A Three-Level Benchmark Dataset for Spatial and Temporal Forensic Analysis of Videos

Naheed Akhtar[1,2], Mubbashar Saddique[3], Paul L. Rosin[4], Xianfang Sun[4], Muhammad Hussain[5], Zulfiqar Habib[1,*]

Abstract

Video Forensics keeps developing new technologies to verify the authenticity of digital videos. The existing datasets have limitations including unrealistic and poor-quality tampering, small size, few types of forgeries, lack of range of video content, different lighting conditions, and a range of camera models. This paper proposes the COMSATS Structured Video Tampering Evaluation Dataset (CSVTED), a three-level benchmark dataset organized by tampering quality and video complexity. This dataset includes a diversity of tampering in the spatial and temporal domains such as frame duplication, deletion, insertion, copy-move, and splicing. The dataset aims to facilitate the evaluation of video forgery detection methods by providing 1047 videos (133 original and 914 tampered), captured by multiple cameras in different lighting conditions (morning, noon, evening, night, fog). To develop the benchmark dataset, videos are tampered with a variable number of duplicated/deleted/inserted frames as well as Event-Object-Person (EOP) based tampering. Special care has been taken to ensure minimal abrupt changes in tampered videos by using Structural Similarity Index Measure (SSIM) and Optical Flow (OF) to determine the optimal positions for duplication/insertion/deletion in the video. Taking into account the direction of motion of objects in the video, these techniques aid in seamlessly integrating the tampered frames while maintaining visual coherence. Furthermore, the videos in CSVTED depict natural scenes after the tampering process and are in the common formats of avi, mp4, or mov. This dataset will be publicly available for researchers in the domain of video forensic analysis.

Keywords Video Forgery · Spatial Forgery · Temporal Forgery · Benchmark Dataset · Copy-move · Splicing.

I. Introduction

A massive amount of digital visual content is being shared on the web by different people from all over the world to convey information, especially on social media networks, such as Facebook, WhatsApp, Twitter, and YouTube. Over 500 hours of videos are uploaded every minute and 5 billion videos are watched every day from YouTube [1]. This tremendous use of digital videos has been followed by a rise of techniques to modify video content, using different editing software. These tools are very simple and enable even inexperienced users to perform unauthorized modification in order to maliciously forge a video sequence, and these videos are often shared on different social networks to spread fake news. Videos may contain sensitive and important information about the event that occurred, and this can be easily tampered with by editing tools like Adobe Premiere Pro, Light Works, and Final Cut Pro X, etc. The attacker aims to create visually realistic video forgeries. It is difficult to detect the changes in the contents of videos with the naked human eye [2-4]. Therefore, it is challenging to differentiate between non-modified videos and their fake counterparts to determine if the video is authentic or tampered. Many fields such as courts, investigation departments, social media networks, and media groups need the verification of videos [5, 6].

Broadly, video forgery detection can be classified into two groups: passive and active video tampering detection. Active techniques involve the use of known traces such as digital signatures or watermarks which are embedded into the content during the acquisition phase when the video is being recorded or later during the transmission of data. Any change in this embedded information indicates tampering. However, this technique may be ineffective if alteration occurs before insertion of the digital signature or watermark [7-9]. In contrast, passive video tampering detection methods are more convenient as they do not rely on encapsulated information (digital signatures, hash values) within the video. Therefore, the passive methods determine the genuineness of a video by exploring the intrinsic features left by the capturing devices or manipulation act in the media [10-12]. Similarly, digital videos can be forged by three different techniques: (a) spatial or object-based (intra-frame) forgery, (b) temporal or frame-based (inter-frame) forgery, and (c) spatio-temporal forgery. In spatial tampering, videos are altered by modifying the content (objects) within the frames, thereby affecting the visual appearance. Objects can be removed, replaced, and added within the frame(s) of a video. In temporal tampering, frames can be removed, inserted, duplicated, or shuffled to hide important information. For instance, frame duplication may be done to extend the time duration of an activity. Spatio-temporal

[*] *Corresponding author: Z. Habib.*

[1,2] N. Akhtar is with Department of Computer Science COMSATS University Islamabad, Lahore Campus, Pakistan and University of Education Lahore, Pakistan (e-mail: sp19-pcs-009@cuilahore.edu.pk; naheedswl@ue.edu.pk).

[3] M. Saddique is with Department of Computer Science and Engineering, University of Engineering and Technology Lahore, Narowal Campus, Pakistan (e-mail: Mubashar.chaudary@gmail.com).

[4] P. L. Rosin and X. Sun is with Department of Computer Science and Informatics, Cardiff University, UK. (e-mail: rosinpl@cardiff.ac.uk; sunx2@cardiff.ac.uk).

[5] M. Hussain is with Department of Computer Science, King Saud University, Riyadh, Saudi Arabia; (e-mail: mhussain@ksu.edu.sa).

[1] Z. Habib is with Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan (e-mail: drzhabib@cuilahore.edu.pk).

tampering involves combination of both spatial and temporal tampering [13] as shown in Fig. 1.

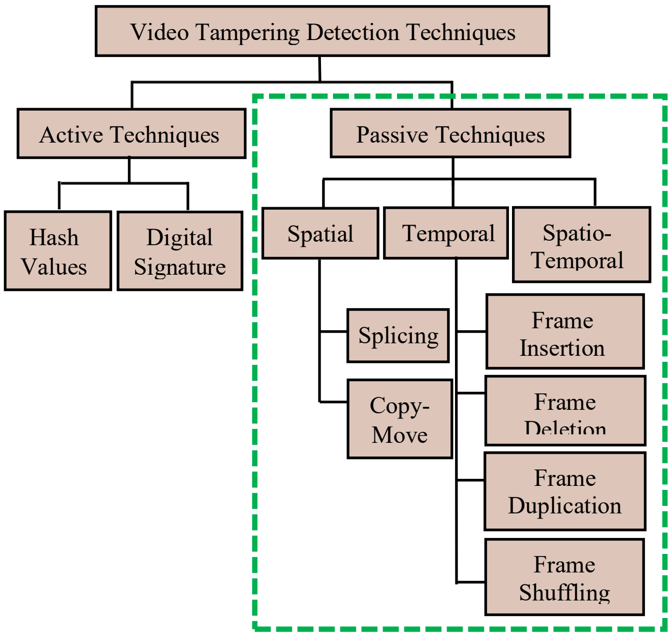


Fig. 1. Categorization of video tampering detection techniques.

Numerous algorithms for video tampering detection have been developed to identify a specific tampering attack by exploiting unique features like motion characteristics [11, 14], noise properties [15], video compression and coding features [16, 17], color models [18], and texture based features [19]. All these algorithms are presented by different researchers and perform well on particular datasets typically developed by them. To assess the performance of developed techniques, it is required to test them on the datasets freely available to the public. Unfortunately, such datasets are not available for other communities/researchers, hindering to evaluate the performance of their algorithms. High accuracies achieved on individual small datasets may give the impression that video tampering detection is a solved problem, which stifles innovation since new research may yield little improvements on saturated datasets. In contrast, a great effort in the area of image forensics [20-24], source device identification [25-27], AI-generated content detection (AICD) and deepfake detection [28-32] has been made to the development of many challenging benchmark datasets, yet no comparable benchmark dataset is available for video forensics.

Video tampering detection, focuses on manipulated real videos where genuine footage is altered like frame deletion, insertion, duplication, copy-move, or splicing. In contrast, AIGC detection deals with entirely synthetic videos generated by deep learning models such as GANs, diffusion models, or neural rendering. Tampering detection is critical in forensics, law enforcement, and media verification, whereas AIGC detection is essential for combating synthetic media, misinformation, and DeepFake threats in cybersecurity. While both fields aim to ensure video authenticity.

Several papers have been published on video datasets, including works by Akbari et al. [33], Singla et al. [34], Sharma and Kanwal [35], Panchal and Shah [36], Ulutas et al. [37], Almansa et al. [38], Al-Sanjary et al. [39], Feng et al. [14], Su et al. [40], Chen et al. [41], Bestagini et al. [42], and Qadir et al. [43]. However, several limitations have been observed in these existing datasets:

- Social media or YouTube are taken as video sources in [36, 39, 44] to develop the video forgery datasets, lacking information about the capturing devices and do not ensure the authenticity of the available videos.
- The datasets do not cover all types of video forgeries in both the spatial and the temporal domains.
- Limited availability of datasets for the research community since many are not made publicly available.
- Tampering is often conducted in an unnatural and unrealistic manner. Since manually producing tampered videos is very time-consuming, most authors synthetically doctored sequences by selecting a fixed position for insertion/deletion/duplication. An example is in Panchal and shah [36] in which insertion, deletion, and duplication of 10, 20, and 30 frames at fixed positions frame 101, frame 201, and frame 301 are performed in every video.
- Poor quality of tampering, such that most of the tampering is clearly detected by the naked eye.
- There is limited variety in the video forgery dataset. Ideally, there should be different resolutions, different frame rates, and different light conditions such as morning, noon, evening, night, and fog.
- Datasets are too small; until now no dataset has been found that has more than 220 tampered videos.

Recognizing the importance of video tampering detection to the computer vision and video processing communities, we have prepared three-level benchmark dataset for spatial and temporal forensic analysis of videos. This CSVTED consists of 1047 videos (including tampered and untampered videos). A distinguish feature of CSVTED is the fact that tampering quality and complexity of every video is computed. The quality of tampering provides insights into the characteristics of different types of manipulations, such as frame deletion, insertion or duplication. Understanding these characteristics can guide the development of more targeted detection methods tailored to specific types of tampering. The highlights of this paper are:

- Existing video tampering datasets have been overviewed.
- CSVTED comprises of all types of spatial and temporal domain tampering including EOP based tampering.
- To achieve more robust and adaptable solutions to tampering detection, the developed algorithms must be assessed across different levels of tampering severity. To overcome this issue, CSVTED is split into 3 levels based on the quality of tampering and video complexity. Furthermore, these aspects have been quantified, which allows us to validate the contents of the tiers of our dataset.
- A detailed description of the proposed structured three-level benchmark dataset along with systematic and natural ways of incorporating the video tampering process is presented.

- A variety of videos having different frame rates, video formats, and lighting conditions in the dataset along with measures of their complexity and quality of tampering are provided.
- A comparison of the proposed dataset with state-of-the-art datasets is also presented.

Audio constitutes a significant element of video content, and in some forensic investigations, audio analysis holds paramount importance. Nonetheless, there are several factors that may lead to the exclusion or lower prioritization of audio when preparing a video forensic dataset. Firstly, video forensic analysis often focuses primarily on the visual aspects of the footage, such as identifying individuals, objects, and events. This is because video evidence can provide valuable information even in the absence of audio, such as facial recognition, license plate identification, or establishing the sequence of events. Secondly, audio analysis requires specialized skills and tools that are distinct from those used in video analysis. Thirdly, audio recordings within video footage may encounter various quality issues, including background noise, distortion, or low volume. These issues can make audio analysis challenging and less reliable, which may discourage its inclusion in some datasets. In any case, audio can provide some clues to detect tampering, but we want users to rely on the images alone, which is why this additional information is not be provided as part of CSVTED. The original 133 videos are captured using 9 different cameras with varying resolutions, frame rates, and lighting conditions such as morning, noon, evening, night, and fog. Videos are tampered with by applying different tampering techniques such as Frame duplication, Frame deletion, Frame insertion, Copy-move, and Splicing. In this way, this dataset contains 133 original and 914 tampered videos. The details are presented in Table A1.

