1090
8. Aparicio-Díaz E, Cumplido R, Pérez Gort ML, Feregrino-Uribe C (2019) Temporal copy-move forgery detection and localization using block correlation matrix. J Intell Fuzzy Syst 36:5023–5035. https://doi.org/10.3233/JIFS-179048
9. Ardizzone E, Mazzola G (2015) A tool to support the creation of datasets of tampered videos. In: International Conference on Image Analysis and Processing. Springer, pp 665–675. https://doi.org/10.1007/978-3-319-23234-8_61
10. Bagiwa MA, Wahab AWA, Idris MYI, Khan S, Choo KKR (2016) Chroma key background detection for digital video using statistical correlation of blurring artifact. Digit Investig 19:29–43. https://doi.org/10.1016/j.diin.2016.09.001
11. Bakas J, Naskar R (2018) A digital forensic technique for inter–frame video forgery detection based on 3D CNN. In: International Conference on Information Systems Security. Springer, pp 304–317
12. Bakas J, Naskar R, Dixit R (2019) Detection and localization of inter-frame video forgeries based on inconsistency in correlation distribution between Haralick coded frames. Multimed Tools Appl 78:4905–4935. https://doi.org/10.1007/s11042-018-6570-8
13. Banerjee D, Chatterjee B, Bhowal P, Bhattacharyya T, Malakar S, Sarkar R (2021) A new wrapper feature selection method for language-invariant offline signature verification. Expert Syst Appl 186:115756. https://doi.org/10.1016/j.eswa.2021.115756
14. Barburiceanu S, Terebes R, Meza S (2021) 3D texture feature extraction and classification using GLCM and LBP-based descriptors. Appl Sci 11(5):2332. https://doi.org/10.3390/app11052332
15. Barnes C, Shechtman E, Finkelstein A, Goldman DB (2009) PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Trans Graph 28(3):24
16. Barua S, Malakar S, Bhowmik S, Sarkar R, Nasipuri M (2017) Bangla handwritten city name recognition using gradient-based feature. In: 5th international conference on Frontiers in intelligent computing: theory and applications. Springer, Singapore, pp 343–352
17. Benford F (1938) The law of anomalous numbers. Proc Am Philos Soc 78(4):551–572. http://www.jstor.org/stable/984802
18. Bestagini P, Milani S, Tagliasacchi M, Tubaro S (2013) Local tampering detection in video sequences. In: 2013 IEEE 15th international workshop on multimedia signal processing (MMSP). IEEE, pp 488–493. https://doi.org/10.1109/MMSP.2013.6659337
19. Bidokhti A, Ghaemmaghami S (2015) Detection of regional copy/move forgery in MPEG videos using optical flow. In: 2015 the international symposium on artificial intelligence and signal processing (AISP). IEEE, pp 13–17. https://doi.org/10.1109/AISP.2015.7123529
20. Bonettini N, Cannas ED, Mandelli S, Bondi L, Bestagini P, Tubaro S (2021). Video face manipulation detection through ensemble of cnns. In: 25th international conference on pattern recognition (ICPR). IEEE, pp 5012–5019
21. Nimble Challenge (2017) https://www.nist.gov/itl/iad/mig/nimble-challenge-2017-evaluation. Accessed 30 Sep 2022
22. Change detection video database (2022) http://changedetection.net/. Accessed 30 Sep 2022
23. Chao J, Jiang X, Sun T (2012) A novel video inter-frame forgery model detection scheme based on optical flow consistency. In: International Workshop on Digital Watermarking. Springer, pp 267–281
24. Chatterjee A, Malakar S, Sarkar R, Nasipuri M (2018) Handwritten digit recognition using DAISY descriptor: a study. In: proceedings of 5th international conference on emerging applications of information technology (EAIT 2018). IEEE, pp 1–4
25. Chen W, Shi YQ (2008) Detection of double MPEG compression based on first digit statistics. In: International Workshop on Digital Watermarking. Springer, pp 16–30
