mathematics

MDPI

Review

Digital Video Tampering Detection and Localization: Review, Representations, Challenges and Algorithm

Naheed Akhtar [1], Mubbashar Saddique [2], Khurshid Asghar [3], Usama Ijaz Bajwa [1], Muhammad Hussain [4] and Zulfiqar Habib [1,*]

[1] Department of Computer Science, COMSATS University Islamabad, Lahore 54000, Pakistan; sp19-pcs-009@cuilahore.edu.pk (N.A.); usamabajwa@cuilahore.edu.pk (U.I.B.)
[2] Department of Computer Science and Engineering, Narowal Campus, University of Engineering & Technology Lahore, Narowal 51600, Pakistan; dr.mubbashar@uet.edu.pk
[3] Department of Computer Science, University of Okara, Okara 56300, Pakistan; khasghar@uo.edu.pk
[4] Department of Computer Science, King Saud University, Riyadh 11564, Saudi Arabia; mhussain@ksu.edu.sa
* Correspondence: drzhabib@cuilahore.edu.pk

Abstract: Digital videos are now low-cost, easy to capture and easy to share on social media due to the common feature of video recording in smart phones and digital devices. However, with the advancement of video editing tools, videos can be tampered (forged) easily for propaganda or to gain illegal advantages—ultimately, the authenticity of videos shared on social media cannot be taken for granted. Over the years, significant research has been devoted to developing new techniques for detecting different types of video tampering. In this paper, we offer a detailed review of existing passive video tampering detection techniques in a systematic way. The answers to research questions prepared for this study are also elaborated. The state-of-the-art research work is analyzed extensively, highlighting the pros and cons and commonly used datasets. Limitations of existing video forensic algorithms are discussed, and we conclude with research challenges and future directions.

Keywords: video tampering detection; passive video forgery; spatial video forensic; temporal video forensic; video tampering; localization

Citation: Akhtar, N.; Saddique, M.; Asghar, K.; Bajwa, U.I.; Hussain, M.; Habib, Z. Digital Video Tampering Detection and Localization: Review, Representations, Challenges and Algorithm. *Mathematics* **2022**, *10*, 168. https://doi.org/10.3390/math10020168

Received: 22 November 2021
Accepted: 17 December 2021
Published: 6 January 2022

1. Introduction

The availability of sophisticated low-cost digital video cameras in mobile phones, gadgets, and a large number of video-sharing websites such as (YouTube, Facebook, and Dailymotion) play an important role in daily life to disseminate and share visual information. The visual data can also serve as powerful evidence before a court of law to verify or support the testimony of a person being questioned. In the presence of sophisticated and user-friendly video-editing software, the genuineness of videos cannot be taken for granted. With advanced editing tools, information manipulation has become easy. Videos can be edited by inserting or deleting objects/events, with good or bad intentions [1].

Videos are not accepted without their forensic reports as a matter of evidence by law enforcement agencies. Every instance of video tampering does not have equal significance, e.g., tampered footage of a pop star is not as harmful as the tampered footage of a crime scene [2]. The film industry benefits from video editing technologies to add virtual reality in scenes. Video evidence is also important for news reporting, intelligence agencies, insurance companies, copywriting, criminal investigations, etc. Forensic analysis of videos and images is the focus of recent research to ensure the authenticity of multimedia content [3]. Such research is never ending due to the progressive advancement in video editing tools.

Progress in video tampering has a significant effect on our society. Although only a few digital video forgeries have been exposed, such instances have eroded public trust in video clips [4].

The objective of video tampering detection is to ensure the authenticity and to expose the potential modifications and forgeries (i.e., to verify whether a video is authentic or

not), and to carry out localization (i.e., to find the forged (tampered) area within the frame, adjacent frame (spatial tampering), or to identify where the frames were inserted, replaced, reordered, or deleted (temporal tampering) in the video). Several techniques have been proposed to authenticate and localize tampering in images [5–7], but these techniques cannot be applied to videos directly due to the following reasons: (a) most videos are encoded and compressed prior to storage and transmission because of presence of a massive volume of data in video frames; (b) these reported techniques have high computational complexity when applied to a video sequence; and (c) temporal tampering such as frame insertion, deletion, duplication, or shuffling in a video cannot be detected by applying any image forgery detection technique. However, many techniques appear in literature specifically for the identification and localization of video tampering. In this study, we present a systematic literature review of the state-of-the-art video tampering detection techniques by highlighting their merits and demerits.

Organization of This Study

The remaining part of this study is organized in the following sections. Section 2 describes this study's distinction from other survey papers. Section 3 elaborates the survey protocol of this study. Section 4 explains the types of video tampering (forgery). Section 5 provides the detail of video forensic detection approaches. Sections 6 and 7 elaborate the state-of-the-art spatial and temporal tampering detection techniques, datasets used, comparison, and limitations. Section 8 concludes the analysis, and challenges are highlighted. In Section 9, future directions are presented, and finally, Section 10 concludes this review.

2. Distinction from Other Surveys

Considering the fact that video tampering detection has been maturely developed and enough research work has been published on passive video detection techniques, a comprehensive analysis on proposed schemes for passive video tampering detection and localization is required to determine future research directions. To our knowledge, this is the first study of systematic literature surveys in the domain of passive video tampering detection.

Several researchers have reviewed video tampering (forgery) detection techniques. Details are shown in Table 1 of the works published so far. A few papers [8,9] published in reputable journals have focused on passive video tampering detection techniques. Journal papers [4,10,11] partially discussed the video techniques and their focus was on image tampering detection techniques. Review papers [12–18] are published in conferences and less reputable journals. Rocha et al. [4] reviewed the video forgery detection and localization issues by discussing two video tampering techniques, but without highlighting the pros and cons of video tampering detection techniques. Moreover, the major emphasis of this review was on image forensics rather than video forensics. Similarly, Milani et al. [11] discussed video acquisition and compression issues only. This review is also a partial and mixed representation of image and video forensic analysis. Pandey et al. [10] presented review on passive techniques of image and video tampering detection but only focused on techniques that are based on noise features. This survey highlighted that the video tampering detection domain is facing issues such as video acquisition, post-processing operations (compression, blurring, noise addition, geometric transformation) and robustness. Sharma et al. [17] reviewed passive techniques, but their discussion was limited to only copy-move attacks on digital videos.

Sitara et al. [9] also analyzed passive tampering detection methods and their limitations, but there is no discussion comparing the accuracy of these methods, which is an important part of our survey paper. Singh et al. in [8] reported that there are few video forgery detection methods that have been evaluated extensively because of the lack of availability of large-scale video forgery datasets and base lines for comparison of different video forgery detection techniques. There is a dire need for a comprehensive collection of videos for advanced evaluation of video forensic techniques; however, they highlighted

the non-availability of a large-scale video forensic dataset only. Tao et al. in [18] and Mizher et al. in [19] reviewed video tampering detection in comprehensive ways, but these papers were published in 2017 and thus, several current state-of-the-art techniques are not considered.

Sharma et al. [20] reviewed existing video forgery detection techniques by their functionality. The review has strength in terms of exploring video forgery detection techniques by their functionality and datasets. Johnston et al. in [21] critically reviewed spatial video forgery detection techniques based on deep learning. Existing video tampering datasets used to evaluate video forgery detection techniques were also reviewed. The researchers highlighted the challenges and trends in video forgery detection in the spatial domain; however, the research gaps in the temporal domain of video forgery detection still need to be explored. In a recent survey, Kaur and Jindal [22] explored the current challenges and trends in the domain of image and video forensics. The review was focused on highlighting the image copy-move and splicing forgeries, and inter- and intra-frame video forgery challenges. Issues regarding benchmarking and datasets were also highlighted. This review presented both image and video forgery issues, but the major focus was on highlighting the issues in the image forensic domain, and few aspects related to video forgery forensic are elaborated. Recently, Shelke and Kasana [23] presented a comprehensive survey on passive techniques for video forgery detection based on features, types of forgeries identified, datasets and performance parameters. Pros and cons of different passive forgery detection techniques are elaborated, along with future challenges. Anti-forensics techniques, deep fake detection in videos and a brief review of existing datasets of video forgery are also included in this survey paper.

Table 1. Summary of survey papers on video forgery detection.

Major Focus	References	Conference /Journal	Publisher	Publication Year	Citations
Image forgery detection with partial focus on video forgery detection	[4]	Journal	ACM	2011	305
	[11]	Journal	Others	2012	277
	[10]	Journal	Elsevier	2016	37
Video forgery detection	[13]	Conference	-	2012	280
	[12]	Journal	Others	2013	7
	[24]	Journal	Others	2013	3
	[14]	Conference	IEEE	2014	38
	[15]	Journal	Others	2015	15
	[16]	Journal	Others	2015	32
	[9]	Journal	Elsevier	2016	85
	[17]	Conference	-	2016	9
	[18]	Conference	-	2017	3
	[19]	Journal	Others	2017	11
	[8]	Journal	Springer	2018	54
	[21]	Journal	Elsevier	2019	16
	[20]	Conference	-	2019	10
	[22]	Journal	Others	2020	6
	[25]	Journal	Springer	2020	2
	[23]	Journal	Springer	2021	2

The above discussion is summarized in Table 1. The surveys published so far in top-ranked journals have focused more on image forgery and partially on video forgery. Some

surveys have only focused on noise-based techniques, some only on spatial tampering detection techniques and some have not discussed the major challenges faced in video forgery detection. Therefore, this study covers all these gaps comprehensively.

Contribution of This Work

A comprehensive survey of passive video tampering detection techniques is presented with the following salient features.

- Almost all published papers in the domain of video forgery/tampering to date are considered to show the overall picture of research contribution in the field.
- To our knowledge, this is the first systematic comprehensive survey to filter out rich research contributions in the domain.
- This survey is categorized based on the proposed methodologies for easy comparison of their performance evaluation and the selection of the most suitable technique.
- This review will be helpful to new researchers regarding the issues and challenges faced by the community in this domain. Moreover, this paper analyzes the research gaps found in the literature that will help future researchers to identify and explore new avenues in the domain of video forensics.

3. Survey Protocol

The objective of this systematic study is to perceive and arrange the strategies, models, methods, and tools that are used to investigate existing video tampering techniques. The procedure of systematic study helps us to analyze the available research in the subject domain. In this study, the guidelines of systematic literature survey [26] are followed and the survey protocol plan of this study is shown in Figure 1. The following subsections elaborate the research questions, search string and inclusion/exclusion criteria, extract data and present their analysis.

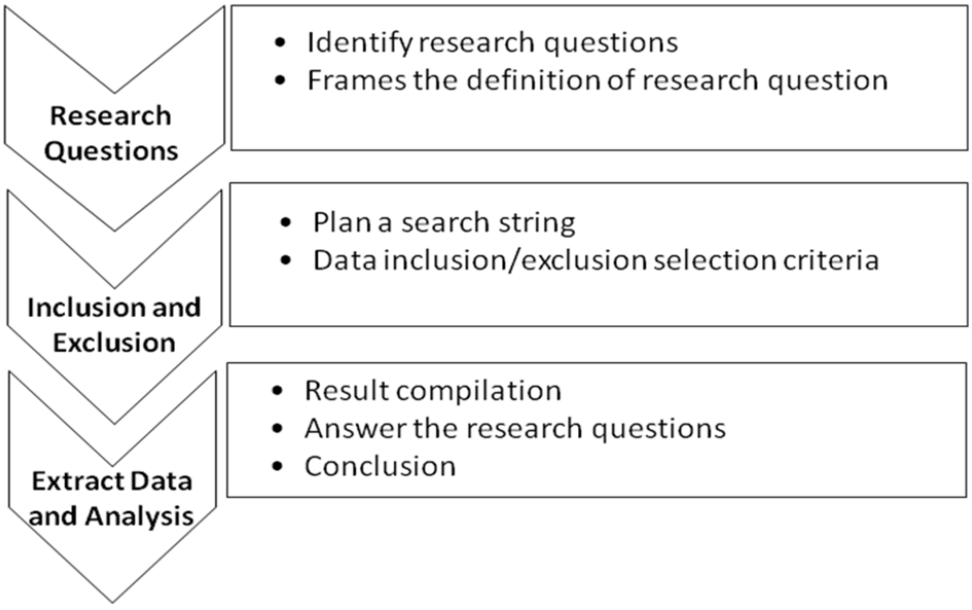

Figure 1. Survey protocol plan.

3.1. Research Questions

The first step of the systematic survey is to define the research questions. Various research questions were formulated to conduct this survey:

- Q1. What are various types of video tampering?
- Q2. What are the various techniques for video tampering detection available in the literature?

- Q3. What are the pros and cons of existing techniques?
- Q4. What are the challenges faced by the researchers?
- Q5. What are the evaluation measures and datasets used to evaluate video tampering detection and localization?

The answers of the first, second, third and fifth question are explained in Sections 4–7. The answer to question 4 is elaborated in Section 8.

3.2. Search Strategy

An efficient search strategy is required to extract the appropriate information and filter out inappropriate studies from the research area. For this purpose, a dynamic search string was prepared, based on research questions, keywords, and alternate words for major keywords. The search string is a combination of "OR" and "AND" Boolean operator, given below.

{(Video forgery) AND (detection OR localization)} OR {(video tampering) AND (detection OR localization} OR {(localization of OR detection of) AND {(video forgery) OR (video tampering)}} OR {review on video forgery} OR {review on video tampering}

The search string was applied to different digital resources, i.e., ACM digital library, Science Director, Springer, Elsevier, IEEE explorer, Google Scholar, and others.

3.3. Research Inclusion/Exclusion Criteria

Firstly, search criteria were set to extract the maximum publications from the selected sources. The publishing years are limited between the years 2007 to 2021. In order to gather more relevant papers, the selection criteria are divided into three steps. In the first step, to remove the duplicate and irrelevant papers, the title of the paper is checked. In the second stage, we read the abstract of the papers obtained in the first stage to select relevant papers to the focused area. At the last stage, we read out the detail of each paper and finalized the papers for this study. A total of 122 papers were selected as the most relevant papers in the domain of passive video forgery. Similarly, a total of 99 research papers were selected for the primary analysis. These papers were selected as they are published in reputable journals or conferences which have more citations. The year-wise details of these papers published in conferences, journals, and others (books and thesis) are shown in Figure 2, which depicts an overall pictorial representation of published papers, books, and theses in the past 15 years on video tampering detection using blind or passive techniques. It highlights that in recent years, passive techniques for video forgery detection are drawing significant attention in the research community. There is a demand in many areas such as judicial forensics, insurance industry, information security, etc., to develop robust, standard, and economically feasible techniques for the detection of a wide variety of tampering in digital videos to overcome these challenges related to passive video forgery detection. Much progress has been achieved over the past few years, but certain important milestones still remain unmet. That is because of the wide range of possible alterations that can be applied to digital content that makes it practically indistinguishable from genuine content. The absence of a universally applicable solution to this problem has gained the attention of the scientific community and researchers. Table 2 represents the published papers on passive (blind) forgery detection techniques that are categorized according to standard journals such as IEEE, Springer, Elsevier and others, or well-known conferences. Concrete and category-wise discussions on the papers presented in Table 2 are given in Sections 6 and 7.