The structure of this paper contains a literature review regarding the development of datasets for video forgery. In Section 2, methods of video forgery are explained. In Section 3 and Section 4, the framework of video tampering and ground-truth values of CSVTED is described. Finally, Section 5 concludes this work.

1.1 Existing Video Forgery Datasets

The availability of an efficient dataset is crucial to conduct any research work. Unfortunately, there are only few datasets exist for video forensics, and they often comprise a limited number of forged videos, possess low-resolution, and were not captured using modern portable devices in different light conditions (morning, noon, evening, and night). For instance, SULFA [43], publicly available on the University of Surrey website, consists of 150 original videos with a low resolution of 320 × 240 pixels and a duration of 10-seconds, with only 5 tampered videos covering spatial and temporal tampering. The existing video forgery datasets are not appropriate due to small size (few videos), and covering few types of forgeries [39, 40, 43, 45, 46], prompting many researchers to develop their personal datasets to conduct experiments. For example, Le et al., 2017 [38] developed a dataset of video in-painting containing 53 tampered videos, aimed to recover the missing area/region in a video frame by exploiting spatio-temporal information from the rest of that video. There is lack of realistic representation in the VIFFD dataset [47], it only incorporates certain scenarios; frames are only inserted at the start or end of the video in frame insertion tampering. If frames are deleted at the start of the video, the deleted frames are replaced by black frames; which is not a practical approach.

The latest video dataset on three types of temporal tampering was introduced by Panchal and Shah [36]. A thorough analysis revealed several drawbacks. Firstly, the video capturing source is unknown for social media and YouTube videos. Secondly, there is no guarantee that the uploaded videos on YouTube are always in their authentic form. Such doubtful videos are not suitable for experiments such as different YouTube videos are used as a source in [36] and [39] to prepare tampered datasets.. Thirdly, fixed locations are selected as start tampering points in all videos for deletion/insertion/duplication. For example, in [36] frames 101, 201, and 301 are fixed for tampering of 10, 20, and 30 frames. Fourthly, the tampering approaches are unnatural or unrealistic. Frames are duplicated/deleted/inserted without considering that the inserted or deleted frames are creating a big jump or gap, giving a strong clue of tampering. Mostly tampering is done by unprofessional people using commonly available video editing tools.

Al Sanjary et al. [39] addressed three types of tampering: copy-move, splicing, and frame duplication in their dataset, which contains 33 original and 26 forged videos (10 copy-move, 10 splicing, and 6 frame duplication). While this dataset is openly available for researchers and encompasses a variety of tampering in both spatial and temporal domains, the main problem with this dataset is the absence of information about the capturing devices, different lighting conditions, and their features. Furthermore, not all types of temporal tampering are covered in this dataset. D'Avino, Cozzolino et al. [48] proposed a method for video splicing detection based on auto-encoder and recurrent neural networks. Their developed dataset comprises of 10 short video clips, specifically designed for splicing detection. The tampering in these videos can be easily detectable with the naked eye. Several datasets such as FaceForensics 2018 [49], VISION 2017 [50], and Newson et al. 2017 [51] have been produced; they primarily focus on a single tampering method such as splicing, motion transfer, or in-painting. The videos in VIDEOSHAM [32] are manipulated using spatial and temporal attacks. This lack of comprehensive coverage underscores the necessity for a sufficiently large dataset for video forensics, containing a heterogeneous set of videos along with ground-truth references.

Datasets containing both original and tampered videos play a vital role in estimating the accuracy of algorithms [14]. The development of a reliable dataset for algorithms of video forensic investigation is important, yet to the best of our knowledge, no large and efficient dataset exists so far in the domain of video forgery. Therefore, a new dataset of tampered videos is developed where each video has undergone a single tampering attack such as frame deletion, frame duplication, and frame insertion. Static cameras are used to capture the videos and are forged with various forgery techniques. The comparison between existing datasets and the developed dataset is presented in Table 1a and Table 1b. The primary objective behind developing the CSVTED is to assess the robustness of video forgery detection algorithms by providing this dataset to the researchers.

2 Types of Video Forgery

Hackers have devised many ways to tamper digital videos; including splicing, copy-move, frame insertion, deletion, duplication, and shuffling. It becomes challenging for digital forensic investigators to differentiate between original and tampered content. Video tampering typically falls into three categories: spatial forgery, temporal forgery, and spatio-temporal forgery [40, 52-55]. In spatial tampering, forgers try to manipulate the pixels within the video frame. Fig. 2 (a) represents an original video; while Fig. 2 (b) illustrates a tampered video in the spatial domain as its pixels P11, P12, P21, and P22 are manipulated. In temporal forgery, forgers try to disrupt the sequence of the frames by adding, removing, reordering, and replacing the frames. Fig. 2(c) illustrates a temporally forged video by deleting and replacing the frames F6, F7, F8 by F1, F2, and F3. Similarly, Fig. 2(d) represents a combination of temporal and spatial tampering where both pixel manipulation and frame sequence disruption occur. In spatial tampering, copy-move and splicing methods are used to forge the video frame and these techniques can also be applied to still images. The root of tampered regions are the processes that are post-performed to forge the video [54]. Fig. 3 provides an example of copy-move forgery, demonstrating how this technique is used to manipulate videos taken from CSVTED. The copy-move technique is applied by copying a specific area from a frame of the video and pasting it on the same frame to conceal the car and fire extinguisher in Fig. 3(a) and 3(b), respectively. All videos in the CSVTED are recorded using different cameras and contain scenes representative of realistic situations, with frame rates ranging from 12.50 to 30.0174 fps, and the resolution of 1920×1080, 1280×720, and 640×480. The duration of the original videos varies from 6.016s to 47.67s.

Another type of spatial tampering technique is splicing in which a specific portion of a frame in a video is copied and pasted into another frame of the same or the other video. Splicing becomes challenging when a moving camera is used to record the video as unstable direction and lighting conditions make it difficult to perform. Maintaining consistent frame rate in the splicing method is also a challenge. The splicing method involves merging various segments of video frames either from the same video or different videos to create seamless boundaries, resulting in a more realistic appearance. Examples of splicing techniques applied to the video tampering dataset are depicted in Fig. 4, showing alterations to the number plate of the car and content displayed on monitor.

In the temporal domain, video tampering involves changing frames through frame insertion, duplication, and deletion [53]. Temporal video tampering can occur at frame, scene/shot, or video. For example, deletion of a complete scene is performed by removing the entire scene from a video. Similarly, inserting or duplicating a video is performed at the video level [56]. Fig. 5 (b)-(d) represents an example of duplicating, deleting, and inserting frames at the frame level. The frame count increases or decreases by modifying the source video in the temporal domain. Frame deletion and insertion examples from CSVTED are illustrated in Fig. 6. In the original video scene (see Fig. 6(a)), the frames (113, 114, 115, and 116) are deleted from a video shown by blank rectangular boxes as in Fig. 6(b). Similarly, the frames-113, 114, 115, and 116 are inserted from some other video (see Fig. 6(c))

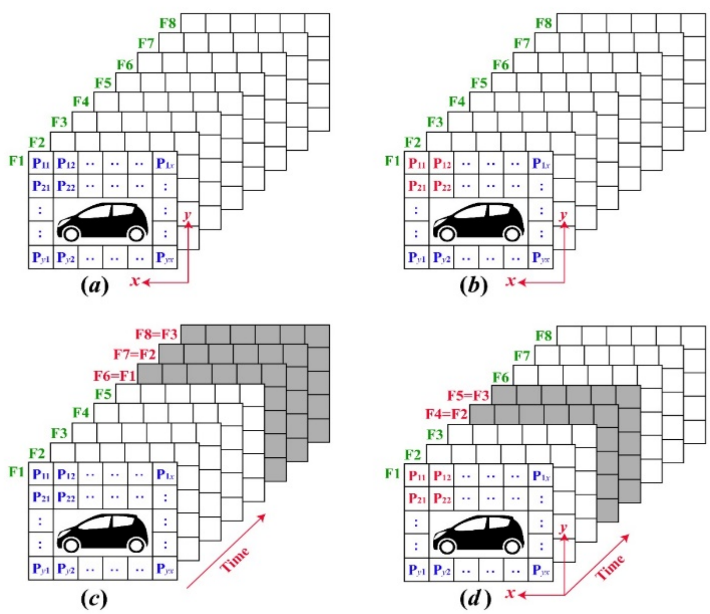


Fig. 2 (a) original video (b) spatially forged video (c) temporally forged video (d) spatio-temporal forged video. Width and height are represented by x and y respectively