26. Chen M, Fridrich J, Goljan M, Lukáš J (2007) Source digital camcorder identification using sensor photo response non-uniformity. In: security, steganography, and watermarking of multimedia contents IX. SPIE, pp 517–528
27. Chen H, Chen Z, Zeng X, Fan W, Xiong Z (2008) A novel reversible semi-fragile watermarking algorithm of MPEG-4 video for content authentication. In: Second International Symposium on Intelligent Information Technology Application. IEEE, pp 37–41. https://doi.org/10.1109/IITA.2008.451
28. Chen C, Shi YQ, Su W (2008) A machine learning based scheme for double JPEG compression detection. In: Proceedings of International Conference on Pattern Recognition (IAPR). IEEE, pp 1–4
29. Chen S, Tan S, Li B, Huang J (2015) Automatic detection of object-based forgery in advanced video. IEEE Trans Circuits Syst Vid Technol 26:2138–2151
30. Chittapur G, Murali S, Anami BS (2019) Video forgery detection using motion extractor by referring block matching algorithm. Int J Sci Technol Res 8:3240–3243

31. Corripio JR, González DMA, Orozco ALS, Villalba LJG, Hernandez-Castro J, Gibson SJ (2013) Source smartphone identification using sensor pattern noise and wavelet transform. In: 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013). pp 1–6. https://doi.org/10.1049/ic.2013.0267
32. Costanzo A, Barni M (2016) Detection of double AVC/HEVC encoding. In: 24th European Signal Processing Conference (EUSIPCO). IEEE, pp 2245–2249. https://doi.org/10.1109/EUSIPCO.2016.7760648
33. 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
34. D'Amiano L, Cozzolino D, Poggi G, Verdoliva L (2018) A patchmatch-based dense-field algorithm for video copy–move detection and localization. IEEE Trans Circuits Syst Vid Technol 29:669–682
35. D'Avino D, Cozzolino D, Poggi G, Verdoliva L (2017) Autoencoder with recurrent neural networks for video forgery detection. IS T Int Symp Electron Imaging Sci Technol 92–99. https://doi.org/10.2352/ISSN.2470-1173.2017.7.MWSF-330
36. Das S, Chatterjee A, Dey S, Saha S, Malakar S (2023) Breast cancer detection from histology images using deep feature selection. In: Proceedings of International Conference on Frontiers in Computing and Systems. Springer, pp 323–330
37. DERF's collection (2022) https://media.xiph.org/video/derf/. Accessed 30 Sep 2022
38. Dey S, Roychoudhury R, Malakar S, Sarkar R (2022) An optimized fuzzy ensemble of convolutional neural networks for detecting tuberculosis from chest X-ray images. Appl Soft Comput 114:108094
39. Dey C, Bose R, Ghosh KK, Malakar S, Sarkar R (2022) LAGOA: learning automata based grasshopper optimization algorithm for feature selection in disease datasets. J Ambient Intell Humaniz Comput 13:3175–3194
40. Dirik AE, Sencar HT, Memon N (2008) Digital single lens reflex camera identification from traces of sensor dust. IEEE Trans Inf Forensic Secur 3:539–552
41. Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, Ferrer, CC (2020) The deepfake detection challenge (dfdc) dataset. arXiv Prepr arXiv200607397
42. Donahue J, Hendricks LA, Guadarrama S, Rohrbach M, Venugopalan S, Saenko K, Darrell T (2015) Long-term recurrent convolutional networks for visual recognition and description. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp 2625–2634
43. Elrowayati AA, Abdullah MFL, Manaf AA, Alfagi AS (2017) Tampering detection of double-compression with the same quantization parameter in HEVC video streams. In: 2017 7th IEEE international conference on control system, Computing and Engineering (ICCSCE). IEEE, pp 174–179
