Multimedia Systems (2018) 24:211–240
DOI 10.1007/s00530-017-0538-9

REGULAR PAPER

Video content authentication techniques: a comprehensive survey

Raahat Devender Singh[1] · **Naveen Aggarwal**[1]

Received: 14 July 2016 / Accepted: 19 January 2017 / Published online: 17 February 2017

Abstract In this digital day and age, we are becoming increasingly dependent on multimedia content, especially digital images and videos, to provide a reliable proof of occurrence of events. However, the availability of several sophisticated yet easy-to-use content editing software has led to great concern regarding the trustworthiness of such content. Consequently, over the past few years, visual media forensics has emerged as an indispensable research field, which basically deals with development of tools and techniques that help determine whether or not the digital content under consideration is authentic, i.e., an actual, unaltered representation of reality. Over the last two decades, this research field has demonstrated tremendous growth and innovation. This paper presents a comprehensive and scrutinizing bibliography addressing the published literature in the field of passive-blind video content authentication, with primary focus on forgery/tamper detection, video re-capture and phylogeny detection, and video anti-forensics and counter anti-forensics. Moreover, the paper intimately analyzes the research gaps found in the literature, provides worthy insight into the areas, where the contemporary research is lacking, and suggests certain courses of action that could assist developers and future researchers explore new avenues in the domain of video forensics. Our objective is to provide an overview suitable for both the researchers and practitioners already working in the field of digital video forensics, and for those researchers and general enthusiasts who are new to this field and are not yet completely equipped to assimilate the detailed and complicated technical aspects of video forensics.

Keywords Video forgery detection · Video tamper detection · Passive-blind video forensics · Video anti-forensics · Video content authentication · Video phylogeny detection · Video re-capture detection · Video up-conversion detection

1 Introduction

The wide-spread proliferation of inexpensive and portable video-capture devices, such as digital cameras and cell phones, combined with the remarkable surge in the use of surveillance cameras, has caused a sudden increase in the amount of digital audio-visual data being generated every single day. All this data not only fulfils a recreational purpose but also serves as a record of events occurring in every corner of the world. The information provided by the contents of digital images and videos forms the basis of several crucial and consequential decisions in the fields of criminal or forensic investigations, intelligence services, politics, and journalism. For instance, during a criminal trial, surveillance footage of the crime, if available, can be admitted as 'video evidence', and its contents are expected to provide a truthful depiction of the event.

A decade ago, digital videos would have been considered infallible, but the wide-spread availability of low cost and easy to use the video-editing software, such as Adobe Premiere, Photoshop, Cinelerra, and Lightworks, and development of specialized forgery techniques [1–7], has led to the realization that this is no longer the case. Even novice individuals are now capable of altering the contents

Communicated by S. Kopf.

✉ Raahat Devender Singh
raahat.singh@hotmail.com

[1] University Institute of Engineering and Technology, Panjab University, Chandigarh, India

Fig. 1 Examples of copy–paste forgeries. **a** and **c** Represent sample frames from original videos [8], and **b** and **d** are their forged versions wherein certain objects have been removed

of digital videos in a manner that renders them practically indistinguishable from genuine content.

There are several different kinds of video forgeries[1], but they all usually belong to one of two categories: inter-frame forgeries or intra-frame forgeries.

(a) *Inter-frame forgeries* These are the kinds of forgeries that affect the sequence of frames in a video in some way. Usually, such forgeries involve removal or insertion of a set of frames from or into a video sequence. Frame replication or duplication is also a kind of inter-frame forgery, where a set of frames are copied and inserted into the same video at another temporal location. Such forgeries can also be referred to as 'inter-frame copy–paste forgeries'.

Another kind of inter-frame forgery is temporal splicing, where frames of two or more different videos are interpolated to generate a new video.

(b) *Intra-frame forgeries* In an intra-frame forgery, the actual contents of individual frames are modified. Copy–paste and upscale-crop are examples of intra-frame forgeries.

1. *Copy–paste forgeries (aka partial manipulation)* In a copy–paste (or copy–move) forgery, an attacker might add or remove an object to or from a scene represented in the video frames. The term 'partial' here basically means that only a small region of the frame undergoes manipulation and the rest of the frame remains untouched.

When an object is removed from a video scene, a technique called inpainting is used to restore the missing or tainted regions in a visually plausible manner. Inpainting can be performed in one of two ways. Either the most coherent blocks from temporally adjacent frames are used to fill in the missing region [Temporal Copy and Paste (TCP)], or the missing regions are filled in with the help of sample textures [Exemplar-Based Texture Synthesis (ETS)]. ETS is generally less suitable for inpainting a video, because it treats each frame individually and that makes it difficult to maintain temporal coherence between successive frames after inpainting. TCP is much more equipped to preserve this temporal coherence.

Green-screening or blue-screen compositing are also examples of intra-frame forgeries.

2. *Upscale-crop* Such forgeries entail cropping the frames of a video to eliminate evidence of occurrence of a crime in the outermost parts of said video, and then enlarging the affected frames so as to maintain consistent resolution across the entire video.

Figures 1, 2, 3 and 4 present some examples to better illustrate the different kinds of video forgeries.

All these examples are demonstrative of the fact that plausible and carefully constructed video forgeries may remain completely inconspicuous to a human viewer. Therefore, in matters where a video constitutes potential evidence, it is vital to ascertain that its contents are in fact an actual and unaltered representation of reality, and since subjective inspection cannot provide adequate assurance, specialized forensic techniques have to be relied upon. These specialized solutions are offered by the research domain known as digital visual media forensics. Basically, digital visual media forensics is concerned with the accomplishment of three main tasks [11], as illustrated in Fig. 5.

To handle the challenge of digital content authentication, the domain of visual media forensics provides a set a tools and techniques which are collectively known as

[1] Technically, a 'forgery' refers to something that is falsely made with the intent to deceive whereas 'tampering' refers to the intentional modification of structure or composition of something that would render it harmful. Being subtly different, in this survey, as in the literature, these terms would be used synonymously.

Fig. 2 Examples of upscale-crop forgeries. **a** and **c** Original video frames [8], which are cropped and resized to eliminate objects at the extremities of the frames, thereby producing the forged frames (**b**) and (**d**), respectively

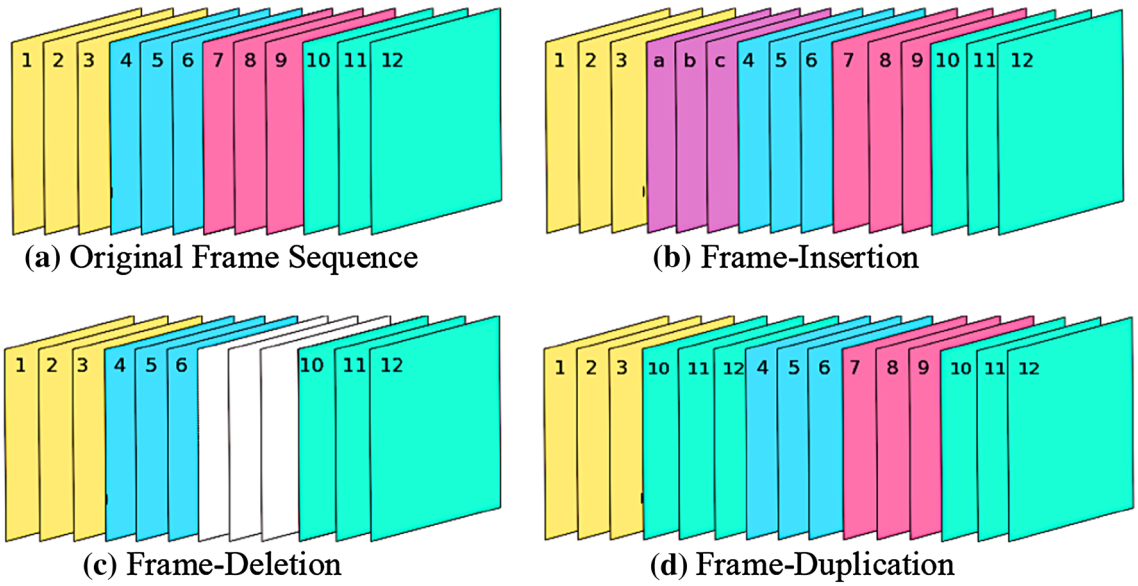

Fig. 3 Inter-frame forgeries. **a** Original frame sequence. **b** Example of frame-insertion forgery, where frames **a, b**, and **c** have been inserted into the video. **c** Example of frame-deletion forgery, where frames 7, 8, and 9 (denoted in *white*) have been removed from the video. **d** Example of frame-duplication forgery, where frames 10, 11, and 12 have been duplicated at another location in the video

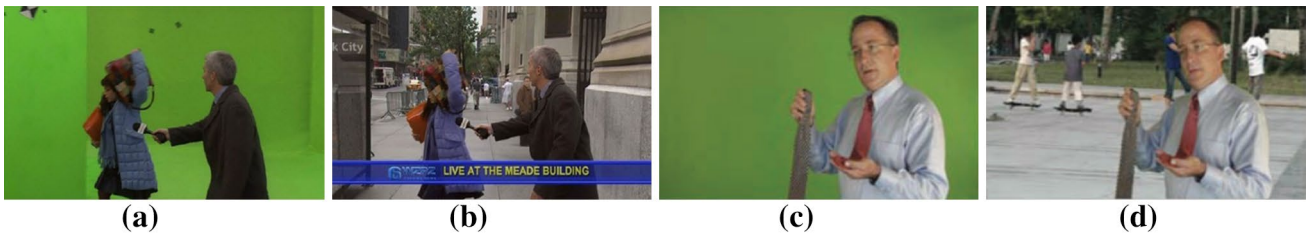

Fig. 4 Examples of green screening **a** and **c** represent sample frames of videos with people engaged in various activities in front of a green screen. **b** and **d** Frames from a new video where these people have been composited into their new surroundings [9, 10]

tamper or forgery detection techniques. These techniques are based on the premise that all post-production content modification operations inevitably disturb the inherent statistical properties of the digital content and leave certain discernible traces (sometimes known as 'forensic fingerprints', 'footprints' or 'forensic artifacts') in the data, which, upon suitable analysis, provide an insight into the possible modifications that the content may have gone through.

The existing forgery detection methods can be categorized in a number of different ways. All the commonly utilized classifications have been listed in Table 1.

For two decades now, visual media forgery detection field has remained at the receiving end of much innovation.

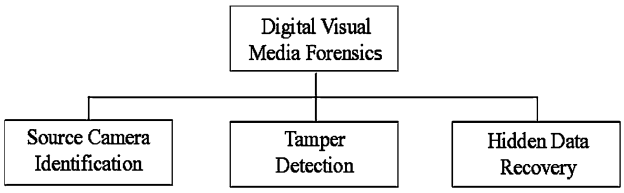

Fig. 5 Objectives of digital visual media forensics

Over the years, a multitude of image forensic techniques have been developed, and have subsequently been analyzed and catalogued in various surveys [12–27]. However, only a few papers [11, 28–31] have appraised the innovations in the field of video forensics.

Although these papers possess some fine qualities of their own, they remain inadequate in some aspects. While in [11], only two video forgery detection techniques have been analyzed (the rest of the paper focuses on image tamper detection) [28–30] and appraise only a few video forensic techniques. Several noteworthy and contemporary contributions have not been discussed or analyzed, and since an effective survey is expected to be a thorough repository of every relevant and notable piece of research and innovation in any given field, if some important video tamper detection techniques remain un-cited, they might eventually remain unidentified. Furthermore, in the absence of such crucial references, it would become difficult to completely understand the current state of affairs of the digital video forensics domain. A recent survey [31] reviews passive video forgery detection techniques and offers quite an informative study of the basics of this field. However, this work too overlooked several important works of merit. The techniques discussed in the survey were found to have been described in a somewhat restricted manner and the nature of test data used by the respective authors, or the qualitative or quantitative performances of the proposed techniques had not been discussed. The absence of such information may negatively affect the extent of comprehension of the subject matter thereof. The study in [31] lacked rigorous analysis; it also lacked thorough exposition of both the video anti-forensics domain and the open issues that need to be tackled in the near future. A noteworthy paper [32] provides an overview of the advancements in the field of information forensics, wherein the authors have examined the subjects of device forensics, embedded fingerprinting, and watermarking, and environmental signatures, such as electric network frequency (ENF), have also analyzed various social factors and behavior dynamics pertaining to the field of forensics. They have provided a very brief outline of the tamper detection and anti-forensic domains, but the investigation was almost entirely limited to digital images, with only a cursory discussion of few forensic features used in the video forgery detection domain.

To overcome the aforementioned impediments, this survey presents a comprehensive and up-to-date assemblage of the most influential contributions in the field of passive-blind video tamper detection and anti-forensics, including several other cognate research fields, such as video up-conversion detection, re-capture detection, and video phylogeny. In this survey, we classify the forensic features used for forgery and video re-capture detection, intimately analyze the available literature, and highlight the favorable features and shortcomings of each of the discussed techniques. Furthermore, we discuss certain open issues that require immediate attention, along with some other long-term goals. Through this discussion, we seek to develop a perspective on the selected domain, which could prove useful to the researches and practitioners working in the field of video forensics to find new utilitarian ideas and to help the video processing community identify novel research challenges. A vital objective of this survey is to provide an overview suitable for both the researchers and practitioners working in the field of digital video forensics and for those researchers and general enthusiasts who are new to this field and are not yet completely equipped to assimilate the detailed and complicated technical aspects of video forensics. To achieve this, it was essential to provide a clear

Table 1 Possible classifications of video tamper detection methods

Classifications	Underlying methodology
Active techniques	These use a known identifying trace such as a signature or watermark that is either embedded into the content at the time of recording or is sent with it to the receiver (also known as intrusive or non-blind techniques)
Passive techniques	These use only the received content to determine its authenticity without the help of any other kind of side-information (also known as non-intrusive or blind techniques)
Intra-frame techniques	Analysis is performed by considering one frame at a time
Inter-frame techniques	Relationships between adjacent frames are considered to detect the forgery
Detection techniques	The presence of tampering is detected but its exact location, either within the frame (spatial localization) or within the video sequence (temporal localization) is not determined
Localization techniques	In addition to detection, tampering is localized as well

and simplistic understanding of the subject matter, which is why in-depth and complicated technical description of any particular method has been avoided in this survey.

Analysis of the literature revealed that due to lack of standardized video forensics databases, researchers had tested their forgery detection technique on videos from different databases, the most frequently used being [8, 33]. The remaining techniques were validated on videos captured and forged by the respective authors themselves. It is important to realize that the nature of the scene being captured and processed can have a significant effect on the outcomes of the various forensic methods. All the available forgery detection techniques are diverse in functionality and have been designed to tackle different forensic challenges, and since they have not been tested on a standardized neutral platform, it becomes difficult to judge their efficaciousness from a universal standpoint. Albeit, analysis of the outcomes of a particular technique with respect to the visual contents and underlying characteristics of test videos used during performance validation can provide valuable information regarding the applicability of that techniques in a specific forensic scenario. Such an analysis, however, would be unable to predict the technique's behavior in an entirely different forensic scenario. Therefore, to provide an idea of how these different forensic solutions would behave in a neutral setting, we have assessed the performances of some forgery detection techniques which we believe to be among the most representative and contemporary advancements documented in the literature. We implemented each of these techniques by adhering to the specifications, assumptions, variable parameters settings, and threshold values suggested by the respective authors, and validated them on test videos form two databases [8, 34]. SULFA [8] provides a large number of original videos and five forged videos, exhibiting intra-frame copy–paste forgeries. The database in [34] was constructed by us, by performing intra-frame copy–paste and inter-frame forgeries on original videos taken from [8]. Figures 1c, d and 2c, d represent some examples of forged content available in [34].

The rest of the paper is organized as follows. Section 2 serves as a compendium of the advancements in the field of video forgery detection, wherein the existing video forgery detection literature has been thoroughly examined. The enumerated forgery detection techniques have been categorized on the basis of the type of forgeries they detect, i.e., inter-frame forgeries and intra-frame forgeries. Then, we discuss various video anti-forensic strategies, and provide an overview of the fields of video up-conversion, phylogeny, and re-capture detection. The best possible fit of the presented references into these categories has been attempted. The quantitative results of our comparative analysis have also been provided in Sect. 2. Section 3 is geared towards the exposition of the various issues and impediments pertaining to the video forensics domain, and the survey is concluded in Sect. 4.

2 Techniques proposed for passive-blind content authentication in videos

Passive-blind techniques use only the received content to determine its authenticity, and work in the complete absence of any kind of identifying trace or embedded information, such as watermarks or signatures. The task of forensic analysis begins with the reconstruction of the processing history of the data under investigation. The first step of this process is the extraction of certain descriptive features from the given video that provide useful hints regarding the authenticity of its contents. Over the years, various researchers have utilized many different kinds of such descriptive features to accomplish the task of forgery detection. These have been presented in Fig. 6.

The evidence of content modification manifests as specific artifacts in one or more of these features. These artifacts are not only discernible but are also unique to each content alteration operation. Therefore, the presence of a particular forensic artifact serves as an evidence of presence of the corresponding post-production content manipulation operation.

In the upcoming sections, we examine various passive-blind content authentication techniques that have been proposed in the literature so far. The forgery detection techniques discussed hereinafter have been organized according to the outline presented in Fig. 7.

2.1 Inter-frame forgery detection techniques

In this section, we present an analysis of the inter-frame forgery detection techniques proposed in the literature. First, we discuss the various frame-insertion, frame-deletion, and frame-duplication/replication detection techniques, followed by the methods suggested for the detection of temporal interpolation.