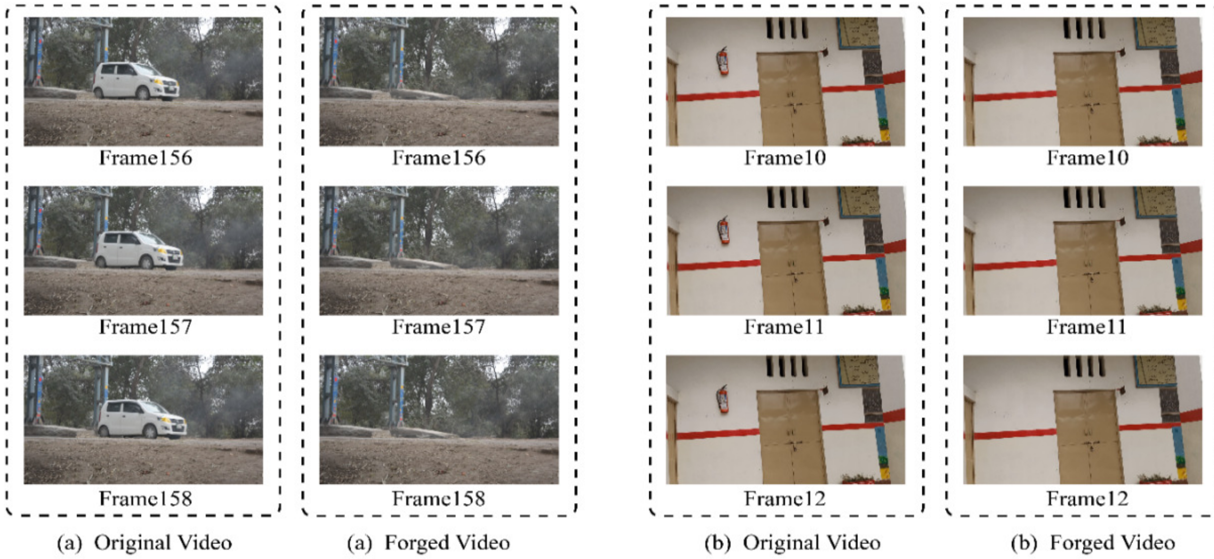


Fig. 3 Example of copy-move tampering

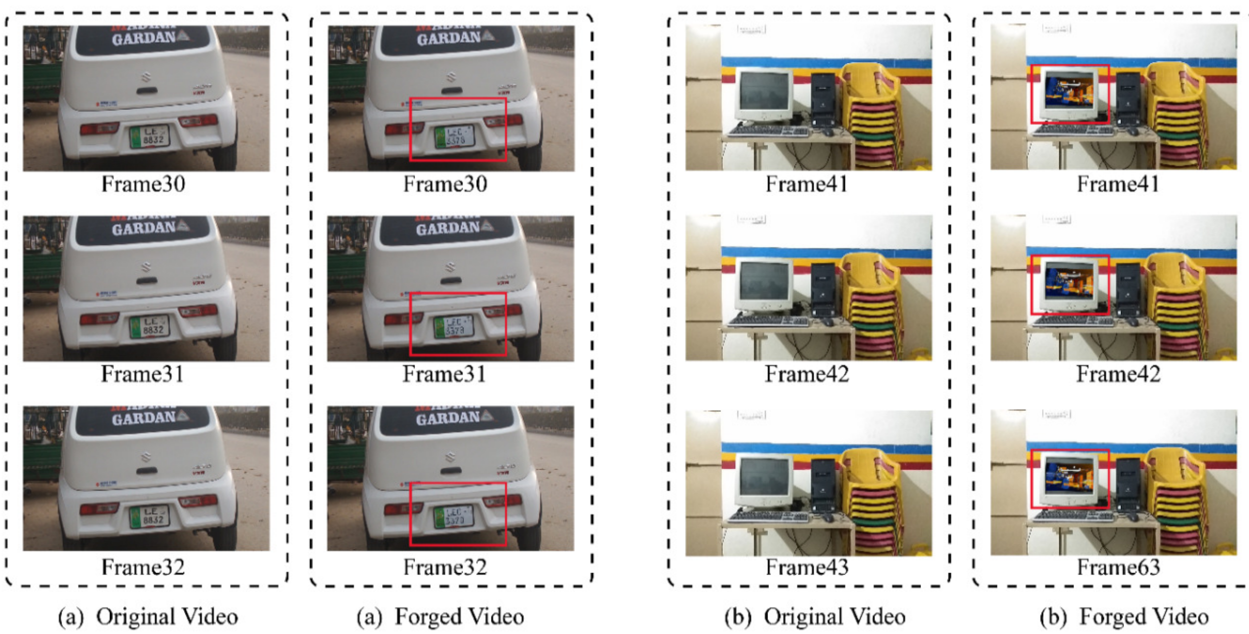
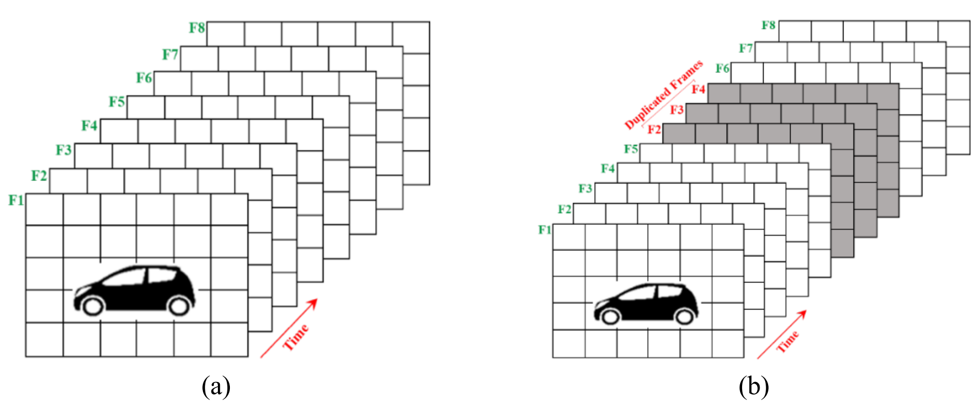
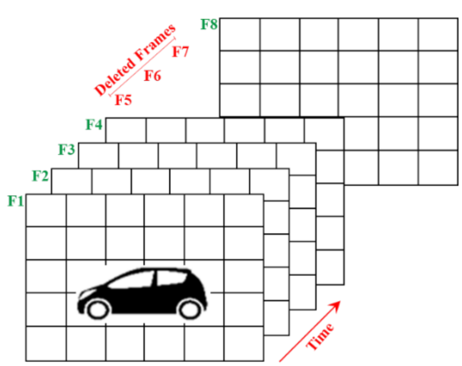
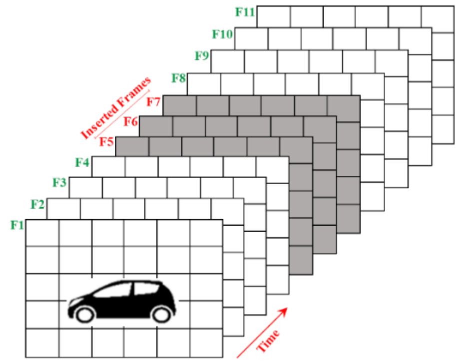


Fig. 4 Example of splicing tampering





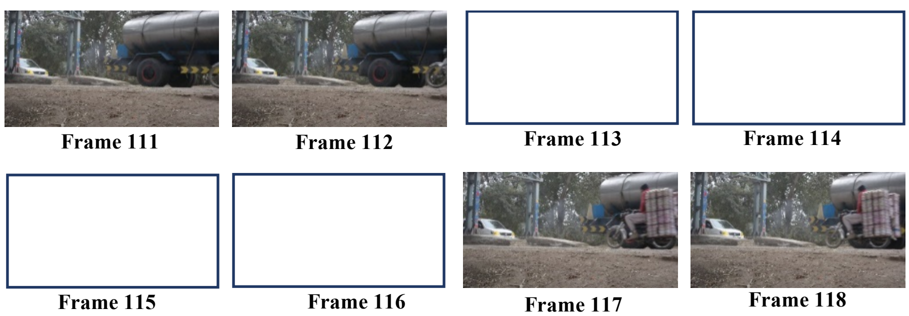
(c)



(d)

Fig. 5 Temporally tampered examples (a) original frames (b) frame duplication (c) frame deletion (d) frame insertion

(a)



(b)

(c)

Fig. 6 Examples of temporal tampering (a) original video (b) frame deletion (c) frame insertion

Table 1a The Comparison of Existing Datasets with The Developed Dataset

Reference	Number of videos	Video length in seconds	Video source	Static/moving camera	Type of video forgery	Scenario
Akbari et al. 2023 [33]	Original: 250 Tampered: 250	11-15s	iPhone and android smart phones	Moving	Object based	N/A
Singla et al, 2023 [34]	Original: 60 Tampered: 966	N/A	iPhone and android smart phone	Static & moving	Frame insertion, deletion, duplication, cloning, splicing and inpainting	Indoor, outdoor
Sharma and Kanwal, 2021 [35]	Original: 200 Tampered: 200	06-15s	Smartphones of ASUS and VIVO brands	Static & moving	Frame duplication	Indoor, outdoor, nature
Panchal and Shah, 2020 [36]	Original: 40 Tampered: 210	06-18s	SULFA, YouTube	Static & moving	Frame insertion, deletion, duplication, and smart tampering	N/A
Ulutas et al., 2018[37]	Original + Tampered: 31	N/A	SULFA and different movie scenes	Static & moving	Frame Duplication	N/A
Le, Almansa et al., 2017 [38]	Tampered: 53	N/A	N/A	Static & moving	Video In-painting	N/A
Al-Sanjary, Ahmed et al., 2016 [39]	Original: 7 Tampered: 26	14-16s	YouTube	Static & moving	Copy–move, swapping frames, Splicing	N/A
Feng, Xu et al., 2016 [14]	Original: 122 Tampered: 732	N/A	YUV files http://trace.eas.asu.edu/yuv/ http://media.xiph.org/video/derf/ ftp://ftp.tnt.uni-hannover.de/pub/svc/testsequences/ http://202.114.114.212/quick_motion/yuv_download.html	Static & moving	Frame Deletion	N/A
Su, Huang et al., 2015 [40, 57]	Total: 20	N/A	SONY DSCP10	Static	Copy–move	N/A
Chen et al., 2015 [41]	Total: 100	11s	Commercial Surveillance Cameras	Static	Object based	N/A
Bestagini, Milani et al., 2013 [42]	Original: 10 Tampered: 10	07-19s	Canon SX220, Nikon S3000 Fujifilm S2800HD	Static	Copy-move	N/A
Qadir, Yahaya et al., 2012 [43]	Original: 166 Tampered: 5	04-18s	SULFA Dataset	Static	Copy-move	N/A
Proposed CSVTED	Original: 133 Tampered: 914	6.016 - 47.67s	Nikon D5200, IP Cam H.265, OPPO F11, Redmi Note 4X, HIKVISION outer, HIKVISION Inner, Nikon L29, Wi-Fi RoboticV380, Samsung Galaxy A51	Static	Splicing, Copy-move, Frame insertion, Frame deletion, Frame duplication	Morning, Noon, Evening, Night, and Fog

Table 1b The Comparison of Existing Datasets with The Developed Dataset

Dataset	Name of Dataset	Tampering in spatial domain	Tampering in temporal domain	Includes all temporal domain tampering	Available in public domain	Number of tampered videos greater than 250	3-Level benchmark dataset based on video complexity
Akbari et al. 2023 [33]	-	YES	NO	NO	NO	NO	NO
Singla et al, 2023 [34]	HTVD	YES	YES	YES	YES	YES	NO
Sharma and Kanwal, 2021 [35]	VLFD	NO	YES	NO	NO	NO	NO
Panchal, Shah et al., 2020 [36]	TDTVD	NO	YES	YES	YES	NO	NO
Ulutas et al., 2018 [37]	Test Database	NO	YES	NO	YES	NO	NO
Le, Almansa et al., 2017 [38]	-	YES	NO	NO	YES	NO	NO
Al-Sanjary, Ahmed et al., 2016 [39]	VTD	YES	YES	NO	YES	NO	NO
Feng, Xu et al., 2016 [14]	-	NO	YES	NO	NO	YES	NO
Chen et al., 2015[41]	SYSU-OBJFORG	YES	NO	NO	NO	NO	NO
Su, Huang et al., 2015 [40]	-	YES	NO	NO	NO	NO	NO
Bestagini, Milani et al., 2013[42]	REWIND PROJECT	YES	YES	NO	YES	NO	NO
Qadir, Yahaya et al., 2012 [43]	SULFA	YES	YES	NO	NO*	NO	NO
Proposed CSVTED	CSVTED	YES	YES	YES	YES	YES	YES

* Given link is not accessible.

3 Methodology

3.1 Basis

We have constructed a three-level benchmark video tampering dataset which has increasing difficulty in detection. The first level contains videos with simple backgrounds and a small number of moving objects. The second level increases the challenge by introducing complications such as complex background, lighting conditions, and random movement of objects. Moving to the third level intensifies the detection challenge by tampering with the complete events, activities or objects in a manner that there will be no jump at the tampering points. The key considerations are as follows:

3.1.1 Challenging Videos:

The benchmark needs to include videos that are likely to be challenging for video tampering detection methods. Revealing weaknesses in the state-of-the-art helps drive research progress. Some frames of videos are presented in Fig. 7.