44. Fadl S, Han Q, Qiong L (2020) Exposing video inter-frame forgery via histogram of oriented gradients and motion energy image. Multidim Syst Signal Process 31:1365–1384
45. Fadl S, Megahed A, Han Q, Qiong L (2020) Frame duplication and shuffling forgery detection technique in surveillance videos based on temporal average and gray level co-occurrence matrix. Multimed Tools Appl 79:17619–17643
46. Fadl S, Han Q, Li Q (2021) CNN spatiotemporal features and fusion for surveillance video forgery detection. Signal Process Image Commun 90:116066. https://doi.org/10.1016/j.image.2020.116066
47. Fayyaz MA, Anjum A, Ziauddin S, Khan A, Sarfaraz A (2020) An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues. Multimed Tools Appl 79:5767–5788
48. Feng C, Xu Z, Jia S, Zhang W, Xu Y (2016) Motion-adaptive frame deletion detection for digital video forensics. IEEE Trans Circuits Syst Vid Technol 27:2543–2554
49. Ganguly S, Mohiuddin S, Malakar S, Cuevas E, Sarkar R (2022) Visual attention-based deepfake video forgery detection. Pattern Anal Appl 25(4):981–992
50. Ganguly S, Ganguly A, Mohiuddin S, Malakar S, Sarkar R (2022) ViXNet: vision transformer with Xception network for deepfakes based video and image forgery detection. Expert Syst Appl 210:118423
51. GRIP (2022) http://www.grip.unina.it/. Accessed 30 Sep 2022
52. Güera D, Delp EJ (2018) Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS). IEEE, pp 1–6
53. Gupta A, Gupta S, Mehra A (2015) Video authentication in digital forensic. In: 2015 international conference on futuristic trends on computational analysis and knowledge management (ABLAZE). IEEE, pp 659–663
54. He P, Jiang X, Sun T, Wang S, Li B, Dong Y (2017) Frame-wise detection of relocated I-frames in double compressed H. 264 videos based on convolutional neural network. J Vis Commun Image Represent 48:149–158

55. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
56. Hodge V, Austin J (2004) A survey of outlier detection methodologies. Artif Intell Rev 22:85–126
57. Hong JH, Yang Y, Oh BT (2019) Detection of frame deletion in HEVC-coded video in the compressed domain. Digit Investig 30:23–31
58. Hsu C-C, Hung T-Y, Lin C-W, Hsu C-T (2008) Video forgery detection using correlation of noise residue. In: 2008 IEEE 10th workshop on multimedia signal processing. IEEE, pp 170–174
59. Huang M, Wang R, Xu J, et al (2015) Detection of double compression for HEVC videos based on the co-occurrence matrix of DCT coefficients. In: International Workshop on Digital Watermarking. Springer, pp 61–71
60. Huang CC, Zhang Y, Thing VLL (2017) Inter-frame video forgery detection based on multi-level subtraction approach for realistic video forensic applications. In: 2017 IEEE 2nd international conference on signal and image processing (ICSIP). IEEE, pp 20–24
61. Huang CC, Lee CE, Thing VLLL (2020) A novel video forgery detection model based on triangular polarity feature classification. Int J Digit Crime Forensic 12:14–34. https://doi.org/10.4018/IJDCF.2020010102
62. Hyun D-K, Ryu S-J, Lee H-Y, Lee H-K (2013) Detection of upscale-crop and partial manipulation in surveillance video based on sensor pattern noise. Sensors 13:12605–12631
63. Hyun D-K, Lee M-J, Ryu S-J, et al (2013) Forgery detection for surveillance video. In: The Era of Interactive Media. Springer, pp 25–36
64. Iglewicz B, Hoaglin D (1993) Volume 16: how to detect and handle outliers. In: the ASQC basic references in quality control: statistical techniques. Quality Press, Welshpool