2.1.1 Detection of frame-insertion/deletion/duplication

2.1.1.1 Use of sensor artifacts for inter-frame forgery detection Recording devices usually leave specific detectable traces in the recorded video. In the past, such artifacts were mostly utilized for the purposes of source camera identification [35–37], but some authors exploit them as a means of tamper detection as well. For this, the basic idea is to determine whether or not all the scenes of the video were recorded using the same camera.

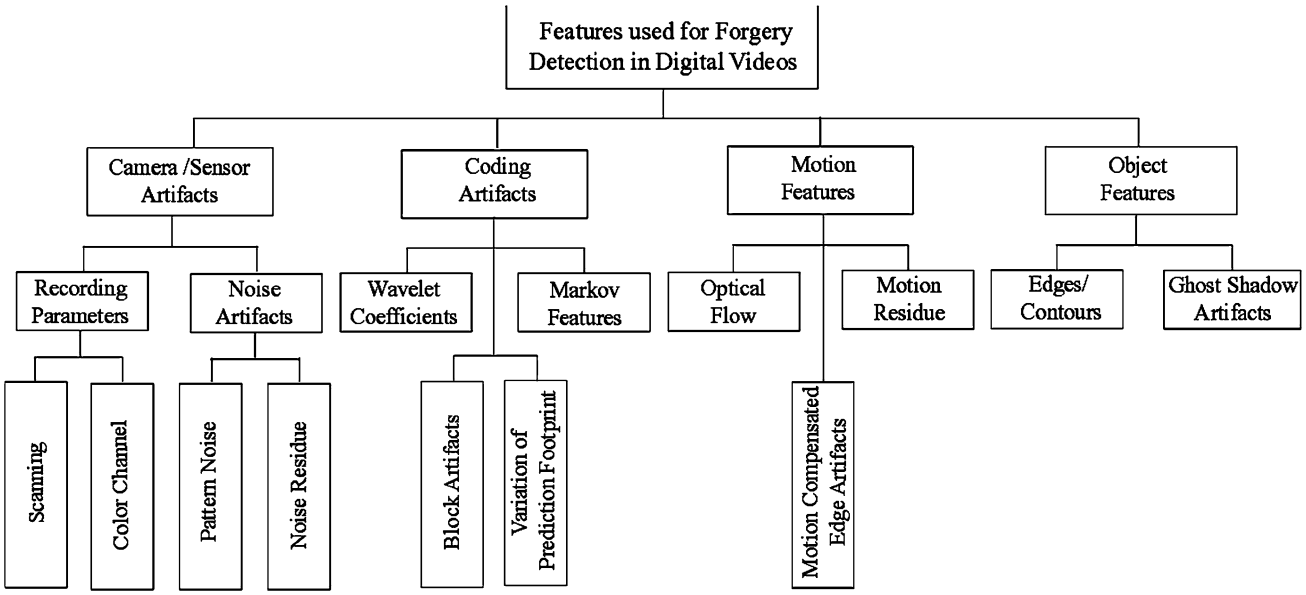

Fig. 6 Different features used for detecting forgeries in digital videos

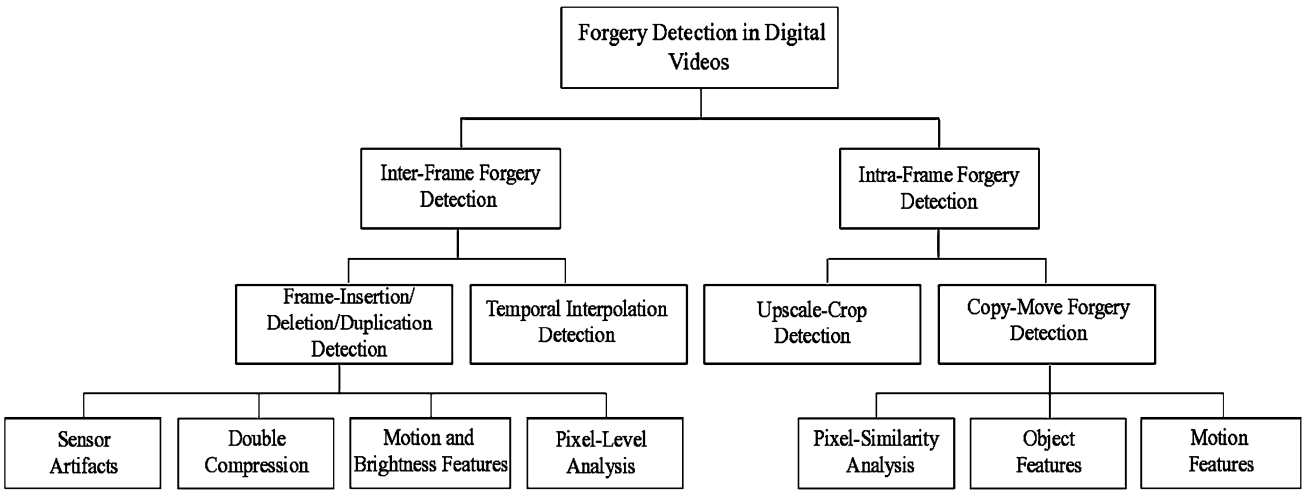

Fig. 7 Categorization of the video forgery detection techniques discussed in the survey

A CCD digital camera introduces readout noise in every frame it records upon readout. Since this noise follows a particular pattern over a sequence of consecutive frames, the authors in [38] suggested that any variation in this pattern could be helpful in detecting forgeries. To detect frame insertion, the authors computed the variance between mean noise (for all the frames in the video) and noise in a particular frame. Frames with high variance from mean value were tagged as not belonging to the original video. To detect object insertion in a frame, correlation patterns between noise in a frame-sub-block and overall noise of the frame were observed, and any inconsistency therein hinted towards the possibility of forged regions. The authors did not provide any information regarding the performance of this technique. However, thorough analysis of its functionality revealed certain drawbacks. First, its method of operation suggested that it would most likely be unable to perform well for compressed videos. Second, it relied on a number of hard thresholds that were determined empirically and, therefore, lacked flexibility. In addition, it was tested on only one self-recorded video sequence, which was not enough to provide an accurate estimation of the scope of applicability of this technique.

The authors in [39] suggested using the artifacts caused by components of imaging pipeline (specifically interlaced scanning) and proposed a motion adaptive algorithm that was capable of detecting and localizing forgeries in interlaced and de-interlaced videos in both spatial and temporal domains. The basis of the algorithm was the detection of disturbances in correlation (for de-interlaced videos) and disturbances in motion between frames (for interlaced videos), which occurred when some sort of tampering was involved. For a de-interlaced un-compressed video captured using a digital video camera, 100% detection accuracy was achieved. However, when this video underwent additional MPEG compression, detection accuracy was 97%, 96.1%, and 93.3% at 9, 6, and 3 Mbps, respectively. No clear results were reported for the three interlaced videos used for testing. Compression artifacts seemed to make it difficult for the algorithm to estimate the de-interlacing correlations, thereby rendering this approach unsuitable for low-quality videos. Noise too was found to be a debilitating factor for this technique. Insufficient validation of the proposed method's effectiveness was another major limitation of this work.

In [40], a fingerprinting-based approach was suggested which utilized the concept of Sensor Pattern Noise (SPN) to determine if all the frames of the given test video were recorded with the same camera, based on sign modifications seen in the number (or content) of the frames. The unique noise patterns of the camcorder used to capture the video were estimated for the initial frames of the video sequence, and were used to identify different types of forgeries. The correlation between SPN of the frame under investigation and a reference SPN was calculated and compared with an empirical threshold to detect frame insertion and replication, and object insertion within a frame. Different tests were carried out on both un-compressed and MPEG compressed videos. Instead of quantitative results, the authors reported some case studies after implementing their algorithm. The results indicated that the system was reliable for un-compressed videos, but algorithm's performance deteriorated significantly in case of compressed videos. The method was claimed to be reliable for videos, whose frames had undergone interpolation but was not found to be very accurate if the SPN had high-frequency components.

Another camera-based approach was proposed in [41] that provided pixel-level authentication of digital videos, and was also capable of localizing any and all inter-frame forgeries. Suspicious regions in a given video (frames recorded using a different camera and inserted in a desired location in the given video) recorded in any scan format were detected by utilizing the inconsistencies in photon shot noise introduced by different acquisition devices. When the method was tested on two un-compressed videos recorded indoors and outdoors, the results were impressive: 97% of forged pixels were located with a false alarm rate of 2.5%. The system's performance suffered severe degradation when two compressed videos were experimented upon. The detection rate dropped to 46% for MPEG-2, 6% for Cinepak, and 39% for H.264 codecs. The reason for performance affliction was evident: since modern codecs make good use of motion information, they effectively remove most of the noise characteristics from a video. Considering the fact that this method was based entirely on the utilization of noise characteristics, its performance dropped as the amount of noise in the test video decreased. The method was applicable only to static-scene videos, such as those recorded by stationary surveillance cameras. In addition, since the method was tested on a very small number of test sequences, its wide-scale utility cannot be determined with certitude.

2.1.1.2 Detection of forgery via detection of double compression The sheer volume of digital data being generated everyday necessitates that this data, videos especially, be compressed before they are stored or transmitted. To modify the content of a compressed video, the forger would first have to convert it into a sequence of frames, perform the forgery (say frame-deletion), and then re-encode this modified sequence. Re-compression or double compression is, therefore, an inevitable consequence of the forgery, and its detection could help detect the presence of the tampering operation.

The first successful probing step in this direction can be attributed to the authors in [42], who worked with MPEG-1, 2 compressed videos. Their algorithm was based on the simple assumption that when an MPEG video was tampered with, it suffered two compressions: first, when the video was being created and, second, when it was re-saved after said tampering. The authors investigated a spatial and a temporal phenomenon. The first one was based on the fact that when a video is re-compressed after removing some of its frames, its Group of Pictures (GOP) patterns get desynchronized (GOP refers to the specific arrangement of I-, B-, and P-frames in a compressed video). This desynchronization causes disturbances in the Discrete Cosine Transform (DCT) coefficient distribution, which introduced easily detectable periodic artifacts in the histograms of double quantized frames. They also exploited the fact that within a GOP, the frames exhibit great correlation and adding or deleting a frame in a GOP increases the error of motion estimation, which also results in detectable periodic spikes. The authors stated that this approach was quite effective in detecting if a video was re-saved after it was recorded. The technique was tested on one self-recorded video, but no quantitative or qualitative results were reported, although it was evident that the presence of noise in the given test

content caused significant degradation in performance. In-depth analysis revealed that the periodic peaks in motion estimation error do not show if the entire GOP or a multiple of GOP is deleted, in which case, the method would be unable to detect any forgery. Furthermore, this method could not determine whether a video was only re-saved (after a harmless viewing, for instance) or if it was re-saved after some sort of tampering. Finally, this algorithm implicitly assumed that the videos were coded via Variable Bit Rate (VBR) coding model [a VBR model sets a fixed Quantization Parameter (QP) for each frame]. The method was thus rendered inapplicable to videos recorded using the Constant Bit Rate (CBR) coding model.

Survey of the literature revealed that another course of action for detecting double compression was the utilization of the first digit law, aka Benford's law [43]. This law was used to develop many double-compression detection approaches for digital images, and usually generated an accuracy of over 90%. When applied to videos, the efficiency of the detector reduced to about 70% [29].

The authors in [44] observed that MPEG compression introduced certain block artifacts in different frames of a video, which persisted even after the video underwent post frame-deletion re-compression. Therefore, they proposed to detect signs of frame deletion by analyzing the changes in temporal patterns of block artifacts (i.e., the degree of change in the Block-Artifact Strength (BAS) of the re-compressed video), which was a function of number of deleted frames and previous GOP structure of the video. 11 forged sequences created from an MPEG-2 test sequence were used for experimentation. The authors claimed that this technique was less sensitive to video content and could discern GOP transformation as well, but no empirical results were reported. Moreover, this technique was applicable to MPEG-2 videos with a fixed GOP structure only. The concept of BAS also presented some constraints. Usually, the effects of change in BAS are discernible as long as the second compression is weaker than the first one. It was only natural that as the strength of the second compression increased, the performance of the system would rapidly deteriorate.

In [45], the authors presented a technique for detecting double quantization, which resulted from either re-compressing an MPEG compressed video or combining videos of dissimilar characteristics. The technique, unlike [42], could detect highly localized tampering by determining if the DCT coefficients of the video frames underwent double compression at any point. Experiments were performed on two sets of digital videos (total 498 sequences) captured using a digital video recorder and downloaded off of the Internet from Microsoft's showcase. The empirical results indicated that the detection rate was highly dependent on the ratio of first and second quantization scales. When the ratio was less than 1.3, True Positive Rate (TPR) was 2.5%. TPR was 41.2% when the ratio was between 1.3 and 1.7, and for ratio higher than 1.7, an average TPR of 99.4% was reported. Evidently, the technique was effective as long as the second compression quality factor was higher than the first one. This requirement limited the effective scope of applicability of this method. In addition, this technique could only work when the quantization scale was kept constant throughout the test sequence.

In [46], forgeries in MPEG-2 encoded videos were detected by examining DCT coefficient distribution. This algorithm was based on the observation that histogram of quantized DCT coefficients of a video suffering from double compression exhibited a convex pattern. On a set of 100 self-captured videos for a bitrate range of 4–8 Mbps, the algorithm generated TN rate in the range of 94.02–93.96% (for originally compressed videos) and TP rate in the range of 86–100% (for double-compressed videos). Unlike [45], which was highly dependent on the quantization scales, the authors in this case suggested controlling the output bitrate, which made this algorithm adaptable to the needs of different kinds of video encoding systems. However, unlike [45], this method could not localize the forgery in the video. It also could not perform well for slow-motion videos.

The work in [47] also focused on detecting frame-based tampering by detecting double compression in MPEG-2 videos. Double-compression detection techniques that relied on temporal artifacts [42, 44] were extremely vulnerable to the influence of encoder parameters. Therefore, instead of basing the frame-addition/deletion detection process on temporal features, the authors suggested using frequency features. It was observed that when an MPEG video was re-compressed after frame addition/removal, some high-frequency components of the re-compressed frames were lost because of the desynchronization of GOPs and the non-linear quantization performed in the coding process. These variations created clearly observable periodic patterns in the energy values of DCT coefficients, which not only helped detect the forgery but locate it as well. A set of 20 test videos taken from Video Quality Experts Group [48] and 30 high-quality DVD videos were used for performance evaluation, but no quantitative results were reported that could help estimate the accuracy or precision of the method. In addition, since the entire method was based on the detection of coding-type change, any forgery that did not introduce such change in the video remained undetected. For instance, this method would not detect removal of the entire GOP or its multiples. Furthermore, the final decision regarding the presence of absence of forgeries was based on a hard threshold, which the authors determined empirically. The method, therefore, had restricted practical applicability.

The methods proposed in [44, 45] suffered significant performance deterioration as the strength of the second compression increased. With the aim of overcoming this limitation, the authors in [49] proposed a robust and distinctive footprint called Variation of Prediction Footprint (VPF). The work in [49] was not directed towards actual forgery detection but suggested certain solutions that could be useful in the field of video forensics. The authors claimed that VPF could not only help detect double compression but also determine the original GOP structure of the video sequence. They believed that estimating the original size of the GOP and assessing the video's processing history was an important step towards forensic analysis of the said video. The technique was tested on a total of 1344 test samples constructed from 14 sequences. For H.264 encoded videos, this system yielded 94% detection rate with 5% false positives. For MPEG-x encoded videos, the detection rate was 80% when 5% false positives were allowed. Although this technique was not exactly built for detecting forgeries, it still laid a solid foundation upon which forgery detection mechanisms could be built. However, before attempting to utilize VPF for forgery detection, it is necessary to understand that this technique would not be able to detect double compression if frames are removed from the beginning of the video, and if double compression is not detected, the forgery could pass undetected as well. In addition, in its present form, this technique works only for videos with a fixed GOP structure.

Another MPEG double-compression-based forgery detection technique was proposed in [50], where abnormalities in DCT coefficient patterns were treated as an indicative of frame insertion/deletion. The authors extracted a 12D feature from the GOPs, which was then used by a serial Support Vector Machine (SVM) to determine original bitrate of the given double-compressed video. 12 MPEG encoded YUV sequences were used for testing and the results indicated 97.9% TNR and 100% TPR. The original bitrate was estimated with 95.83% accuracy. The authors observed that the technique's detection performance was relatively poorer for videos with lower bitrate, because larger quantization scale required a coarser quantization process, which the technique was not equipped to handle.

In the same year, another similar technique was proposed in [51]. The novelty of this technique was its ability to detect transcoded videos, i.e., videos that had been doubly compressed using two different compression standards. The periodicity introduced in the reconstructed non-zero DCT coefficients after MPEG-2 compression has been a thoroughly utilized forensic artifact. The authors further observed that after an MPEG-2 video was transformed to MPEG-4 video, the previous MEPG-2 compression traces generated new periodicities that were distinctly observable in the histograms of the reconstructed DCT coefficients. An SVM was used to train and test the classifier that could label each video as being originally MPEG compressed or being transcoded. The algorithm was tested on videos in YUV format taken from VQEG [48]. The authors presented the results in the form of Receiver Operating Characteristics (ROC) curves and stated that perfect results were obtained in case of low bitrates. These curves also demonstrated that as the target output bitrate increased, the detection performance declined. The reason can be attributed to the fact that increased bitrates led to smaller MPEG-2 quantization scale factors, which in turn made the periodicities in histograms weaker and, therefore, harder to detect. In addition to not being able to locate the forgery, this technique worked for MPEG-2 to MPEG-4 transcoding only. It also assumed that transcoding was always suggestive of tampering.