3.1.2 Range of Difficulty:

The benchmark should include tampered videos covering a range of levels of difficulty, to better assess the performance of video tampering detection algorithms, i.e., showing when they work, and when they fail. Excessively high detection accuracy on a simple dataset indicates that performance on this dataset has saturated and that the problem has been successfully resolved, which might discourage new researchers from striving for better models. Conversely, if the benchmark is overly challenging, it may discourage users and restrict the adoption within the community, again discouraging research progress. The first level should be attainable by the majority of existing methods.

3.1.3 Large Number of Videos

The benchmark dataset on video tampering should be as large as possible, having enough variety to be representative. Four main factors explain the need for a large dataset. First, within the video tampering detection community, where deep learning techniques are prevalent, large benchmark datasets are essential for training effective models. Second, the video forgery dataset should encompass a wide range of scenarios, including different resolutions, varying frame rates, and lighting conditions, such as morning, noon, evening, night, and fog. Third, extensive datasets should comprehensively cover all types of video forgeries across spatial and temporal domains. Last, a substantial dataset motivates authors and researchers to develop, generalize, and cross-validate their algorithms using a benchmark that encompasses a diverse range of manipulated videos.

This approach discourages the development of individual

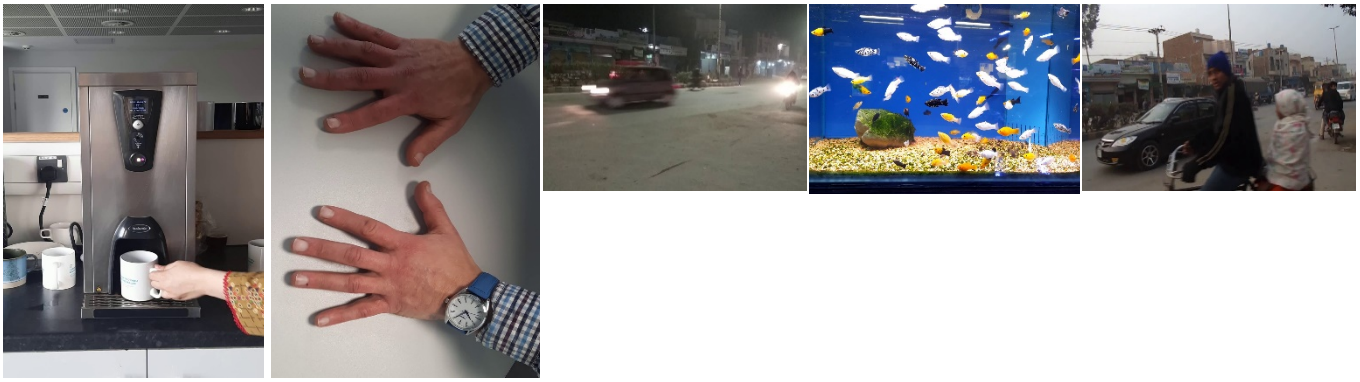


Fig. 7 Sample frames from CSVTED

small datasets where algorithms may exhibit exceptional performance.

Due to the unavailability of benchmark datasets, researchers often create different small datasets. It is suggested to create a dataset sufficiently large that can be treated in its entirety. However, striking a balance between dataset size and coverage of target domains poses a challenge. The need for a large benchmark dataset is emphasized by the argument made earlier. Yet, it is also essential that the benchmark encompasses a thorough representation of the target domains. To reconcile these considerations, a decision was made to include a large number of diverse videos for all categories of inter-frame tampering. This choice seeks to achieve both a sizable benchmark and a demonstration of diverse content, addressing the tension between the two factors.

3.1.4 Gap Between Levels

The difficulty gap between level n and level n + 1 should not be too large since we desire fine granularity of what conditions cause algorithms to fail, while also avoiding excessive number of levels that could render the benchmark unwieldy. The inclusion of three levels in CSVTED provides more varied and challenging content. However, there remains scope for further levels that cover more complex scenes (e.g., multiple people/objects, complex motion, zoom in and zoom out).

3.1.5 Variety of Video Sources, Compression Levels, and Resolution

Video quality, compression, and brightness are some of the major challenges in video forensics. Low-resolution footage and variations in brightness can mislead the analysis [58]. To address these challenges, videos should come from a variety of sources to ensure that a variety of cameras, lighting conditions, backgrounds, and varied levels of professionalism of the photographers are included. We have recorded videos using a variety of cameras as well as taking videos from different CCTV cameras installed indoors and outdoors. To ensure the effectiveness of video tampering detection algorithms, they must be capable of handling videos with varying resolutions, frame rates, and formats. Therefore, during the development of the benchmark, videos with different resolutions, frame rates, compression levels, and formats are intentionally included. This diverse selection of videos serves as a comprehensive testbed for evaluating the algorithms' performance across a wide range of real-world scenarios.

3.2 Development of Inter-Frame Tampering Dataset

During the creation of datasets with inter-frame tampering, the straightforward removal of predefined frame numbers from each sequence raises questions about whether this action results in discernible effects [59]. Rather than selecting random or predefined fixed positions to insert, delete, or duplicate selected frames, a more sophisticated approach is sought to create a tampered video with imperceptible jumps. The primary rationale behind the tentative choice of start and end points is to perform purposeful duplication, insertion, or deletion of certain objects, individuals, or events within the content. Subsequently, the algorithm identifies the most suitable clip to insert, ensuring optimal alignment with the video's existing content for insertion or duplication purposes. In the absence of these parameters, an entirely automated system might inadvertently select and manipulate the static segments of the video, either deleting them or placing that content within them. The objective is to identify a set of frames that can be inserted into the video in a way that remains undetectable to the naked eye. To achieve this, the Structural Similarity Index Measure (SSIM) and Optical Flow (OF) are utilized to determine the optimal matching positions for the selected video segment. To consider the direction of motion of objects in the video, OF aids in integrating the tampered frames while maintaining visual coherence. The SSIM image similarity measure is a combination of three factors computed in local windows: correlation, luminance, and contrast [60]. It has shown superior performance compared to mean squared error (MSE) and its derivate in terms of accuracy [61, 62]. It combines accuracy of prediction, simplicity of computation, and intuitiveness of design. The framework of the video forgery process is illustrated in Fig. 9 and algorithms for frame duplication, frame deletion, and frame insertion tampering are given below.

Algorithm 1: The preparation of frame duplication tampered video from a given untampered video

Input: Frames F_i of untampered video V along with tentative start S and endpoint E of frames to be duplicated.

Output: Tampered video with frame duplication tampering

Procedure:

1. For each frame A from S-r to S+r and B from E-r to E+r, where r=20:
   i. For each consecutive frame pair {C, D} of video V, such that C,D < A or C,D > B:
      a. Compute the optical flow $OF_{C-1, C}$, $OF_{C, A}$, $OF_{A, A+1}$
      b. Compute the optical flow $OF_{C, D}$, $OF_{D, B}$, $OF_{B, B+1}$
      c. Compute SSIM of colored frames: $SSIM_{C,A}$ and $SSIM_{B,D}$
      d. Q_{CA1} = SSIM ($OF_{C-1, C}$, $OF_{C, A}$) + SSIM ($OF_{C, A}$, $OF_{A, A+1}$)
      e. Q_{BD1} = SSIM ($OF_{C, D}$, $OF_{D, B}$) + SSIM ($OF_{D, B}$, $OF_{B, B+1}$)
      f. $Q_{CA} = Q_{CA1} + SSIM_{C,A}$ and $Q_{BD} = Q_{BD1} + SSIM_{B,D}$
      g. Min (Q_{CA}, Q_{BD}) → Q
   ii. Max(Q) → List
2. Max(List) → Best Match
3. Frames A and B at Best Match→ Selected Frames
4. Insert the Selected Frames after Best Match to form a tampered video

Algorithm 2: The preparation of frame deletion tampered video from a given untampered video

Input: Frames F_i of untampered video V along with tentative start S and endpoint E of frames to be deleted.

Output: Tampered video with frame deletion tampering

Procedure:

1. For each frame A from S-r to S+r and B from E-r to E+r, where r=20:
   a. Compute the optical flow $OF_{A-1, A}$, $OF_{A, B}$, $OF_{B, B+1}$
   b. Compute SSIM of colored frames: $SSIM_{A,B}$
   c. Q_{AB1} = SSIM ($OF_{A-1, A}$, $OF_{A, B}$) + SSIM ($OF_{A, B}$, $OF_{B, B+1}$)
   d. $Q_{CA} = Q_{AB1} + SSIM_{A,B}$
   e. Q_{CA} → List
2. Max(List) → Best Match
3. Frames A and B at Best Match→ Selected Frames
4. Delete all frames between Selected Frames (excluding A and B) to form a tampered video

Algorithm 3: The preparation of frame insertion tampered video from a given untampered video

Input: Frames F_i of untampered video V_1 along with tentative start S and endpoint E of frames to be inserted in video V_2 with frames G_i.

Output: Tampered video with frame insertion tampering)

Procedure:

1. For each frame A from S-r to S+r and B from E-r to E+r of video V1, where r=20:
   i. For each consecutive frame pair {C, D} of video V_2,:
      a. Compute the optical flow $OF_{C-1, C}$, $OF_{C, A}$, $OF_{A, A+1}$
      b. Compute the optical flow $OF_{C, D}$, $OF_{D, B}$, $OF_{B, B+1}$
      c. Compute SSIM of colored frames: $SSIM_{C,A}$ and $SSIM_{B,D}$
      d. Q_{CA1} = SSIM ($OF_{C-1, C}$, $OF_{C, A}$) + SSIM ($OF_{C, A}$, $OF_{A, A+1}$)
      e. Q_{BD1} = SSIM ($OF_{C, D}$, $OF_{D, B}$) + SSIM ($OF_{D, B}$, $OF_{B, B+1}$)
      f. $Q_{CA} = Q_{CA1} + SSIM_{C,A}$ and $Q_{BD} = Q_{BD1} + SSIM_{B,D}$
      g. Min (Q_{CA}, Q_{BD}) → Q
   ii. Max(Q) → List
2. Max(List) → Best Match
3. Frames A and B at Best Match→ Selected Frames
4. Insert the Selected Frames after Best Match to form a tampered video

3.3 Development of Spatial Tampering Dataset

For development of the spatial tampering dataset, the initial step involves importing videos into Adobe Premiere Pro for tampering. The original videos encompass a timeline as per video length and a single layer in animation window. The animation window facilitates frame-by-frame processing, allowing the identification of undesired objects. Upon detection, the unwanted objects are eliminated using the Clone Stamp tool in Photoshop, which involves copying pixels from nearby similar areas and pasting them over the objects. To enhance the similarity of the forged area with its surroundings, adjustments are made to the brightness and contrast. After its completion, the newly tampered area is duplicated and copied into a new video layer for further tampering actions. Care is taken to ensure that the pasted regions blend seamlessly with the original video's appearance on the timeline. Employing a combination of subtractive masking and the aforementioned processes, After Effects Photoshop CS was utilized to address unwanted objects. These methods effectively remove undesirable objects while preserving the integrity of the remaining regions. Finally, the fully tampered video is exported from the tool. For more specific information regarding the spatial tampering dataset, please see Table A2.