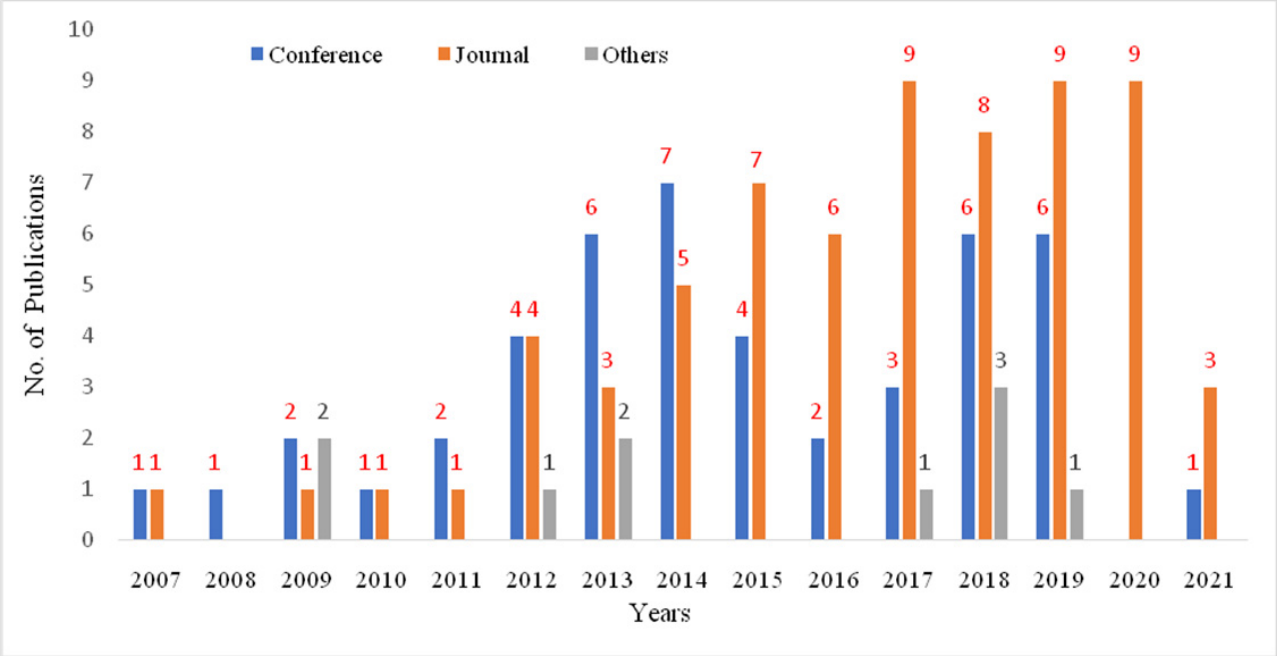

Figure 2. Motion detail of year-wise publications in conferences, journals and others.

Table 2. Summary of research papers on passive video forgery detection published in different journals and conference papers.

Years	IEEE	Springer	Elsevier	Other Journals	Conferences	Total
2007	[27]	-	-	-	[28]	2
2008	-	-	-	-	[29]	1
2009	[30]	-	-	-	[31–33]	4
2010	[34]	-	-	-	[35]	2
2011	-	-	-	-	[36,37]	2
2012	[38,39]	-	[40]	-	[41–43]	6
2013	-	-	[44]	-	[45–51]	8
2014	-	-	[52,53]	[54]	[55–60]	9
2015	[61]	[62–64]	[1]	[65,66]	[67–69]	10
2016	[70]	[71]	[72,73]	-	[74]	5
2017	-	[75,76]	[77,78]	[79–81]	[82,83]	9
2018	[84–86]	[87,88]	[89]	-	[90–96]	13
2019	[97]	[98–100]	[21,101,102]	[103–105]	[20,106–109]	15
2020	[110]	[111–113]	-	[22,114–117]	-	9
2021	-	-	[118]	[119,120]	[121]	4
Total	12	14	14	16	43	99

4. Types of Video Tampering (Forgery)

Videos are usually tampered in the following ways: (a) tampering in the spatial domain, (b) tampering in the temporal domain, (c) spatio-temporal tampering and (d) re-projection [2,122]. Details of spatial, temporal and spatio-temporal tampering are

highlighted in Figure 3. In this figure, F_i represents the ith frame, where $I = 1, 2, \ldots n$, P_{HW} is the pixel intensity, and H and W are frame height and width, respectively. F'_I is the manipulated ith frame and P'_{HW} is the manipulated pixel intensity. A forger can tamper source videos spatially (i.e., spatial forgery) by manipulating a block of pixels within a video frame or in adjacent video frames, as shown in Figure 3b. Furthermore, as presented in Figure 3c, source videos can be tampered with respect to time (i.e., temporal forgery) by disturbing the frame sequence through replacement, reordering, addition, and removal of video frames. Lastly, Figure 3d shows video tampering by combining both spatial and temporal domains (i.e., spatio-temporal forgery). Re-projection means recording a movie from the theatre screen by which the forger violates the copyright law.

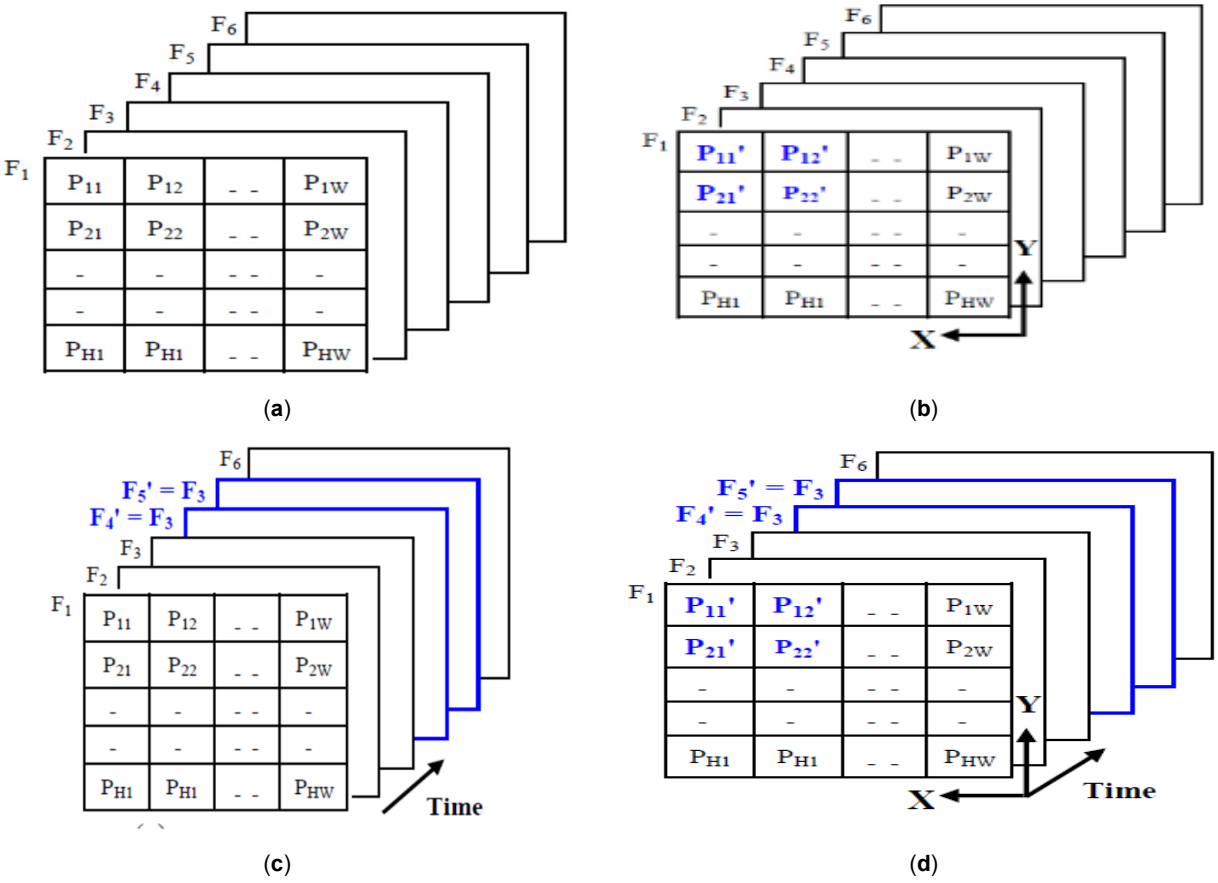

Figure 3. Engineering of video tampering (forgery). (**a**) Actual video, (**b**) spatially tampered video, (**c**) temporally tampered video, (**d**) spatio-temporal tampered video.

5. Video Tampering Detection

Video tampering detection approaches can be broadly classified into active and passive (blind) [4,11,13–16], as shown in Figure 4 and described in the following subsections.

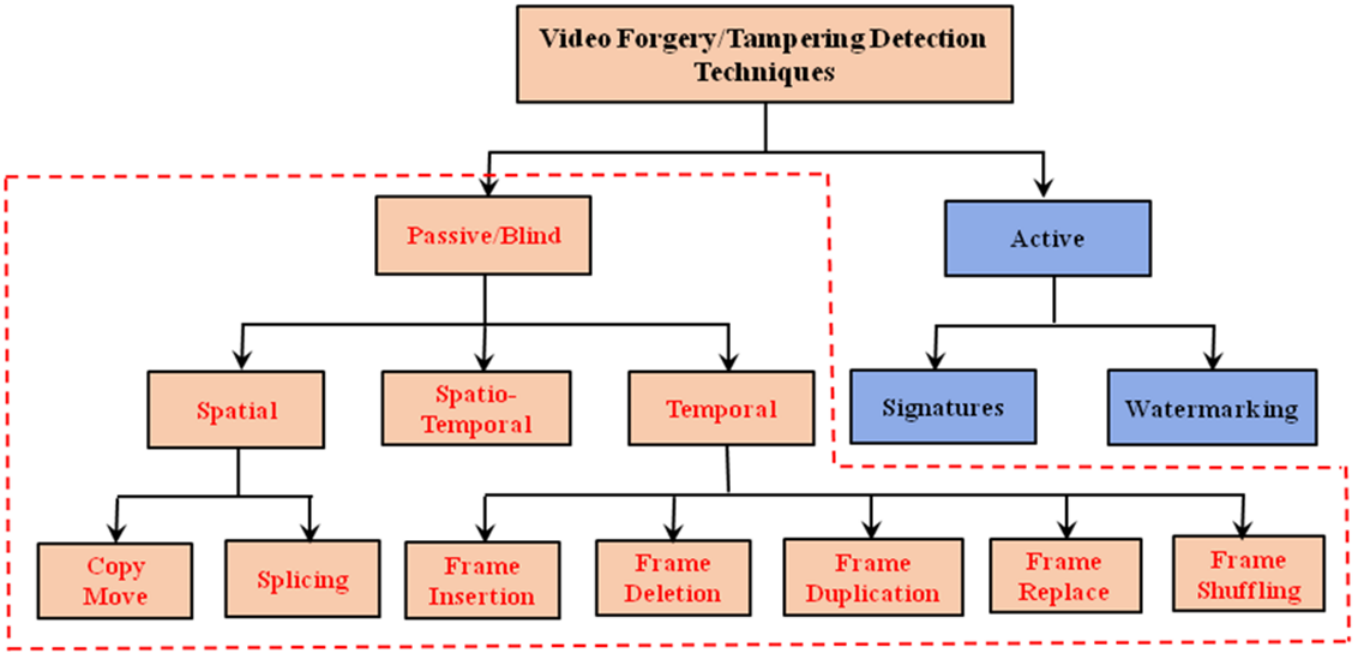

Figure 4. Categories of video tampering detection techniques.

5.1. Active Approaches

The active approaches can be further divided into two categories based on approaches to watermarks and digital signatures [123]. There are several kinds of watermarks. Fragile and semi-fragile watermarks are used to detect video forgery [124,125]. Fragile watermarking works by inserting invisible information into the video. If an attempt is made to modify the contents of the video, that invisible information (watermark) is also altered, and hence, forgery is detected. Semi-fragile watermarking is less sensitive to change as compared to fragile watermarking. For both the fragile and semi-fragile techniques, a watermark must be inserted when the video has been recorded, which makes active techniques dependent on both algorithmic and hardware implementation [2]. All capturing devices may not have the capability to embed digital signatures or water marks. If this information is embedded intentionally in videos after the acquisition phase, this method may fail in situations where tampering is carried out before inserting the signature or watermark. Since most of the videos reported in datasets for experiments, evaluation of video forgery detection and localization have no prior information about their watermark or signature, our survey is focused on passive techniques instead of active techniques, which are highlighted in the red dotted box in Figure 4.

5.2. Passive Approaches

Passive video tampering detection techniques do not require any prior information that is embedded in videos, such as digital watermarks or signatures. These techniques work by exploiting traces left in the frames of the video due to tampering and cannot be seen with the naked eye. However, the statistical properties are changed during the tampering process. Due to the change in statistics, the inconsistencies of different features such as noise, residues, texture, abnormalities in optical flow (OF), etc., can be used in passive approaches. Furthermore, whenever forensic analysis is required of any video, the source video is not available and forensic experts must make decisions based on current (under observation) video. In this case, active techniques are not workable and passive techniques are the best choice. Passive approaches are further divided into spatial and temporal tampering detection techniques, which are discussed in Sections 6 and 7, respectively.

6. Review of Spatial (Intra-Frame) Video Tampering Detection Techniques

Different types of information (artifacts or footprints) are available to forensic experts for the detection of spatial tampering and localization. According to this information, the methods are categorized into the following categories, shown in Figure 5: (i) methods based on deep learning, (ii) methods based on camera source features, (iii) methods based on pixels and texture features, (iv) methods based on SVD (Singular Value Decomposition), (v) methods based on compression features and (vi) methods based on statistical features. These categories are discussed in the following subsections.

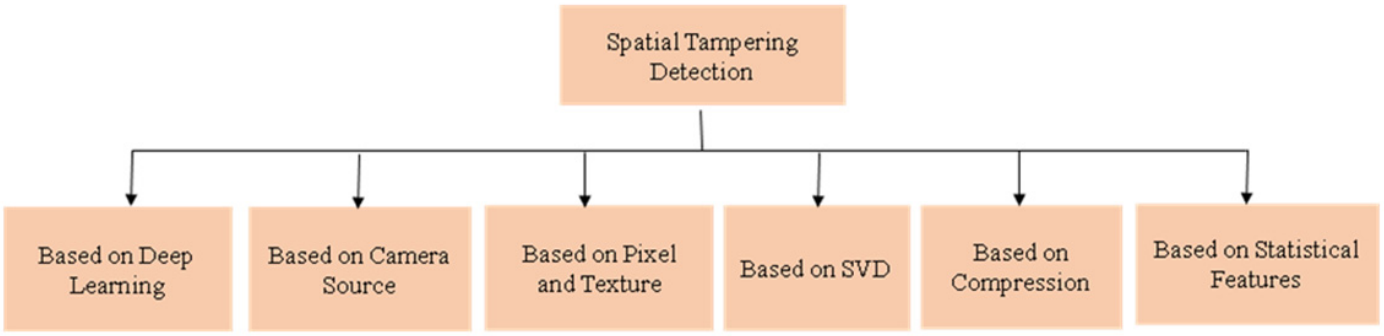

Figure 5. Categories of spatial tampering detection methods.

6.1. Methods Based on Deep Learning

Deep learning is a sub-domain of machine learning based on neural networks. Problem-specific, complex, high-dimensional features can be extracted with the help of deep learning techniques, which are helpful for classification tasks. Zampoglou et al. [106] applied Q4 and Cobalt forensic filters with pre-trained ResNet and GoogLeNet networks for the detection of spatial video forgery. Two datasets, Dev1 and Dev2, are used to evaluate the method.