65. Ismael Al-Sanjary O, Ahmed AA, Sulong G (2016) Development of a video tampering dataset for forensic investigation. Forensic Sci Int 266:565–572. https://doi.org/10.1016/j.forsciint.2016.07.013
66. IVY LAB (2022) http://ivylabdb.kaist.ac.kr/base/dataset/data.php. Accessed 30 Sep 2022
67. Jia S, Xu Z, Wang H et al (2018) Coarse-to-fine copy-move forgery detection for video forensics. IEEE Access 6:25323–25335
68. Jiang X, Wang W, Sun T, Shi YQ, Wang S (2013) Detection of double compression in MPEG-4 videos based on Markov statistics. IEEE Signal Process Lett 20:447–450
69. Jiang X, He P, Sun T, Wang R (2019) Detection of double compressed HEVC videos using GOP-based PU type statistics. IEEE Access 7:95352–95363
70. Johnston P, Elyan E (2019) A review of digital video tampering: from simple editing to full synthesis. Digit Investig 29:67–81
71. Joshi V, Jain S (2020) Tampering detection and localization in digital video using temporal difference between adjacent frames of actual and reconstructed video clip. Int J Inf Technol 12:273–282
72. Kancherla K, Mukkamala S (2012) Novel blind video forgery detection using markov models on motion residue. In: Asian Conference on Intelligent Information and Database Systems. Springer, pp 308–315
73. Kang X, Liu J, Liu H, Wang ZJ (2016) Forensics and counter anti-forensics of video inter-frame forgery. Multimed Tools Appl 75:13833–13853
74. Kaur H, Jindal N (2020) Image and video forensics: a critical survey. Wirel Pers Commun 112:1–22
75. Kaur H, Jindal N (2020) Deep convolutional neural network for graphics forgery detection in video. Wirel Pers Commun 112:1763–1781. https://doi.org/10.1007/s11277-020-07126-3
76. Kaur R, Kaur EJ (2016) Video forgery detection using hybrid techniques. International Journal of Advanced Research in Computer and Communication Engineering 5(12):112–117. https://doi.org/10.17148/IJARCCE.2016.51221
77. Kay SM (1993) Fundamentals of statistical signal processing: estimation theory. Prentice Hall PTR
78. Kerekes RA, Vijaya Kumar BVK (2006) Correlation filters with controlled scale response. IEEE Trans Image Process 15(7):1794–1802. https://doi.org/10.1109/TIP.2006.873468
79. Kharat J, Chougule S (2020) A passive blind forgery detection technique to identify frame duplication attack. Multimed Tools Appl 79(11–12):8107–8123
80. Kingra S, Aggarwal N, Singh RD (2016) Video inter-frame forgery detection: a survey. Indian J Sci Technol 9(44):1–9
81. Kingra S, Aggarwal N, Singh RD (2017) Video Inter-frame Forgery Detection Approach for Surveillance and Mobile Recorded Videos Int J Electr Comput Eng 7(2):831
82. Kirchner M, Johnson C (2020) Spn-cnn: boosting sensor-based source camera attribution with deep learning. In: IEEE International Workshop on Information Forensics and Security (WIFS). IEEE, pp 1–6
83. Kohli A, Gupta A, Singhal D (2020) CNN based localisation of forged region in object-based forgery for HD videos. IET Image Process 14(5):947–958

84. Kono K, Yoshida T, Ohshiro S, Babaguchi N (2018) Passive video forgery detection considering Spatio-temporal consistency. In: International Conference on Soft Computing and Pattern Recognition. Springer, pp 381–391
85. Korshunov P, Marcel S (2018) Deepfakes: a new threat to face recognition? Assessment and detection. arXiv Prepr arXiv181208685
86. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Communications of the ACM 60(6):84–90
87. Kumar V, Gaur M (2022) Deep feature based forgery detection in video using parallel convolutional neural network: VFID-net. Multimed Tools Appl 81(29):42223–42240