3.4 Levels

Evaluating tampering detection algorithms is challenging due to the presence of various confounding factors that can influence evaluation measure scores. These factors include scene complexity, image resolution and quality, camera characteristics, compression artifacts, and frame rate. Ideally, to obtain a more precise assessment of an algorithm's strengths and limitations, these factors should be distinguished during evaluation. In this study, we focus on one of the most critical distinguishing characteristics between strong and weak algorithms i.e., video complexity. By categorizing our dataset into three distinct complexity levels, we aim to explicitly analyze the impact of video complexity on algorithm performance independently of the other factors. CSVTED has been partitioned into three benchmark tiers, categorized by the degree of tampering quality. Multiple levels also increase its usefulness. The multiple levels allow the dataset to present multiple levels of challenge, which in turn allows researchers to focus on an individual level of CSVTED. That is, if level 3 is perceived as too difficult, then the researcher could restrict testing to the lower levels.

A proficient form of tampering remains imperceptible to the naked eye and exhibits minimal abrupt changes. The quality of tampering is measured using the histograms of the SSIM maps across each of the three channels (RGB) at the start and end of tampering. A high quality tampering is characterized by high SSIM scores between the original and tampered frames across the point of tampering in a video. Various versions of SSIM, such as contrast-weighted SSIM, (CTW-SSIM), saliency-weighted SSIM (SW-SSIM), distortion-weighted SSIM (DW-SSIM), multiscale SSIM (MS-SSIM), and information content-weighted SSIM (IW-SSIM) have been employed by various researchers [61, 63]. We devise W^2F for the purpose of detecting small changes that exist over extensive regions. In this context W corresponds to weight/intensity and F represents the frequency in histograms of the SSIM maps. W^2F also helps in dividing videos into different levels based on tampering quality and can be calculated as follows:

$$W^2F = \left(\frac{\min(V_{start-1}, V_{end})}{Average\left(\sum_{i=1}^{start-2} V_i + \sum_{j=end+1}^{N-1} V_j\right)} \right), \qquad (1)$$

where 'start' and 'end' represent the start and end of a tampered region in a video and V is given below.

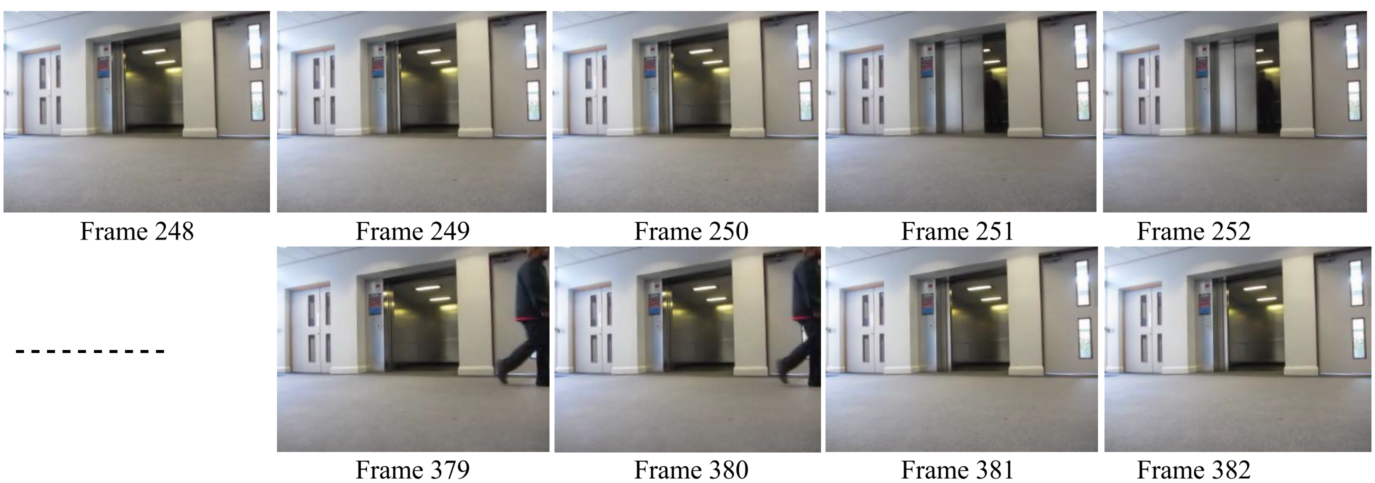
$$V = Average\left(\left(\sum_k w_k^2 f_k\right)_{R,G,B}\right), \qquad (2)$$

where

$$w, f = Histogram\left(SSIM_{Map}(F_i, F_{i+1})\right)_{R,G,B}, \forall i = 1, \dots N \qquad (3)$$

Here $SSIM_{Map}$ gives the SSIM value for each pixel corresponding to frames F_i and F_{i+1} across each channel.

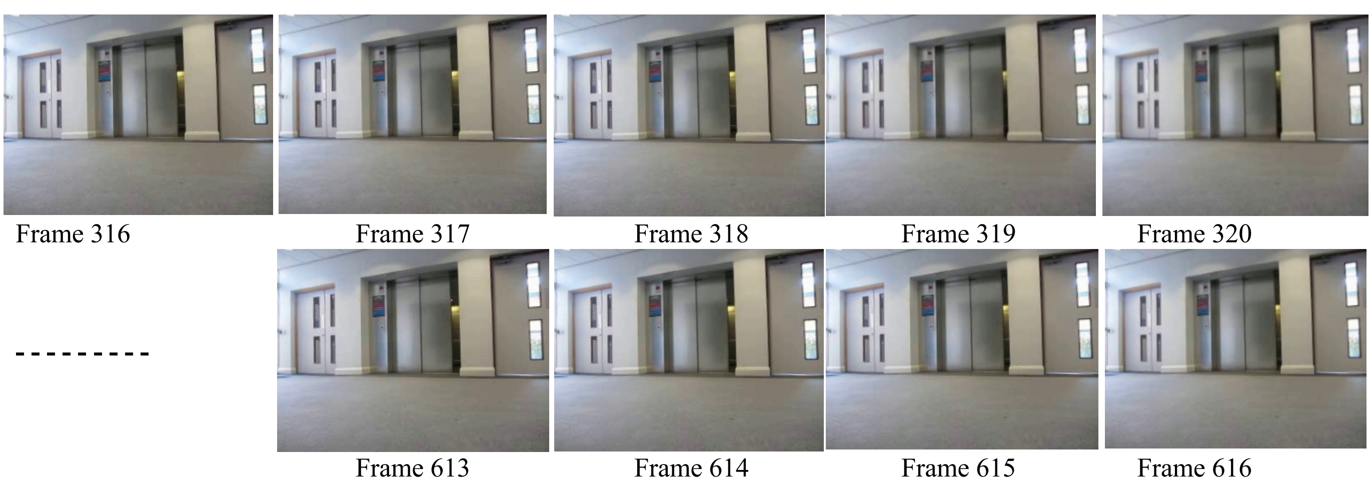
An example illustrating the quality of tampering is demonstrated in Fig. 8. Fig. 8(a) represents the EOP tampering performed by [36] where Frame-251 to Frame-380 are the tampered frames. This tampering is easily discernible due to the abrupt change in content at the start and end (i.e., Frame-251 and Frame-380) of the tampering points, resulting in low tampering quality value of 0.8701. Whereas, Fig. 8(b) demonstrates the tampering performed by the proposed method. Frame-319 to Frame-614 represent the tampered portion where tampering is not detectable at the start or end of the tampering region, resulting in a high tampering quality value of 0.9543.



Frame 248 Frame 249 Frame 250 Frame 251 Frame 252

Frame 379 Frame 380 Frame 381 Frame 382

(a)



Frame 316 Frame 317 Frame 318 Frame 319 Frame 320

Frame 613 Frame 614 Frame 615 Frame 616

(b)

Fig. 8 EOP based frame duplication tampering is performed by (a) the existing method (b) the proposed method

During tampering, the complexity of the video is a significant factor. The complexity of the video includes its content, visual elements, and motion, which can make it more challenging to detect alterations. Complex scenes with a lot of details, fast movement of objects, or varying lighting conditions can provide more chances for tampering to go unnoticed. An attempt is made to measure video complexity, inspired by the image complexity measure [64]. Therefore, complexity estimate will be:

$$image\ complexity\ (IC) = \frac{RMS\_Error}{Compression\_Ratio}, \quad (4)$$

resulting from jpg compression of the image.

The quality factors are selected heuristically with Level 1 ($W^2F \geq 0.95$) indicating high-quality tampering, Level 2 ($0.5 \leq W^2F < 0.95$) indicating medium quality and Level 3 ($W^2F < 0.5$) indicating tampering of low quality. Tables 2-4 provide details of EOP based duplication, deletion, and insertion tampering while Table 5 offers a concise overview of the three-level benchmark dataset based on the tampering quality and video complexity i.e., High Complexity (HC) if value≥0.6, Medium Complexity (MC) if 0.45≤value<0.6, Low Complexity (LC) if value<0.45).

Table 2 Tampering Quality and Complexity of EOP Based Duplication Tampering

Video Name	Tampering Quality	Complexity
Aquarium 01.mp4	1.0043	0.4572
CCTV Moving Bikes.mp4	0.9760	0.4994
Laces_knot.mp4	0.7565	0.5123
Man_Crossing.mp4	0.9318	0.5154
Original.avi	1.0074	0.6138
Picture.AVI	1.0142	0.5387
Professor Hands 01.mp4	0.9087	0.6211
Professor Hands 02.mp4	0.9300	0.6813
Real Varifocal Bullet.avi	0.9507	0.4850
RedBuilding 01.mp4	0.8968	0.4957
RedBuilding 02.mp4	0.9231	0.465
RedBuilding 03.mp4	0.9196	0.4654
Traffic.MOV	0.9420	0.5636
Train Track 01.mp4	1.0172	0.4555
UTurn 02.mp4	0.9070	0.5891
Water 02.mp4	0.9282	0.5300
Water Glass.mp4	0.8955	0.5605
Waterfall 02.mp4	0.9967	0.3976
Waterfall 03.mp4	0.6728	0.3817
Train Track 02.mp4	1.0096	0.4541