The Dev1 dataset contains 30 authentic and 30 tampered videos while Dev2 contains 86 pairs of videos having 44 k and 134 k frames. The accuracy achieved on the union of Dev1 and Dev2 is 85.09%, and mean average precision is 93.69%. Yao et al. [79] used a CNN (Convolutional Neural Network) to extract complex high-dimensional features and used the absolute difference between consecutive frames to reduce the temporal redundancy, a max pooling layer is introduced to minimize the computational complexity, and a high-pass filter layer is placed to boost the residual left during the tampering process. One hundred authentic and one hundred forged videos are used to train and test the method. This method has achieved forged frame accuracy (FFACC), pristine frame accuracy (PFACC), frame accuracy, precision, recall and F1 scores of 89.90%, 98.45%, 96.79%, 97.31%, 91.05% and 94.07%, respectively. Kono et al. [94] combined a CNN and recurrent neural network to detect video forgery. The authors also developed their own dataset of 89 forged videos, named Inpainting-CDnet2014, and a dataset of 34 forged videos, named Modification Database. The method obtained an area under curve (AUC) of 0.977 and an equal error rate (EER) of 0.061 was achieved. Avino et al. [81] performed detection using auto-encoders and a recurrent neural network. The authors used only 10 videos for experiments. The receiver operating curve (ROC) was obtained to investigate the performance of the method. Kaur et al. [116] developed an inter-frame forgery detection method based on a Deep Convolutional Neural Network (DCNN). The method classifies the forged and authentic video frames on the basis of correlation. The system was evaluated on REWIND and GRIP video datasets and achieved 98% accuracy. The method has significant accuracy; however, there is a need for cross validation to ensure the generalization. Aditi et al. [114] developed a spatiotemporal video forgery detection and localization technique based on CNN. Video frames are detected as tampered or authentic using temporal CNN; latterly, the forgery in video frames is located using spatial CNN. Motion residual is used to train the model. The method was evaluated on SYSU-OBJFORG dataset and achieved comparable

results. Although the method has significant performance, there is still a need for cross data validation.

The algorithms of this class give high-dimensional features and achieve suitable accuracy; however, the small size of tampering cannot be detected by employing the algorithms developed so far.

6.2. Methods Based on Camera Source

During court proceedings, when a video is presented as proof, there is a need to identify the camera that recorded the presented video. In such situations, source camera-based features are used by forensic experts. If a source camera exists, then active forgery detection techniques will be used. Otherwise, camera-based features such as fixed pattern noise (FPN), photo response non-uniformity noise (PNRU) and sensor pattern noise (SPN) are calculated from the presented video and used for forgery detection. Different authors have used camera noise characteristics to detect the spatial (object-based) forgeries [29,33,34]. Hsu et al. [29] detected forgery by calculating the noise residual from high-frequency bands, wavelet coefficients and Bayesian classifiers. The authors used three videos; each one had 200 frames with a still background and each was captured using a JVC GZ-MG50TW digital camcorder. The frame rate is 30 fps. Video resolution of each frame and bitrate is 720 $\times$ 480 pixels and 8.5 Mbps, respectively. The recall, precision, miss rate and false positive rates are 96%, 55%, 32% and 4%, respectively. This study did not localize the forged region, the dataset is limited, and videos are prepared under a controlled environment. Kobayashin et al. [33] detected forged regions by determining the inconsistencies between noise characteristics of different video frames. In this work, a Point Grey Flea digital camera was used with 128 grayscale frames. The frame rate and resolution are 30 fps and 640 $\times$ 480 pixels, respectively. During recording, the camera and object are stationary. The recall and precision are 94% and 75%, respectively. Furthermore, the authors used a limited video dataset to test the proposed technique and did not detect the temporal forgery. The proposed algorithm worked only for grayscale videos. For detection of region tampering in videos, a technique based on utilization of extrinsic camera parameters was developed by Hu, Ni et al. in [126]. At first step, each frame of the video is divided into different regions, followed by the computation of extrinsic parameters from these regions of frames. Then differences between these parameters are calculated. Lastly, a threshold is selected to identify the tampering. Fayyaz et al. [113] developed a video tampering detection method based on sensor noise patterns of video frames. The noise patterns were extracted using denoising video frames; latterly, noise patterns were averaged to detect sensor noise patterns. Locally adaptive DCT (Discrete Cosine Transform) was used to determine the sensor noise patterns. Finally, the correlation of noise residues of different video frames was computed to detect authentic or forged video. The method was evaluated using noise pattern-based dataset and achieved suitable results, but these results depend upon the physical properties of the source device.

The algorithms of this category although performed well but are dependent on the hardware.

6.3. Methods Based on Pixels and Texture Features

A basic element of the frame (image) is called a pixel. The color model of the frame (image) is defined based on the number of bits per pixel. Various color models are used in digital media, such as RGB (Red-Green-Blue), YC_bC_r (Y is the luminance, blue and red chroma components are C_b and C_r, respectively), HSI (Hue-Saturation-Intensity), CMY (Cyan-Magenta-Yellow), etc. Different types of information (such as color, gamma, intensity, hue, contrast, etc.) can be calculated mathematically from these color models. Several types of features (such as HOG (Histogram of Oriented Gradients), LBP (Local Binary Pattern), etc.) that are based on pixels can be calculated to detect the passive forgery [52]. Subramanyam et al. [41] exploited compression features and Histogram of Oriented Gradients (HOG) to detected spatial forgery. In this approach, the authors used

6000 frames from 15 different videos for spatial forgery and 150 GOPs (Groups of Pictures) of size 12 frames each for temporal forgery. The original video is compressed at 9 Mbps using MPEG-2 video codec. Spatial tampering is carried out by copying and pasting regions of size 40 × 40 pixels, 60 × 60 pixels and 80 × 80 pixels in the same and different frames. Detection accuracy (DA) is 80%, 94% and 89% for 40 × 40 pixels, 60 × 60 pixels and 80 × 80 pixels blocks, respectively. This technique detected spatial forgery with better accuracy, but training and testing are performed on a small dataset. There are certain limitations of this algorithm, i.e., it failed to detect forgery when post-processing operations such as scaling and rotation were applied to forged regions. Moreover, this technique was unable to localize the forged regions. Al-Sanjary et al. [107] exploited inconsistency in optical flow to detect and localize the copy-move forged region. This study used nine videos to test the method and achieved 96% accuracy. The performance of the method is not sufficient in high-resolution videos.

The algorithms of this class are simple, and length of feature vectors is small. However, these algorithms do not perform well when various post-processing operations are applied to hide the forgery.

6.4. Methods Based on SVD

SVD is a factorization technique that extracts geometric features. This algorithm is widely used to detect the copy-move tampering due to its invariant nature of scaling and rotation. Su et al. [64] extracted features from a difference between frames using the K-SVD (K-Singular Value Decomposition) algorithm. Features are then randomly projected to reduce their dimension. K-means clustering is applied to the reduced features to detect spatial forgery. In total, 700 videos were prepared using SONY DSC-P10 at 25 fps and acquired at 3 Mbps for experimentation. Videos were forged using the Mokey 4.1.4 tool. The accuracy, precision and recall rates for this approach are 89.6%, 89.9% and 90.6%, respectively. This approach did not localize the forged regions. The algorithms of this category, although they have simple and small feature vectors, they cannot work for all types of post-processing operations.

6.5. Methods Based on Compression

Storage space requirements can be optimized with the compression of videos. During the compression process, different types of artifacts are acquired, such as quantization, properties of a group of pictures (GOP), motion vector, etc. These artifacts can also be used for the detection of spatial forgery. Labartino et al. [46] explored video frames using Double Quantization (DQ) to detect the spatial forgery. This method worked on assumption that the video is forged (by changing the contents of a group of frames) before the second compression takes place. In [69], Tan et al. developed an approach for automatic identification of object-based forgery in videos encoded with advanced video encoding standards based on its GOP structure. Video clips of two categories are used; one category is pristine frames and the second is double compressed frames, which have undergone re-compression after manipulation. CC-PEV feature extractor extracts feature that are used by an ensemble classifier to classify the frame as pristine or forged based on double compression. The final decision was made on the basis if all I- and P/B-frames of at least one GOP are forged, in which case, that video clip is considered as forged. The evaluation was performed on the SYSU-OBJFORG dataset, but this dataset is not publicly accessible to the research community. The proposed approach achieved 80% accuracy.

Bakas et al. in [95] presented a forensic solution to detect and localize double compression-based forgery in MPEG videos by exploiting its I-frames. They introduced CNN architecture that exploits the fact that double compression introduces specific artifacts in the DCT coefficients of the I-frames of an MPEG video. The model was tested on 20 YUV sequences in CIF of size 352 × 288 pixels taken from the video TRACE library, available online at http://trace.eas.asu.edu/yuv (accessed on 20 November 2021). They achieved detection

and localization accuracy of 90% and 70%, respectively. This method has high computational complexity.

This class of algorithms depends upon the inherent attributes of cameras instead of estimating the actual inconsistencies and discontinuities in tampered videos that occurred during the forgery.

6.6. Methods Based on Statistical Features

Tone, context, and texture are the major parts of any frame (image). During the spatial video tampering process, the texture of the video is changed, which is always present in the frame (image). The statistical features can be utilized for the illustration of this texture [127,128]. Object-based video forgery [27,49,53] has been detected with the statistical features by many researchers. Richao et al. [53] employed statistical features for the detection of spatial tampering. First, four moments of the wavelet and average gradient of each color channel are calculated. These features are feed-forwarded to SVM for training of the model to classify the forged and original videos. A set of twenty videos having a resolution of 200 $\times$ 240 pixels was utilized for conducting the experiment. The accuracy and AUC are attained as 95% and 0.948, respectively. An outcome of 85.45% is represented by the ROC curve. These results are obtained on limited dataset and no experiment was performed on videos having different compression ratios. Su et al. [86] detected duplicated regions by using exponential Fourier moments (EFMs) and tampered regions were localized by utilizing the adaptive parameter-based compression tracking algorithm. This method achieved detection accuracy of 93.1%.

The algorithms of this class are based on statistical features. The feature vectors of these methods are small in length as compared to other categories of algorithms but are unable to detect forgery in presence of different types of post-processing operations. A summary of different spatial forgery techniques is shown in Table 3.

6.7. Discussion and Analysis of Spatial Video Tampering Detection Techniques

It is not easy task to work with videos as images due to their unique set of complexities. A main limitation of many state-of-the-art approaches is the lack of cross dataset validation or validation on realistically forged videos. Every technique presented in the literature is designed to deal with one type of forgery. As of now, there is no universal tool for video tampering detection. Hence, to provide a real, practically applicable solution to forgery detection and localization challenges, a comprehensive, economically feasible and versatile forensic system is needed, which is a combination of different kinds of video forgery detection techniques, where each specialized technique is responsible for detecting the types of forgery it has been developed to tackle. In comparison to the image forensic domain, the video forensic domain is seriously under-underdeveloped, and research in this field is required.

Table 3. Summary of spatial tampering (forgery) detection techniques.

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Deep Learning							
2019 [106]	Q4 + Cobalt forensic filters + GoogLeNet+ ResNet networks	30 authentic and 30 forged videos for dataset Dev1 86 pairs of videos containing 44 k and 134 k frames for dataset Dev2	85.09%	-	93.69%	-	Cannot detect a small size of forgery
2017 [79]	CNN + absolute difference of consecutive frames + high pass filter layer	100 authentic and 100 forged videos	98.45%	91.05%	97.31%	F1 Score 94.07% FFACC 89.90%	Cannot detect a small size of forgery
2018 [94]	CNN + recurrent neural network	89 forged videos, named Inpainting-CDnet2014 34 forged videos, named Modification Database	-	-	-	AUC 0.977 and EER 0.061	Cannot work well in presence of different types of object modifications
2017 [81]	Auto-encoder + recurrent neural network	10 authentic and 10 forged videos	-	-	-	ROC	Cannot work in presence of different types of post-processing operations (scaling, rotation, translation)
Methods Based on Source Camera Features							
2008 [29]	Noise residual + Bayesian classifier	Three videos with 200 frames Camera: JVC GZ-MG50TW Frame rate is 30 fps, Resolution 720 × 480 pixels, Bit Rate 8.5 Mbps	-	96%	55%	FPR 4%. Miss Rate 32%	Has no robustness to quantization noise Technique is hardware-dependent Forged regions are not localized Performance is measured relatively on a tiny dataset

Table 3. *Cont.*

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Source Camera Features							
2009 [33]	Noise characteristics	128 grayscale frames 30 fps Resolution 640 × 480 pixels Compressed by Huffyuv, lossless compression Codec	-	94%	75%	-	The algorithm is hardware-dependent Limited to spatial forgery only Dataset was relatively small
Methods Based Pixels and Texture Features							
2012 [41]	HOG features + matching module	6000 frames from 15 different videos for spatial forgery 150 GOPs of size 12 frames each for temporal forgery Original video is compressed at 9 Mbps using MPEG-2 video codec Forgery is performed by copying and pasting regions of size 40 × 40, 60 × 60 and 80 × 80 pixels in the same and different frames	94% for 60 × 60 pixels	-	-	-	Forgery is dependent on block size Has no robustness to a geometric operation such as large scaling. The algorithm is unable to localize the forged regions Only 12 videos are used for experimentation
2013 [45]	Motion residual + correlation	120 videos Resolution 320 × 240 pixels with 300 frames	90%	-	-	AUC 0.92	Experiments are performed only on 10, 15 and 20 percent compression rates Relatively poor accuracy with a compression rate exceeding up to 30% and more
2018 [86]	Exponential Fourier Moments fast compression tracking algorithm	Video download from internet and SULFA dataset	93.1%	-	-	-	Does not work on different compression rates
2019 [107]	Optical flow	3 videos from SULFA and 6 videos of video tampering dataset (VTD)	96%				The performance of the method is not suitable in high-resolution videos

Table 3. *Cont.*

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on SVD							
2015 [64]	K-SVD + K-Means	Camera: SONY DSC-P10 Seven handmade videos Frame rate 25 fps Bitrate is 3 Mbps, Forged video are generated by Mokey 4.1.4 developed by the Imagineer Systems	89.6%	90.5%	89.9%	-	Has not experimented on different compression rates Forged objects are not localized Dataset is small
Methods Based on Compression							
2013 [46]	Double Quantization (DQ)	Download video from http://media.xiph.org/video/derf/ (accessed on 22 November 2021)	-	-	-	AUC 0.8. ROC	Works on assumption that the video is forged (by changing the contents of a group of frames) before the second compression take place
2018 [95]	CNN +DCT	Video TRACE library available online at: http://trace.eas.asu.edu/yuv (accessed on 16 November 2021)				Detection 90%, Localization 70%	High computational complexity
Methods Based on Statistical Features							
2013 [49]	Correlation coefficients + saliency-guided region segmentation	One video with 75 frames and Resolutions 360 × 240 pixels	High accuracy is claimed without statistical measure				Only one video is used for experimentation
2014 [53]	Moment + average gradient + SVM	20 videos Resolution 320 × 240 pixels	95%	-	-	AUC 0.948 ROC	Dataset is small

7. Review of Temporal (Inter-Frame) Video Tampering Detection Techniques

Forgers tamper a video temporally by inserting, duplicating, deleting or swapping frames. State-of-the-art temporal tampering (forgery) detection algorithms have been proposed [12,28,31,34,40,43,44,51,53–56,65,66,129]. These methods are analyzed in this review. The algorithms used to detect the temporal forgery can be divided into the following categories, as shown in Figure 6: (i) methods based on statistical features, (ii) methods based on a frequency domain, (iii) methods based on residual and optical flow, (iv) methods based on pixels and texture features, (v) methods based on deep learning and (vi) others. The detail of each category is explained in the following subsections.