88. Kumar V, Gaur M (2022) Multiple forgery detection in video using inter-frame correlation distance with dual-threshold. Multimed Tools Appl 81:43979–43998
89. LASIESTA dataset (2022) https://computervisiononline.com/dataset/1105138810. Accessed 30 Sep 2022
90. Li C-T (2010) Source camera identification using enhanced sensor pattern noise. IEEE Trans Inf Forensic Secur 5:280–287
91. Li ZH, Jia R, Zhang ZZ, Liang, XY, Wang JW (2017) Double HEVC compression detection with different bitrates based on co-occurrence matrix of PU types and DCT coefficients. In: ITM web of conferences. EDP Sciences, p 01020
92. Li Q, Wang R, Xu D (2018) Detection of double compression in HEVC videos based on TU size and quantised DCT coefficients. IET Inf Secur 13(1):1–6
93. Li Y, Chang MC, Lyu S (2019) Exposing AI created fake videos by detecting eye blinking. In: Proceedings of the 2018 IEEE Int Work on Inf Forensics Secur (WIFS). IEEE, pp 1–7. https://doi.org/10.1109/WIFS.2018.8630787
94. Li Y, Yang X, Sun P, et al (2020) Celeb-DF: a large-scale challenging dataset for DeepFake forensics. In: Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit (CVPR). IEEE, pp 3204–3213. https://doi.org/10.1109/CVPR42600.2020.00327
95. Liang X, Li Z, Yang Y, Zhang Z, Zhang Y (2018) Detection of double compression for HEVC videos with fake bitrate. IEEE Access 6:53243–53253
96. Lin CS, Tsay JJ (2014) A passive approach for effective detection and localization of region-level video forgery with spatio-temporal coherence analysis. Digit Investig 11:120–140. https://doi.org/10.1016/j.diin.2014.03.016
97. Lin X, Li JH, Wang SL, Liew AWC, Cheng F, Huang XS (2018) Recent advances in passive digital image security forensics: a brief review. Engineering 4(1):29–39
98. Liu H, Li S, Bian S (2014) Detecting frame deletion in H. 264 video. In: international conference on information security practice and experience. Springer, pp 262–270
99. Liu Y, Huang T, Liu Y (2018) A novel video forgery detection algorithm for blue screen compositing based on 3-stage foreground analysis and tracking. Multimed Tools Appl 77:7405–7427
100. Long C, Smith E, Basharat A, Hoogs A (2017) A c3d-based convolutional neural network for frame dropping detection in a single video shot. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE, pp 1898–1906
101. López RR, Luengo EA, Orozco ALS, Villalba LJG (2020) Digital video source identification based on Container's structure analysis. IEEE Access 8:36363–36375
102. Lucas BD, Kanade T (1981) Iterative image registration technique with an application to Streo vision. In: 7th international joint conference on articficial intelligence (IJCAI). HAL, pp 674–679
103. Ma R, Chen J, Su Z (2010) MI-SIFT: Mirror and inversion invariant generalization for SIFT descriptor. In: 2010 ACM International Conference on Image and Video Retrieval (CIVR 2010). ACM Digit Lib, pp 228–235
104. Mandelli S, Bestagini P, Tubaro S, et al (2018) Blind detection and localization of video temporal splicing exploiting sensor-based footprints. In: 2018 26th European signal processing conference (EUSIPCO). IEEE, pp 1362–1366
105. Matern F, Riess C, Stamminger M (2019) Exploiting visual artifacts to expose deepfakes and face manipulations. In: 2019 IEEE winter applications of computer vision workshops (WACVW). IEEE, pp 83–92
106. Mizher MA, Ang MC, Mazhar AA, Mizher MA (2017) A review of video falsifying techniques and video forgery detection techniques. Int J Electron Secur Digit Forensic 9:191–208
107. Mohiuddin S, Malakar S, Sarkar R (2021) Duplicate frame detection in forged videos using sequence matching. In: Proceedings of 3rd international conference on computational intelligence in communications and business analytics (CICBA-2021). Springer, pp 29–41