The work in [52] was directed towards detecting multiple compressions (up to three) in a video sequence. Multiple SVMs were trained by exploiting Benford's law [43], and then tested on the statistics of quantized DCT coefficients to retrace the compression steps. The system was tested on 12 H.264/AVC encoded videos and the detection accuracy varied with respect to the number of compressions undergone by the video. The technique could detect a single compression with 100% accuracy, two compressions with 73.9% accuracy, and the third one with accuracy of 77.8%. The shortcomings of this technique were the assumptions that were imposed on its functionality. First, the QPs used during compression were not entirely random but instead uniformly distributed over a particular interval. Second, the QPs between consecutive compressions had to differ by at least two units, otherwise, the technique would not be able to detect the compression. The authors also conceded that this technique was entirely ineffective against any anti-forensic approach directed towards deliberately hiding the traces of compression. The work itself was pioneering in nature since none of the previous works in the literature attempted detection of more than two compressions in a given video sequence. However, further innovation is required to make this technique widely applicable, and that too with improved accuracy of re-compression detection.

A technique for detecting frame deletion with the help of machine learning was proposed in [53]. The first step was feature vector extraction, where the process of feature selection was based on several important observations. Since frame-deletion affects prediction residuals and percentage of intra/inter-coded MBs, mean and standard deviation of these residuals and percentages were used as suitable features. The fact that frame-deletion causes degradation of video quality led to Peak Signal-to-Noise Ratio (PSNR) being used as another feature. Finally, the author

observed that in videos encoded using CBR model, frame deletion manifested in terms of increased quantization scales. The extracted feature vector first underwent dimensionality reduction with the help of spectral regression, followed by use of logistic regression (L Reg), K-Nearest Neighbor (KNN), and SVM to detect forgeries. Performance evaluation was performed on 36 MPEG-2 encoded QCIF sequences. Average TPR using SVM, L Reg, and KNN was 94.3%, 94.8%, and 95.6%, respectively, while the False Alarm Rate (FAR) ranged from 3.8% to 4.4%. The technique worked well regardless of the number of deleted frames. Unlike [42], the methodology in [53] could successfully distinguish re-encoded videos with and without frame deletion, with average TPR of 96% (SVM), 95.1% (L Reg), and 92.8% (KNN). Unlike [39], it could handle videos coded using both VBR and CBR models. However, despite its effectiveness, it could not localize the forgery. Furthermore, the technique worked only if the number of deleted frames was not a multiple of the length of a GOP.

The methodology in [54] dealt with detection of double compression in MPEG-4 encoded videos, using Markov statistics as the distinguishing features. The basic idea was to treat every GOP as a detection unit, followed by extracting Markov features and classifying each GOP as singly or doubly compressed using Fisher's Linear Discriminant (FLD) analysis [55]. For experimentation, a data set containing 5040 sample clips constructed from 30 YUV sequences in CIF resolution was used. For different values of the first and second quantization scales (Q1 and Q2), different classification results were obtained. When Q2 was an odd multiple of Q1, classification accuracy ranged from 50.1–52.4%. When Q2 was an even multiple of Q1, this range was 92.7–99.3%. The authors also compared Markov features with other features commonly used to detect double compression, such as First Digit Distribution (FDD) feature, as used in [56], and DCT histogram features, as used in [45, 50]. These comparisons indicated that whenever the two quantization scales differed by only a small value, the first-order statistics failed to detect forgeries, while the second-order statistics (Markov features) possessed strong discriminative power. For instance, when Q2 was an even multiple of Q1, while the classification accuracy using FDD ranged from 53.5–54.1%, this range was 92.7–99.3% for Markov statistics. This contribution was of considerable significance, given the fact that nowadays a large proportion of digital and surveillance cameras support the MPEG-4 codec. However, the detection accuracy of this technique was found to be extremely dependent on the ratio of Q1 and Q2.

Along the lines of the system proposed in [51], another video forensic technique capable of detecting frame addition/deletion in transcoded videos was proposed in [57]. The authors adopted the scheme proposed in [49] and modified it to detect any misalignments in the frame structure of a test video as a result of frame deletion/addition. This method could detect double encoding even if a set of leading frames were deleted. This method had an additional advantage of being able to effectively locate the forgery. The method was suitable for H.264 encoded videos too, unlike [51] which worked for MPEG videos only. The modified methodology was also able to estimate the number of deleted frames. 14 YUV videos in CIF resolution were used for constructing a test set containing 10,206 test clips for evaluating the performance of the proposed method. MPEG-2, 4 & H.264 encoders were used to compress the videos. While the average frame-removal detection accuracy was 84%, for frame-addition detection, this value was 79%. The only other VPF based scheme proposed in the literature so far, [58], generated better detection results than this scheme, but only as long as the two quantization scales were carefully monitored and controlled. In addition, while [58] was applicable to MPEG-2 encoded video only, [57] was suitable for MPEG-2, 4 and H.264 encoded videos. However, even though [57] was a viable method with a wide range of applicability, certain limitations need to be pointed out here. First, applicability of this method was limited to videos compressed using CBR model. Second, the method could be applied to videos with a fixed GOP structure only. Third, forgery localization was not as precise as that achieved by, say, motion vector schemes. Fourth, manipulations involving removal/insertion of entire GOPs could not be detected, because in that case, the exact periodicity of the original signal is re-established. Finally, the authors assumed, as in [51], that double compression is always indicative of tampering, which, from a forensics point of view, was a rather unsound assumption.

It may seem like a convenient and effective idea to detect forgeries by detecting double compression, but this concept suffers from a dire shortcoming; the presence of double compression is not always indicative of a forgery. This fact was brought to light by the authors in [29] who stated that multiple compressions were a poorly explored topic and that it was risky to make any assumptions regarding the authenticity of the digital content merely on the basis of the presence of double compression. Their claim was supported by the simple fact that the (perfectly legitimate) digital content available on the Internet usually undergoes more than one compression. The authors thereof also proposed a technique to accurately discern the nature of the double JPEG compression in [29], but it was not extended to videos.

It is important to realize that in a real-life scenario, application of a forensic scheme and interpretation of its results are different from what it seems to be in theory. Deciding whether or not a technique will be viable is not something that can be generalized; the actual case context

defines which methods could be useful. Evidence provided by any forensic technique must in fact be analytically judged before reaching any conclusions regarding content authenticity. For instance, for an ordinary video (such as one downloaded off of the Internet or one recorded on a mobile phone), the presence of double-compression artifacts may not be suspicious. A doubly compressed surveillance footage on the other hand must not be considered innocuous, simply because once recorded, surveillance footage is supposed to remain unaltered in every way, which means that double compression should normally not occur in such a scenario. If a given footage shows signs of double compression, it would indicate the presence of some sort of unauthorized modification. Therefore, the presence of double-compression (or multiple compression) artifacts would serve as the first and quite possibly the most important sign of tampering in surveillance footage, while in any other scenario, double-compression artifacts may be of the least importance.

2.1.1.3 Motion and brightness feature-based inter-frame forgery detection techniques Detection of frame addition/deletion can also be achieved with the help of motion-compensated edge artifacts (MCEA). This artifact was utilized in [59] and later in [60]. The advantage of MCEA is that its calculation requires no information, whatsoever, regarding the original video.

The technique in [59] introduced the idea of using MCEA to detect frame-deletion and determine the location of such tampering. In general, whenever consecutive frames are deleted from a video sequence, temporal correlations between adjacent frames decrease. MCEA energy was found to have close association with temporal correlation and thus was useful in measuring this change. An impact factor (calculated using MCEA energy values) effectively indicated the GOP that suffered from frame-deletion. Experimentation was done on 25 MPEG-2 videos, but quantitative results were presented in the form of impact factor values rather than accuracy rates. The authors claimed that as the number of deleted frames increased, so did the detection accuracy. Examination of the technique reveled that it was likely to be ineffective for videos devoid of rapid motion. In addition, the technique worked for videos with fixed GOP structures only and could not detect any forgery if the number of frames deleted were a multiple of this GOP length.

The scheme suggested in [60] was another attempt at using MCEA to detect frame-deletion. This was basically an improved version of [59], where MCEA difference between adjacent P-frames was exploited to see if any periodic spikes appeared in the frequency spectrum after application of DFT. These spikes indicated that the frame sequence within a GOP had been disturbed somehow, either due to frame-deletion or addition. This method could also predict the original GOP structure of the video, as long as the minimum distance between two P-frames remained the same. Observing the frequency spectrum in the Fourier domain eliminated the need of a hard threshold, which was a necessity in [59]. Testing was performed on four CIF and QCIF videos, and the results were presented in the form of Fast Fourier Transform (FFT) spectrums. This technique worked for videos with fixed GOP structures only and could not detect any forgery if the number of frames deleted were a multiple of this GOP length.

In both [59, 60], the influence of B-frames on the MCEA of P-frames was not investigated. Meanwhile, H.264/AVC codec would also prove to be a challenge for these techniques, since this codec contains some new features, such as integer transform and multiple intra/inter prediction modes, which neither [59] nor [60] were equipped to handle.

A novel motion-based approach for detecting forgeries in videos was proposed in [61], where motion-based features were extracted by modeling motion between the adjacent frames of a video sequence using Markov models. The first step was obtaining a base frame by applying collusion on adjacent frames. Then, a small window was centered on each frame and an average function was applied to all the frames within that window to capture the motion information between those frames. Motion residue was obtained from the difference between estimated and actual motion and a Markov model was applied to this residue. Finally, pattern recognition was performed with the help of an SVM. The authors reported an average detection accuracy of about 87%. The technique was found to be computationally intensive, and its performance was not validated sufficiently.

An inter-frame forgery detection method was proposed in [62], which was based on detecting optical-flow inconsistencies that resulted from frame addition/deletion. Optical flow represents the distribution of perceptible velocities of objects in a video frame or an image. Using a small moving window and calculating optical flow between the first and last frames inside the window (for detecting frame-insertion), and between every pair of adjacent frames (for detecting frame-deletion), the authors were able to detect highly inconsistent optical-flow values. For performance evaluation, they constructed one test video database using TRECVID Content Based Copy Detection (CBCD) scripts and two databases using OpenCV function library (on visual studio 2010). For frame-insertion detection, the recall and precision rates of 95.4% and 95.3% were reported (for videos created using CBCD scripts). These rates were 94.3% and 97.9% (for videos created using OpenCV function). The performance dropped as the number of inserted frames decreased. For detecting frame deletion, recall and

precision rates of 85.7% and 89.4% were reported. Again, these rates dropped as number of deleted frames decreased. It was evident that the method was less likely to be effective in case of insertion or removal of small number of frames. In addition, since the frame-deletion detection mechanism used a single threshold for all videos, any attempt to increase the recall rate increased the false alarm rate as well. For the frame-addition detection scheme, an adaptive threshold was used. However, while a low threshold ensured high recall, it also ascertained more false alarms. Furthermore, the absence of large motion in the test video could cause the system to miss some forgeries. In addition, calculation of optical-flow values is a very complex and computationally intensive process. It is highly prone to errors as well, which could eventually compromise the forgery detection accuracy of this technique.

Another optical-flow-based forgery detection technique was suggested in [63], where the authors first modeled the probability distributions of optical-flow variations for untampered video sequences by a Gaussian distribution. Any abnormality in the flow variations was considered to be an anomaly, and a statistical inference test (Grubb's test) was used to assign an anomaly score to the optical-flow patterns of every test video. This score depended on the degree to which that pattern exhibited anomalous behavior. Finally, to detect inter-frame forgeries, three cut-off thresholds (one each for frame insertion, frame deletion, and frame duplication) were applied to the anomaly score, which helped select the anomalies. The method was tested on a total of 160 test clips, where all these test clips were produced from two original MPEG-2 encoded videos, taken from TRECVID's surveillance event detection data set [64]. Average accuracies for frame-deletion, insertion and duplication detection were reported to be 75%, 85%, and 82.5%, respectively. For forgery localization, corresponding accuracies of 96.9%, 100% and 86.2% were reported. It was observed that the detection performance degraded as the bitrates of the test videos decreased. Although this was quite a novel forensic scheme, it suffered from certain limitations. First, the videos constructed for frame-insertion detection tests were not created in a plausible manner. Secondly, the parameters for the Gaussian distribution and the thresholds used during forgery detection were entirely dependent on the visual content of the given video, and were inherently unsuitable for any other test video exhibiting essentially different characteristics. This technique was formally tested on test clips created from only two videos, which did not offer the degree of content diversity that is required to ensure the wide-spread applicability of a forensic scheme. Furthermore, the technique was suitable for videos MPEG-2 encoded videos only.

An inter-frame forgery detection and localization scheme based on a novel concept of velocity field consistency was presented in [65]. The concept of velocity field is closely associated to particle image velocimetry, where the general idea is to estimate the displacement between adjacent video frames caused by time separation. The authors computed Velocity Field Intensity (VFI) and Relative Factor (RF) sequences for the given video, and showed that any inter-frame forgery ultimately caused distortions in the VIF sequences, which manifested as discernible peaks in the frequency spectra of RF sequences. This scheme was evaluated on a data set consisting of 120 test sequences constructed from 4 MPEG-2 encoded TRECVID videos. Videos suffering from frame-deletion were detected with an accuracy of 85% while those suffering from frame-duplication were detected with an average accuracy of 80%. This scheme could detect tampered videos with an accuracy of 96.3% with 10 false positives (without being able to distinguish frame-deletion from frame-duplication). However, as the quantization scale of second compression increased from 1 to 3, detection accuracy dropped from 80% to 62.5%. Aside from the excessive dependence on the quantization scales, another major drawback of this scheme was that every false detection was automatically reported as a frame-deletion forgery. Even though the performance of this scheme was far from desirable, this work represents a novel step in the video forensic domain, and although it was only tested on MPEG-2 encoded videos, it demonstrates potential to be extended to the MPEG-4 and H.264 domain.

A frame-deletion detection scheme specifically designed for H.264 encoded videos was proposed in [66], where the authors introduced a feature called Sequence of Average Residual of P-frames (SARP) and demonstrated that in case of frame-deleted videos, SARP exhibited periodicities when analyzed in the time domain. In the frequency domain, these periodicities resulted in characteristic spikes at particular locations in the Discrete Time Fourier Transform (DTFT) spectrum. These locations were then used to pinpoint the position of the deleted frames in the given sequence. Experimental validation was performed on a set of YUV test sequences from the video trace library [33] which were encoded using H.264 codec, and on another set of videos from the consumer digital video library [67]; average detection accuracy of 92.08% was reported. The authors showed that by combining information acquired from both time and frequency domains, they were not only able to overcome the various limitations faced by the methodology of [42], but achieved better detection results as well. However, the final separation of original videos from the tampered ones was based entirely on a hard threshold, which did not offer the desired level of flexibility to suggest an extensive scope of this method, especially for videos with essentially different visual content and characteristics from those used during performance validation. In

addition, the authors assumed a fixed GOP structure for the test videos.

In [68], a Blockwise Brightness Variance Descriptor (BBVD) was proposed to help detect frame insertion and deletion. The basic idea here was that in equal time intervals, the ratio of BBVD will display disturbances in case of videos suffering from inter-frame forgeries. The video was first divided into overlapping sub-sequences, and then all the frames of every sub-sequence were divided into 4×4 blocks. Ratios of BBVDs in every group were then computed and any irregularities therein, above a pre-determine threshold, were considered to be an indicative of the presence of forgery. Number of peaks in the BBVD error sequence helped differentiate fame-insertion forgery from frame-deletion forgery. For performance validation, the authors acquired original test videos from the recognition of human actions database [69] and created a total of 220 forged test sequences with the help of TRECVID CBCD scripts. For frame-insertion detection (with number of inserted frames >25), average recall and precision rates of 94.2% and 86.3% were reported. These rates were 89.2% and 79.4% for frame-insertion localization. If the number of inserted frames was less than 25, 83.5% recall rate and 75.8% precision rate could be achieved. The authors did not report any results for frame-deletion detection. This technique worked only for video recorded with a stationary camera and could detect a forgery only if the number of inserted or deleted frames were more than 10. Furthermore, this technique also required that only one kind of forgery be present in the given video at one time (either frame insertion or removal) and that the forgery be performed only once. All these assumptions rendered this technique unsuitable for utilization in real-life forensic scenarios.

In [70], a frame-duplication and deletion detection technique was proposed. To detect duplicated frames, the authors performed frame differencing at spatial ROIs to compute mean square values of motion energy. The entropy of difference between every successive frame and average object area was used as features to train an SVM classifier to detect videos suffering from frame deletion. The authors, however, did not provide any information regarding the database used for testing, neither did they provide any clearly comprehensible quantitative results, thereby making it difficult to evaluate the forensic significance and utility of this technique.

In [71], the authors suggested an improved version of the methodology proposed in [66], where, in addition to the periodic artifacts caused by frame deletion in SARP, they analyzed the magnitudes of the P-frame prediction error as well. However, unlike [66], this technique functioned entirely in the frequency domain. The authors then analyzed a possible anti-forensic method in which the traces of forgery could be hidden by explicitly increasing the P-frame prediction error of the forged sequence. Then, a counter anti-forensic approach was suggested, where the authors estimated the actual prediction error and compared it with the prediction error obtained from the video under consideration. The technique's performance was independent of the motion estimation algorithm used during the initial compression of the videos. This technique could detect both frame insertion and deletion and but was unable to localize the forgery.