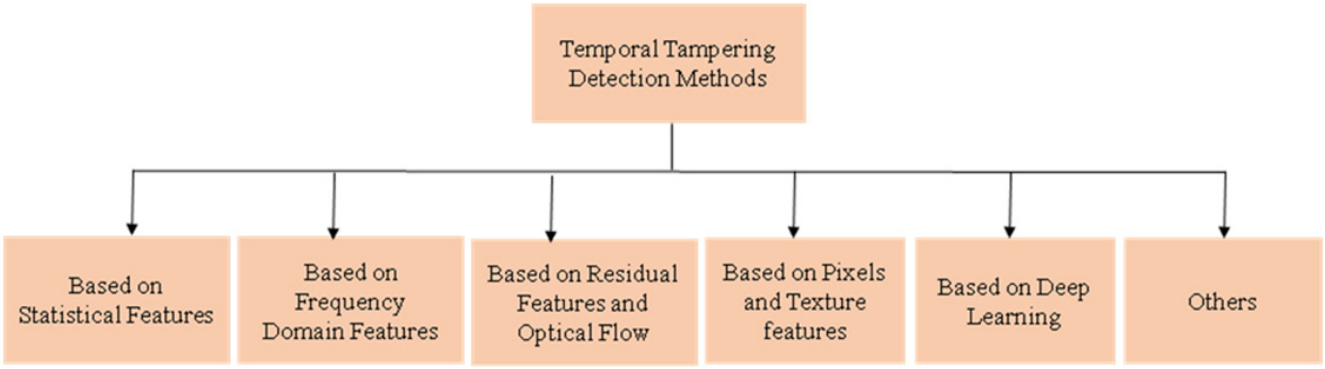

Figure 6. Categories of temporal tampering detection methods.

7.1. Methods Based on Statistical Features

When a forger tampers a video, its statistical properties are disturbed, and by investigating these properties, the tampered video is detected. Wang et al. [28] used a correlation between frames of a video to detect duplicated frames by using accuracy and false positive rates as evaluation measures. The algorithm was evaluated using only two videos recorded by SONY-HDR-HC3 having 10,000 frames each. One video sequence is recorded by placing the camera on a tripod and keeping it stationary throughout video recording, and a second video is recorded with a hand-held moving camera. Average detection accuracy of 85.7% and 95.2% is achieved for stationary and moving cameras, respectively, while the false positive rates were 0.06 and zero for stationary and moving cameras, respectively. The algorithm was evaluated on a very small dataset and is unable to detect forged videos when forged by means of frame insertion and deletion process. Wang et al. [54] identified forgery by calculating Consistency of Correlation Coefficients of Gray Values (CCCoGV) between frames and used SVM for classification. This technique did not localize the forged region and the video dataset is also limited. The technique did not produce results for different compression rates. The accuracy for 25 frames insertion and deletion is 96.21%; for 100 frames insertion and deletion, it is 95.83%. Singh et al. [98] exploited the mean of each DCT vector of every frame and correlation coefficients to detect the duplicated frames and duplicated regions. Accuracy of 96.6% and 99.5% was achieved for detection of duplicated regions and frames, respectively. This method requires high computational time and is not able to detect a smaller number of duplicated frames and smaller duplicated regions.

Huang et al. [117] proposed the Triangular Polarity Feature Classification (TPFC) framework to detect frame insertion and deletion forgeries from videos. Input video was divided into overlapped small groups of frames. Each frame was divided into blocks, and latterly, Block-Wise Variance Descriptor (BBVD) was applied on groups of frames to compute the ratio of BBVD. Finally, to classify a video as authentic or forged, gross error detection from probability theory was employed. The framework was evaluated on 100 videos and achieved 98.26% recall and 95.76% precision. The framework also achieved 91.21% localization accuracy. Although the framework has reasonable results,

cross validation is not explored, which is the ultimate way to expose the strength and weaknesses of any video forgery detection system.

The algorithms of this class are based on statistical features and have a feature vector of small length, but they are not able to detect forgery in the presence of different types of compressions.

7.2. Methods Based on Frequency Domain Features

Discrete Cosine Transformation (DCT), Discrete Wavelet Transformation (DWT) and Fast Fourier Transformation (FFT) are widely used to transform into frequency domain before extraction of features. These techniques are used to verify the small changes. Su et al. [31] utilized Motion-Compensated Edge Artifact (MCEA) and DCT on GOP for detection of video forgery by means of frame deletion. In this research work, five videos, "Bus", "Stefan", "Foreman", "Mother-daughter" and "Flower" were used. TM5 (Test Model 5) was selected as the standard MPEG-2 codec. Consecutive frames in the range of 3, 6 and 9 are deleted from the original video sequences. Videos sequences are encoded on a constant bit-rate ranging from 3 Mbits/s to 9 Mbits/s. Dong et al. [40] also used MCEA to detect the frame deletion based forgery. FFT spikes were used after double MPEG compression. In this study, four videos, "carphone", "container", "hall" and "mobile" with CIF and QCIF format were used. The third, sixth, ninth, twelfth and fifteenth frames are deleted and saved with 15 GOPs. The dataset used in this study is limited in size and the localization of the deleted frames was not exercised. Jaiswal et al. [12] extracted features through DCT, DFT and DWT from Prediction Error Sequence (PES) techniques and classification is performed through SVM and Ensemble-based classifier. This algorithm is unable to detect which frames underwent post-processing operations, such as geometrical transformations. Huang et al. [89] fused audio channels for video forgery detection, where discrete packet decomposition and analysis of singularity points of audio are used to locate forged points. Features are extracted by perceptual hash and Quaternion Discrete Cosine Transform (QDCT) to locate the forgery position in the video. The proposed technique is evaluated by creating a database of forged videos, which are taken from SULFA (Surrey University Library for Forensic Analysis), Open Video Project digital video collection (OV) and self-recorded videos. Precision and recall rates without fine detection were 0.83 and 0.80, respectively, and with fine detection, these rates were 0.9876 and 0.9867, respectively. The restriction is that an audio file is required with the video, which is not always available. Wang et al. [115] proposed a video forgery detection method based on Electronic Network Frequency (ENF). The cubic spline was used to generate the suitable datapoints of ENF signals. The forgery in a video was located using phase continuity interruption, which was observed using correlation between adjacent datapoints of ENF signals. The method has sufficient performance while detecting video forgery in terms of frame deletion, duplication, and insertion. The method is evaluated on limited dataset.

The algorithms based on frequency domain features, i.e., DCT, FFT and DWT, are simple, and the length of the feature vector is small. However, these algorithms are hardware-dependent because the noise is used as a clue for forgery.

7.3. Methods Based on Residual and Optical Flow

Optical flow is a technique that can be calculated by estimating the apparent velocities of movement of brightness patterns from a frame of videos. Similarly, motion residual can also be calculated to estimate the motion in a video [130]. These characteristics can also be useful to detect modifications in a video. Shanableh et al. [44] extracted features based on prediction residuals, a percentage of intra-coded macro-blocks, quantization scales and reconstruction quality of a video. Feature dimension is reduced using Spectral Regression Discriminant Analysis (SRDA). K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Logistic Regression are used to detect the accuracy of the algorithm. The author used 36 video sequences for testing the proposed work with deletion of 1 to 10 frames. The true positive rates of 94% and 95.4% were claimed using SVM classifier

with CBR and VBR, respectively, and false positive rates of 5.5% and 8.2% were achieved by using SVM classifier with CBR and VBR, respectively. The algorithm was tested on limited compression rates. Chao et al. [43] detected frame insertion and deletion by using the fluctuation characteristics of optical flow. In this study, test videos are taken from KTH database and TRECVID Content-Based Copy Detection (CBCD) scripts are used for insertion of frames. Similarly, the CBCD script is used for the deletion of frames. This research detected both types of forgery but has not been tested on different compression ratios. The recall and precision are 95.43% and 95.34%, respectively. Feng et al. [55] proposed an algorithm based on the total motion residual of video frames to detect the frame deletion point. The algorithm is tested on 130 raw YUV tampered videos and made with 5, 10, 15, 20, 25 and 30 deleted frames. True positive and true negative rates were 90% and 0.8%, respectively. The algorithm localized the deletion point but did not consider different compression ratios. Fluctuation features were developed by Feng et al. [70] based on frame motion residual to identify frame deletion points (FDP). Post-processing is used to eliminate minor interferences (sudden lighting change, focus vibration, frame jitter). The proposed technique is evaluated on quick and slow-motion videos to detect frame deletion. The TPR (true positive rate) is 90% if 30 or more than 30 frames are deleted. Performance decreases if the number of frames deleted is lower. This approach is not effective for videos with slow-motion content. Kingra et al. [76] proposed a hybrid technique capable of detecting frame insertion, deletion and duplication exclusively. Multiple features generated by optical flow (OF) and prediction residual (PR) are combined to identify frame base tampering under some threshold. The proposed algorithm was tested on surveillance videos having static background and self-recorded mobile videos. The detection and localization accuracy were 83% and 80%, respectively. This technique can deal individually with frame insertion, deletion, duplication and localization, but did not give satisfactory performance for video sequences that have high illumination. Thorough analysis revealed certain drawbacks. First, this technique was developed for videos having fixed GOP structure and it fails when a whole GOP or its multiples undergo some tampering attack. Second, it is dependent on the number of thresholds that were selected empirically, so there is a lack of flexibility. Third, the model was tested on self-created video sequences that were not sufficient to provide a precise estimation of the applicability of this technique in real scenarios. Jia et al. [85] also used optical flow sum consistency for the detection of duplicated frames in the video. This study used 115 videos to test the proposed algorithm, which are tampered with 10, 20 and 40 duplicated frames. Poor performance is achieved on videos made by a static camera. Joshi et al. [99] exploited frame prediction error and optical flow to classify the authentic and forged videos. Although this method achieved accuracy of 87.5%, it cannot work well for videos shorter than 7 s.

The algorithms of this class are also simple, and the length of feature vector is small; however, they are not able to work on different types of compression rates.

7.4. Methods Based on Pixel and Texture

Texture is an important property of the images that can be used for different types of classification and identification problems. For texture analysis, the pixels are the basic unit. Various texture descriptors are available in the literature that can be used for various tasks. During the tampering process, the texture of the frames of a video is also disturbed and several authors used texture features to detect the tampering in a video. Zhang et al. [66] used quotients of correlation coefficients among sequential Local Binary Pattern(LBP)-coded frames as features and correlation to detect the insertion and deletion of frames. This approach can detect if forgeries exist or not, but it cannot differentiate between frame deletion and insertion forgery. Performance reduces if small numbers of frames are inserted or deleted. Additionally, the forged region is not localized. The precision and recall rates are 88.16% and 85.80%, respectively. The proposed work was not tested for videos compressed at different compression rates. Liao and Huang [48] extracted Tamura texture features, which are based on contrast, orientation and roughness of a video frame and combined

into a 3D feature vector. Euclidean distance is calculated to find the duplicate frame of all feature vectors of all the frames of a video. The method was tested on 10 videos captured using stationary and moving hand-held cameras having a resolution of 640 × 480 pixels and a frame rate of 25–30 fps. The method obtained precision of 99.6%. This method is weak to detect highly similar and duplicated frames having slow sharpness changes. Zhao et al. [88] proposed an algorithm that is divided into two stages. In the first stage, HSV (Hue-Saturation-Value) color histograms are calculated for each frame in a video shot, and similarities between histograms are compared for the detection and localization of tampered frames. Once the forged position is obtained, in the second stage, the candidate frames are double checked by extracting features through SURF (Speeded Up Robust Features) and FLANN (Fast Library for Approximate Nearest Neighbors) matching as a similarity analysis. This method used 10 video shots of different lengths. The precision, recall and accuracy are used as evaluation measures. The method gives suitable results, but only on a small dataset of 10 shots and does not work on grayscale videos. Bakes et al. [100] used Harlalick features of a gray-level co-occurrence matrix (GLCM) for detection of insertion, duplication and deletion of frames. This study used 30 videos tampered with the insertion, deletion and duplication of 10, 20, 30, 40 and 50 frames. Precision, recall and F1 score are used to evaluate the method. The main benefit of the proposed approach is that it does not depend on the size/structure of GOP and the number of deleted frames. However, this method requires a high execution time and cannot detect frame shuffling forgery. Furthermore, it does not work well in the presence of different compression ratios.

Kharat et al. [112] proposed a video forgery detection and localization method based on motion vector, Scale Invariant Feature Transform (SIFT). The forged video frames were identified using motion vector. SIFT features were computed to compare forged frames. Lastly, RANSAC was utilized to localize the forged region. This method was evaluated both on compressed and uncompressed videos. The method achieved overall 99.8% detection accuracy (DA), which is better as compared to other methods. The method was evaluated on 20 videos downloaded from YouTube. It has reasonable performance on duplicate frame detection and localization; however, the method was evaluated on limited authentic and forged videos. Fadl et al. [111] proposed a framework to detect duplicated and shuffled frames based on temporal average and gray-level co-occurrence matrix. The framework achieved 99% precision even in the presence of post-processing operations with high false positives due to weak boundaries of duplicated frames. The method was evaluated on SULFA and LASIESTA datasets. Shelke and Kasana [120] proposed a passive algorithm that utilizes entropy-based texture features, correlation consistency between entropy coded frames and abnormal point detection to detect as well as localize multiple inter-frame forgeries. A dataset of 30 original and 30 forged videos was prepared by using original videos from SULFA, REWIND and VTL. This dataset is not publicly available. Although detection and localization accuracies are 97% and 96.6% in the case of multiple forgeries, this accuracy is attained on a small dataset of 60 videos.

The techniques in the category produced suitable results; however, these methods have long features length and complexity is high.

7.5. Methods Based on Deep Learning

The use of deep learning in the domain of computer vision encourages researchers and scientists to employ deep learning and machine learning models in the domain of video forensics.

In the past few years, deep learning-based methods such as CNN have attained great success in the domain of image processing and computer vision. The reason is that deep neural networks are capable of extracting problem-specific and complex high-dimensional features to efficiently represent the information needed. Deep learning-based approaches have been used recently in many fields, such as camera model identification [131], steganalysis [132], image manipulation detection [133], image copy-move forgery detection [134] and so on. I3D and Siamese(Resnet152) are used for feature extraction, frame duplication

detection and localization in videos by Long et al. [109]. Duplicated frames are distinguished from original frames by an inconsistency detector using I3D. Evaluation was performed on self-recorded iPhone videos, VIRAT [135], and Media Forensics Challenge dataset (MFC18), which is not publicly available. Accuracy of 81% and 84% is obtained in case of iPhone and VIRAT videos while the MCC (Matthews Correlation Coefficient) scores for MFC-dev and MFC-eval set were 0.66 and 0.36, respectively. This technique is capable of detecting just one type of temporal tampering; other manipulation tasks are not carried out, such as frame dropping, frame shuffling, frame rate variations, and effect of various video codecs on algorithm accuracy. Zampoglou et al. [106] explored the potential of two novel filters based on DCT and video requantization error. The output of these filters is used to train deep learning model CNN to discriminate authentic videos from tampered. The model is evaluated on two datasets, one is provided by the NIST 2018 Media Forensics Challenge, and the second is InVID Fake Video Corpus. The accuracy is 85% when training and testing are performed on the same MFC dataset and 60% when testing is performed on the videos of the FVC dataset. Availability of annotated data is one major requirement in this approach, and localization is not addressed. Johnston et al. [136] developed a framework using a CNN for tampering detection which extracted features from authentic content and utilized them to localize the tampered frames and regions. The CNN was trained to estimate quantization parameters, deblock setting and intra/inter mode of pixel patches from an H.264/AVC sequence with suitable accuracy. These features are used for localization of tampered regions in singly and doubly compressed videos having different bitrates. Fadl et al. [118] proposed a system for inter-frame forgery detection where a video is divided into video shots then spatial and temporal information is fused to create a single image of each shot. A pre-trained 2D-CNN model is used for efficient spatiotemporal feature extraction. Then, the structural similarity index (SSIM) is applied to produce deep learning features of a whole video. Finally, they used 2D-CNN and RBF Multiclass Support Vector Machine (RBF-MSVM) to detect temporal tampering in the video. To evaluate the performance of the proposed model, they created their own dataset containing 13135 videos containing three types of forged videos under different conditions by using original videos from VRAT, SULFA, LASIESTA and IVY datasets and achieved TPRs of 0.987, 0.999 and 0.985 for the detection of inter-frame forgery, namely, frame deletion, insertion, and duplication, respectively. Techniques based on deep learning are data-driven (i.e., requiring a large volume of data), and they have the capability to automatically learn high-dimensional features required to detect tampering in the video.