108. Mohiuddin S, Ganguly S, Malakar S, Kaplun D, Sarkar R (2022) A feature fusion based deep learning model for Deepfake video detection. In: International Conference on Mathematics and its Applications in new Computer Systems (MANCS-2021). Springer, pp 197–206

109. Nguyen HH, Yamagishi J, Echizen I (2019) Capsule-forensics: using capsule networks to detect forged images and videos. In: 2019 IEEE international conference on acoustics, Speech and Signal Processing (ICASSP-2019). IEEE, pp 2307–2311
110. NTHU Forensics project (2022) http://www.ee.nthu.edu.tw/cwlin/forensics/forensics.html. Accessed 30 Sep 2022
111. Panchal HD, Shah HB (2020) Video tampering dataset development in temporal domain for video forgery authentication. Multimed Tools Appl 79:24553–24577. https://doi.org/10.1007/s11042-020-09205-w
112. Pandey RC, Singh SK, Shukla KK (2014) Passive copy-move forgery detection in videos. In: 2014 international conference on computer and communication technology (ICCCT). IEEE, pp 301–306
113. Park JY, Lim JH, Kim GS, Won CS (2002) Invertible semi-fragile watermarking algorithm distinguishing MPEG-2 compression from malicious manipulation. In: 2002 digest of technical papers. International conference on consumer electronics (IEEE cat. No. 02CH37300). IEEE, pp 18–19
114. Paul A, Pramanik R, Malakar S, Sarkar R (2022) An ensemble of deep transfer learning models for handwritten music symbol recognition. Neural Comput & Applic 34:10409–10427
115. Pramanik R, Dey S, Malakar S, Mirjalili S, Sarkar R (2022) TOPSIS aided ensemble of CNN models for screening COVID-19 in chest X-ray images. Sci Rep 12(1):15409
116. Qadir G, Yahaya S, Ho AT (2012) Surrey university library for forensic analysis (SULFA) of video content. In: IET Conference on Image Processing (IPR 2012). http://sulfa.cs.surrey.ac.uk/index.php. Accessed 30 Sep 2022
117. Recognition of human actions (2022) https://www.csc.kth.se/cvap/actions/. Accessed 30 Sep 2022
118. REWIND Database (2022) https://sites.google.com/site/rewindpolimi/downloads/datasets. Accessed 30 Sep 2022
119. Rocha A, Scheirer W, Boult T, Goldenstein S (2011) Vision of the unseen: current trends and challenges in digital image and video forensics. ACM Comput Surv 43(4):1–42
120. Rossler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2019) Faceforensics++: learning to detect manipulated facial images. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. IEEE, pp 1–11
121. Roy S, Bhattacharya A, Sarkar N, Malakar S, Sarkar R (2020) Offline hand-drawn circuit component recognition using texture and shape-based features. Multimed Tools Appl 79:31353–31373
122. Roy S, Sarkar D, Malakar S, Sarkar R (2021) Offline signature verification system: a graph neural network based approach. J ambient Intell Humaniz Comput 1–11. https://doi.org/10.1007/s12652-021-03592-0
123. Saddique M, Asghar K, Bajwa UI, Hussain M, Habib Z (2019) Spatial video forgery detection and localization using texture analysis of consecutive frames. Advances in Electrical and Computer Engg 19(3):97–108
124. Saddique M, Asghar K, Bajwa UI, Hussain M, Aboalsamh HA, Habib Z (2020) Classification of authentic and tampered video using motion residual and parasitic layers. IEEE Access 8:56782–56797. https://doi.org/10.1109/ACCESS.2020.2980951
125. Sah AK, Bhowmik S, Malakar S, Sarkar R, Kavallieratou E, Vasilopoulos N (2018) Text and non-Text recognition using modified HOG descriptor. In: Proceedings of IEEE Calcutta Conference (CALCON 2017). IEEE, pp 64–68. https://doi.org/10.1109/CALCON.2017.8280697