In [72], a hybrid forensic system for the detection of frame-insertion, removal and replication was developed. For the detection of frame-insertion and removal, the system detected irregularities introduced in the brightness gradient component of optical flow by post-production frame-tampering. Frame-replication forgeries were detected and localized by analyzing the abnormalities in the prediction residual patterns of forged videos. For performance validation, ten original MJPEG and H.264/AVC encoded videos from SULFA [8] and ten original YUV format videos from video trace library [33] were acquired. These videos were then tampered with using different combinations of frame-insertion, removal and replication, and were re-compressed and transcoded before and after the forgeries using FFmpeg [129], to obtain a data set consisting of 480 test videos. The system reportedly detected frame-insertion, frame-removal and frame-replication forgeries with average accuracy rates of 98.2%, 98.6% and 98.3%, respectively. The system was capable of identifying the presence of forgeries in MJPEG, MPEG-2, MPEG-4 and H.264/AVC encoded videos, regardless of the number and/or location of the tampered frames. It was independent of heuristically determined thresholds and its performance remained unaffected by removal of entire GOPs or multiples of GOPs. The frame-replication detection technique of the proposed system missed no forgeries but generated a few false alarms, specifically in case of videos that contained (non-replicated) visually similar frames. Moreover, the system was not capable of detecting the presence of multiple post-production compressions.

In the forensic system proposed in [74], the authors first detected the presence of double compression by calculating normalized histogram of the most significant digits of all DCT coefficients of the I-frames. An SVM was then used to distinguish singly compressed videos from double-compressed ones. For inter-frame forgery detection, time-domain analysis of Mean Absolute of Residual Errors (MARE) of P-frames was performed. The authors demonstrated that MARE sequences for tampered videos exhibited discernible peaks, which were different from the peaks generated in case of original videos. Empirically determined thresholds were used to distinguish the different kinds of peaks in the MARE sequences. Then, another set of thresholds was applied to the magnitudes of the peaks to

determine if the given video was forged or authentic. The final decision regarding content authenticity was performed with the help of a decision-fusion scheme, where several rules were proposed to combine the evidence provided by the double-compression and frame-deletion detection schemes. For experimentation, 22 YUV test videos in CIF and QCIF formats were acquired from video trace library [33], and were then singly and doubly compressed using MPEG-2 encoder. Tampered videos were constructed by removing the same number of frames from the same location in each of the test videos. Average detection accuracies of 87.1% for double-compression detection and 83.39% for frame-deletion detection were reported. Analysis of the system's functionality revealed that the double-compression detection accuracy decreased with increase in the quantization scale of the first compression, and the frame-deletion detection accuracy decreased when the ratio of the second and first quantization scales increased. Furthermore, the authors assumed a fixed GOP structure of the test videos, and the system could detect deletion of entire GOPs only.

2.1.1.4 Pixel-level analysis-based techniques In [75], the authors detected frame duplication by observing pixel-level anomalies. They proposed a coarse-to-fine process to determine similarity between given query clip and candidate clip to determine the presence of duplicate clips and to localize their positions. First, to detect duplication in temporal domain, difference in histograms of successive frames in RGB color space was calculated. Next, actual spatial correlations between the corresponding frames of the query clip and candidate clip were calculated to ascertain that those frames were in fact identical and not just errors caused by noise. Experiments conducted on four test videos revealed that their method generated no false alarms and missed detections were caused in only one video. The accuracy of the system for localizing the duplicate frames in four test sequences was reported to be 100%. However, testing on only four videos does not ensure practical suitability of this method. Furthermore, the entire method was based on several empirically determined hard thresholds, which limited its scope of applicability even more. In addition, it was found that this method would be unable to detect any forgery if the frames were shuffled before being pasted at another location within the video.

An inter-frame forgery detection scheme was suggested in [76], where the authors used Consistency of Correlation Coefficients of Gray Values (CCCoGV) as a forensic feature, and stated that while for original videos, CCCoGV remained consistent, any post-production disturbance in the frame sequence caused this value to demonstrate abnormalities. An SVM was used to perform the final classification. For quantitative analysis of their approach, the authors used five data sets (one containing original videos and four containing tampered ones), each consisting of 598 videos with still background and a little camera movement. Average detection accuracy of 99% (for frame-insertion detection) and 95% (for frame-deletion detection) was reported. Although the authors did not mention the compression standards used to create the videos, a basic analysis of this scheme's functionality revealed that strong compression would ultimately cause severe degradation in its detection accuracy.

2.1.2 Detection of temporal interpolation and frame-rate up-conversion

Another way to forge a video is via temporal slicing, i.e., interpolating two or more different videos to generate a new video. If the source videos do not share the same frame rate, they have to be temporally interpolated before they can be spliced together. This is done with the help of an operation known as Frame-Rate Up-Conversion (FRUC), wherein new frames are generated with the help of the existing ones and are inserted into the original video, thereby increasing its frame rate.

The interpolation detection method suggested in [77] worked on the principle that whenever an attacker tried to perform frame interpolation while simultaneously trying to minimize the resultant temporal artifacts, it was done with the help of motion-compensated interpolation. However, motion-compensated interpolation left characteristic and quite detectable footprints of its own. The authors were able to suggest a system that worked for un-compressed and slightly compressed (e.g., H.264) videos (such as television broadcast videos) and achieved promising results, even when used on only a subset of the frames. Exact quantitative results were not reported, but the authors admitted that the performance of their system was not good for MPEG-2 encoded videos, because such compression was much more vigorous than H.264/AVC encoding. Moreover, the system functioned well on small-sized spatial windows, which allowed this detector to be used as a possible tool for detecting copy–paste forgery attacks. However, the number of observed interpolated frames had to be large enough for the system to detect forgeries successfully.

An edge intensity-based frame-rate up-conversion detection technique was proposed in [78], where after computing edge intensities form all the video frames, an adaptive threshold was calculated using the Kaufmann Adaptive Moving Average (KAMA) technique. This threshold was then used to distinguish original fames from up-converted ones and estimate the original frame rate as well. A set of 15 un-compressed YUV sequences taken from the video trace library [64] were up-converted to different target rates, and were also lossy-compressed using H.264/AVC

encoder. Overall, for a total of 300 test sequences, an average detection of 95.4% was reported.

The authors in [79] observed that most frame-rate up-conversion techniques introduced visually discernible periodic artifacts into the texture regions of the affected frames. This observation led the authors to develop a two-stage blind detection method that based on the frame-level analysis of a feature called Average Texture Variation (ATV). In the first stage of the method, ATV values for every frame were computed to obtain an ATV curve of candidate video. In the second stage, each ATV curve was further processed to detect the periodic artifacts, which were considered to be the evidence of frame-rate up-conversion operation. This technique could localize the position of the interpolated frames and help estimate the original frame rate of the video as well. Exhaustive testing on videos from the video trace library [64] and xiph.org video test media library [80] demonstrated that the technique was quite effective for the detection of common frame-rate up-conversion techniques, such as motion-compensated and linear frame averaging; an average detection accuracy of 96% was reported.

2.2 Intra-frame forgery detection techniques

In this section, we analyze various methods suggested for detection of copy–move and upscale-crop forgeries in digital videos.

2.2.1 Copy–move forgery detection techniques

2.2.1.1 Pixel-similarity and correlation analysis-based techniques The techniques that detect copy–move or copy–paste forgeries generally proceed by looking for similarities or correlations between regions of successive video frames or regions of the same frame, which in theory should not have anything in common (primarily because they have different origins or have different time/place associated with their lineage).

To detect copy–move forgeries, the authors in [81] computed spatial and temporal correlation coefficients to identify and locate resemblance between separate parts of the video. For detecting region duplication, the accuracy of 100% (for stationary-camera videos) and 81.2% (for moving-camera videos) was reported for MPEG compressed videos. This performance affliction was attributed to the fact that compression artifacts were more pronounced in the presence of motion in the video. Performance was observed to drop with decrease in the region size (e.g., at 9 Mbps bitrate, when the region size dropped from 256×256 pixels to 64×64 pixels, accuracy dropped from 100% to 35% for stationary-camera videos and from 87.8% to 39.4% for moving-camera videos). When tested for frame-duplication detection in two un-compressed videos recorded using a digital video camera, the method's accuracy was reported to be 84.2% (for stationary-camera videos) and 100% (for moving-camera videos). For compressed videos, the results were 85.7% (for stationary-camera videos) and 95.2% (for moving-camera videos). The method was robust to compression as long as the compression rate was large enough (for instance, the detection rate dropped from 100% for an un-compressed moving-camera video to 86.8% for a video compressed at 3 Mbps bitrate). Moreover, the method had a very high time complexity. It also generated false alarms for frames with areas of uniform texture, such as the sky. Insufficient validation was another limiting factor.

A forgery detection and localization method was suggested in [82] that was based on the fact that correlation between temporal noise residue suffers a significant change whenever a region is forged (noise residue for every frame is calculated by subtracting the actual frame from its denoised version). Detecting this change led to the detection of the forged region. Experimental tests performed on three surveillance-like videos captured using a digital camcorder generated the following results. For videos suffering from temporal copy–paste attacks, the recall rate was 55%, precision rate was 96.6%, and the false positive rate was 3.3% with a miss rate of 44.2%. For synthetically inpainted frames, these values were 74.5%, 92.8%, 7.2%, and 25.4%, respectively. This resulted in an average accuracy ranging from 70% to 82%. Too high or too low illumination was the main cause of the false positives. Being a noise-based method, lossy encoded and poor quality videos like those streamed on low-bandwidth Internet were problematic, since the excessive amount of noise in such videos caused the residue calculations to became highly unreliable. Furthermore, many missed detections were caused when the noise intensities of the original and manipulated regions differed significantly, and the method failed to calculate the noise residues accurately. The method was also sensitive to quantization noise. In addition, since correlation of noise residue was highly unstable for moving-camera videos, the method was rendered ineffective in such scenarios. Though this method could have been useful for frame-replication detection, the authors did not perform experiments in this direction.

Another technique for copy–move detection was proposed in [83, 84] which was based on the hypothesis that correlation attributes of pixel sub-blocks within as well as between the frames were bound to be disarranged by tampering attacks, such as double compression, retouching, or resampling. The authors extracted noise residue and quantization residue features from adjacent frames and then performed correlation analysis using Canonical Correlation Analysis (CCA), Cross-modal Factor Analysis (CFA), and Latent Semantic Analysis (LSA). Acute observation of such disturbances helped the technique to differentiate

the fingerprints of a genuine video from those of a tampered one. For a set of video clips taken from low-bandwidth Internet streamed movies, the technique reportedly achieved accuracy of 92.1% (for CFA), 91.8% (for CCA) and 91.2% (for LSA). The accuracy of this technique was found to be significantly better than that of [82], which was also a noise-residue-based approach. This improved performance can be attributed primarily to the utilization of two residue features (noise and quantization) rather than a single noise-residue feature. By not relying on noise residues entirely, this technique was able to overcome the drawbacks of [82] specifically, its inapplicability to lossy encoded and poor quality videos. The techniques, however, was still sensitive to significant illumination variations, and was applicable to videos with static backgrounds only.

An ETS detection technique based on fuzzy theory was proposed in [85], where the authors observed that even though the effects of ETS inpainting were quite imperceptible, they introduced certain unnatural similarities into frame block pairs, thereby revealing the presence of a forgery. The first step was conversion of the video into a sequence of frames, followed by a block-matching process, performed within a Region of Suspicion (ROS), to determine the degree to which sub-blocks of frames were similar to one another. By working on a portion of the frame instead of the entire frame, the technique was able to maintain a good balance between performance and computational complexity. Fuzzy set theory was used to handle the uncertainty associated with similarity-measure calculations. Although the authors did not present any quantitative results with which to judge the performance of the technique, it was observed that the detection performance was extremely dependent on a number of parameters that required careful empirical setting. Inapplicability to compressed videos was yet another limitation of this technique.

The forgery detection and localization method of [86] was similar in functionality to [81] but was fully automatic, and the location of the tampered frames was not considered to be a priori information. It was a two-step algorithm, where first, frame-level manipulations were detected by computing and analyzing zero-motion video residual, which was obtained by subtracting pixels occupying same spatial position on consecutive frames. Then, to detect duplicated content along the temporal dimension of the video, the authors cross-correlated small 3D spatio-temporal blocks of the frames. The presence of high-correlation indicated the location of the identical content. The method was completely unsupervised and to judge its performance, and 120 realistic H.264 encoded sequences taken from a standard data set SULFA [8] were used. Originally, all the videos were slightly compressed, but to test the method's robustness to compression, the authors deliberately re-compressed certain test sequences. These latter videos were referred to as 're-compressed videos', the former were called 'not re-compressed videos'. For frame-level manipulation detection, average detection accuracy of 84.2% (for not re-compressed videos) was reported. For video-based attacks, the detector correctly identified duplicated blocks on 90% of the not re-compressed videos. If re-compressed videos were also considered, 87% of the blocks were reported to have been detected correctly. The method was found to be tolerant of mild compression only.

In [87], the authors proposed an approach for detecting and localizing region-level forgeries in videos. The approach was designed to detect irregularities caused by inpainting in spatio-temporal coherence between consecutive frames. The video was first divided into sets of frames, where the effect of motion could be considered negligible in each set. Then spatio-temporal slices were extracted from each set of frames and coherence was computed between every one of these slices. Pairs of slices exhibiting unnaturally high coherence (caused by TCP inpainting) or abnormally low coherence (caused by ETS inpainting) were classified as those belonging to frames suffering from copy–paste forgery. This approach was tested on 18 self-captured video sequences. For un-compressed videos, average precision, recall, and accuracy rates for TCP inpainting detection were reported to be 93.6%, 80.2%, and 85.5%, respectively. These values were 75.7%, 78.1%, and 74.3% for ETS inpainting detection. For JPEG compressed frames (quality factor 90), the algorithm's precision, recall, and accuracy dropped to 89.9%, 77.4%, and 81.4% (TCP inpainting) and 78.5%, 76%, and 74.8% (ETS inpainting). The results further degraded in case of MPEG and WMV compressions (70% for MPEG-2, 71% for MPEG-4, and 70.3% for WMV-9 for TCP inpainting and 73.5% for MPEG-2, 75% for MPEG-4, and 75.3% for WMV-9 for ETS inpainting, at 9 Mbps bitrate), suggesting that the method worked best for completely un-compressed videos. The decision regarding the presence of high or low coherence was based on empirically thresholds, which further limits its scope.

An object removal detection and localization technique was suggested in [88]. It utilized Scale Invariant Feature Transform (SIFT) coupled with k-NN matching to detect spatial copy-paste forgeries, and noise-residue cross correlation to detect temporal copy–paste forgeries. This technique was tested on 150 test videos, some of which were the authors' own and others were taken from SULFA [8]. The authors reported an average detection accuracy of 99%, for spatial forgery detection. For temporal forgery detection, average detection accuracy of 98% was reported. For frame-duplication detection, average accuracy of 98% was reported. Though the technique performed well for the test videos under consideration, it suffered significant performance degradation as the compression strength increased.

In [10], a method to detect blue-screen compositing was proposed. The authors suggested that although blue-screen compositing is generally used in applications such as weather forecasting on TV, it has the potential to be used as a forgery scheme. In this method, the authors computed quantized DCT coefficients of the foreground and background of the video separately. Then, a similarity measure called pattern distance was calculated from the histograms of the DCT coefficients, which was then compared to a threshold to classify the given video as original or composite. 21 self-recorded MPEG encoded CBR video sequences were to evaluate the performance of this method, and for bitrates ranging from 5 to 9 Mbps, detection accuracies in the range 70–100% were reported.

A rotation-invariant copy-move forgery detection method was proposed in [89], where a specialized 3D version of PatchMatch [90] was devised. Previous instances of use of 3D versions of PatchMatch include tasks such as stereo matching [91] and video inpainting [92]. The authors used Zernike moments as forensic features, and then used PatchMatch to generate a dense approximation of Nearest Neighbor Field (NNF) to search for copy-pasted regions within video frames. Experimental evaluation was performed on a set of 10 test videos taken from the REWIND data set [93]. For these original test videos, average TPR of 48.7% and average False Positive Rate (FPR) of 0.02% were reported (for forgery localization). The authors then re-compressed these test video using H.264/AVC encoder with QP 10, 15, and 20, GOP was fixed at 150. TPR and FPR for re-encoded videos were 46.1% and 0.12% (for QP 10), 29.4% and 0.06% (for QP 20). The technique produced a very small number of false alarms, but the detection accuracy was significantly lower than a desirable standard. The performance was found to be highly dependent on the amount and kind of motion present in the videos; larger motion led to lower detection accuracy. Nevertheless, this technique could detect forgeries even if the copy-pasted regions underwent post-processing operations, such as rotation and scaling. The technique could achieve substantial improvement in performance by proper utilization of motion information.

2.2.1.2 Object features based copy-paste detection techniques Another way to expose copy–paste forgeries is to focus on detecting the artifacts that arise after an object has been removed from a video frame.

The work in [94] was a novel technique that could identify a forged video by detecting ghost shadow artifacts, which arise when moving objects are removed from video frames. The video was first segmented into static background and moving foreground via bock matching. Then, the moving foreground was used to construct a foreground mosaic. The moving foreground was also used to compute an Absolute Difference Frame (ADI), from which isolated regions were removed with the help of morphological operators (erosion and dilation). The presence of any discrepancies between the foreground mosaic and the binary ADI was considered to be evidence of forgery. Ten real-world videos were used to judge the performance of the technique and the results suggested that it was resilient to MPEG-2 compression and re-compression. H.264/AVC encoded videos, however, could not be handled efficiently. In addition, the technique worked only for videos with stationary background. The forgery localization process was found to be imprecise as well.