7.6. Others

Some other techniques are also proposed that cannot be categorized. Patel et al. [65] detected temporal forgery based on the EXIF (Extended Image Format) image tag. By analyzing the difference between consecutive frames of the video, the authors successfully identified the tampered region by using the EXIF tag. Although this method localized the forged region, a large database of EXIF tags is required. Gironi et al. [56] used the Variation of Prediction Footprint (VPF) tool with some changes for detecting the frame insertion and deletion. VPF tools are also used for detecting whether the video is encoded or not [42]. This method works for different compression ratios, but it cannot detect frame manipulations when the attacker deletes/inserts a whole group of pictures (GOP). Moreover, the accuracy is 91% but the dataset for training and testing is limited. To overcome the false detections caused by optical flow features and video jitter noise in inter-frame forgery, Pu et al. [119] proposed a novel framework for the detection of inter-frame forgery from the videos with severe brightness changes and jitter noises. A new OF algorithm was introduced to extract stable features of texture changes. It was based on intensity normalization to reduce the impact of illumination noises, and motion entropy to detect jitter noises. Different thresholds are defined for motion entropy to determine whether a video is jittery or not. Experiments were performed on 200 videos taken from three publicly available datasets: SULFA, the CDNET video library and VFDD video lab. Accuracy of 89% was obtained.

Huang et al. [121] proposed a novel cross-modal system that can detect and localize forgery attacks in each frame of live surveillance videos. They prepared their own dataset by collecting multimodal data of half an hour in total. For intra-frame attack, Faster-RCNN is used to detect and crop a human object out and then replace it with the corresponding blank background segment. Forgery detection accuracy of 95% was found on their test data. No cross-dataset validation has been carried out. The algorithms discussed in this section used different methods for feature extraction and classification. Significant temporal forgery techniques in the literature are summarized in Table 4.

7.7. Discussion and Analysis of Temporal Video Tampering Detection Techniques

There exist many models that exploit unique features in videos, such as motion features, noise features, video compression and coding features, color models and GLCM-based features. There are a few limitations of the current strategies, which opens doors for future researchers to conquer these constraints. The existing models are exclusively designed to identify specific types of temporal tampering and operate with some assumptions on selected data. Therefore, the methods developed for a specific type of tampering are incapable of addressing real practical applications due to the diversity in traces left by each type of tampering. There is a serious lack of an efficient approach for the detection of all kinds of video tampering in this domain. Moreover, existing methods are unable to detect tampering if a video has undergone multiple types of tampering attacks.

Many investigators have performed experiments on synthetically doctored videos. While many temporal tampering detection techniques work well on a selected set of videos, they fail to achieve such performances on other unknown video datasets. Moreover, we could not compare the accuracy of these methods because they are evaluated on their own custom-built datasets that satisfy their research assumptions and constraints. In most studies, the efficiency is not reported. Therefore, developing a robust technique for video temporal tampering detection which is capable of detecting all types of temporal tampering and localizing the tampered region is still a cutting-edge research area of video forensics.

Table 4. Summary of temporal tampering (forgery) detection techniques.

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Statistical Features							
2007 [28]	Correlation coefficient	Camera: SONY-HDR-HC3 Two videos with 10,000 frames One video, the camera placed on a tripod and kept stationary throughout Second video hand-held moving camera is used 3, 6 and 9 Mbps bit rate	85.7% and 95.2%	-	-	FPR 6%	Method has not worked to detect the deletion of frames Dataset is small
2013 [44]	KNN + logistic regression + SVM +SRDA	36 video sequences were used with deletion of 1 to 10 frames	-	-	-	TPR 94% FPR 5.5%	Has not worked on a localization of forgery Only detects frame deletion Dataset for training is small
2014 [54]	Correlation Coefficients + SVM	598 videos with a frame rate of 25 Five types of videos in the database Original videos 25 frames inserted 100 frames inserted 25 frames deleted 100 frames deleted	96.21%	-	-	-	Has not worked on a localization of forgery Method is not applied on different compression levels
2019 [98]	Correlation Coefficient + DCT	24 videos are taken from SULFA 6 videos are downloaded from internet	99.5% and 96.6%	99%	100%	F1 99.4% and F2 99.1%	Cannot detect a smaller number of duplicated frames Not able to detect small, duplicated regions
2014 [59]	Consistency of velocity + Cross-Correlation	120 self-created videos	90%	-	-	-	Forge region is not localized Dataset is small

Table 4. *Cont.*

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Frequency Domain Features							
2009 [31]	MCEA + DCT	5 videos 3, 6, or 9 consecutive frames are deleted from the original	Impact Factor α is used				Has not worked on localization of forgery Dataset is small
2012 [40]	MCEA + FFT spikes used after double MPEG compression	4 videos with CIF and QCIF format 3rd, 6th, 9th, 12th and 15th frames are deleted Save with the same GOP = 15	The quantitative measure was not used				Has not worked on localization of forgery Dataset is small
2018 [89]	Quaternion Discrete Cosine Transform (QDCT) feature	SULFA: 101 videos OV (Open Video Project Analysis digital video collection): 14 videos Self-Recorded: 124 videos	-	98.47%	98.76%	-	Audio file is required with videos Poor localization No evaluation on unknown dataset
Methods Based on Residual and Optical Flow							
2014 [55]	Motion residual	130 raw YUV videos tampered and made a video by deleting 5, 10, 15, 20, 25, 30 frames	-	-	-	TPR 90% FAR 0.8%	Only localized frame deletion point No work on frame insertion
2016 [73]	Variation of prediction residual (PR) and the number of intra macro-blocks (NIMBs)	Self-created video	-	81%	88%	F1 score 84%	This method failed when the size of deleted frames was small, and video was in slow motion
2016 [72]	Motion residual+ wavelet	22 YUV raw video	92.73%	-	-	ROC	Did not work well on low compression rate
2013 [43]	Optical flow	TRECVID Content-Based Copy Detection (CBCD) scripts are used with 3000 for frame insertion and deletion in KTH database	-	95.43%	95.34%	-	Has not localized the forge region

Table 4. *Cont.*

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Residual and Optical Flow							
2016 [70]	Optical Flow + IPE (Inra-Prediction Elimination) Process	Group 1: 44 YUV raw files with slow motion content Group 2: 78 YUV raw files with quick motion 5, 10,15, 20, 25, 30 frames are deleted	-	-	-	True Positive Rate 90%	Not applicable to slow-motion videos False alarm rate is high for long video sequences No machine learning scheme is applied
2017 [76]	OF gradient + PR (Prediction Residual) Gradient	Raw videos taken from DIC Punjab University (videos of surveillance camera and Xperia Z2 mobile) Tampered frames: 1% to 6%	Detection accuracy 83% Localization accuracy 80%	-	-	-	Performance decreases when applied on videos having high illumination
2018 [85]	Correlation coefficient + optical flow	Downloaded 115 videos and self-forged with 10, 20, 40 duplicated frames	-	5.5%	98.5%	-	Poor performance on videos taken from static cameras
2019 [99]	Frame prediction error + optical flow	200 videos	87.5%				Cannot work well for videos less than 7 s long
Methods Based on Pixels and Texture Features							
2013 [12]	DCT+ DFT + DWT from Prediction Error Sequence (PES) + SVM, ensemble-based Classifier	20 videos Resolution 176 × 144	-	-	-	ROC	Limited to detect frame deletion Has not localized the forged region Dataset is small for training and testing
2013 [48]	Tamura texture + Euclidean Distance	10 videos captured using stationary and moving hand-held cameras Resolution 640 × 480 Frame rate 25–30 fps	-	-	99.6%	-	Weak to detect highly similar frames Weak to detect duplicate frames if sharpness changes slowly

Table 4. *Cont.*

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Pixels and Texture Features							
2015 [66]	Local Binary Pattern (LBP) + correlation	599 videos with a frame rate of 25 Five types of videos in the database Original videos 25 frames inserted 100 frames inserted 25 frames deleted 100 frames deleted	-	85.80%	88.16%	-	Forge region is not localized Not tested on different compression rates
2018 [88]	SVD + Euclidean distance	10 videos	99.01%	100%	98.07	-	Does not work on grayscale videos
2019 [100]	Harlalick features	30 videos from different sources with 10, 20, 30, 40, 50 frame insertion, deletion		96%	86%	F1 score 91%	Does not work in presence of compression
Methods Based on Deep Learning							
2019 [109]	13D + ResNet network	Media Forensics Challenge dataset (MFC18) 231 videos in MFC-Eval and 1036videos in MFC-Dev, static camera raw videos from VIRAT: 12 videos Self-recorded iPhone 4 videos: 17 videos Videos of length 0.5 s, 1 s, 2 s, 5 s and 10 s are inserted into same source video	-	-	-	AUC 99.9%	Performance is degraded in presence of multiple sequences of the duplicated frames in a video

Table 4. *Cont.*

References	Methods	Dataset	Evaluation Measures				Limitations/Issues
			Accuracy	Recall	Precision	Others	
Methods Based on Deep Learning							
2019 [106]	CNN	NIST 2018, Media Forensics Challenge2 for the video manipulation detection task, 116 tampered and 116 original 35 real and 33 fake videos are taken from InVID Fake Video Corpus	85%	-	-	-	Labeled video data are required Localization is not completed
2020 [136]	CNN	Face Forensics VTD Dataset provided by [81]	-	-	-	MCC: 0.67 F1: 0.81	Only proposed for videos that have fixed GOP size and still background It can only deal with single type of tampering
2021 [118]	2D-CNN + SSIM + RBF-MSVM	Raw videos taken from VRAT, SULFA, LASIESTA and IVY datasets				TPR in ins, del anddup forgery are: 0.999, 0.987, 0.985	Cross dataset evaluation was not performed on unknown dataset
Others							
2014 [56]	Variation of Prediction Footprint	14 videos Resolution 352 × 288 pixels 1250 frames 100, 300, 700 frames removed and inserted	91%	-	-	-	Has not worked to localize the forged object Dataset is small

8. Research Challenges

Given our analyses of the existing literature on passive video tampering techniques, this field of research faces the following challenge.

8.1. Benchmark Dataset

Performance of every recognition system depends on its training, testing and evaluation. The dataset is the key for proper training, testing and evaluation for any proposed algorithm. To the best of our knowledge, existing video forgery datasets are not appropriate due to being small in size and lacking post-processing operations such as rotation, scaling, blurring, compression, etc. [137]. The details of existing datasets for video forensic analysis are presented in Table 5. Many researchers have developed their own datasets [70,73,85] to conduct experiments for inter-frame forgery detection, but these datasets are not available for other communities/researchers to evaluate the performance of the proposed algorithms. This portrays video tampering detection as a solved problem on specific, self-created, small datasets with high accuracy, which may discourage other researchers from publishing their work with less accuracy. In this regard, a great effort has been made for image forensics and source device identification [138]. On the contrary, no benchmark dataset is available for video forensics. To prepare tampered videos manually is a highly time-consuming process, so many authors used synthetically doctored videos for their experiments, such as Panchal et al. in [139].

Therefore, a benchmark dataset for proper training and testing needs to be developed that could give an unbiased and neutral platform for comparison of various techniques with existing state-of-the-art video tampering (forgery) detection techniques.

8.2. Performance and Evaluation

Most video forgery algorithms are based on camera source identification; therefore, the results can be negatively affected by increasing the number of cameras. Moreover, the camera source identification methods are noted to be dependent on intrinsic camera hardware features such as lens and charge-couple device (CCD) sensor characteristics that can degrade performance of the algorithm. Video double compression artifacts add difficulty to the localization of the video forgery, especially when the video being analyzed is compressed by a low-quality factor, which is seen in most of the recent methods. Similarly, video forgery detection depends on post-processing operations such as edge blurring, compression, noise, scaling, rotation, etc., and can cause high false positives. Most of the existing methods on video forgery detection have no resistance to such post-processing operations. All these aspects degrade the performance of the techniques.

The existing methods are evaluated with different metrics; that's why they can't be compared with each other. Thus, there is a need for standard evaluation measures based on inconsistent lighting and correlation between pixels, so that comparisons can be easily carried out between different algorithms.

8.3. Automation

Existing methods of video forgery detection and localization are not fully automated and require human interpretation, which results in poor accuracy.

Table 5. Details of existing video forgery datasets.

Dataset Name and Reference	Number of Videos	Video Length (in s)	Video Source	Static/Moving Camera	Type of Video Forgery	Scenario (Mor, Eve, Night, Fog)	Available in Public Domain
TDTVD, Panchal, Shah et al., 2020 [139]	Original: 40 Tampered: 210	6–18 s	SULFA, YouTube	Static and moving	Frame insertion, deletion, duplication, and smart tampering	N/A	Yes
Pu, Huang et al., 2021 [119]	Original + Tampered: 200	N/A	CDNET Video Library, SULFA, VFDD Video Library (Video Forgery Detection Database of South China University of Technology Version 1.0)	Static and moving	Frame deletion, insertion, replacement, and copy-move	N/A	No
Shelke and Kasana 2021 [120]	Original: 30 Tampered: 30	N/A	SULFA, REWIND, and VTL	Static and moving	Frame insertion, deletion, duplication, and frame splicing	N/A	No
Test Database, Ulutas et al. [140]	Original +Tampered: 31		SULFA and different movie scenes	Static and moving	Frame duplication	N/A	Yes
Le, Almansa et al., 2017 [141]	Tampered: 53	N/A	N/A	Static and moving camera	Video in-painting	N/A	Yes
VTD dataset, Al-Sanjary, Ahmed et al., 2016 [142]	Original: 7 Tampered: 26	14–16 s	YouTube	Static and moving camera	Copy-move, swapping frames, splicing	N/A	Yes

Table 5. *Cont.*

Dataset Name and Reference	Number of Videos	Video Length (in s)	Video Source	Static/Moving Camera	Type of Video Forgery	Scenario (Mor, Eve, Night, Fog)	Available in Public Domain
Feng, Xu et al., 2016 [70]	Original: 122 Tampered: 732	N/A	YUV files http://trace.eas.asu.edu/yuv/ http://media.xiph.org/video/derf/ ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/ * http://202.114.114.212/quick_motion/yuv_download.html *, (accessed on 16 November 2021)	Static and moving camera	Frame deletion	N/A	No
SYSU-OBJFORG, Chen et al. [61]	Total: 100	11 s	Commercial Surveillance Cameras	Static camera	Object-based	N/A	No
Su, Huang et al., 2015 [64]	Original + Tampered: 20	N/A	SONY DSCP10	Static camera	Copy-move	N/A	No
REWIND PROJECT, Bestagini et al., 2013 [45]	Original: 10 Tampered: 10	7–19 s	Canon SX220, Nikon S3000, Fujifilm S2800HD	Static camera	Copy-move	N/A	Yes
SULFA Dataset, Qadir, Yahaya et al., 2012 [137]	Original: 166 Tampered: 5	4–18 s	Canon SX220, Nikon S3000 Fujifilm S2800HD	Static camera	Copy-move	N/A	No *

* Given link is not accessible.