126. Sarkar S, Ghosh M, Chatterjee A, Malakar S, Sarkar R (2019) An advanced particle swarm optimization based feature selection method for tri-script handwritten digit recognition. In: Proceedings of Second International Conference on Computational Intelligence, Communications, and Business Analytics. Springer Singapore, pp 27–28
127. Sawant R, Sabnis M (2018) A review of video forgery and its detection. J Comput Eng 20(2):1–4
128. Shanableh T (2013) Detection of frame deletion for digital video forensics. Digit Investig 10:350–360
129. Shaw SS, Ahmed S, Malakar S, Garcia-Hernandez L, Abraham A, Sarkar R (2021) Hybridization of ring theory-based evolutionary algorithm and particle swarm optimization to solve class imbalance problem. Complex Intell Syst 7:2069–2091
130. Shaw SS, Ahmed S, Malakar S, Sarkar R (2021) An ensemble approach for handling class imbalanced disease datasets. In: Proceedings of International Conference on Machine Intelligence and Data Science Applications. Springer, pp 345–355
131. Shelke NA, Kasana SS (2021) A comprehensive survey on passive techniques for digital video forgery detection. Multimed Tools Appl 80:6247–6310
132. Shelke NA, Kasana SS (2022) Multiple forgery detection and localization technique for digital video using PCT and NBAP. Multimed Tools Appl 81:22731–22759
133. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. In: 3rd international conference on learning representations (ICLR 2015). https://doi.org/10.48550/arXiv.1409.1556

134. Singh RD, Aggarwal N (2015) Detection of re-compression, transcoding and frame-deletion for digital video authentication. In: 2nd international conference on recent advances in Engineering Computational Sciences (RAECS 2015). IEEE, pp 1–6
135. Singh RD, Aggarwal N (2017) Detection of upscale-crop and splicing for digital video authentication. Digital Investigation 21:31–52
136. Singh RD, Aggarwal N (2018) Video content authentication techniques: a comprehensive survey. Multimedia Systems 24:211–240. https://doi.org/10.1007/s00530-017-0538-9
137. Singh G, Singh K (2019) Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation. Multimed Tools Appl 78:11527–11562. https://doi.org/10.1007/s11042-018-6585-1
138. Singh G, Singh K (2022) Chroma key foreground forgery detection under various attacks in digital video based on frame edge identification. Multimed Tools Appl 81:1419–1446
139. Su L, Li C (2018) A novel passive forgery detection algorithm for video region duplication. Multidim Syst Sign Process 29:1173–1190. https://doi.org/10.1007/s11045-017-0496-6
140. Su L, Li C, Lai Y, Yang J (2017) A fast forgery detection algorithm based on exponential-Fourier moments for video region duplication. IEEE Trans Multimed 20:825–840
141. Su L, Luo H, Wang S (2019) A novel forgery detection algorithm for video foreground removal. IEEE Access 7:109719–109728. https://doi.org/10.1109/access.2019.2933871
142. Subramanyam AV, Emmanuel S (2013) Pixel estimation based video forgery detection. In: 2013 IEEE international conference on acoustics, Speech and Signal Processing. IEEE, pp 3038–3042
143. Sun T, Wang W, Jiang X (2012) Exposing video forgeries by detecting MPEG double compression. In: IEEE international conference on acoustics, Speech and Signal Processing (ICASSP 2012). IEEE, pp 1389–1392
144. Suwajanakorn S, Seitz SM, Kemelmacher-Shlizerman I (2017) Synthesizing obama: learning lip sync from audio. ACM Transactions on Graphics (ToG) 36(4):1–13