Another novel concept was presented in [95], where manipulated videos were identified by detecting physically improbable trajectories of solid objects in the video. The objective was to create a 3D model of the parabolic trajectories of objects in free flight (for instance, the trajectory of a basketball being thrown towards the basket) and its corresponding 2D projection onto the image plane, and then weed out inconsistencies from a geometric point of view. The performance of this method was not contingent upon the presence or absence of compression artifacts or quality of the video. In addition, the method was insensitive to resolution of the video and post-processing operations. For a set of 13 fake and 11 authentic test videos created by the authors and 3 videos downloaded from YouTube, the method was shown be quite efficacious. The authors did not report quantitative results in their paper.

In [96], an object-based tampering detection mechanism was proposed which was based on the observation that object-based manipulations always left certain splicing traces in the video frames. This was due to the fact that object removal was generally followed by some kind of inpainting technique, which inevitably caused inconsistencies near the object boundary or the boundary areas. The first step was the detection of motion objects in a frame to locate object boundaries. Afterwards, wavelet transform was used to extract features that could represent the forgery traces. These features served as inputs to the SVM that classified each frame as original or forged. On a set of 20 self-captured test videos in AVI and WMV formats, average TN, TP, and accuracy rates of 98.2%, 94.4%, and 97.4%, respectively, were achieved. The method worked only for videos with static backgrounds. In addition, it was unfit for videos with very little object motion, because in such cases, it became difficult to extract splicing traces. One possible future direction for this methodology could be the exploration of more powerful discriminating features, such as motion trajectories.

2.2.1.3 Motion feature-based copy–paste detection techniques A motion-residue-based forgery detection technique was proposed in [97], the novelty of which was that

to extract forensic features from motion-residue information, it utilized feature extractors originally built for image steganalysis. The authors first used collusion operators to compute motion residues and then extracted seven steganalytic features, namely, CC-PEV, SPAM, CDF, CF, SRM, CC-JRM, and J + SRM, from these residues to model inter-frame and intra-frame properties of pristine (completely unmanipulated), forged, and double-compressed frames. An ensemble classifier, whose decision was based on majority voting from several base learners, was used to classify every frame as pristine, forged, or double-compressed. This technique was tested on the SYSU-OBJFORG data set (which is not yet publically available), consisting of 100 H.264/MPEG-4 encoded static surveillance camera videos. The authors report a comprehensive assemblage of quantitative results in their paper. Overall, pristine frames could be distinguished from double-compressed frames with an accuracy of 99%, and pristine and forged frames were also distinguished with an accuracy of 99%. Double-compressed frames were correctly classified with an accuracy of 95.8%, and forged frames were correctly classified with an accuracy of 85.3%. The best results were reported to have been generated with the help of J + SRM. However, in case of low-bitrate videos, CC-PEV produced the most desirable results. Being a feature set of low dimensionality, CC-PEV was quite robust to reduced bitrates. An important advantage of this technique was that it was applicable to videos with variable to GOP structures, as opposed to majority of the previous innovations that worked effectively as long as the GOP structure of the test video remained fixed. This technique was not capable of localizing the forged regions within the frames; this remains the subject of the authors' future work.

The work in [58] addressed the issue of localizing intra-frame forgeries in MPEG-2 videos. This method used the VPF based scheme proposed in [49]. First, the authors estimated the size of original GOPs using VPF scheme, followed by performing a Double Quantization (DQ) analysis. This consisted of observing certain periodic trends in DCT coefficient histograms and pinpointing the manipulated regions of the frame. Results of the experiments performed on 7 test sequences were presented in the form of ROC curves which demonstrated that for a given first quantizer scale factor (Q1), low values of second quantizer scale factor (Q2) resulted in high localization accuracy. On the other hand, by increasing Q1, the performance of the system could be raised. For instance, if Q2 was kept at a constant value of 2, detection accuracy of about 98% was achieved for different values of Q1. Further analysis of this method revealed that not only was it ineffective against inter-frame forgeries, its applicability was limited to MPEG-2 videos coded using VBR model only.

The authors in [98] utilized optical-flow inconsistencies to detect region-level copy–move forgeries. Every video frame was first divided into suspicious and authentic regions, and then, Optical-Flow Variation Factors (OFVFs) were computed for all the regions separately. Any and all characteristic peaks arising in the OFVF of a given region were considered to be evidence of forgery. Peak periodicity and auto-correlation analysis were then used to localize the forgery. This method was tested on 10 pairs of original and tampered MPEG videos from the REWIND database [93]. Original frames were correctly identified with a TPR of 86% (with 8% FPR), and tampered frames were correctly identified with 85% TPR (with 1.4% FPR). Overall, the method detected forgeries with an average accuracy of 89.4%. This method worked only for videos with a fixed GOP, and could not locate the forgery if the displacement of the forged region was not a multiple of this GOP length. It was also highly sensitive to the ROI selection process. Peak periodicity analysis, a key operation in this method, was also found to be unreliable in case of GOPs with large motion.

2.2.2 *Upscale-crop detection technique*

Another simple way of video content manipulation consists of enlarging the frames of a video and then cropping them to remove evidence of some incriminating event in the outermost part of the frames. The presence of such forgeries can be detected by looking for traces of resampling, because whenever a video's frames are cropped and enlarged, they undergo a process called resampling (specifically up-sampling) so as to maintain consistent resolution across all the frames of said video.

An upscale-crop and partial manipulation detection method was suggested in [99]. The authors observed that since the resampling introduces certain statistical correlations in the given content, its presence could be detected by looking for these correlations. The authors exploited SPN as the forensic feature and analyzed the variations in the correlation properties of reference SPN and SPN of up-scaled frames. The method was tested on a total of 1920 forged sequences that were constructed from 120 H.264 encoded RGB and infrared self-captured test videos. As long as the scaling and quality factors were closely supervised and controlled, this technique generated a TNR of 100% and TPR greater than 98%. In case of partial manipulation detection, for region sizes in the range 100sq. pixels to 150sq. pixels, detection accuracy of 100% (for dynamic scene videos) and accuracy in the range 94.2–100% (for static-scene videos) were reported. This method was observed to be robust against not only RGB and infrared videos but also handled compressed videos effectively. It worked for both dynamic and static-scene videos, recorded using both static and moving cameras.

Despite its innovative nature and high accuracy, this technique was found to be exceedingly dependent on a large number of content-dependent parameters and thresholds that required extremely careful empirical setting. Moreover, the copy–paste forgery detection scheme was specifically designed to detect superimposed timestamps on video frames, and although the authors tested this technique on a variety of tampered videos, these forgeries were created by copying objects from other videos and then pasting them randomly into the target videos. This forgery creation procedure is inherently inaccurate, since creation of a plausible forgery requires that tampered region of the frame blends into the rest of the frame seamlessly, so as to maintain visual coherence within the frame and throughout the video. When target objects are taken from other videos, it becomes virtually impossible to maintain the required level of consistency because of the visible distortions caused by factors, such as minute variations in pose and scale of the objects, differences in lighting, and shadows caused by illumination variations, and deformation and occlusion.

Upscale-crop forgery is not as common as other kinds of forgeries but is equally consequential. As compared to the image forensics domain, the field of video-resampling detection is dangerously under-populated, and it is evident that this research field is in need of substantial innovation. Motivation derived from the image-resampling domain could be conducive to further advancements in this field.

2.3 Amalgamation of multiple forensic tools

All the techniques discussed so far have been customized to handle at most two kinds of forgeries. No individual technique can be expected to single-handedly detect all forms of malicious manipulations that a video could suffer from. However, the idea of consolidating several techniques and using then together could induce further advancement in the domain of multi-utility video forensic mechanisms.

A doctoral thesis [100] was an early attempt at combining several forensic tools into one. This amalgamated approach detected tampering in interlaced and de-interlaced videos utilizing camera artifacts, as in [39]. Since this method was inherently dependent on the use of motion, it was unsuitable for detecting manipulations in those regions of the video that exhibited no motion. The technique could detect the possibility of frame insertion/removal by identifying artifacts introduced by MPEG double compression [42]. The detection rate was observed to be highly dependent on the ratio of the first and second-quantization scales. Evidently, the technique was effective only in those cases where the second compression scale was greater than the first. In addition, the technique worked best on videos with mostly static backgrounds captured using stationary cameras. Duplicate frames were detected by identifying similarities in the spatial and temporal correlations within and between the frames [79]. To identify re-captured videos, the method looked for distortions in camera skew, which occur when the re-capture device is placed off-axis with respect to the screen on which the video is being originally projected [101]. All the experiments were performed on videos recorded using digital video cameras and the results can be viewed from the previous sections that discuss these individual techniques in more detail. Furthermore, in addition to formulating tamper detection methods, the author also discussed possible anti-forensic techniques for each of these methods, which could prove to be highly beneficial for the development of counter anti-forensic schemes.

2.4 Anti-forensic and counter anti-forensic strategies

Advances in the forensics domain have been analogically mirrored by the exploration in the anti-forensics domain. Simply stated, anti-forensic techniques consist of adapting and re-modeling the forgery process so as to make the unauthorized alterations inconspicuous to tamper detection methods. That is, a malevolent adversary could alter the digital contents in such a manner that any traces of such alterations would remain well disguised and therefore highly difficult to detect or locate.

Review of the literature revealed that several researchers have proposed anti-forensic strategies that have been shown to deceive forensic schemes designed to detect forgeries in digital images [102–111].

The efforts directed towards constructing anti-forensic techniques specifically designed for videos can be attributed to [112–114]. In [112], the authors presented a method that could trick the frame-insertion/removal detection technique proposed in [42] (which detected missing frames from a video sequence by analyzing spikes in the prediction error sequences generated with the help of I-, P-, and B frames). The authors observed that by raising prediction errors of certain frames to the values expected in the spikes, the forgery could be camouflaged, since peaks in the error due to actual desynchronization of the I-, P-, and B frames would no longer remain distinguishable. The trick was simple and worked effectively, but paid a great cost in the form of coding efficiency. After performing this forgery, some of the frames had to be re-encoded at a bitrate that was considerably higher than the one used initially.

The work in [113] demonstrated further advancements in the field of digital forensics and anti-forensics. The authors had already provided considerable insights into the mechanics of interplay between a forger and a forensic investigator in [105–108] and in [113], mathematical models to highlight effects of frame addition/deletion in the P-frame prediction error sequences were proposed.

The authors developed two automatic frame-addition/deletion techniques that worked for video encoded using codecs that used both fixed length and variable length GOPs. They also developed anti-forensic method to hide evidence of frame addition/deletion. This method was an extension of the technique proposed in [112]. Next, they discerned certain fingerprints that make anti-forensic manipulations more susceptible to identification. This knowledge helped them to modify their previous anti-forensic method in such a way that it became more immune to detection via these fingerprints. To evaluate the performance of their forensic and anti-forensic techniques, 36 QCIF MPEG videos were used. From the ROC curves, it could be seen that the technique was able to detect frame addition/deletion with at least 85% probability of detection at less than 5% FAR (for fixed length GOPs) and with at least 90% probability of detection at less than 10% FAR (for variable length GOPs). Their anti-forensic method was able to fool the frame-deletion detection technique almost completely (for a probability of false alarms under 80%, the anti-forensic approach generated a susceptibility of 0.7 or more).

The sheer amount of innovation presented in this work could easily raise it to the standard of a landmark in the fields of digital forensics and anti-forensics. That being said, certain limitations need to be highlighted. First, the detection results were reported for the case when the number of deleted frames was a multiple of sub-GOP lengths only. Second, the authors assumed that frame-deletion began from an I- or P-frame only. Third, all the experiments were carried out on test videos that were coded using VBR mode only, which makes it difficult to determine the extent of the method's applicability to CBR coded videos. The method was sensitive to noise as well.

It was shown in [115, 116] that anti-forensic strategies are not completely immune to forgery detection algorithms. The authors thereof conducted several experiments and came to the conclusion that getting rid of all the traces of the entire compression process (undergone during the forgery process) is a very challenging task. As a result, anti-forensic methods have an inevitable tendency of leaving their own slightly detectable traces behind. The authors studied the anti-forensic technique of [106] and proposed a method to counter it. Albeit originally designed for digital images, this work possesses the potential to be extended to videos.

Pioneering steps in the field of video counter anti-forensics were taken in [114], where a method to counter the anti-forensic method of [113] was proposed. The method of [114] was based on the following observation: to successfully deceive a forensic investigator, it is important for an anti-forensic approach to keep the amount of distortion introduced into the modified video under an acceptable limit. This implies that the anti-forensically altered frame would have to be similar to the original frame. Therefore, it was reasoned that by re-compressing the suspicious video (that has been modified anti-forensically), the true prediction error sequence could be acquired which would be approximately equal to the prediction error sequence of a video that has not been altered anti-forensically. Thus, by estimating the actual prediction error, both frame-deletion and the use of anti-forensics could be detected simultaneously. The authors then suggested an anti-forensic frame-deletion technique, which was simple yet quite clever: they added new frames in place of the deleted ones. The new frames to be inserted into a GOP were copied from the GOP itself. 220 QCIF sequences in YUV format were used for testing the method's performance. It was observed that this method could detect videos tampered via the anti-forensic method proposed in [113] with a maximum accuracy of 95%. Frame-deletion detection accuracy was also reported to be 95%, but if the frames were deleted anti-forensically using the authors' own technique, detection accuracy dropped to 49%.

The authors in [117] proposed their own set of anti-forensic and counter anti-forensic schemes. The anti-forensic scheme proceeded as follows. Once the given video was decoded, information regarding the Macro-Block (MB) types of the frames before and after the target frames (i.e., the frames that were going to be modified) were recorded. After the forgery, these recorded MB types served as a reference that enabled the forger to limit the coding modes of the targeted frames. The targeted frames were then encoded once, and the quantization indices of these frames were stored. Then, in the second round of encoding, these indices, along with the MB-types, were adjusted and recorded, so that a genuine-looking edited video could be created. The authors then suggested using the de-blocking filter of H.264/AVC to detect forgery in this anti-forensically created manipulated video. Furthermore, they stated that the relationship of QP with video bitrate could also be used to expose anti-forensically created forgeries. De-blocking filters and intra-prediction in H.264/AVC would cause difficulties for this method, and it would also be unable to perform if videos were to be transcoded using the same rate control method.

Both anti-forensics and counter anti-forensics have demonstrated their tendency to be very compelling research subjects, and further advancements in this domain may be expected in the near future.

2.5 Video bitrate up-conversion detection

In this section, we discuss notable advances in the field of video bitrate up-conversion. Bitrate up-conversion refers to the process of fraudulently increasing the bitrates of videos. Bitrate is an important criterion for judging the

quality of a digital video; high bitrates are usually associated with better audio-visual quality. By deliberately increasing the bitrate of an originally low-bitrate video, low-quality videos can be made to appear as high-quality ones, and can therefore be used by exploitative individuals to gain increased commercial profits. Another reason for increasing the bitrate of a video could be to hide the evidence of a forgery.

The authors in [118] presented an algorithm dedicated towards detecting videos exhibiting bitrate up-conversion, and estimating the original bitrate of such videos. Although this algorithm was not designed to detect forgeries, it could still serve as a utilitarian pre-processing step that could provide useful hints regarding content authenticity. This algorithm was an extension and improvement of the previous algorithms proposed by the same authors in [119, 120]. While the technique in [119] was evaluated on MPEG-2 encoded videos only, in [120], the authors worked on detecting up-converted videos only, without actually determining the original bitrates of these up-converted videos.

In [118], the authors observed that bitrate up-conversion could not be accomplished without first re-compressing the video. Therefore, re-quantization artifacts, motion vectors, and prediction errors were expected to provide useful clues for detecting fake bitrates. The authors used generalized Benford's law [43], and some similarity measures in the DCT frequency domain to analyze differences in the features of original high bitrate videos and fake ones. The final classification was performed with the help of a binary SVM classifier. Performance evaluation was performed on 295 MPEG-2 and H.264 encoded test clips in CIF, QCIF, and VGA resolutions constructed from a single test sequence. Fake bitrates were detected with average rates of 97.5%, 97.3%, and 98.5% for CIF, QCIF, and VGA resolutions, respectively, for MPEG-2 encoded videos. For H.264 encoded videos, the average detection rates of 91% (CIF), 92.7% (QCIF), and 95.2% (VGA) were reported.

All these works represent innovative steps in the field of video forensics via fake bitrate detection and original bitrate estimation, and could benefit from two suggestions regarding possible future directions. First, advanced features, such as quantization artifacts, feature-curve inflexion points, and some other video properties, such as adaptive rate control encoding schemes, could help improve the performance of a bitrate up-conversion detection technique. Second, at present, these techniques work only for videos re-compressed using the same codec. By adapting to the needs of videos re-compressed using different encoding standards, the scope of these forensic solutions can be widened.