8.4. Localization

Video forgery detection makes a user aware of if the video is authentic or not, but when a user knows which part of the video is forged, the trustworthiness of forgery detection systems will increase. To determine the accurate location of video tampering is another big challenge. Some of the developed approaches are capable of localizing the tampered region in a video, but accuracy rates were inadequate; furthermore, in many studies, little attention has been paid to localizing the tampered region. Moreover, no remarkable results have been observed in existing methods to localize the traces of forged regions in tampered videos. As existing methods have not modeled the structural changes properly, this occurred in videos after spatial forgery. Due to these reasons, the accuracy of localizing the forged region is still a challenge.

8.5. Robustness

An algorithm is known to be robust if it detects and localizes every type of forgery in general and not specifically on a certain dataset. Most of the reported algorithms have high accuracy on certain datasets on which they are evaluated but not in general, which makes it difficult to perform comparative analyses among existing techniques. An important limitation of existing methods is the lack of sufficient validation of standardized datasets. Thus, there is a need to establish benchmarks for the detection and localization of all types of forgery in videos by ensuring high accuracy so that it would be appropriate to deploy in real practical applications.

9. Future Directions

A standard dataset may be developed to benefit the research community to train, test and evaluate their algorithms. Video forgery may be detected and localized in the following ways. The whole process of video tampering detection and localization is elaborated in Figure 7. Initially, features can be extracted through different multi-resolution techniques, namely, local binary pattern (LBP) [143], Weber's law descriptor (WLD) [144] and discriminative robust local binary pattern (DRLBP) [145]. Complementary features can then be integrated from these techniques to gather more discriminative features. Principle component analysis (PCA) is likely to be used for selecting the most suitable or unique features out of the extracted features [146]. These selective features can then be passed to an SVM to classify the video as forged or authentic [147].

Edges are tampering artifacts and give better representation of the objects. The edge irregularity caused by tampering can be noticed in chrominance channels. The YCbCr color model was used by Muhammad et al. in [148] as a pre-processing step to extract features from Cb and Cr channels to represent the structural changes. The reason to extract features using Cb and Cr components is to gather discriminative features which represent the information of edges caused by tampering, because edges appeared sharply in the Cb or Cr channel. Although LBP gives texture information, it failed to retrieve edge information. Since DRLBP and WLD contain both edge and texture information and produce discriminative features to represent the clues of forgery, more accurate results are expected than LBP in detecting video tampering in the spatial domain. Similarly, the spatial/temporal forged region can be localized by using either block-based or clustered-based techniques.

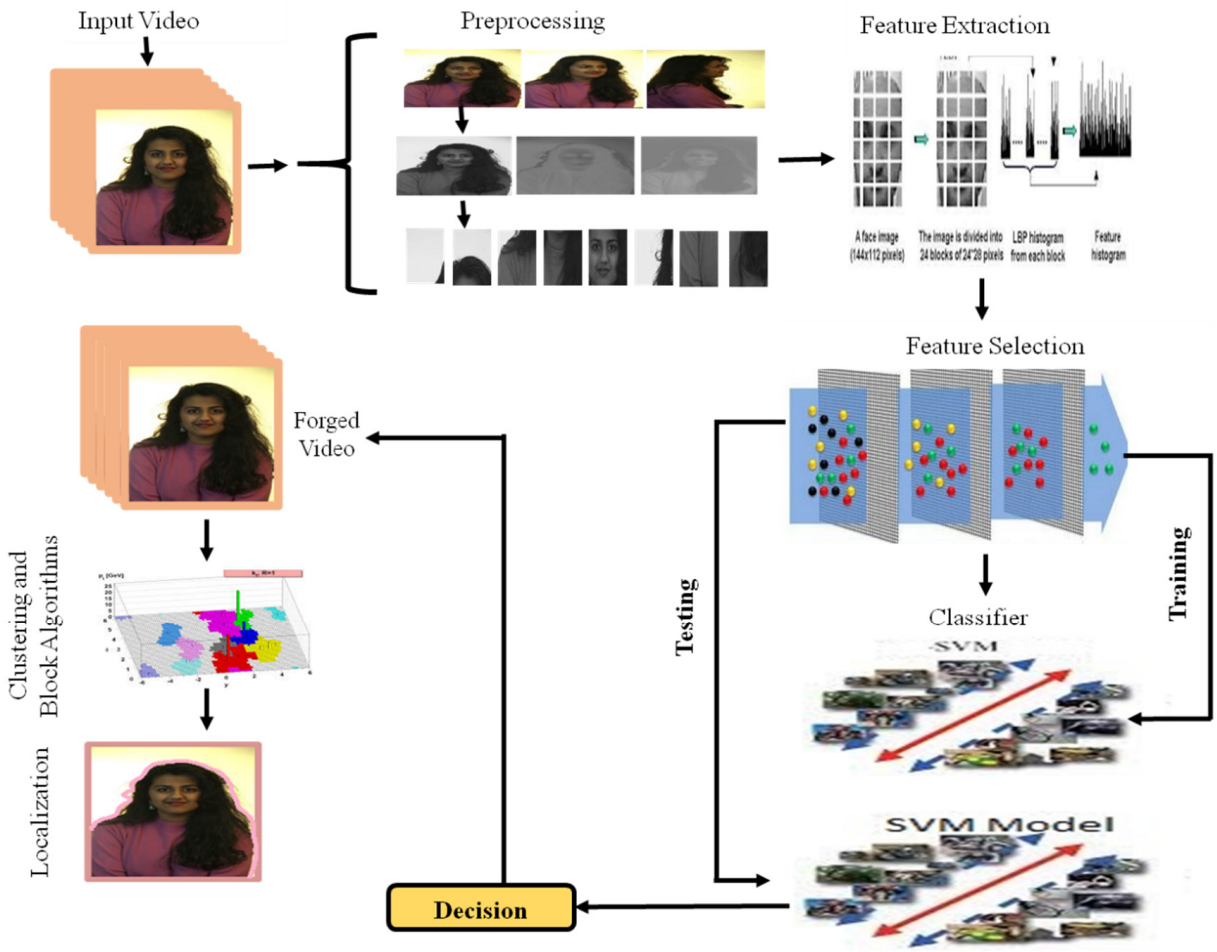

Figure 7. Process of video tampering detection and localization.

Efficiency is another major concern due to the high volume of video frames under observation. For better accuracy and efficiency, Convolutional Neural Network (CNN)-based algorithms such as deep learning (DL), auto encoder or deep belief networks (DBN) can also be evaluated [149] due to their success in artificial intelligence (AI) tasks such as image recognition [150], speech recognition [151] and natural language processing (NLP) [152].

Deep learning [153] has inspired other machine learning techniques to foresee the activity of potential drug molecules [154], reconstruct brain circuits [155], online particle detection [156], predict the effects of mutations in non-coding DNA on gene expression and disease [157], and many other applications. CNN [158] is specialized as fully connected layers and is also easy to train. Major technology companies including Google, Facebook, Yahoo!, Twitter, Microsoft, and IBM have used CNN-based algorithms.

CNN on the large scale is not extremely fast; therefore, CNN-based hardware chips are developed by NVIDIA, Mobil eye, Intel, Qualcomm, and Samsung to reduce the training time. For better efficiency, we also need to think about the extreme learning machine (ELM). ELM not only achieves state-of-the-art results but also shortens the training time from days (spent by deep learning) to several minutes without scarifying the accuracy. Extreme learning is successfully performed in applications such as soft-sensing in the complex chemical process [159], face recognition [160] and many more.

Transfer learning [161,162] is another topic of ongoing interest in the machine learning community. It is the process of the improvement of learning in a new task where training data are limited through the transfer of knowledge from a related task that has already been learned. This shortage of training data can be due to several reasons, such as data being fitful, costly to collect and label or being unavailable. Many applications of machine

learning are successfully applied transferring learning for image classification [163], human activity classification [164], event classification from a video [165], software prediction [166], multi-language text classification [167] and many others. Since the benchmarked forged video datasets are not available, a learning system for video tampering analysis can be developed through transfer learning techniques by using existing partially or closely related learning models.

10. Conclusions

Digital video forensics is still in its infancy and the reliability of digital video as a reference in court is questionable due to tampering (forgery). Numerous video editing tools such as Adobe's (Premier and After Effect), GNU Gimp, Premier and Vegas are readily available to tamper videos. Several techniques have been proposed in the literature to detect tampering, and they all suffer from their share of limitations. In this study, we carried out a systematic review of digital video forgery detection techniques and provided answers to the research questions guiding this work. The existing passive video forgery detection and localization techniques are categorized into spatial and temporal techniques. These spatial and temporal techniques are further categorized based on their features. We performed in-depth investigations of methods, their comparative analysis and the merits and demerits of each category, and we debated challenges extracted from video forensics literature. The review of related work illustrates that various features can be exploited to detect and localize forgery. LBP, frame motion residual, noise features, SURF and optical flow give suitable detection accuracy, but their performance is reduced due to presence of illumination, static scenes, tampering of small number of frames, video quality and variable GOP sizes. Even though techniques based on deep learning are convincing, few researchers have adopted it due to the unavailability of large video forgery datasets. Secondly, the detection of inter-frame forgeries has been addressed exclusively, highlighting the need to establish benchmarks for detection and localization of all kinds of temporal tampering in videos by ensuring high accuracy. Thirdly, to the best of our knowledge, no work is available in the public domain that can detect tampering if a video has undergone multiple types of tampering attacks. The detection of multiple types of tampering in a video is an area of research that needs to be explored. Fourthly, manually producing tampered videos is very time-consuming task, which is why most researchers performed their experiments on synthetically doctored video sequences. Finally, an important limitation of existing methods is the lack of sufficient validation on standardized datasets.

Author Contributions: Conceptualization, M.H., Z.H. and U.I.B.; Data curation, N.A., M.S. and K.A.; Formal analysis, N.A., M.S. and K.A., Funding acquisition, Z.H.; Investigation, N.A. and M.S.; Methodology, N.A. and M.S.; Project administration, Z.H.; Resources, Z.H.; Supervision, M.H., Z.H. and U.I.B.; Writing—original draft, N.A. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the PDE-GIR project, which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skldowska-Curie grant agreement no. 778035.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: There is no conflict of interest with respect to the research.

References

1. Su, P.-C.; Suei, P.-L.; Chang, M.-K.; Lain, J. Forensic and anti-forensic techniques for video shot editing in H. 264/AVC. *J. Vis. Commun. Image Represent.* **2015**, *29*, 103–113. [CrossRef]
2. Wang, W. *Digital Video Forensics in Dartmouth College*; Computer Science Department: Hanover, NH, USA, 2009.
3. Pan, X.; Lyu, S. Region duplication detection using image feature matching. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 857–867. [CrossRef]
4. Rocha, A.; Scheirer, W.; Boult, T.; Goldenstein, S. Vision of the unseen: Current trends and challenges in digital image and video forensics. *ACM Comput. Surv.* **2011**, *43*, 1–42. [CrossRef]
5. Lee, J.-C.; Chang, C.-P.; Chen, W.-K. Detection of copy–move image forgery using histogram of orientated gradients. *Inf. Sci.* **2015**, *321*, 250–262. [CrossRef]
6. Zhao, Y.; Wang, S.; Zhang, X.; Yao, H. Robust hashing for image authentication using Zernike moments and local features. *IEEE Trans. Inf. Forensics Secur.* **2013**, *8*, 55–63. [CrossRef]
7. Asghar, K.; Habib, Z.; Hussain, M. Copy-move and splicing image forgery detection and localization techniques: A review. *Aust. J. Forensic Sci.* **2016**, *49*, 1–27. [CrossRef]
8. Singh, R.D.; Aggarwal, N. Video content authentication techniques: A comprehensive survey. *Multimed. Syst.* **2018**, *24*, 211–240. [CrossRef]
9. Sitara, K.; Mehtre, B. Digital video tampering detection: An overview of passive techniques. *Digit. Investig.* **2016**, *18*, 8–22. [CrossRef]
10. Pandey, R.C.; Singh, S.K.; Shukla, K.K. Passive forensics in image and video using noise features: A review. *Digit. Investig.* **2016**, *19*, 1–28. [CrossRef]
11. Milani, S.; Fontani, M.; Bestagini, P.; Barni, M.; Piva, A.; Tagliasacchi, M.; Tubaro, S. An overview on video forensics. *APSIPA Trans. Signal Inf. Process.* **2012**, *1*, 1–18. [CrossRef]
12. Jaiswal, S.; Dhavale, S. Video Forensics in Temporal Domain using Machine Learning Techniques. *Int. J. Comput. Netw. Inf. Secur.* **2013**, *5*, 58. [CrossRef]
13. Bestagini, P.; Fontani, M.; Milani, S.; Barni, M.; Piva, A.; Tagliasacchi, M.; Tubaro, S. An overview on video forensics. In Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, Romania, 27–31 August 2012.
14. Wahab, A.W.A.; Bagiwa, M.A.; Idris, M.Y.I.; Khan, S.; Razak, Z.; Ariffin, M.R.K. Passive video forgery detection techniques: A survey. In Proceedings of the 10th International Conference on Information Assurance and Security (IAS), Okinawa, Japan, 28–30 November 2014.
15. Al-Sanjary, O.I.; Sulong, G. Detection of video forgery: A review of literature. *J. Theor. Appl. Inf. Technol.* **2015**, *74*, 208–220.
16. Sowmya, K.; Chennamma, H. A Survey On Video Forgery Detection. *Int. J. Comput. Eng. Appl.* **2015**, *9*, 18–27.
17. Sharma, S.; Dhavale, S.V. A review of passive forensic techniques for detection of copy-move attacks on digital videos. In Proceedings of the 3rd International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 22–23 January 2016.
18. Tao, J.; Jia, L.; You, Y. Review of passive-blind detection in digital video forgery based on sensing and imaging techniques. In Proceedings of the International Conference on Optoelectronics and Microelectronics Technology and Application. International Society for Optics and Photonics, Shanghai, China, 5 January 2017.
19. Mizher, M.A.; Ang, M.C.; Mazhar, A.A.; Mizher, M.A. A review of video falsifying techniques and video forgery detection techniques. *Int. J. Electron. Secur. Digit. Forensics* **2017**, *9*, 191–208. [CrossRef]
20. Sharma, H.; Kanwal, N.; Batth, R.S. An Ontology of Digital Video Forensics: Classification, Research Gaps & Datasets. In Proceedings of the 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 11–12 December 2019.
21. Johnston, P.; Elyan, E. A review of digital video tampering: From simple editing to full synthesis. *Digit. Investig.* **2019**, *29*, 67–81. [CrossRef]
22. Kaur, H.; Jindal, N. Image and Video Forensics: A Critical Survey. *Wirel. Pers. Commun.* **2020**, *112*, 67–81. [CrossRef]
23. Shelke, N.A.; Kasana, S.S. A comprehensive survey on passive techniques for digital video forgery detection. *Multimed. Tools Appl.* **2021**, *80*, 6247–6310. [CrossRef]
24. Parmar, Z.; Upadhyay, S. A Review on Video/Image Authentication and Temper Detection Techniques. *Int. J. Comput. Appl.* **2013**, *63*, 46–49. [CrossRef]
25. Alsmirat, M.A.; Al-Hussien, R.A.; Al-Sarayrah, W.a.T.; Jararweh, Y.; Etier, M. Digital video forensics: A comprehensive survey. *Int. J. Adv. Intell. Paradig.* **2020**, *15*, 437–456. [CrossRef]
26. Kitchenham, B.; Brereton, O.P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering–a systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [CrossRef]
27. Wang, W.; Farid, H. Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Trans. Inf. Forensics Secur.* **2007**, *2*, 438–449. [CrossRef]
28. Wang, W.; Farid, H. Exposing digital forgeries in video by detecting duplication. In Proceedings of the 9th Workshop on Multimedia & Security, New York, NY, USA, 20–21 September 2007.
29. Hsu, C.-C.; Hung, T.-Y.; Lin, C.-W.; Hsu, C.-T. Video forgery detection using correlation of noise residue. In Proceedings of the IEEE 10th Workshop on Multimedia Signal Processing, Cairns, Australia, 8–10 October 2008.