Table 3 Tampering Quality and Complexity of EOP Based Deletion Tampering

Video Name	Tampering Quality	Complexity
Aquarium 01.mp4	0.6504	0.4540
Bicycle 02.mp4	0.9597	0.4654
CCTV Blue Bus.mp4	0.8386	0.4336
CCTV Delivery boy.mp4	0.8686	0.4866
CCTV Moving Bikes.mp4	0.7224	0.5005
Laces_knot.mp4	0.7663	0.5124
Man_Crossing.mp4	0.8424	0.5150
Picture.AVI	1.0207	0.5375
Professor Hands 01.mp4	0.9096	0.6209
Professor Hands 02.mp4	0.8963	0.6862
RedBuilding 01.mp4	0.7630	0.5024
Road Crossing 03.mp4	0.5338	0.5062
Traffic.MOV	0.9684	0.5699
Train Track 01.mp4	0.9266	0.4540
Turn 01.mp4	0.3350	0.5913
UTurn 03.mp4	0.7972	0.5914
Water 02.mp4	0.8495	0.5256
Water Glass.mp4	0.9369	0.6168
Waterfall 02.mp4	0.7541	0.4035
Waterfall 03.mp4	0.7514	0.3806

Table 4 Tampering Quality and Complexity of EOP Based Insertion Tampering

Video Name	Tampering Quality	Complexity
Aquarium03.mp4	0.6197	0.4519
Bicycle Lane 01a.mp4	0.9103	0.4679
Bicycle Lane 01b.mp4	0.9132	0.4667
Bus Stop 03.mp4	0.6498	0.4247
CCTV Moving Cycle.mp4	0.8875	0.5007
CCTV Red Car 02.mp4	0.9130	0.4503
Laces_knot.mp4	0.7210	0.5136
Lake Side 03.mp4	0.2318	0.4399
Man_Crossing.mp4	0.7126	0.5143
Park 03.mp4	0.1230	0.4598
Pigeons 01.mp4	0.3072	0.6055
Queen Street 02.mp4	0.2999	0.5112
RedBuilding 01.mp4	0.4425	0.4937
RedBuilding 02.mp4	0.6736	0.4754
Street view 04.mp4	0.8845	0.3911
Traffic 01.mp4	0.5048	0.4739
Train Track 05.mp4	0.8939	0.4543
UTurn 03.mp4	0.8269	0.5881
Uni Road 01.mp4	0.7633	0.4637
Water 06.mp4	0.9201	0.5282

Table 5 Detail of CSVTED, High Complexity (HC), Medium Complexity (MC), Low Complexity (LC)

Tampering Type	No. of Videos	Levels	Number of Videos with complexity level

Frames Duplication	311	Level 01: 42	(HC: 02 MC: 35 LC: 05)
		Level 02: 219	(HC: 40 MC: 123 LC: 56)
		Level 03: 50	(HC: 08 MC: 27 LC: 15)
Frames Insertion	313	Level 01: 14	(HC: 0 MC:14 LC: 0)
		Level 02: 116	(HC: 03 MC: 91 LC: 22)
		Level 03: 183	(HC:24 MC: 101 LC: 58)
Frames Deletion	280	Level 01: 32	(HC: 06 MC:20 LC: 06)
		Level 02: 195	(HC: 30 MC: 102 LC:63)
		Level 03: 53	(HC: 02 MC: 36 LC: 15)

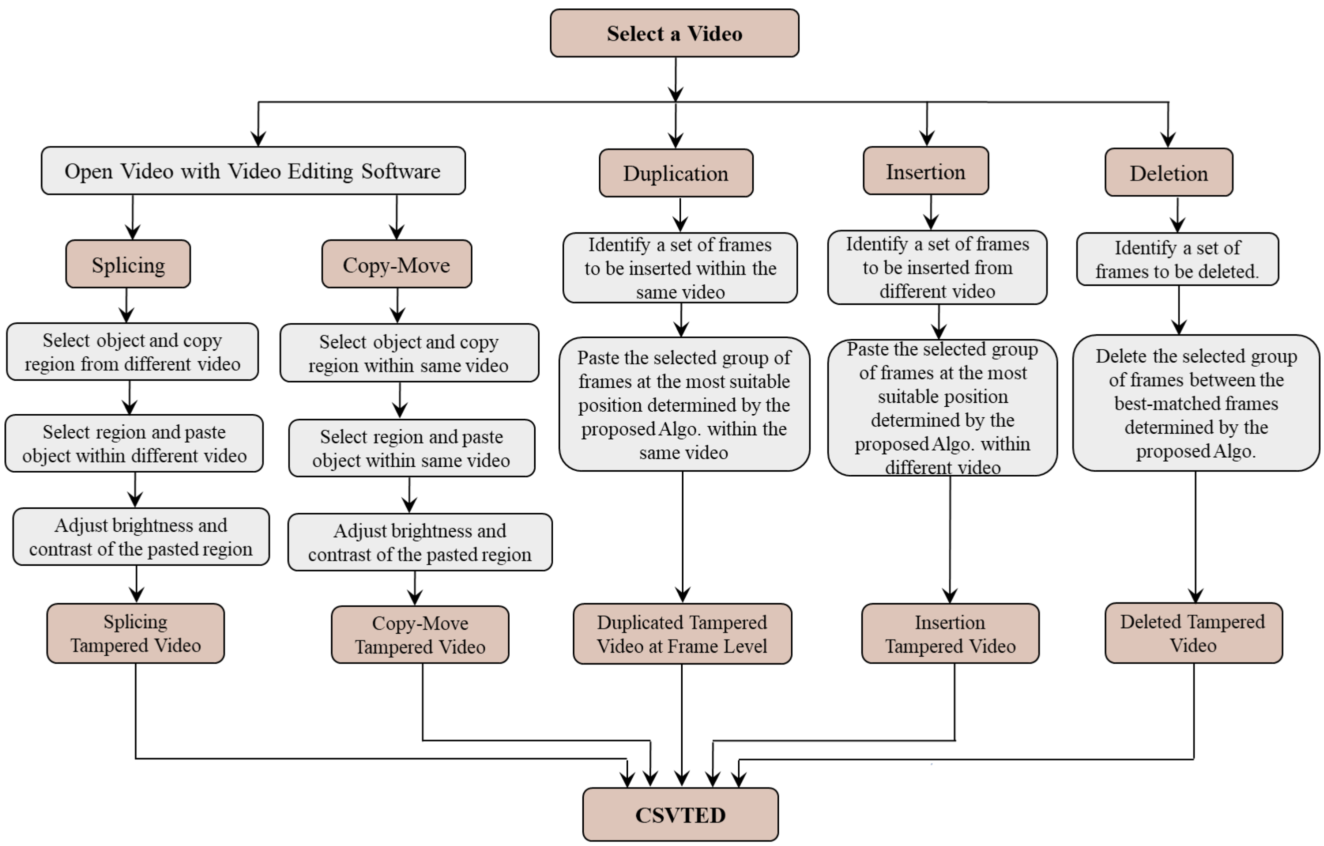


Fig. 9 The framework of the video forgery process

(a) Original (b) Tampered (c) Deduct

Fig. 10 Deducted ground truth

4 Ground Truth

The purpose of providing the ground truth of the dataset is to provide detailed information and explanation concerning each type of video forgery in the developed dataset, which then enables video forensic researchers to quantitatively evaluate the performance of their algorithms. Different types of tools are used for the video forgery in the dataset developed in this paper. It contains the largest number of forged videos compared to existing datasets of video tampering, with a duration from 6.0160s to 47.67s. Along with tampered videos, information about a set of tampered frames is also provided in the dataset that is collected while modifying the original videos. There are two types of footage, i.e., untampered and tampered in each shot, which contains the video name, duration, total number of frames, forged frames in each video, and the percentage of tampering. Researchers can use this information to evaluate the accuracy of their methods. The calculation of the forged percentage depends on the size of the forged area in each frame. In spatial tampering, the percentage of the forged area in a frame is calculated by resolution of the frame. For example, if the frame consists of 10×10 pixels and a 3×4 area is forged, then 12% of the frame size is tampered. In temporal tampering, the ratio of manipulated frames and the number of total frames in the original video is calculated to compute the percentage of the tampered frames of a video. Fig. 10 represents an 8.5% tampering.

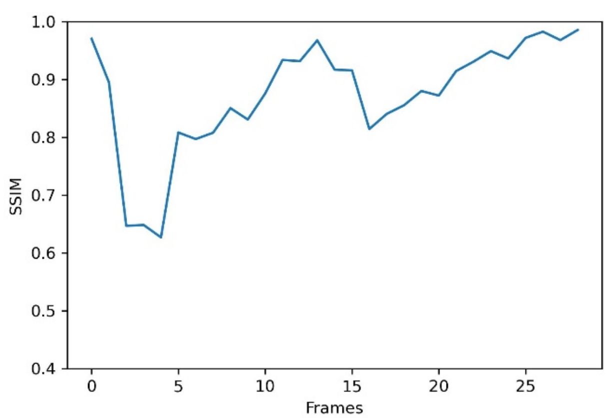
Furthermore, CSVTED contains 1047 videos (both original and forged), 133 of which are not modified to evaluate the ability of algorithms to recognize the original videos, which are shown in Table A1. The remaining 914 videos are divided into five tampering categories: (1) frame insertion, (2) frame deletion; (3) frame duplication, (4) copy-move; and (5) splicing. In the ground truth, complete information related to the original and forged videos is provided.

An important contribution is the 3 levels, and the use of quantitative measures of complexity & quality to confirm the appropriateness of the videos assigned to each tier. Together, the development of innovative video forgery datasets and the quantification of video quality and complexity represent key pillars in advancing the field of multimedia forensics. These metrics not only enable accurate quantification of video fidelity but also guide to develop more effective forgery detection algorithms by capturing subtle alterations indicative of tampering.

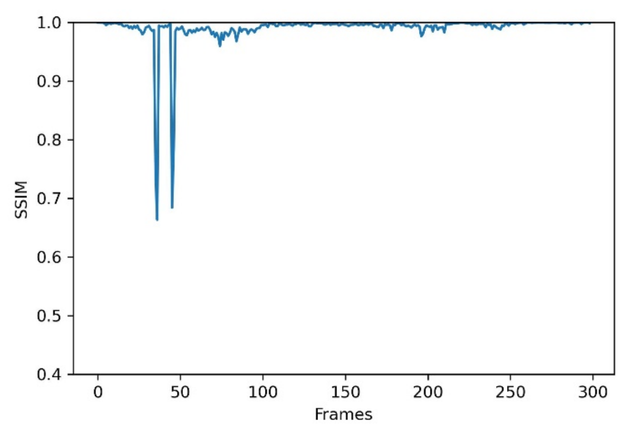
5 Evaluation

Inter-frame tampering detection is conducted using existing algorithms across three different datasets. Two of these datasets, SULFA+VIRAT+LASIESTA+IVY and TDTVD, are sourced from previous research [36, 65], while the third dataset, CSVTED, is the proposed dataset of this study.