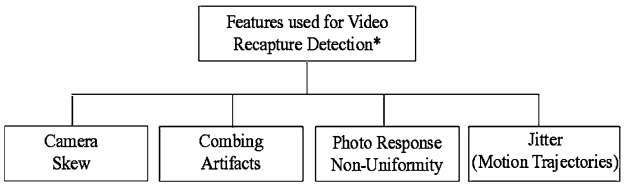

Fig. 8 Various features used for detecting re-captured videos. *In the literature, video re-capture detection is treated as a sub-task of forgery detection. However, since a re-captured video may or may not be a forgery, these features have been catalogued separately from the forgery detection features presented in Fig. 6

2.6 Technique for detecting video phylogeny

If two videos have the same content but are different in terms of attributes, such as size, resolution, and color, they are called 'near-duplicates' of one another. Given a set of near-duplicate videos, one might be interested in finding out the reason behind generating one video from another and to understand the causal association among these videos.

This problem was first presented for images and was called 'image phylogeny' [121] or 'image dependencies' [122]. In the context of videos, it is investigated as 'video phylogeny'. The first (and by far the only) published work that dealt with video phylogeny was [123]. The method involves finding a sturdy and informative dissimilarity function capable of comparing the given duplicates and extracting the differences that lie hidden within their content. The calculated dissimilarities were then arranged in the form of a tree, known as Video Phylogeny Tree (VPT). Experiments were conducted on 16 video commercials downloaded from YouTube, which were used to create a total of 265 test videos. The quantitative metrics used to evaluate the reconstructed tree and compare it to the ground truth were: Root, Leaves, Edges, and Ancestry. Out of the five algorithms proposed by the authors, the best one was able to locate the root of the phylogeny tree, i.e., the video from which the entire set originated, with an accuracy of 91% and could accurately categorized the leaves 77.7% of the time. In comparison with the ground truth, the algorithm could find 65.8% correct connections (edges) correct ancestry information was identified 70.4% of the time.

Given that this research field is still comparatively new, increased interest and further innovation in this domain is expected in the near future.

2.7 Video re-capture detection techniques

Re-acquisition or re-capture refers to the activity of capturing videos that are being reproduced on display monitors or projected on screens. According to [124], the challenge of

video re-capture detection is necessary to the field of digital forensics, since re-capture often indicates the existence of some previous tampering activity. Figure 8 presents the different features that have been used for detecting re-captured videos.

In [101], a notable technique for distinguishing original videos from re-captured ones was presented. For this, the authors utilized the concept of multiple view geometry. Videos re-captured with a camera placed off-axis with respect to the screen on which the video was being projected were discovered by observing discrepancies in the inherent parameters of the camera. The authors conducted several tests on simulated videos, and achieved good results. For completely noise-free sequences, instances of re-projection were correctly detected 84.9% of the times. In the presence of noise, detection accuracy varied from 85% with 0.3% false alarms to 88% with 0.4% false alarms. The authors performed only one test on a real-world video sequence and the reason was the immense complexity of the skew estimation process involved in such a scenario.

The work in [124] presented a re-capture detection technique, the basis of which was to analyze the high-frequency jitter that was introduced when a handheld camera was used to re-capture a video. The authors observed that this jitter (generated by shaky movements of the hand while holding the camera) resulted in high-frequency 2D motion fields that were almost uniform. By tracking features of the video and deriving correlation between these high-frequency components of the feature's trajectories, a trained model was then used to classify the videos as original or re-captured. The results demonstrated that the technique was indeed quite effective.

The algorithm in [125] addressed the challenge of determining if a given image might be a screenshot re-captured from an interlaced video. To accomplish this, the authors utilized the combing artifact of screenshots as an indicative of interlaced video re-capture. Combing artifacts are generated due to motion when even and odd scan lines are weaved together to create an interlaced video, and is considered to be one of the most representative feature of such videos. The algorithm was tested on TV programs and videos recorded using a camcorder. The detection accuracies of 4500 input screenshots from 20 MPEG-2, MPEG-4, and H.264 format videos were 98.1%, 97.3%, and 97.8%, respectively. Detection accuracies of 1,500 JPEG, TIFF, and BMP format input screenshots were 97.7%, 97.8%, and 97.9%, respectively. The average detection accuracy of the algorithm was 97.8%.

The authors in [126] detected video re-capture by exploiting the Photo Response Non-Uniformity (PRNU) of selected video shots. PRNU is a stochastic fingerprint unique to all image or video sensors. The method proceeded with determining which shots were recorded with a specific camcorder. Connections were made between shots that gradually led to the separation of original shots from the re-captured ones. For performance testing, the authors used four digital camcorders to re-capture ten original videos (generating a total of 40 re-captured videos). They also tested their technique on H.264/MPEG-4 AVC encoded videos. On an average, 100% detection accuracy was reported. This technique, however, did not account for post-re-capture geometric distortions in a given test video. If such a distortion completely destroys or partly damages the PRNU information, this technique would be rendered ineffective. In addition, the detection rate of the technique was found to be highly dependent on the quality of the video and the scaling factor used.

The same authors proposed another re-capture detection technique in [127]. The authors observed how geometric primitives, such as straight lines, get distorted while passing through the series of re-capture process. Based on this observation, they derived a mathematical curve model for a straight line distorted after single capture and then extended that model for re-captured lines. They tested their model to automatically extricate distorted straight lines for categorization and applied that information to synthetic video sequence.

A method to detect videos re-captured from an LCD monitor was suggested in [128]. This method had some advantages over the previously proposed techniques in the literature. Unlike [101], where camera was placed off-axis with the screen, the camcorder used in this approach was perfectly aligned with the screen, making the task of re-capture detection considerably harder (since there was no way to rely on geometrical inconsistencies to provide a solution). In addition, the authors kept their camcorder fixed on a tripod, unlike in [124], where handheld cameras produced jitter, which provided an important cue regarding the nature of the video. The idea was to use ghosting artifacts that were produced as a result of the lack of synchronization between the camera and the monitor. After experiments performed on 18 test videos, it was observed that the maximum accuracy of the method was 94% and the minimum was 89%, giving an average of 91%.

2.8 Comparative analysis of video forgery detection techniques

The forgery detection techniques proposed in the literature had all been validated on different databases consisting of videos of distinct characteristics and displaying diverse range of visual contents. The basal dissimilarities between the test videos can have significant effect on the outcomes of the given forensic method, and while the results reported in a particular work can provide an idea about the usefulness of the corresponding technique in a certain forensic

Table 2 Comparison of average accuracies (%) of copy–paste detection techniques at different bitrates and QFs

Forgery detection technique	Bitrates (Mbps)	Quality factors		
		60	80	100
De et al. [38]	3	54.9	62.9	74.3
	6	56.8	65.6	79.1
	9	58.7	70.5	82.0
Hsu et al. [82]	3	66.8	76.2	83.9
	6	69.3	79.1	86.6
	9	72.5	80.8	88.3
Goodwin and Chetty [84]	3	75.1	82.3	88.3
	6	78.4	85.2	90.6
	9	80.7	87.0	93.5
Bestagini et al. [86]	3	69.9	75.9	82.6
	6	71.3	77.6	85.2
	9	73.5	80.3	89.1
Lin and Tsay [87]	3	58.7	65.3	72.1
	6	60.1	69.4	74.2
	9	62.8	71.3	76.7
Zhang et al. [94]	3	62.0	72.5	85.3
	6	65.3	75.4	87.2
	9	70.1	78.3	88.7
Bidokhti and Ghaemmaghami [98]	3	54.8	65.3	76.1
	6	58.2	68.7	79.9
	9	62.5	73.3	82.1

Table 3 Effect of number of inserted/deleted/duplicated frames on the average accuracies (%) of inter-frame forgery detection techniques at different QFs

Forgery detection technique	QF	Number of deleted/inserted/duplicated frames		
		30	65	100
De et al. [38]	100	71.9	82.3	86.1
	80	65.8	69.3	72.3
	60	57.9	60.5	63.5
Chao et al. [62]	100	86.6	88.5	91.6
	80	74.6	78.3	81.7
	60	68.9	69.5	71.1
Lin et al. [75]	100	83.9	85.4	88.0
	80	77.7	80.3	82.2
	60	70.8	72.0	75.5
Wang et al. [63]	100	79.1	82.9	85.3
	80	72.0	74.2	76.4
	60	67.1	69.6	70.5
Singh and Aggarwal [72]	100	95.6	97.2	99.3
	80	94.1	96.5	98.9
	60	92.0	94.4	96.3

scenario, they cannot help predict its behavior in a completely different setting.

In an attempt to determine the extent of applicability of different forensic schemes, we perform comparative analysis of some forgery detection techniques, which we believe to be among the most innovative and representative advancements documented in the literature.

2.8.1 *Experimental environment*

All these techniques were implemented by adhering to the specifications, assumptions, variable parameters settings, and threshold values suggested by the respective authors, and were validated on test videos form two databases [8, 34]. All the test videos were originally MJPEG and H.264/AVC encoded, with resolution 320×240 pixels, and had been recorded at 30 FPS. While the forged sequences in [8] were MJPEG and H.264/AVC encoded as well, the forged sequences in [34] were encoded to MPEG-2 and H.264/AVC formats using *FFmpeg* [129], after the forgery. Overall, a total of 530 test videos were used during comparative analysis. All these videos exhibit several simple and complex life-like scenarios, and represent both indoor and outdoor scenes. All the forgeries have been created in a plausible manner so as to simulate realistic forensic scenarios. Experimentation was performed in MATLAB v. R2015b (8.6.0.267246), and the results reported in Tables 2 and 3 represent the average detection accuracies obtained.

2.8.2 *Comparative analysis of copy–paste forgery detection techniques*

We analyzed the performances of the following copy–paste detection techniques: the noise-based approaches proposed in [38, 82], the noise and quantization residue-based scheme of [84], motion-residue-based approach proposed in [86], the pixel-coherence analysis technique[2] suggested in [87], the object-based technique suggested in [94], and the optical-flow-based method proposed in [98].

Table 2 presents a comparative summary of the outcomes, as a function of various compression quality factors (QF) and bitrates.

2.8.3 *Comparative analysis of inter-frame forgery detection techniques*

We analyzed the performances of the following inter-frame forgery detection techniques: the pixel-correlation based approach proposed in [75], the noise-based method

[2] Pseudo-codes of these techniques are available in the respective papers.

of [38], and the optical-flow-based schemes presented in [62, 72, 632, 72]. Table 3 presents a comparative summary of the outcomes as a function of various QFs and number of inserted/deleted/duplicated frames. For these tests, the bitrate of the videos was fixed at 5 Mbps. Note that the aforementioned techniques have been designed to detect different kinds of inter-frame forgeries. While the technique in [35, 59] detect frame-insertion and deletion, that in [75] detects frame duplication. The methods proposed in [63, 72] detect all three kinds of inter-frame forgeries. The results in Table 3 pertain to the outcomes of the experiments performed with respect to the specific kinds of forgeries each of these techniques detect.

To ascertain the efficacy of a forgery detection technique, it needs to be tested on plausible forgeries that simulate real-world forensic scenarios. For a video recorded at 25 or 30 FPS, removal of even one second of would require deletion of at least 25–30 frames. Removal of a slow moving object would require even more frames to be deleted. The same rationale is valid in case of frame insertion or duplication. Therefore, the forged videos used during this set of experiments were created by inserting, duplicating, or removing 30–100 frames into or from the original videos.

Tables 2 and 3 represent the results we obtained after testing the selected techniques in a neutral setting. We can observe a notable depreciation in the detection accuracies of some of these techniques in comparison with the performances reported in the respective papers. This is mainly due to the fact that these techniques were based on several content-dependent parameters and thresholds, and their performances were negatively affected when the basal attributes of test videos changed. Those techniques that did not rely too heavily on the visual contents or characteristics of the test data generated comparable results [70, 77, 79, 89].

3 Open issues and future challenges

The astonishing growth in the creation and use of multimedia data in today's world demands a parallel if not superior progress in the field of digital data forensics.

Over the years, image forensics has come a long way from simple spatial-analysis-based forgery detection [130] to more untraditional tamper detection approaches, such as fuzzy-fusion theory [131] and Extreme Machine Learning (EML)-based watermarking [132].

Videos, inherently not being as easy to work with as images, present their own unique set of difficulties. A lot has been achieved over the past few years, but certain milestones still remain to be reached. The research spheres that suffer from some unresolved issues have been summarized in the following.

3.1 Inefficacious management of videos with variable GOP structures

The performances of a vast majority of the methodologies discussed in this survey depend on the codecs used for compressing the video sequences. Most of the video encoders use fixed GOP structures, because they are easier to implement, and therefore, many of the techniques in the literature work under the assumption that the test video under investigation consists of GOPs with fixed number of frames [42, 44, 47, 49, 53, 57, 59, 60, 66, 74, 89, 98]. However, prevalent compression standards, such as H.264/MPEG-4 AVC, use adaptive GOP structures, and the lengths of these GOPs can be up to 250 frames (depending on the amount and frequency of change in the video content). Consequently, such techniques may fail entirely for videos encoded using H.264/MPEG-4 codecs. Furthermore, algorithms that exploit abnormal changes in motion or noise residue, and relocated I-frames are also unsuitable for this encoding standard, on account of the adaptive GOP structure. Evidently, there is a dire need for more powerful algorithms that are free from the assumption of fixed GOP structures, so that they can not only work effectively for video encoded using modern codecs but can also overcome another persistent constraint faced by several contemporary forensic solutions, i.e., the inability to detect removal of entire GOPs or multiples of GOP lengths.

3.2 Inadequate experimentation on realistically doctored video sequences

A major shortcoming of many state-of-the-art methodologies is that they lack adequate validation on realistically tampered videos. Fabricating manually forged videos is highly time-intensive and thus most of the authors ran experiments on synthetically doctored sequences. Verifying the integrity of such material is sometimes quite easy and could be achieved with just a simple visual examination.

The method in [77] was tested on TV broadcast videos to ascertain the effectiveness of the detector in a real-life situation. The results were acceptable for slight compression only. Surveillance-like test videos with static backgrounds were used for testing in [83], but the detection accuracy rates were inadequate. The method in [99] was experimented on RGB and infrared videos, and although good detection accuracy was reported, the method of creating partially manipulated frames was found to be inherently erroneous, as discussed earlier in Sect. 2.2.2. Desirable detection accuracy was reported in [97], where the proposed algorithm was tested on videos captured by static

surveillance cameras. The only techniques in the literature that generated satisfactory results for realistic tampered videos taken from a standard data set SULFA [8] were [72, 86, 88].

3.3 Lack of multifaceted forensic systems

Digital video forensics is often considered to be still in its primitive stages. Digital forgery detection is a very complicated task and the absence of a universally applicable solution exacerbates the situation. Every technique proposed in the literature has been exclusively designed to tackle specific kinds of forgeries or tampering attacks. However, in a real-life scenario, the authenticity of a digital video needs to be established without any prior knowledge regarding the kind of forgery it might be suffering from. A video may have undergone multiple tamper attacks. Therefore, to provide a real-world solution to the forgery detection challenge, a multifaceted forensic system is required, which is composed of multiple forgery detection techniques and all the specialized constituent techniques are responsible for detecting the kind of forgeries that they have been designed to handle.

The first (and by far the only) remedial step in this direction was taken in [100], where several forensic techniques developed by the authors over the years were combined into one composite forensic system. In spite of various limitations and unpractical assumptions which restricted this system's wide-spread applicability, this pioneer work still demonstrated its ability to serve as a foundation for further advances towards development of comprehensive tamper detection systems.

In a real-life situation, where the forensic examiner is oblivious to the nature of forgeries present in the given test video (if present at all), and stringent conditions typical of a laboratory experiment are not enforceable, a comprehensive forensic system could be of immeasurable value. With the help of such a system, the investigator could inspect the video, analyze the forensic evidence provided by each of the constituent forensic technique, and then reach a decision regarding the veracity of the given content, based on whether or not the video showed signs of any of the possible forgeries it was tested for.

In the near future, integrity verification schemes, such as those based on the concept of utilizing the information provided by Video Event Data Recorders (VEDRs) [133], can also be incorporated into the existing passive forgery detection structures to develop all the more robust and reliable forensic investigation mechanisms.

Furthermore, to handle epistemic uncertainties in the decision-making process, mathematical theories, such as Dempster–Shafer theory of evidence [134, 135], can be utilized to effectively consolidate incomplete, inaccurate or even contradictory evidence provided by multiple features or artifacts, and establish the content's authenticity with a higher degree of conviction. Decision-fusion schemes have been successfully implemented in the digital image forensics domain [136–139], and finally, the suitability of Dempster-Shafer Theory (DST) has also been explored [140–142]. The results are encouraging; it has been demonstrated that when tested on realistic data sets, DST outperforms several decision-fusion schemes, including logical disjunction-based and SVM-based fusion approaches [141, 142]. DST has also been shown to produce favorable results in the counter anti-forensics domain [143]. The field of video forensics would undeniably benefit from similar advancements.

3.4 Insufficient anti-forensic and counter anti-forensic strategies

As stated earlier in Sect. 2.4, while the field of image anti-forensics has received its fair share of attention, similar interest in video anti-forensics domain has not been observed. References [71, 112–114, 117] remain the only works that deal with this issue. Moreover, the algorithm proposed in [110] is the only attempt at countering the anti-forensic technique proposed in [113]. The authors in [117] suggested a counter anti-forensic strategy that was specifically designed to counter their own anti-forensic technique. In [71], the authors proposed certain forgery detection techniques, and then suggested possible anti-forensic and suitable counter anti-forensic schemes.

Such lack of innovation in this field is surprising and causes us to become skeptical of the resilience of the techniques surveyed in this paper towards clever modifications.