30. Shih, T.K.; Tang, N.C.; Hwang, J.-N. Exemplar-based video inpainting without ghost shadow artifacts by maintaining temporal continuity. *IEEE Trans. Circuits Syst. Video Technol.* **2009**, *19*, 347–360. [CrossRef]
31. Su, Y.; Zhang, J.; Liu, J. Exposing digital video forgery by detecting motion-compensated edge artifact. In Proceedings of the IEEE International Conference on Computational Intelligence and Software Engineering, CiSE, Wuhan, China, 11–13 December 2009.
32. Zhang, J.; Su, Y.; Zhang, M. Exposing digital video forgery by ghost shadow artifact. In Proceedings of the First ACM Workshop on Multimedia in Forensics, Beijing, China, 23 October 2009.
33. Kobayashi, M.; Okabe, T.; Sato, Y. Detecting video forgeries based on noise characteristics. In *Advances in Image and Video Technology*; Springer: Tokyo, Japan, 2009; pp. 306–317.
34. Kobayashi, M.; Okabe, T.; Sato, Y. Detecting forgery from static-scene video based on inconsistency in noise level functions. *IEEE Trans. Inf. Forensics Secur.* **2010**, *5*, 883–892. [CrossRef]
35. Chetty, G. Blind and passive digital video tamper detection based on multimodal fusion. In Proceedings of the 14th WSEAS International Conference on Communications, Corfu, Greece, 23–25 July 2010.
36. Goodwin, J.; Chetty, G. Blind video tamper detection based on fusion of source features. In Proceedings of the IEEE International Conference on Digital Image Computing Techniques and Applications (DICTA), Noosa, Australia, 6–8 December 2011.
37. Stamm, M.C.; Liu, K.R. Anti-forensics for frame deletion/addition in MPEG video. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011.
38. Conotter, V.; O'Brien, J.F.; Farid, H. Exposing digital forgeries in ballistic motion. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 283–296. [CrossRef]
39. Stamm, M.C.; Lin, W.S.; Liu, K.R. Temporal forensics and anti-forensics for motion compensated video. *IEEE Trans. Inf. Forensics Secur.* **2012**, *7*, 1315–1329. [CrossRef]
40. Dong, Q.; Yang, G.; Zhu, N. A MCEA based passive forensics scheme for detecting frame-based video tampering. *Digit. Investig.* **2012**, *9*, 151–159. [CrossRef]
41. Subramanyam, A.; Emmanuel, S. Video forgery detection using HOG features and compression properties. In Proceedings of the IEEE 14th International Workshop on Multimedia Signal Processing (MMSP), Banff, AB, Canada, 17–19 September 2012.
42. Vazquez-Padin, D.; Fontani, M.; Bianchi, T.; Comesaña, P.; Piva, A.; Barni, M. Detection of video double encoding with GOP size estimation. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Tenerife, Spain, 2–5 December 2012.
43. Chao, J.; Jiang, X.; Sun, T. A novel video inter-frame forgery model detection scheme based on optical flow consistency, in Digital Forensics and Watermaking. In Proceedings of the 11th International Workshop, IWDW 2012, Shanghai, China, 31 October–3 November 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 267–281.
44. Shanableh, T. Detection of frame deletion for digital video forensics. *Digit. Investig.* **2013**, *10*, 350–360. [CrossRef]
45. Bestagini, P.; Milani, S.; Tagliasacchi, M.; Tubaro, S. Local tampering detection in video sequences. In Proceedings of the 15th International Workshop on Multimedia Signal Processing (MMSP), Pula, Italy, 30 September–2 October 2013.
46. Labartino, D.; Bianchi, T.; De Rosa, A.; Fontani, M.; Vazquez-Padin, D.; Piva, A.; Barni, M. Localization of forgeries in MPEG-2 video through GOP size and DQ analysis. In Proceedings of the 15th International Workshop on Multimedia Signal Processing, Pula, Italy, 30 September–2 October 2013.
47. Li, L.; Wang, X.; Zhang, W.; Yang, G.; Hu, G. Detecting removed object from video with stationary background. In Proceedings of the International Workshop on Digital Forensics and Watermarking, Taipei, Taiwan, 1–4 October 2013.
48. Liao, S.-Y.; Huang, T.-Q. Video copy-move forgery detection and localization based on Tamura texture features. In Proceedings of the 6th International Congress on Image and Signal Processing (CISP), Hangzhou, China, 16–18 December 2013.
49. Lin, C.-S.; Tsay, J.-J. Passive approach for video forgery detection and localization. In Proceedings of the Second International Conference on Cyber Security, Cyber Peacefare and Digital Forensic (CyberSec2013), The Society of Digital Information and Wireless Communication, Kuala Lumpur, Malaysia, 4–6 March 2013.
50. Subramanyam, A.; Emmanuel, S. Pixel estimation based video forgery detection. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vancouver, BC, Canada, 26–31 May 2013.
51. Wang, W.; Jiang, X.; Wang, S.; Wan, M.; Sun, T. Identifying video forgery process using optical flow, in Digital-Forensics and Watermarking. In Proceedings of the 11th International Workshop, IWDW 2012, Shanghai, China, 31 October–3 November 2012; Springer: Berlin/Heidelberg, Germany, 2013; pp. 244–257.
52. Lin, C.-S.; Tsay, J.-J. A passive approach for effective detection and localization of region-level video forgery with spatio-temporal coherence analysis. *Digit. Investig.* **2014**, *11*, 120–140. [CrossRef]
53. Richao, C.; Gaobo, Y.; Ningbo, Z. Detection of object-based manipulation by the statistical features of object contour. *Forensic Sci. Int.* **2014**, *236*, 164–169. [CrossRef]
54. Wang, Q.; Li, Z.; Zhang, Z.; Ma, Q. Video Inter-Frame Forgery Identification Based on Consistency of Correlation Coefficients of Gray Values. *J. Comput. Commun.* **2014**, *2*, 51. [CrossRef]
55. Feng, C.; Xu, Z.; Zhang, W.; Xu, Y. Automatic location of frame deletion point for digital video forensics. In Proceedings of the 2nd ACM Workshop on Information Hiding and Multimedia Security, Salzburg, Austria, 11–13 June 2014.
56. Gironi, A.; Fontani, M.; Bianchi, T.; Piva, A.; Barni, M. A video forensic technique for detecting frame deletion and insertion. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014.

57. Liu, H.; Li, S.; Bian, S. Detecting frame deletion in H. 264 video. In Proceedings of the International Conference on Information Security Practice and Experience, Fuzhou, China, 5–8 May 2014.
58. Pandey, R.C.; Singh, S.K.; Shukla, K. Passive copy-move forgery detection in videos. In Proceedings of the International Conference on Computer and Communication Technology (ICCCT), Allahabad, India, 26–28 September 2014.
59. Wu, Y.; Jiang, X.; Sun, T.; Wang, W. Exposing video inter-frame forgery based on velocity field consistency. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 4–9 May 2014.
60. Chen, S.; Tan, S.; Li, B.; Huang, J. Automatic detection of object-based forgery in advanced video. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *26*, 2138–2151. [CrossRef]
61. Zheng, L.; Sun, T.; Shi, Y.-Q. Inter-frame video forgery detection based on block-wise brightness variance descriptor. In Proceedings of the International Workshop on Digital Watermarking, Tokyo, Japan, 3 June 2014.
62. Jung, D.-J.; Hyun, D.-K.; Lee, H.-K. Recaptured video detection based on sensor pattern noise. *EURASIP J. Image Video Process.* **2015**, *2015*, 40. [CrossRef]
63. Kang, X.; Liu, J.; Liu, H.; Wang, Z.J. Forensics and counter anti-forensics of video inter-frame forgery. *Multimed. Tools Appl.* **2015**, *75*, 1–21. [CrossRef]
64. Su, L.; Huang, T.; Yang, J. A video forgery detection algorithm based on compressive sensing. *Multimed. Tools Appl.* **2014**, *74*, 6641–6656. [CrossRef]
65. Patel, H.C.; Patel, M.M. An Improvement of Forgery Video Detection Technique using Error Level Analysis. *Int. J. Comput. Appl.* **2015**, *111*. [CrossRef]
66. Zhang, Z.; Hou, J.; Ma, Q.; Li, Z. Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames. *Secur. Commun. Netw.* **2015**, *8*, 311–320. [CrossRef]
67. Bidokhti, A.; Ghaemmaghami, S. Detection of regional copy/move forgery in MPEG videos using optical flow. In Proceedings of the International symposium on Artificial intelligence and signal processing (AISP), Mashhad, Iran, 3–5 March 2015.
68. D'Amiano, L.; Cozzolino, D.; Poggi, G.; Verdoliva, L. Video forgery detection and localization based on 3D patchmatch. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Torino, Italy, 29 June–3 July 2015.
69. Tan, S.; Chen, S.; Li, B. GOP based automatic detection of object-based forgery in advanced video. In Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), Hong Kong, China, 6–19 December 2015.
70. Feng, C.; Xu, Z.; Jia, S.; Zhang, W.; Xu, Y. Motion-adaptive frame deletion detection for digital video forensics. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *27*, 2543–2554. [CrossRef]
71. Yang, J.; Huang, T.; Su, L. Using similarity analysis to detect frame duplication forgery in videos. *Multimed. Tools Appl.* **2014**, *75*, 1793–1811. [CrossRef]
72. Aghamaleki, J.A.; Behrad, A. Inter-frame video forgery detection and localization using intrinsic effects of double compression on quantization errors of video coding. *Signal Process. Image Commun.* **2016**, *47*, 289–302. [CrossRef]
73. Yu, L.; Wang, H.; Han, Q.; Niu, X.; Yiu, S.; Fang, J.; Wang, Z. Exposing frame deletion by detecting abrupt changes in video streams. *Neurocomputing* **2016**, *205*, 84–91. [CrossRef]
74. Mathai, M.; Rajan, D.; Emmanuel, S. Video forgery detection and localization using normalized cross-correlation of moment features. In Proceedings of the IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), Santa Fe, NM, USA, 6–8 March 2016.
75. Liu, Y.; Huang, T.; Liu, Y. A novel video forgery detection algorithm for blue screen compositing based on 3-stage foreground analysis and tracking. *Multimed. Tools Appl.* **2017**, *77*, 7405–7427. [CrossRef]
76. Kingra, S.; Aggarwal, N.; Singh, R.D. Inter-frame forgery detection in H. 264 videos using motion and brightness gradients. *Multimed. Tools Appl.* **2017**, *76*, 25767–25786. [CrossRef]
77. Singh, R.D.; Aggarwal, N. Detection and localization of copy-paste forgeries in digital videos. *Forensic Sci. Int.* **2017**, *281*, 75–91. [CrossRef]
78. Fadl, S.M.; Han, Q.; Li, Q. Authentication of Surveillance Videos: Detecting Frame Duplication Based on Residual Frame. *J. Forensic Sci.* **2017**, *63*, 1099–1109. [CrossRef]
79. Yao, Y.; Shi, Y.; Weng, S.; Guan, B. Deep learning for detection of object-based forgery in advanced video. *Symmetry* **2017**, *10*, 3. [CrossRef]
80. Bozkurt, I.; Bozkurt, M.H.; Ulutaş, G. A new video forgery detection approach based on forgery line. *Turk. J. Electr. Eng. Comput. Sci.* **2017**, *25*, 4558–4574. [CrossRef]
81. D'Avino, D.; Cozzolino, D.; Poggi, G.; Verdoliva, L. Autoencoder with recurrent neural networks for video forgery detection. *Electron. Imaging* **2017**, *2017*, 92–99. [CrossRef]
82. Huang, C.C.; Zhang, Y.; Thing, V.L. Inter-frame video forgery detection based on multi-level subtraction approach for realistic video forensic applications. In Proceedings of the IEEE 2nd International Conference on Signal and Image Processing (ICSIP), Singapore, 4–6 August 2017.
83. Al-Sanjary, O.I.; Ghazali, N.; Ahmed, A.A.; Sulong, G. Semi-automatic Methods in Video Forgery Detection Based on Multi-view Dimension. In Proceedings of the International Conference of Reliable Information and Communication Technology, Johor, Malaysia, 23–24 April 2017.