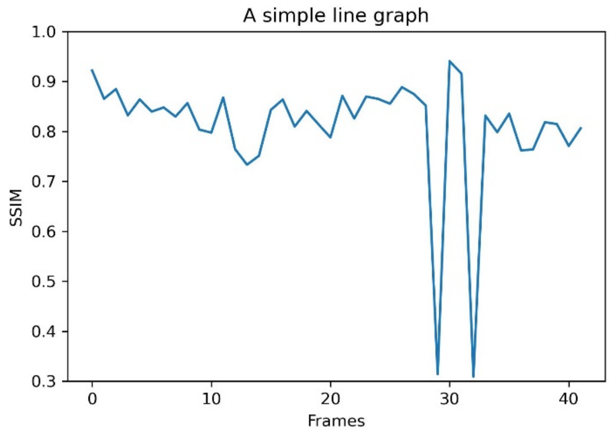
The tampering detection capability of the selected algorithms is assessed through quantitative evaluation and visualized on randomly selected tampered video samples, as illustrated in Fig. 11. It represents that the detection algorithm proposed in [65] is unable to identify tampering on well prepared realistic tampered video samples. Additionally, the performance comparison of two algorithms, measured in terms of detection accuracy, is presented in Fig. 12. The detection algorithm proposed in [65] achieves an accuracy of 99%, 77.27% on SULFA+VIRAT+LASIESTA+IVY and TDTVD respectively, which is high as compared to the detection accuracy on proposed CSVTED. The results also indirectly indicate that the proposed CSVTED contains more diverse and challenging scenarios of video tampering.



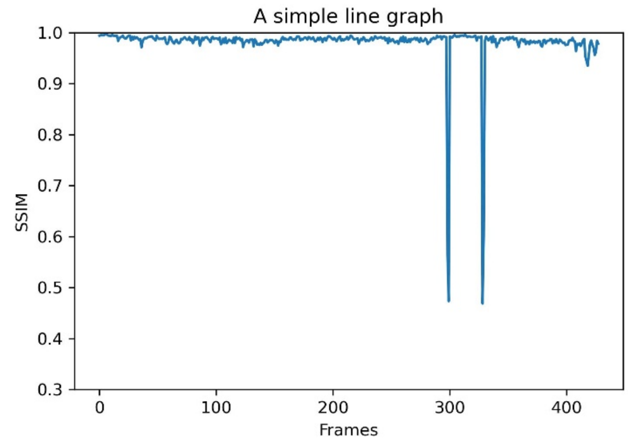
(a) Frame insertion from 150 to 160- Undetected



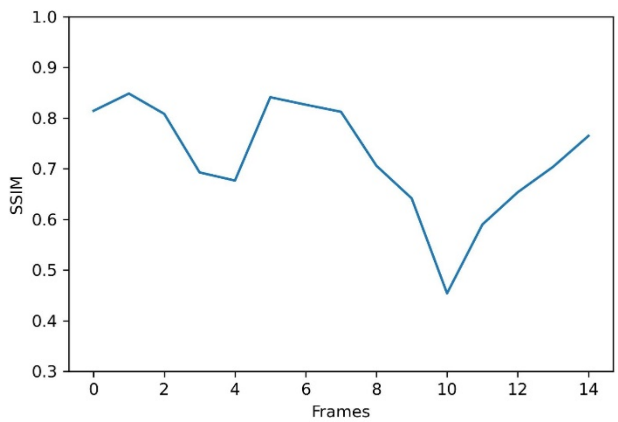
(b) Frame insertion from 150 to 160- Detected



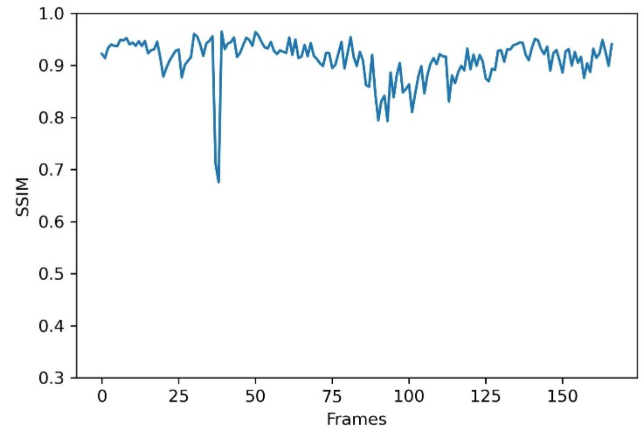
(c) Frame insertion from 43 to 75- Detected



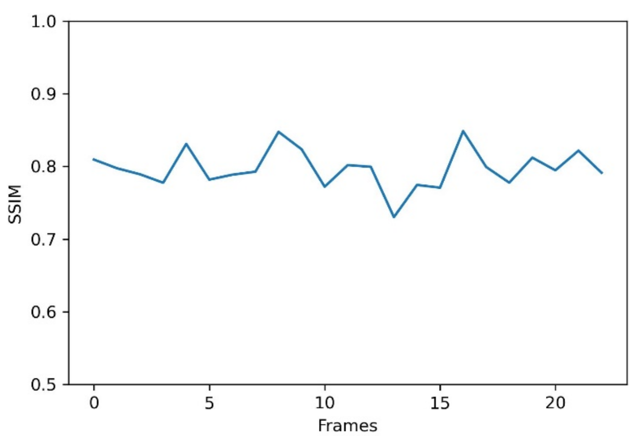
(d) Frame insertion from 43 to 75- Detected



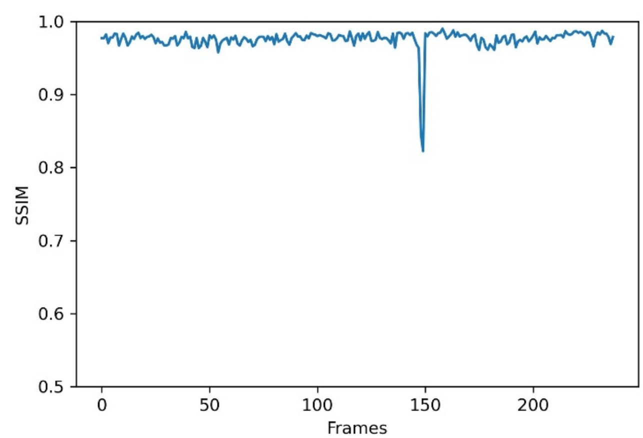
(e) Frame Deletion at 39 - Undetected



(f) Frame Deletion at 39 - Undetected



(g) Frame Deletion at 150 - Undetected



(h) Frame Deletion at 150 - Detected

Fig. 11 Inter-frame tampering detection results on random sample taken from proposed CSVTED. Subfigures (a, c, e, g), and (b, d, f, h) represents the detection results obtained using [65, 66], respectively

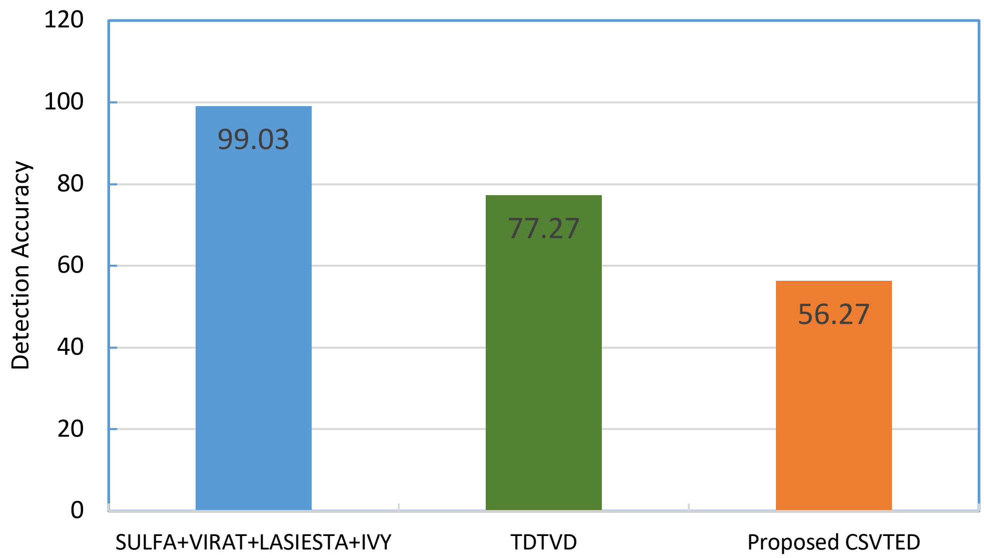


Fig. 12 Inter-frame tampering detection results of [65] on SULFA+VIRAT+LASIESTA+IVY dataset, TDTVD dataset, and proposed CSVTED

6 Conclusion

A three level benchmark dataset of tampered videos CSVTED is comprehensively presented in this paper for the forensic analysis of videos. An important contribution is the 3 levels, and the use of quantitative measures of complexity & quality to confirm the appropriateness of the videos assigned to each tier. Analyzing the quality of tampering provides insights into the characteristics of different types of manipulations, such as frame insertion, deletion, duplication or EOP based tampering. Understanding these characteristics can guide the development of more targeted detection methods tailored to specific types of tampering. All 1047 videos (133 original and 914 tampered videos) in CSVTED are recorded using different digital cameras. To the best of our knowledge, this is the largest tampered video library among the existing video forgery datasets. The primary purpose of developing and designing CSVTED is to standardize the existing methods of video forgery detection, including frame insertion, deletion, duplication, splicing, and copy-move. The frame rate of videos in the dataset varies between 12.50 to 30.0174 fps and the duration of videos from 6.016s to 47.67s. Different cameras are used to record the videos in different realistic situations and the developed dataset also contains videos from smartphones and other digital devices. Extensive experiments have been conducted to assess the performance of the proposed dataset. Furthermore, a comparison with widely used tampered video datasets, as presented in the paper, highlights that the proposed dataset is the largest freely accessible high-definition dataset to date This dataset will be available publicly. Detailed ground truth information is also given for each developed tampered video for authentication of tampering detection algorithms. This library will be very helpful for researchers in evaluating their algorithms.

Author Contributions: Conceptualization, Paul Rosin, Xianfang Sun, Muhammad Hussain and Zulfiqar Habib; Data curation, Naheed Akhtar, Mubbashar Saddique; Formal analysis, Naheed Akhtar; Methodology, Naheed Akhtar and Mubbashar Saddique; Supervision, Paul Rosin, Xianfang Sun, Muhammad Hussain and Zulfiqar Habib; Writing – original draft, Naheed Akhtar and Mubbashar Saddique.

Funding and/or Conflicts of interests/Competing interests: The authors declare that they have no funding, conflict and competing interests.