Here, we would like to suggest a viable anti-forensic strategy. The forensic techniques presented in Sect. 2.1.1.2 detected frame manipulation by looking for evidence of double compression. However, if a forger alters the encoded video directly and creates no second-quantization spatial and/or temporal artifacts, the research in [42, 46, 50, 113, 119] would be rendered ineffectual. Furthermore, the anti-forensic approach proposed in [111] seems to have the potential to be molded into a potent digital video anti-forensic tool.

At this point, our understanding of the dynamics of forensic and anti-forensic evolution equips us to make the following observation: the tendency of anti-forensic operations to leave their own detectable traces in the digital content causes forgers to compromise between complete removal of evidence of forgery and introduction of new evidence of anti-forensic manipulation. Likewise, since forensic investigators have to limit the likelihood of false alarms, they end up compromising between forgery detection accuracy and accuracy of detecting anti-forensic

operations. This knowledge could prove to be highly useful for elevating the level of innovation in the field of counter anti-forensics.

3.5 Complete oversight of the audio component of digital videos

Although the visual contents of digital videos help us form opinions and take decision in several sectors of litigation and criminal justice (among other areas), the role of audio data in the decision-making process cannot be ignored. In all the video forgery detection techniques proposed in the literature so far, the audio aspect of digital videos has not been taken under consideration. All these techniques focus entirely on the visual content of the test sequences, and even though the video forensics domain has achieved several milestones by doing so, the audio component of videos, whenever present, is certainly capable of providing valuable clues regarding content authenticity.

For instance, one of the forged test sequences[3] available in SULFA [8] shows a street scene, where a vehicle, which was originally moving across the frame on the street, had been removed. In the forged sequence, no easily discernible clue of the missing vehicle can be noticed, except for the sound that the vehicle made when it passed by on the street. Now, this may very well have been an oversight on the part of the forger, but it does prove a very important point: audio data, if available in the test video, must not be ignored in a situation where forgery detection is the key objective. This particular domain of forgery detection is in immediate need of attention. There are several noteworthy developments in the field of audio authentication (details of which can be obtained from the survey [144]), which are currently concentrating on detection of audio manipulations in audio recordings, recordings of phone conversations, and commercially distributed audio files (songs and music albums). If combined with frame-based analyses that are generally performed in the video forensics domain, such techniques (or their more flexible and potent modified versions) could certainly help further the cause of video content authentication.

3.6 Overall lack of vigor, potentially induced by lack of standardized databases

In all the techniques surveyed in this paper, the detection performances were not subjected to as severe analysis as is commonly observed in other fields of digital video processing. When it came to strict inspection and comparisons of the detection accuracies of the systems, a general lack of eagerness was observed, which indicates that a research field as vital as this one warrants a far greater exposure and commitment than it is currently receiving.

Perhaps, the main reason for this inadequate enthusiasm was the lack of a large-scale video library or standard data set that could provide a neutral platform for unbiased comparison of various forgery detection techniques. To the best of the authors' knowledge, the only such publically available data set is SULFA [8]. However, even SULFA was found to be inadequate, not only because it is still in its early stages of development but also because it provides only a few forged videos exclusively constructed for the purpose of tamper detection. Most of the videos are dedicated towards tasks, such as source camera identification. In [97], the authors tested their technique on a self-created database called SYSU-OBJFORG, although this database has not been made publically available yet. For advanced video forensic evaluation, a comprehensive collection of tampered videos similar to the image forensic databases of [145–148] is required.

4 Conclusion

This paper presents a repository of information regarding the kinds of tamper attacks a video can suffer from and a comprehensive source of references for the passive-blind techniques proposed for detecting such attacks. The domain of video anti-forensics and counter anti-forensics has also been explored. Along with the analysis of each technique's most consequential drawbacks and practical advantages, the limitations that need to be overcome in the long-term perspective and some open issues that require immediate attention have also been discussed. Some important interconnected research domains, such as video up-conversion, phylogeny, and re-capture detection, have been overviewed as well. We believe that this work could not only prove useful to the researches and developers working in the video forensics domain to find new utilitarian ideas and identify novel research challenges, but also motivate new researchers to partake in this tremendously exciting research domain of incalculable worth.

References

1. Kwatra, V., Schödl, A., Essa, I., Turk, G., Bobick, A.F.: Graph cut textures image and video synthesis using graph cuts. ACM Trans. Graph. **22**(3), 277–286 (2003)
2. Pèrez, P., Gangnet, M., Blake, A.: Poisson image editing. ACM Trans. Graph. (SIGGRAPH'03). **22**(3), 313–318 (2003)

[3] This test sequence is available in the SULFA database under the name "van_car".

3. Criminisi, A., Pèrez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Trans. Image Process. **13**(9), 1200–1212 (2004)
4. Shen, Y., Lu, F., Cao, X., Foroosh, H.: Video completion for perspective camera under constrained motion. In Proceedings of 18th IEEE International Conference on Pattern Recognition (ICPR'06). Hong Kong, China, pp. 63–66. (2006)
5. Komodakis, N., Tziritas, G.: Image completion using efficient belief propagation via priority scheduling and dynamic pruning. IEEE Trans. Image Process. **16**(11), 2649–2661 (2007)
6. Patwardhan, K.A., Sapiro, J., Bertalmio, M.: Video inpainting under constrained camera motion. IEEE Trans. Image Process. **16**(2), 545–553 (2007)
7. Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Trans. Graph. (SIGGRAPH'07) **26**(3), 1–7 (2007)
8. Qadir, G., Yahaya, S., Ho, A.T.S.: Surrey University Library for Forensic Analysis (SULFA) of video content. In IET Conference on Image Processing (IPR '12). London, UK, pp. 1–6. (2012). http://sulfa.cs.surrey.ac.uk/. Accessed 23 Mar 2016
9. [Online]. http://imgur.com/gallery/4zeEy. Accessed 3 July 2016
10. Xu, J., Yu, Y., Su, Y., Dong, B., You, X.: Detection of blue screen special effects in videos. In: Proceedings of International Conference on Medical Physics and Biomedical Engineering, Beijing, China, 1316–1322 (2012)
11. 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. **43**(4), 26 (2011)
12. Farid, H.: Digital doctoring: How to tell the real from fake. Significance. **3**(4), 162–166. (2006)
13. Ng, T.-T., Chang, S.-F., Lin, C.-Y., Sun, Q.: Passive-blind image forensics. In: Zeng, W., Yu, H., Lin, C.-Y. (eds.) Multimedia security technologies for digital rights. Elsevier, Hawthorne (2006)
14. Lanh, T.V., Chong, K.S., Emmanuel, S., Kankanhalli, M.S.: A survey on digital camera image forensic methods. In Proceedings of IEEE International Conference on Multimedia and Expo (ICME'07). Beijing, China, pp. 16–19 (2007)
15. Luo, W., Qu, Z., Pan, F., Huang, J.: A survey of passive technology for digital image forensics. Front. Comput. Sci. China **1**(2), 166–179 (2007)
16. Zhang, Z., Ren, Y., Ping, X.J., He, Z.Y., Zhang, S.Z.: A survey on passive-blind image forgery by doctor method detection. In Proceedings of 7th International Conference on Machine Learning and Cybernetics. Kunming, China, pp. 3463–3467 (2008)
17. Sencar, H.T., Memon, N.: Overview of state-of-the-art in digital image forensics, part of indian statistical institute platinum jubilee monograph series titled 'statistical science and interdisciplinary research', pp. 1–20. World Scientific Press, Japan (2008)
18. Mahdian, B., Saic, S.: Blind methods for detecting image fakery. In: Proceedings of 42nd Annual IEEE International Carnahan Conference on Security Technology. Prague, Czech Republic, pp. 280–286 (2008)
19. Farid, H.: A survey of image forgery detection. IEEE Signal Process. Mag. 2(26), 16–25 (2009)
20. Christlein, V., Riess, C., Angelopoulou, E.: A Study on features for the detection of copy-move forgeries. In: Sicherheit, F.C. Freiling (ed.) Gesellschaft fur Informatik e.V., Bonn, Belin, Germany, pp. 105–116 (2010)
21. Granty, R.E.J., Aditya, T.S., Madhu, S.S.: Survey on passive methods of image tampering detection. In: Proceedings of IEEE International Conference on Communication and Computational Intelligence (INCOCCI'10). Erode, India. pp. 431–436 (2010)
22. Mahdian, B., Saic, S.: A bibliography on blind methods for identifying image forgery. Signal Process. Image Commun. **25**(6), 389–399 (2010)
23. Poisel, R., Tjoa, S.: Forensics investigations of multimedia data: a review of the state-of-the-art. In Proceedings of 6th International Conference on IT Security Incident Management and IT Forensics (IMF'11). Stuttgart, Germany, pp. 48–61 (2011)
24. Birajdar, G.K., Mankar, V.H.: Digital image forgery detection using passive techniques: A survey. Digit. Investig. **10**(3), 226–245 (2013)
25. Qazi, T., Hayat, K., Khan, S., Madani, S.A., Khan, I.A., Kołodziej, J., Li, H., Lin, W., Yow, K.C., Xu, C.-Z.: Survey on blind image forgery detection Tanzeela. IET Image Process. 1–11 (2013)
26. Ansari, M.D., Ghrera, S.P., Tyagi, V., Pixel based image forgery detection: a review. IETE J. Educ. Taylor Francis **55**(1), 40–46 (2014)
27. Asghar, K., Habib, Z., Hussain, M.: Copy-move and splicing image forgery detection and localization techniques: a review. Aust. J. Forensic Sci. (2016). doi:10.1080/00450618.2016.1153711
28. Milani, S., Fontani, M., Bestagini, P., Barni, M., Piva, A., Tagliasacchi, M., Tubaro, S.: An overview on video forensics. APSIPA Trans. Signal Inf. Process. **1**(1), 1–18 (2012)
29. 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 10th International Conference on Information assurance and security, Okinawa, Japan, pp. 29–34 (2014)
30. Joshi, V., Jain, S., Tampering detection in digital video e a review of temporal fingerprints based techniques. In: Proceedings of 2nd International Conference on Computing for sustainable global development, New Delhi, India, pp. 1121–1124 (2015)
31. Sitara, K.K., Mehtre, B.M.: Digital video tampering detection: an overview of passive techniques. Digit. Investig. **18**, 8–22 (2016)
32. Stamm, M.C., Wu, M., Liu, K.J.R, Information Forensics: An Overview of the first decade. Access IEEE. **1**, 167–200 (2013)
33. Video Trace Library [Online]. http://trace.eas.asu.edu/. Accessed 7 July 2016
34. [Online]. https://1drv.ms/f/s!Aj8xYEfdOdJ-i0xnQ5YSlUpZl-HrT. Accessed 2 May 2016
35. Kurosawa, K., Kuroki, K., Saitoh, N.: CCD fingerprint method-identification of a video camera from videotaped images. In: Proceedings of IEEE International Conference on Image Processing, Kobe, Japan, pp. 537–540 (1999)
36. Lukáš, J., Fridrich, J., Goljan, M.: Digital camera identification from sensor pattern noise. IEEE Trans. Inf. Forensics Secur. **1**(2), 205–214 (2006)
37. Goljan, M., Chen, M., Comesaña, P., Fridrich, J.: Effect of compression on sensor-fingerprint based camera identification. Electron. Imag. 1–10 (2016)
38. De, A., Chadha, H., Gupta, S.: Detection of forgery in digital video. In: Proceedings of 10th World Multi Conference on Systems, Cybernetics and Informatics. V, pp. 229–233 (2006)
39. Wang, W., Farid, H.: Exposing digital forgeries in interlaced and deinterlaced video. IEEE Trans. Inf. Forensics Secur. **2**(3), 438–449 (2007)
40. Mondaini, N., Caldelli, R., Piva, A., Barni, M., Cappellini, V.: Detection of malevolent changes in digital video for forensic applications. In: Delp, E.J., Wong, P.W., (eds.) Proceedings of SPIE Conference on Security, Steganography and Watermarking of Multimedia Contents. Vol. 6505, No. 1 (2007)
41. Kobayashi, M., Okabe, T., Sato Y.: Detecting forgery from static-scene video based on inconsistency in noise level

functions. IEEE Trans. Inf. Forensics Secur. **5**(4), 883–892 (2010)
42. Wang, W., Farid, H.: Exposing digital forgeries in video by detecting double MPEG compression. In: Voloshynovskiy, S., Dittmann, J., Fridrich, J.J. (eds.) Proceedings of 8th Workshop on Multimedia and Security (MM&Sec'06). ACM Press, New York, NY, pp. 37–47 (2006)
43. Fu, D., Shi, Y.Q., Su., W.: A generalized Benford's law for jpeg coefficients and its applications in image forensics. In: Delp, E.J., Wong, P.W., (eds.) Proceedings of SPIE Security, Steganography and Watermarking of Multimedia Contents IX. Vol. 6505, San Jose, CA, pp. 39–48 (2007)
44. Luo, W., Wu, M., Huang, J.: MPEG recompression detection based on block artifacts. In: Delp, E.J., Wong, P.W., Dittmann, J, Memon, N.D., (eds.) Proceedings of SPIE Security, Forensics, Steganography, and Watermarking of Multimedia Contents X. Vol. 6819, San Jose, CA (2008)
45. Wang, W., Farid, H: Exposing digital forgeries in video by detecting double quantization. In: Proceedings of 11th ACM Workshop on Multimedia and Security. ACM Press, New York, NY, pp. 39–48 (2009)
46. Su, Y., Xu, J.: Detection of double-compression in MPEG-2 videos. In: Proceedings of 2nd International Workshop on Intelligent Systems and Applications. Vol. 1, no. 4, pp. 22–23 (2010)
47. Su, Y., Nie, W., Zhang, C.: A frame tampering detection algorithm for MPEG videos. In Proceedings of 6th IEEE Joint International Information Technology and Artificial Intelligence Conference, Chongqing, China vol. 2, pp. 461–464 (2011)
48. [Online]. http://www.its.bldrdoc.gov/vqeg/. Accessed 13 Apr 2016
49. Vázquez-Padín, D., Fontani, M., Bianchi, T., Comesana, P., Piva, A., Barni, M.: Detection of video double encoding with GOP size estimation. In: Proceedings on IEEE International Workshop on Information Forensics and Security, Tenerife, Spain, 151 (2012)
50. Sun, T., Wang, W., Jiang, X.: Exposing video forgeries by detecting mpeg double compression. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, pp. 1389–1392 (2012)
51. Xu, J., Su, Y., You, X.: Detection of video transcoding for digital forensics, In: Proceedings of International Conference on Audio, Language and Image Processing, Shanghai, China, pp. 160–164 (2012)
52. Milani, S., Bestagini, P., Tagliasacchi, M., Tubaro, S.: Multiple compression detection for video sequences. In: Proceedings of 14th IEEE International Workshop on Multimedia Signal Processing, Banff, AB, pp. 112–117 (2012)
53. Shanableh, T.: Detection of frame deletion for digital video forensics. Digit. Investig. **10**, 350–360 (2013)
54. Jiang, X., Wang, W., Sun, T., Shi, Y.Q., Wang, S.: Detection of double compression in MPEG-4 videos based on Markov statistics. IEEE Signal Process. Lett. **20**(5), 447–450 (2013)
55. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Ann. Eugen. **7**(2), 179–188 (1936)
56. Chen, W., Shi, Y.Q.: Detection of double MPEG compression based on first digit statistics. Digit. Watermarking **5450**, 16–30 (2009)
57. Gironi, A., Fontani, M., Bianchi, T., Piva, A., Barni, M.: A video forensic technique for detecting frame deletion and insertion. In: Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, Florence, Italy, pp. 6267–6271 (2014)
58. Labartino, D., Bianchi, T., Rosa, A.D., Fontani, M., Vázquez-Padín, D., Piva, A., Barni, M.: Localization of forgeries in MPEG-2 video through GOP size and DQ analysis. In: Proceedings of IEEE 15th International Workshop on Multimedia Signal Processing, Pula, Italy. pp. 494–499 (2013)
59. Su, Y., Zhang, J., Liu, J.: Exposing digital video forgery by detecting motion-compensated edge artifact. In: Proceedings of International Conference on Computational Intelligence and Software Engineering, Wuhan, China. Vol. 1, no. 4, pp. 11–13 (2009)
60. Dong, Q., Yang, G., Zhu, N.: A MCEA based passive forensics scheme for detecting frame-based video tampering. Digit. Invest **9**(2), 151–159 (2012)
61. Kancherla, K., Mukkamal, S.: Novel blind video forgery detection using Markov models on motion residue. Intell. Inf. Database Syst. **7198**, 308–315 (2012)
62. Chao, J., Jiang, X., Sun, T.: A novel video inter-frame forgery model detection scheme based on optical flow consistency. Digit. Forensics Watermarking. **7809**, 267–281 (2013)
63. Wang, W., Jiang, X., Wang, S., Meng, W.: Identifying video forgery process using optical flow, Digital Forensics and Watermarking. pp. 244–257. Springer, Berlin Heidelberg, (2014)
64. TREC Video Retrieval Evaluation [Online]. http://trecvid.nist.gov/. Accessed 15 Apr 2016
65. Wu, Y., Jiang, X., Sun, T., Wang, W.: Exposing video inter-frame forgery based on velocity field consistency. In: Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing, Florence, Italy, pp. 2693–2697 (2014)
66. Liu, H., Li, S., Bian, S.: Detecting frame deletion in H.264 video. In: Proceedings of 10th International Conference, ISPEC Fuzhou, China, pp. 262–270 (2014)
67. Consumer digital video library [Online]. http://www.cdvl.org/. Accessed 9 July 2016
68. Zheng, L., Sun, T., Shi, Y-Q.: Inter-frame video forgery detection based on blockwise brightness variance descriptor. In: Proceedings of 13th International Workshop on Digital-forensics and watermarking, Taipei, Taiwan, 2014, Revised Selected Papers. Springer International Publishing, pp. 18–30 (2015)
69. Recognition of human actions database. [Online]. http://www.nada.kth.se/cvap/actions. Accessed 3 May 2016
70. Gupta, A., Gupta, S., Mehra, A.: Video authentication in digital forensic. in Proceedings of International Conference on Futuristic Trends on computational analysis and knowledge management (ABLAZE), Noida, India. pp. 659–663 (2015)
71. Kang, X., Liu, J., Liu, H., Wang, Z.J., Forensics and counter anti-forensics of video inter-frame forgery. Multimed Tools Appl. **75**(21), 1–21 (2015)
72. Singh, R.D., Aggarwal, N.: Detection of Re-Compression, Transcoding and Frame-Deletion for Digital Video Authentication. In: Proceedings of 3rd International Conference on Recent Advances in Engineering and Computer Sciences. Chandigarh India, pp. 1–6 (2016)
73. Change detection video database. [Online]. http://changedetection.net/. Accessed 1 June 2016
74. Aghamaleki, J.A., Behrad, A.: Malicious inter-frame video tampering detection in MPEG videos using time and spatial domain analysis of quantization effects, Multimed Tools and Applications, pp. 1–27 (2016)
75. Lin, G.-S., Chang, J.-F., Chuang, F.-H.: Detecting frame duplication based on spatial and temporal analyses. In: Proceedings of 6th IEEE International Conference on Computer Science and Education (ICCSE'11), SuperStar Virgo, Singapore, pp. 1396–1399 (2011)
76. 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. **2**(4), 51–57 (2014)
77. Bestagini, P., Battaglia, S., Milani, S., Tagliasacchi, M., Tubaro, S.: Detection of temporal interpolation in video sequences. In: Proceedings of IEEE International Conference