84. D'Amiano, L.; Cozzolino, D.; Poggi, G.; Verdoliva, L. A patchmatch-based dense-field algorithm for video copy-move detection and localization. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 669–682. [CrossRef]
85. Jia, S.; Xu, Z.; Wang, H.; Feng, C.; Wang, T. Coarse-to-fine copy-move forgery detection for video forensics. *IEEE Access* **2018**, *6*, 25323–25335. [CrossRef]
86. Su, L.; Li, C.; Lai, Y.; Yang, J. A Fast Forgery Detection Algorithm Based on Exponential-Fourier Moments for Video Region Duplication. *IEEE Trans. Multimed.* **2018**, *20*, 825–840. [CrossRef]
87. Su, L.; Li, C. A novel passive forgery detection algorithm for video region duplication. *Multidimens. Syst. Signal Process.* **2018**, *29*, 1173–1190. [CrossRef]
88. Zhao, D.-N.; Wang, R.-K.; Lu, Z.-M. Inter-frame passive-blind forgery detection for video shot based on similarity analysis. *Multimed. Tools Appl.* **2018**, *77*, 25389–25408. [CrossRef]
89. Huang, T.; Zhang, X.; Huang, W.; Lin, L.; Su, W. A multi-channel approach through fusion of audio for detecting video inter-frame forgery. *Comput. Secur.* **2018**, *77*, 412–426. [CrossRef]
90. Al-Sanjary, O.I.; Ahmed, A.A.; Jaharadak, A.A.; Ali, M.A.; Zangana, H.M. Detection clone an object movement using an optical flow approach. In Proceedings of the IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE), Penang, Malaysia, 28–29 April 2018.
91. Guo, C.; Luo, G.; Zhu, Y. A detection method for facial expression reenacted forgery in videos. In Proceedings of the Tenth International Conference on Digital Image Processing (ICDIP 2018), International Society for Optics and Photonics, Shanghai, China, 11–14 May 2018.
92. Bakas, J.; Naskar, R. A Digital Forensic Technique for Inter–Frame Video Forgery Detection Based on 3D CNN. In Proceedings of the International Conference on Information Systems Security, Funchal, Purtugal, 22–24 January 2018.
93. Antony, N.; Devassy, B.R. Implementation of Image/Video Copy-Move Forgery Detection Using Brute-Force Matching. In Proceedings of the 2nd International Conference on Trends in Electronics and Informatics (ICOEI), Tamil Nadu, India, 11–12 May 2018.
94. Kono, K.; Yoshida, T.; Ohshiro, S.; Babaguchi, N. Passive Video Forgery Detection Considering Spatio-Temporal Consistency. In Proceedings of the International Conference on Soft Computing and Pattern Recognition, Porto, Purtugal, 13–15 December 2018.
95. Bakas, J.; Bashaboina, A.K.; Naskar, R. Mpeg double compression based intra-frame video forgery detection using cnn. In Proceedings of the International Conference on Information Technology (ICIT), Bhubaneswar, India, 19–21 December 2018.
96. Afchar, D.; Nozick, V.; Yamagishi, J.; Echizen, I. MesoNet: A Compact Facial Video Forgery Detection Network. *arXiv* **2018**, arXiv:1809.00888.
97. Fadl, S.M.; Han, Q.; Li, Q. Inter-frame forgery detection based on differential energy of residue. *IET Image Process.* **2019**, *13*, 52–528. [CrossRef]
98. Singh, G.; Singh, K. Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation. *Multimed. Tools Appl.* **2019**, *78*, 11527–11562. [CrossRef]
99. Joshi, V.; Jain, S. Tampering detection and localization in digital video using temporal difference between adjacent frames of actual and reconstructed video clip. *Int. J. Inf. Technol.* **2019**, *78*, 11527–11562. [CrossRef]
100. Bakas, J.; Naskar, R.; Dixit, R. Detection and localization of inter-frame video forgeries based on inconsistency in correlation distribution between Haralick coded frames. *Multimed. Tools Appl.* **2018**, *78*, 4905–4935. [CrossRef]
101. Sitara, K.; Mehtre, B. Differentiating synthetic and optical zooming for passive video forgery detection: An anti-forensic perspective. *Digit. Investig.* **2019**, *30*, 1–11. [CrossRef]
102. Hong, J.H.; Yang, Y.; Oh, B.T. Detection of frame deletion in HEVC-Coded video in the compressed domain. *Digit. Investig.* **2019**, *30*, 23–31. [CrossRef]
103. Aparicio-Díaz, E.; Cumplido, R.; Pérez Gort, M.L.; Feregrino-Uribe, C. Temporal Copy-Move Forgery Detection and Localization Using Block Correlation Matrix. *J. Intell. Fuzzy Syst.* **2019**, *36*, 5023–5035. [CrossRef]
104. Saddique, M.; Asghar, K.; Mehmood, T.; Hussain, M.; Habib, Z. Robust Video Content Authentication using Video Binary Pattern and Extreme Learning Machine. *IJACSA* **2019**, *10*, 264–269. [CrossRef]
105. Saddique, M.; Asghar, K.; Bajwa, U.I.; Hussain, M.; Habib, Z. Spatial Video Forgery Detection and Localization using Texture Analysis of Consecutive Frames. *Adv. Electr. Comput. Eng.* **2019**, *19*, 97–108. [CrossRef]
106. Zampoglou, M.; Markatopoulou, F.; Mercier, G.; Touska, D.; Apostolidis, E.; Papadopoulos, S.; Cozien, R.; Patras, I.; Mezaris, V.; Kompatsiaris, I. Detecting Tampered Videos with Multimedia Forensics and Deep Learning. In Proceedings of the International Conference on Multimedia Modeling, Thessaloniki, Greece, 8–11 January 2019.
107. Al-Sanjary, O.I.; Ahmed, A.A.; Ahmad, H.; Ali, M.A.; Mohammed, M.; Abdullah, M.I.; Ishak, Z.B. Deleting Object in Video Copy-Move Forgery Detection Based on Optical Flow Concept. In Proceedings of the IEEE Conference on Systems, Process and Control (ICSPC), Melaka, Malaysia, 13–14 December 2019.
108. Cozzolino Giovanni Poggi Luisa Verdoliva, D. Extracting camera-based fingerprints for video forensics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–20 June 2019.
109. Long, C.; Basharat, A.; Hoogs, A.; Singh, P.; Farid, H. A Coarse-to-fine Deep Convolutional Neural Network Framework for Frame Duplication Detection and Localization in Forged Videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–20 June 2019.

110. Saddique, M.; Asghar, K.; Bajwa, U.I.; Hussain, M.; Aboalsamh, H.A.; Habib, Z. Classification of Authentic and Tampered Video Using Motion Residual and Parasitic Layers. *IEEE Access* **2020**, *8*, 56782–56797. [CrossRef]
111. Fadl, S.; Megahed, A.; Han, Q.; Qiong, L. Frame duplication and shuffling forgery detection technique in surveillance videos based on temporal average and gray level co-occurrence matrix. *Multimed. Tools Appl.* **2020**, *79*, 1–25. [CrossRef]
112. Kharat, J.; Chougule, S. A passive blind forgery detection technique to identify frame duplication attack. *Multimed. Tools Appl.* **2020**, *79*, 8107–8123. [CrossRef]
113. Fayyaz, M.A.; Anjum, A.; Ziauddin, S.; Khan, A.; Sarfaraz, A. An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues. *Multimed. Tools Appl.* **2020**, *79*, 5767–5788. [CrossRef]
114. Kohli, A.; Gupta, A.; Singhal, D. CNN based localisation of forged region in object-based forgery for HD videos. *IET Image Process.* **2020**, *14*, 947–958. [CrossRef]
115. Wang, Y.; Hu, Y.; Liew, A.W.-C.; Li, C.-T. ENF Based Video Forgery Detection Algorithm. *Int. J. Digit. Crime Forensics (IJDCF)* **2020**, *12*, 131–156. [CrossRef]
116. Kaur, H.; Jindal, N. Deep Convolutional Neural Network for Graphics Forgery Detection in Video. *Wirel. Pers. Commun.* **2020**, *14*, 1763–1781. [CrossRef]
117. Huang, C.C.; Lee, C.E.; Thing, V.L. A Novel Video Forgery Detection Model Based on Triangular Polarity Feature Classification. *Int. J. Digit. Crime Forensics (IJDCF)* **2020**, *12*, 14–34. [CrossRef]
118. Fadl, S.; Han, Q.; Li, Q. CNN spatiotemporal features and fusion for surveillance video forgery detection. *Signal Process. Image Commun.* **2021**, *90*, 116066. [CrossRef]
119. Pu, H.; Huang, T.; Weng, B.; Ye, F.; Zhao, C. Overcome the Brightness and Jitter Noises in Video Inter-Frame Tampering Detection. *Sensors* **2021**, *21*, 3953. [CrossRef]
120. Shelke, N.A.; Kasana, S.S. Multiple forgeries identification in digital video based on correlation consistency between entropy coded frames. *Multimed. Syst.* **2021**, *34*, 1–14. [CrossRef]
121. Huang, Y.; Li, X.; Wang, W.; Jiang, T.; Zhang, Q. Towards Cross-Modal Forgery Detection and Localization on Live Surveillance Videos. *arXiv* **2021**, arXiv:1206.4660.
122. Bennett, E.P.; McMillan, L. Video enhancement using per-pixel virtual exposures. *ACM Trans. Graph.* **2005**, *24*, 845–852. [CrossRef]
123. Hernandez-Ardieta, J.L.; Gonzalez-Tablas, A.I.; De Fuentes, J.M.; Ramos, B. A taxonomy and survey of attacks on digital signatures. *Comput. Secur.* **2013**, *34*, 67–112. [CrossRef]
124. Chen, H.; Chen, Z.; Zeng, X.; Fan, W.; Xiong, Z. A novel reversible semi-fragile watermarking algorithm of MPEG-4 video for content authentication. In Proceedings of the Second International Symposium on Intelligent Information Technology Application, Shanghai, China, 20–22 December 2008.
125. Di Martino, F.; Sessa, S. Fragile watermarking tamper detection with images compressed by fuzzy transform. *Inf. Sci.* **2012**, *195*, 62–90. [CrossRef]
126. Hu, X.; Ni, J.; Pan, R. Detecting video forgery by estimating extrinsic camera parameters. In Proceedings of the International Workshop on Digital Watermarking, Tokyo, Japan, 7–10 October 2015.
127. Haralick, R.M.; Shanmugam, K.; Dinstein, I.H. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **1973**, *6*, 610–621. [CrossRef]
128. Aminu Mustapha, B. Passive Video Forgery Detection Using Frame Correlation Statistical Features/Aminu Mustapha Bagiwa. Ph.D. Thesis, University of Malaya, Kuala Lumpur, Malaysia, 2017.
129. Yu, J.; Srinath, M.D. An efficient method for scene cut detection. *Pattern Recognit. Lett.* **2001**, *22*, 1379–1391. [CrossRef]
130. Kancherla, K.; Mukkamala, S. Novel blind video forgery detection using markov models on motion residue. In *Intelligent Information and Database Systems*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 308–315.
131. Bondi, L.; Baroffio, L.; Güera, D.; Bestagini, P.; Delp, E.J.; Tubaro, S. First steps toward camera model identification with convolutional neural networks. *IEEE Signal Process. Lett.* **2016**, *24*, 259–263. [CrossRef]
132. Xu, G.; Wu, H.-Z.; Shi, Y.-Q. Structural design of convolutional neural networks for steganalysis. *IEEE Signal Process. Lett.* **2016**, *23*, 708–712. [CrossRef]
133. Bayar, B.; Stamm, M.C. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, Vigo Galicia, Spain, 20–22 June 2016.
134. Rao, Y.; Ni, J. A deep learning approach to detection of splicing and copy-move forgeries in images. In Proceedings of the IEEE International Workshop on Information Forensics and Security (WIFS), Abu Dhabi, United Arab Emirates, 4–7 December 2016.
135. Oh, S.; Hoogs, A.; Perera, A.; Cuntoor, N.; Chen, C.-C.; Lee, J.T.; Mukherjee, S.; Aggarwal, J.; Lee, H.; Davis, L. A large-scale benchmark dataset for event recognition in surveillance video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Washington, DC, USA, 20–25 June 2011.
136. Johnston, P.; Elyan, E.; Jayne, C. Video tampering localisation using features learned from authentic content. *Neural Comput. Appl.* **2020**, *32*, 12243–12257. [CrossRef]
137. Qadir, G.; Yahaya, S.; Ho, A.T. Surrey university library for forensic analysis (SULFA) of video content. In Proceedings of the IET Conference on Image Processing (IPR), London, UK, 3–4 July 2012.
138. Shullani, D.; Al Shaya, O.; Iuliani, M.; Fontani, M.; Piva, A. A dataset for forensic analysis of videos in the wild. In Proceedings of the International Tyrrhenian Workshop on Digital Communication, Palermo, Italy, 18–20 September 2017.

139. Panchal, H.D.; Shah, H.B. Video tampering dataset development in temporal domain for video forgery authentication. *Multimed. Tools Appl.* **2020**, *79*, 24553–24577. [CrossRef]
140. Ulutas, G.; Ustubioglu, B.; Ulutas, M.; Nabiyev, V.V. Frame duplication detection based on bow model. *Multimed. Syst.* **2018**, *24*, 549–567. [CrossRef]
141. Le, T.T.; Almansa, A.; Gousseau, Y.; Masnou, S. Motion-consistent video inpainting. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017.
142. Al-Sanjary, O.I.; Ahmed, A.A.; Sulong, G. Development of a video tampering dataset for forensic investigation. *Forensic Sci. Int.* **2016**, *266*, 565–572. [CrossRef]
143. Guo, Z.; Zhang, L.; Zhang, D. A completed modeling of local binary pattern operator for texture classification. *Image Process. IEEE Trans.* **2010**, *19*, 1657–1663.
144. Hussain, M.; Muhammad, G.; Saleh, S.Q.; Mirza, A.M.; Bebis, G. Image forgery detection using multi-resolution Weber local descriptors. In Proceedings of the IEEE International Conference on Computer as a Tool (EUROCON), Zagreb, Croatia, 1–4 July 2013.
145. Satpathy, A.; Jiang, X.; Eng, H.-L. LBP-based edge-texture features for object recognition. *IEEE Trans. Image Process.* **2014**, *23*, 1953–1964. [CrossRef]
146. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
147. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [CrossRef]
148. Muhammad, G.; Al-Hammadi, M.H.; Hussain, M.; Bebis, G. Image forgery detection using steerable pyramid transform and local binary pattern. *Mach. Vis. Appl.* **2014**, *25*, 985–995. [CrossRef]
149. Chen, J.; Kang, X.; Liu, Y.; Wang, Z.J. Median Filtering Forensics Based on Convolutional Neural Networks. *Signal Process. Lett. IEEE* **2015**, *22*, 1849–1853. [CrossRef]
150. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Montréal, QC, Canada, 3–6 December 2012.
151. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.-R.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Process. Mag. IEEE* **2012**, *29*, 82–97. [CrossRef]
152. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
153. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
154. Ma, J.; Sheridan, R.P.; Liaw, A.; Dahl, G.E.; Svetnik, V. Deep neural nets as a method for quantitative structure–activity relationships. *J. Chem. Inf. Model.* **2015**, *55*, 263–274. [CrossRef]
155. Helmstaedter, M.; Briggman, K.L.; Turaga, S.C.; Jain, V.; Seung, H.S.; Denk, W. Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **2013**, *500*, 168–174. [CrossRef]
156. Xiong, H.Y.; Alipanahi, B.; Lee, L.J.; Bretschneider, H.; Merico, D.; Yuen, R.K.; Hua, Y.; Gueroussov, S.; Najafabadi, H.S.; Hughes, T.R. The human splicing code reveals new insights into the genetic determinants of disease. *Science* **2015**, *347*, 1254806. [CrossRef]
157. Leung, M.K.; Xiong, H.Y.; Lee, L.J.; Frey, B.J. Deep learning of the tissue-regulated splicing code. *Bioinformatics* **2014**, *30*, i121–i129. [CrossRef]
158. Le Cun, B.B.; Denker, J.S.; Henderson, D.; Howard, R.E.; Hubbard, W.; Jackel, L.D. Handwritten digit recognition with a back-propagation network. In Proceedings of the Advances in Neural Information Processing Systems, Lakewood, CO, USA, 26–29 June 1990.
159. Peng, D.; Xu, Y.; Wang, Y.; Geng, Z.; Zhu, Q. Soft-sensing in complex chemical process based on a sample clustering extreme learning machine model. *IFAC-PapersOnLine* **2015**, *48*, 801–806. [CrossRef]
160. Peng, Y.; Wang, S.; Long, X.; Lu, B.-L. Discriminative graph regularized extreme learning machine and its application to face recognition. *Neurocomputing* **2015**, *149*, 340–353. [CrossRef]
161. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
162. Zhang, Z. Detect Forgery Video by Performing Transfer Learning on Deep Neural Network. Ph.D. Thesis, Sam Houston State University, Huntsville, TX, USA, 2019.
163. Duan, L.; Xu, D.; Tsang, I. Learning with augmented features for heterogeneous domain adaptation. *arXiv* **2012**, arXiv:1206.4660.
164. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* **2013**, *36*, 537–556. [CrossRef]
165. Chen, L.; Duan, L.; Xu, D. Event recognition in videos by learning from heterogeneous web sources. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 23–28 June 2013.
166. Nam, J.; Kim, S. Heterogeneous defect prediction. In Proceedings of the 10th Joint Meeting on Foundations of Software Engineering, Bergamo, Italy, 30 August–4 September 2015.
167. Prettenhofer, P.; Stein, B. Cross-language text classification using structural correspondence learning. In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Uppsala, Sweden, 11–16 July 2010.