145. TREC Video Retrieval Evaluation (2022) http://trecvid.nist.gov/. Accessed 30 Sep 2022
146. Verdoliva L (2020) Media forensics and deepfakes: an overview. IEEE J Sel Top Signal Process 14:910–932
147. Video Inpainting Under Camera Motion (2022) http://www.tc.umn.edu/~patw0007/video-inpainting/. Accessed 30 Sep 2022
148. Video Motion Interpolation for Special Effect (2022) http://member.mine.tku.edu.tw/www/TSMC09/. Accessed 30 Sep 2022
149. Video Trace Library (2022) http://trace.eas.asu.edu/. Accessed 30 Sep 2022
150. Wang W, Farid H (2006) Exposing digital forgeries in video by detecting double MPEG compression. In: Proceedings of the 8th workshop on multimedia and security. ACM Digital Library, pp 37–47
151. Wang W, Farid H (2009) Exposing digital forgeries in video by detecting double quantization. In: proceedings of the 11th ACM workshop on multimedia and security, pp 39–48
152. Wang W, Jiang X, Wang S, Wan M, Sun T (2013) Identifying video forgery process using optical flow. In: International Workshop on Digital Watermarking. Springer, pp 244–257
153. Wu Y, Jiang X, Sun T, Wang W (2014) Exposing video inter-frame forgery based on velocity field consistency. In: IEEE international conference on acoustics, speech and signal processing (ICASSP 2014). IEEE, pp 2674–2678
154. Xiaozhong P, Haoming W (2012) The detection method of image regional forgery based DWT and 2DIMPCA. In: Advanced Materials Research. Trans Tech Publications Ltd, pp 692–696
155. Xu J, Yu Y, Su Y, Dong B, You X (2012) Detection of blue screen special effects in videos. Phys Procedia 33:1316–1322
156. Xu J, Su Y, Liu Q (2013) Detection of double MPEG-2 compression based on distributions of DCT coefficients. Int J Pattern Recognit Artif Intell 27:1354001
157. Xu Q, Sun T, Jiang X, Dong Y (2017) HEVC double compression detection based on SN-PUPM feature. In: International Workshop on Digital Watermarking. Springer, pp 3–17
158. Yao Y, Shi Y, Weng S, Guan B (2018) Deep learning for detection of object-based forgery in advanced video. Symmetry (Basel) 10(1):1–10
159. Yin P, Yu HH (2001) Classification of video tampering methods and countermeasures using digital watermarking. In: Proceedings of International Symposium on the Convergence of IT and Communications. SPIE, pp 239–246
160. YouTube [HD] (2022) https://www.youtube.com/watch?v=66Ob1aJedHc&t=14s. Accessed 30 Sep 2022
161. Yu L, Yang Y, Li Z et al (2019) HEVC double compression detection under different bitrates based on TU partition type. EURASIP J Image Vid Process 2019(1):1–12
162. Yu Y, Yao H, Ni R, Zhao Y (2020) Detection of fake high definition for HEVC videos based on prediction mode feature. Signal Process 166:107269

163. Yuan Z, Chen H, Li T, Liu J, Wang S (2021) Fuzzy information entropy-based adaptive approach for hybrid feature outlier detection. Fuzzy Sets Syst 421:1–28
164. YUV Dataset (2022) http://www.trace.eas.asu.edu/yuv/index.html. Accessed 30 Sep 2022
165. Zhang K, Zhang L, Yang M-H (2014) Fast compressive tracking. IEEE Trans Pattern Anal Mach Intell 36(10):2002–2015
166. Zhang Z, Hou J, Ma Q, Li Z (2015) Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames. Secur Commun Netw 8:311–320
167. Zhao D-N, Wang R-K, Lu Z-M (2018) Inter-frame passive-blind forgery detection for video shot based on similarity analysis. Multimed Tools Appl 77:25389–25408
168. Zheng L, Sun T, Shi Y-Q (2014) Inter-frame video forgery detection based on block-wise brightness variance descriptor. In: International Workshop on Digital Watermarking. Springer, pp 18–30

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