on Acoustics, Speech and Signal Processing (ICASSP'13). Vancouver, BC (2013)
78. Yao, Y., Yang, G., Sun, X., Li, L.: Detecting video frame-rate up-conversion based on periodic properties of edge-intensity. J. Inf. Secur. Appl. **26**, 39–50 (2016)
79. Xia, M., Yang, G., Li, L., Li, R., Sun, X.: Detecting video frame rate up-conversion based on frame-level analysis of average texture variation. Multimed. Tools Appl. **72**(1), 1–23 (2016)
80. Xiph.org Video Test Media (derf's collection) [Online]: http://media.xiph.org/video/derf/. Accessed 2 May 2016
81. Wang, W., Farid, H.: Exposing digital forgeries in video by detecting duplication. In: Kundur, D., Prabhakaran, B, Dittmann, J., Fridrich, J.J. (eds.) Proceedings of 9th ACM workshop on Multimedia & Security (MM&Sec'07), ACM Press, New York, NY, pp. 35–42 (2007)
82. Hsu, C.-C., Hung, T.-Y., Lin, C.-W., Hsu, C.-T.: Video forgery detection using correlation of noise residue. In: Proceedings of 10th IEEE Workshop on Multimedia Signal Processing. Cairns, Australia, pp. 170–174 (2008)
83. Chetty, G., Blind and passive digital video tamper detection based on multimodal fusion. In: Proceedings of 14th WSEAS International Conference on Communications. Corfu, Greece, pp. 109–117 (2010)
84. Goodwin, J., Chetty, G., Blind video tamper detection based on fusion of source features. In: Proceedings of IEEE International Conference on Digital image computing techniques and applications (DICTA), Noosa, QLD, pp. 608–613 (2011)
85. Das, S., Darsan, G., Shreyas L., Devan, D.: Blind Detection Method for Video Inpainting Forgery. Int. J. Comput. Appl. **60**(11), 33–37 (2012)
86. Bestagini, P., Milani, S., Tagliasacchi, M., Tubaro, S.: Local tampering detection in video sequences. In: Proceedings of 15th IEEE International Workshop on Multimedia Signal Processing. Pula, pp. 488–493 (2013)
87. 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. **1**(2), 120–140 (2014)
88. Pandey, R.C., Singh, S.K., Shukla, K.K.: Passive copy-move forgery detection in videos. In: Proceedings of IEEE 5th International Conference on Computer and Communication Technology, Allahabad, India. pp. 301–306 (2014)
89. D'Amiano, L., Cozzolino, D., Poggi, G., Verdoliva, L.: Video forgery detection and localization based on 3D patchmatch. In: Proceedings of IEEE International Conference on Multimedia expo workshops (ICMEW) Turin, Italy, pp. 1–6 (2015)
90. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B., Patchmatch: A randomized correspondence algorithm for structural image editing, ACM Trans. Graph. **28**(3) (2009)
91. Bleyer, M., Rhemann, C., Rother, C.: Patchmatch stereo–stereo matching with slanted support windows. In: Proceedings of British Machine Vision Conference, pp. 1–11 (2011)
92. Newson, A., Almansa, A., Fradet, M., Gousseau, Y., Pérez, P.: Towards fast, generic video inpainting. In: Proceedings of the 10th European Conference on Visual Media Production (2013)
93. REWIND Database. [Online]. https://sites.google.com/site/rewindpolimi/downloads/datasets/videocopy-move-forgeries-dataset. Accessed 18 May 2016
94. Zhang, J., Su, Y., Zhang, M.: Exposing digital video forgery by ghost shadow artifact. In: Proceedings of 1st ACM Workshop on Multimedia in Forensics (MiFor'09). ACM Press, New York, NY, pp. 49–54 (2009)
95. Conotter, V., O'Brien, J.F., Farid, H.: Exposing digital forgeries in ballistic motion. In: IEEE Transactions on Information Forensics and Security, Part 2, vol. 7, no. 1, pp. 283–296 (2012)
96. Richao, C., Gaobo, Y., Ningbo, Z.: Detection of object-based manipulation by the statistical features of object contour. Forensic Sci. Int. **236**, 164–169 (2014)
97. Chen, S., Tan, S., Li, B., Huang, J.: Automatic detection of object-based forgery in advanced video. In: IEEE Transactions on Circuits., Systems and Video Technology, Vol. 99 (2015)
98. Bidokhti, A., Ghaemmaghami, S.: Detection of regional copy/move forgery in MPEG videos using optical flow. In: International symposium on Artificial intelligence and signal processing (AISP), Mashhad, Iran, pp. 13–17 (2015)
99. Hyun, D.-K., Ryu, S.-J., Lee, H.-Y., Lee, H.-K.: Detection of upscale-crop and partial manipulation in surveillance video based on sensor pattern noise. Sensors **13**, 12605–12631 (2013)
100. Wang, W.: Digital Video Forensics. PhD Dissertation. Department of Computer Science. Dartmouth College, Hanover, New Hampshire (2009)
101. Wang, W., Farid, H.: Detecting re-projected video. In: Information Hiding, Solanki, K., Sullivan, K., Madhow, U. (eds.) Lecture Notes in Computer Science, Vol. 5284, pp. 72–86, Springer, Berlin (2008)
102. Kirchner, M., Böhme, R.: Hiding traces of resampling in digital images. IEEE Trans. Inf. Forensics Secur. **3**(4), 582–592 (2008)
103. Kirchner, M., Böhme, R.: Synthesis of color filter array pattern in digital images. In: Proceedings of SPIE-IS&T Electronic Imaging: Media Forensics and Security, 7254 (2009)
104. Cao, G., Zhao, Y., Ni, R., Tian, H.: Anti-forensics of contrast enhancement in digital images. In: Proceedings of 12th ACM Workshop on Multimedia and Security, Rome, Italy. pp. 25–34 (2010)
105. Stamm, M.C., Liu, K.J.R.: Wavelet-based image compression anti-forensics, In: Proceedings of 17th IEEE International Conference on Image Processing (ICIP'10). Hong Kong, China, pp. 1737–1740 (2010)
106. Stamm, M.C., Tjoa, S.K., Lin, W.S., Liu, K.J.R.: Anti-forensics of JPEG compression. In: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP'10). Dallas, TX, pp. 1694–1697 (2010)
107. Stamm, M.C., Tjoa, S.K., Lin, W.S., Liu, K.J.R.: Undetectable image tampering through JPEG compression anti-forensics. In Proceedings of IEEE International Conference on Image Processing (ICIP'10). Hong Kong, China, pp. 2109–2112 (2010)
108. Stamm, M.C., Liu, K.J.R.: Anti-forensics of digital image compression. IEEE Trans. Inf. Forensics Secur. **6**(3), 1050–1065 (2011)
109. Goljan, M., Fridrich, J., Chen, M.: Defending against fingerprint copy attack in sensor-based camera identification. IEEE Trans. Inf. Forensics Secur. **6**(1), 227–236 (2011)
110. Böhme, R., Kirchner, M.: Counter-Forensics: Attacking Image Forensics. In: Sencar, H.T., Memon, N. (eds.) Digital image forensics, pp. 327–366. Springer, New York (2013)
111. Fan, W., Wang, K., Cayere, F., et.al.: A variational approach to JPEG anti-forensics. In: Proceedings of IEEE 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP'13),Vancouver, Canada, pp. 3058–3062 (2013)
112. Stamm, M.C., Liu, K.J.R.: Anti-forensics for frame deletion/addition in mpeg video. In: Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP'11). Prague, Czech Republic, pp. 1876–1879 (2011)
113. Stamm, M.C., Lin, W.S., Liu, K.J.R.: Temporal forensics and anti-forensics for motion compensated video. IEEE Trans. Inf. Forensics Secur. **7**(4), 1315–1329 (2012)
114. Liu, J., Kang, X.: Anti-Forensics of Video Frame Deletion. [Online] http://www.paper.edu.cn/download/downPaper/201407-346 (2014). Accessed 9 July 2016
115. Valenzise, G., Tagliasacchi, M., Tubaro, S.: The cost of jpeg compression anti-forensics. In: Proceedings of IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP'11). Prague, Czech Republic, pp. 1884–1887 (2011)
116. Valenzise, G., Nobile, V., Tagliasacchi, M., Tubaro, S.: Countering jpeg anti-forensics. In: Benoit Macq and Peter, Schelkens (eds.) Proceedings of IEEE 18th International Conference on Image Processing (ICIP'11). Brussels, Belgium, pp. 1949–1952 (2011)
117. 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. **29**, 103–113 (2015)
118. Bian, S., Luo, W., Huang, J.: Exposing fake bit rates video and estimating original bit rates. IEEE Trans. Circuits Syst. Video Technol. **24**(12), 2144–2154 (2014)
119. Bian, S., Luo, W., Huang, J.: Exposing fake bitrate video and its original bitrate. In: Proceeding of IEEE International Conference on Image Processing. pp. 4492–4496 (2013)
120. Bian, S., Luo, W., Huang, J.: Detecting video frame-rate upconversion based on periodic properties of inter-frame similarity. Multimed. Tools Appl. **72**(1), 437–451 (2014)
121. Dias, Z., Rocha, A., Goldenstein, S.: First steps toward image phylogeny. In: Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS'10). Seattle, pp. 1–6 (2010)
122. Rosa, A.D., Uccheddu, F., Costanzo, A., Piva, A., Barni, M.: Exploring image dependencies: a new challenge in image forensics. In: Proceedings of SPIE Conference on Media Forensics and Security II. San Jose, CA. (2010)
123. Dias, Z., Rocha, A., Goldenstein, S.: Video phylogeny: recovering near-duplicate video relationships. In: Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS'11). Iguacu Falls, Brazil, pp. 1–6. (2011)
124. Visentini-Scarzanella, M., Dragotti, P.L.: Modelling radial distortion chains for video recapture detection. In Proceedings of IEEE 15th International Workshop on Multimedia Signal Processing (MMSP'13). Pula, Croatia, pp. 412–417 (2013)
125. Lee, J.-W., Lee, M.-J., Oh, T.-W., Ryu, S.-J., Lee, H.-K., Screenshot identification using combing artifact from interlaced video. In: Proceedings of 12th ACM Workshop on Multimedia and Security (MM&Sec'10). ACM Press, New York, NY, pp. 49–54 (2010)
126. Jung, D.-J., Hyun, D.-K., Ryu, S.-J., Lee, J.-W., Lee, H.-Y., Lee, H.-K., Detecting re-captured videos using shot based photo response non-uniformity. In: Proceedings of 10th International Conference on Digital-Forensics and Watermarking (IWDW'12). Springer-Verlag, Berlin, Heidelberg, pp. 281–291 (2012)
127. Visentini-Scarzanella, M., Dragotti, P.L.: Video jitter analysis for automatic bootleg detection. In: Proceedings of IEEE 14th International Workshop on Multimedia Signal Processing. Banff, AB, pp. 101–106 (2012)
128. Bestagini, P., Visentini-Scarzanella, M., Tagliasacchi, M., Dragotti, P.L., Tubaro, S.: Video recapture detection based on ghosting artifact analysis. In: Proceedings of IEEE International Conference on Image Processing (ICIP'13). Melbourne, VIC, pp. 4457–4461 (2013)
129. FFmpeg [Online]. https://www.ffmpeg.org/. Accessed 1 June 2016
130. Akao, Y., Kobayashi, K., Sugawara, S., Seki, Y.: Discrimination of inkjet-printed counterfeits by spur marks and feature extraction by spatial frequency analysis. In: Proceedings of SPIE Conference on Optical Security and Counterfeit Deterrence Techniques IV, San Jose, CA, pp. 129–137 (2002)
131. Chetty, G., Singh, M.: Nonintrusive image tamper detection based on fuzzy fusion. Int. J. Comput. Sci. Netw. Secur. **10**, 86–90 (2010)
132. Mishra, A., Goel, A., Singh, R., Chetty, G., Singh, L.: A novel image watermarking scheme using extreme learning machine. In: Proceeding International Joint Conference on Neural Networks, Brisbane, Australia. pp. 1–6 (2012)
133. Song, J., Lee, K., Lee, W.Y., Lee, H.: Integrity verification of the ordered data structures in manipulated video content. Digit. Investig. **18**, 1–7 (2016)
134. Dempster, A. P.: Upper and lower probabilities induced by a multivalued mapping. Ann Math. Stat. **38**(2), 325–339 (1967)
135. Shafer, G.: A mathematical theory of evidence, Princeton University Press, USA (1976)
136. Bayram, S., Avcibas, I., Sankur, B., Memon, N.: Image manipulation detection. J. Electron. Imag. **15**(4), 041102-041102-17 (2006)
137. Hsu, Y.-F., Chang, S.-F.: Statistical fusion of multiple cues for image tampering detection. In: Proceedings of IEEE 42nd Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA. pp. 1386–1390 (2008) doi:10.1109/ACSSC.2008.5074646
138. Zhang, P., Kong, X.: Detecting Image Tampering Using Feature Fusion. In: Proceedings of IEEE International Conference on Availability, Reliability and Security (ARES'09), Fukuoka, Japan. pp. 335–340 (2009) doi:10.1109/ARES.2009.150
139. Cozzolino, D., Gragnaniello, D., Verdoliva, L.: Image forgery detection through residual-based local descriptors and block-matching. In: Proceedings of IEEE International Conference on Image Processing, Paris, France. pp. 5297–5301 (2014) doi:10.1109/ICIP.2014.7026072
140. Hu, D., Wang, L., Zhou, Y., Jiang, X., Ma, L.: D-S Evidence Theory based Digital Image Trustworthiness Evaluation model. In: Proceedings on IEEE International Conference on Multimedia Information Networking and Security, Hubei, China. pp. 85–89 (2009) doi:10.1109/MINES.2009.154
141. Fontani, M., Bianchi, T., De Rosa, A., Piva, A., Barni, M.: A Dempster-Shafer framework for decision fusion in image forensics. In: Proceedings of IEEE International Workshop on Information Forensics and Security (WIFS'11), Iguacu Falls, SA. pp. 1–6 (2011) doi:10.1109/WIFS.2011.6123156
142. Fontani, M., Bianchi, T., De Rosa, A., Piva, A., Barni, M.: A Framework for Decision Fusion in Image Forensics based on Dempster-Shafer Theory of Evidence. In: IEEE transactions on Information Forensics and Security. 8, 4. pp. 593–607 (2013) doi:10.1109/TIFS.2013.2248727
143. Fontani, M., Bonchi, A., Piva, A., Barni, M.: Countering anti-forensics by means of data fusion. In: Proceedings of SPIE Conference on Media Watermarking, Security, and Forensics (2014). doi:10.1117/12.2039569
144. Gupta, S., Cho, S., Kuo, C.-C.J.: Current trends and future development in audio authentication. Multimed. Forensics Secur. Intell. **19**, 50–59 (2012)
145. Columbia Image Splicing Detection Evaluation Dataset. [Online]. http://www.ee.columbia.edu/ln/dmvv/downloads/AuthSplicedDataSet/AuthSplicedDataSet.htm. Accessed 3 June 2016
146. CASIA Tampered Image Detection Evaluation Database. [Online]. http://forensics.idealtest.org:8080. Accessed 30 Mar 2016
147. CFReDS—Computer Forensic Reference Data Sets, [Online]. http://www.cfreds.nist.gov/. Accessed 17 May 2016
148. Tralic, D., Zupancic, I., Grgic, S., Grgic, M., CoMoFoD—New Database for Copy-Move Forgery Detection. In: Proceedings of 55th International Symposium ELMAR, Zadar, Croatia, pp. 49–54 (2013) [Online]. http://www.vcl.fer.hr/comofod/download.html. Accessed 18 July 2016