APPENDIX

Table A1 Summary of Original Videos

No.	Video Name	Duration	Resolution	Total Frames	Frame Rate	Camera Model	Scenario
1.	Aquarium 01	20	1920×1080	600	30	Samsung Galaxy A51	Noon
2.	Aquarium 02	14	1920×1080	420	30	Samsung Galaxy A51	Noon
3.	Aquarium 03	20	1920×1080	600	30	Samsung Galaxy A51	Morning
4.	Bicycle 01	24.66	1920×1080	740	30	Samsung Galaxy A51	Morning
5.	Bicycle 02	20	1920×1080	600	30	Samsung Galaxy A51	Morning
6.	Bicycle 03	33.33	1920×1080	1000	30	Samsung Galaxy A51	Morning
7.	Bicycle Lane 01	20	1920×1080	600	30	Samsung Galaxy A51	Morning
8.	Bicycle Lane 02	28	1920×1080	840	30	Samsung Galaxy A51	Morning
9.	Black Car	7.9675	1280×720	239	29.9967	Nikkon L29	Evening
10.	Bus Stop 01	19	1920×1080	570	30	Samsung Galaxy A51	Noon
11.	Bus Stop 02	16	1920×1080	480	30	Samsung Galaxy A51	Noon
12.	Bus Stop 03	20	1920×1080	600	30	Samsung Galaxy A51	Noon
13.	Bus Stop 04	13	1920×1080	390	30	Samsung Galaxy A51	Morning
14.	Calling Person(M)	7.918	1280×720	198	25.0253	IP Cam H.265	Morning
15.	Canal View	7.28	1920×1080	215	29.5324	Oppo F11	Morning
16.	CCTV Bikes in Fog	10.1	1280×720	150	14.881	HIKVISION Turbo HD Outer	Fog
17.	CCTV Blue Bus	7	1920×1080	210	30	Samsung Galaxy A51	Noon
18.	CCTV Customer Footage(E)	6.896	1280×720	102	14.8256	HIKVISION Turbo HD Inner	Evening
19.	CCTV Customer Shoes(M)	7.291	1280×720	109	14.9725	HIKVISION Turbo HD Inner	Morning
20.	CCTV Delivery boy	28	1920×1080	840	30	Samsung Galaxy A51	Noon
21.	CCTV Moving Bike(E)	10.402	1280×720	155	14.9039	HIKVISION Turbo HD Outer	Evening
22.	CCTV Moving Bikes	26.33	1920×1080	790	30	Samsung Galaxy A51	Noon
23.	CCTV Moving Cycle	20	1920×1080	600	30	Samsung Galaxy A51	Noon
24.	CCTV Moving Dogs	27	1920×1080	810	30	Samsung Galaxy A51	Noon
25.	CCTV Moving person(N)	8.173	1280×720	122	14.951	HIKVISION Turbo HD Inner	Night
26.	CCTV Moving Vehicles(M)	11.099	1280×720	166	14.985	HIKVISION Turbo HD Outer	Morning
27.	CCTV Still Bike(M)	06.1227	1280×720	91	14.80	HIKVISION Turbo HD Outer	Morning
28.	CCTV Red Car 01	8.66	1920×1080	260	30	Samsung Galaxy A51	Noon
29.	CCTV Red Car 02	11	1920×1080	330	30	Samsung Galaxy A51	Noon
30.	CCTV RoadTurn 02	29.33	1920×1080	880	30	Samsung Galaxy A51	Noon
31.	CCTV Scootie 02	33.33	1920×1080	1000	30	Samsung Galaxy A51	Noon
32.	CCTV Street Footage(N)	7.894	1280×720	117	14.8477	HIKVISION Turbo HD Outer	Night
33.	CCTV Walking Persons	47.67	1920×1080	1430	30	Samsung Galaxy A51	Noon

34.	Children	8.18	640×480	132	16.2592	Wifi Robotic Camera V380	Morning
35.	CPU	6.0160	1280×720	73	12.37	Redmi Note 4x	Night
36.	Customer Dealing(E)	11.679	1280×720	292	25.024	IP Cam H.265	Evening
37.	Door Opening(N)	10.031	1280×720	251	25.022	IP Cam H.265	Night
38.	Flying Birds	8.398	1280×720	250	29.7682	Redmi Note 4x	Morning
39.	Light Bulb	8.8747	1280×720	263	29.7688	Redmi Note 4x	Morning
40.	Highway 02	18	1920×1080	540	30	Samsung Galaxy A51	Evening
41.	Highway 03	24	1920×1080	720	30	Samsung Galaxy A51	Evening
42.	Laces_knot	18.97	1920×1080	569	30	Samsung Galaxy A51	Noon
43.	Lake Side 01	13.67	1920×1080	410	30	Samsung Galaxy A51	Noon
44.	Lake Side 02	14.33	1920×1080	430	30	Samsung Galaxy A51	Noon
45.	Lake Side 03	14.67	1920×1080	440	30	Samsung Galaxy A51	Noon
46.	Lake Side 04	25	1920×1080	750	30	Samsung Galaxy A51	Noon
47.	Lake View 01	20	1920×1080	600	30	Samsung Galaxy A51	Noon
48.	Lake View 02	43.33	1920×1080	1300	30	Samsung Galaxy A51	Noon
49.	Lake View 03	43.5	1920×1080	1305	30	Samsung Galaxy A51	Noon
50.	Man_Crossing	22	1920×1080	660	30	Samsung Galaxy A51	Noon
51.	Moving Fish 01	23.33	1920×1080	700	30	Samsung Galaxy A51	Noon
52.	Moving Fish 02	7.03	1920×1080	211	30	Samsung Galaxy A51	Noon
53.	Moving Fish 03	11	1920×1080	330	30	Samsung Galaxy A51	Noon
54.	Moving Fish 04	12	1920×1080	360	30	Samsung Galaxy A51	Noon
55.	Moving Trees 01	16.67	1920×1080	500	30	Samsung Galaxy A51	Noon
56.	Moving Vehicle	9.989	1920×1080	295	29.5342	Oppo F11	Evening
57.	Moving Vehicles 01	7.67	1920×1080	230	30	Samsung Galaxy A51	Noon
58.	Moving Vehicles 02	6.67	1920×1080	200	30	Samsung Galaxy A51	Noon
59.	Name Plate	7.7227	1920×1080	227	29.5895	Oppo F11	Evening
60.	No- Plate	8.7467	1280×720	260	30	Nikkon L29	Evening
61.	Office View	8.08	640×480	202	25.1238	Wifi Robotic Camera V380	Night
62.	Oil Tanker	9.1733	1920×1080	228	25	Nikkon D5200	Morning
63.	Park 01	40	1920×1080	1200	30	Samsung Galaxy A51	Noon
64.	Park 02	21	1920×1080	630	30	Samsung Galaxy A51	Noon
65.	Park 03	18	1920×1080	540	30	Samsung Galaxy A51	Noon
66.	Park 04	15	1920×1080	450	30	Samsung Galaxy A51	Noon
67.	Passing man 01	10.03	1920×1080	301	30	Samsung Galaxy A51	Noon
68.	Passing man 02	8.87	1920×1080	266	30	Samsung Galaxy A51	Noon
69.	Pegions 01	17.5	1920×1080	525	30	Samsung Galaxy A51	Noon
70.	Pegions 02	15	1920×1080	450	30	Samsung Galaxy A51	Noon
71.	Pegions 03	11.67	1920×1080	350	30	Samsung Galaxy A51	Noon
72.	Picture	8.8615	1280×720	266	30.0174	Nikkon L29	Night
73.	Plants	8.28	1920×1080	207	25	Nikkon D5200	Evening
74.	Professor Hands 01	23.33	1920×1080	700	30	Samsung Galaxy A51	Noon
75.	Professor Hands 02	16.83	1920×1080	505	30	Samsung Galaxy A51	Noon
76.	Queen Street 01	14	1920×1080	420	30	Samsung Galaxy A51	Morning
77.	Queen Street 02	10.2	1920×1080	306	30	Samsung Galaxy A51	Morning
78.	RedBuilding 01	26	1920×1080	780	30	Samsung Galaxy A51	Morning
79.	RedBuilding 02	14	1920×1080	420	30	Samsung Galaxy A51	Morning

80.	RedBuilding 03	12.9	1920×1080	387	30	Samsung Galaxy A51	Morning
81.	Road Crossing 01	41.67	1920×1080	1250	30	Samsung Galaxy A51	Morning
82.	Road Crossing 02	41.67	1920×1080	1250	30	Samsung Galaxy A51	Morning
83.	Road Crossing 03	25	1920×1080	750	30	Samsung Galaxy A51	Morning
84.	Road Repairing 01	28.2	1920×1080	846	30	Samsung Galaxy A51	Noon
85.	Road Repairing 02	33.4	1920×1080	1002	30	Samsung Galaxy A51	Noon
86.	Road Repairing 04	11.33	1920×1080	340	30	Samsung Galaxy A51	Noon
87.	School Painting	10.24	1920×1080	255	25	Nikkon D5200	Morning
88.	Scootie 01	43.33	1920×1080	1300	30	Samsung Galaxy A51	Noon
89.	Shop Banners	8.043	1280×720	109	13.5711	Redmi Note 4x	Night
90.	Signal 01	12.4	1920×1080	372	30	Samsung Galaxy A51	Noon
91.	Signal 02	13	1920×1080	390	30	Samsung Galaxy A51	Morning
92.	Signal 03	10.4	1920×1080	312	30	Samsung Galaxy A51	Morning
93.	Signal 04	13.33	1920×1080	400	30	Samsung Galaxy A51	Morning
94.	Signal 05	14.167	1920×1080	425	30	Samsung Galaxy A51	Morning
95.	Slides	7.339	1280×720	218	29.7554	Redmi Note 4x	Evening
96.	Standing Person	8.2133	1280×720	245	29.9976	Nikkon L29	Morning
97.	Street Night View 01	25	1920×1080	750	30	Samsung Galaxy A51	Night
98.	Street Night View 02	28	1920×1080	840	30	Samsung Galaxy A51	Night
99.	Street Night View 03	11.367	1920×1080	341	30	Samsung Galaxy A51	Night
100.	Street view 01	27	1920×1080	810	30	Samsung Galaxy A51	Evening
101.	Street view 02	16.2	1920×1080	486	30	Samsung Galaxy A51	Evening
102.	Street view 03	10.67	1920×1080	320	30	Samsung Galaxy A51	Evening
103.	Street view 04	22	1920×1080	660	30	Samsung Galaxy A51	Evening
104.	Street View(N)	7.178	1920×1080	212	29.535	Oppo F11	Night
105.	Swings 03	29.5	1920×1080	885	30	Samsung Galaxy A51	Noon
106.	Swings 04	8.867	1920×1080	266	30	Samsung Galaxy A51	Noon
107.	Traffic	7.56	1920×1080	189	25	Nikkon D5200	Night
108.	Traffic 01	7.4	1920×1080	222	30	Samsung Galaxy A51	Morning
109.	Traffic 02	27	1920×1080	810	30	Samsung Galaxy A51	Morning
110.	Traffic 03	13	1920×1080	390	30	Samsung Galaxy A51	Morning
111.	Traffic 04	14.67	1920×1080	440	30	Samsung Galaxy A51	Morning
112.	Train Track 01	20	1920×1080	600	30	Samsung Galaxy A51	Morning
113.	Train Track 02	34	1920×1080	1020	30	Samsung Galaxy A51	Morning
114.	Train Track 03	18	1920×1080	540	30	Samsung Galaxy A51	Morning
115.	Train Track 04	28	1920×1080	840	30	Samsung Galaxy A51	Morning
116.	Train Track 05	14	1920×1080	420	30	Samsung Galaxy A51	Morning
117.	Turn 01	25.8	1920×1080	774	30	Samsung Galaxy A51	Morning
118.	Turn 02	17	1920×1080	510	30	Samsung Galaxy A51	Morning
119.	Turn 03	19.33	1920×1080	580	30	Samsung Galaxy A51	Morning
120.	Uni Road 01	21	1920×1080	630	30	Samsung Galaxy A51	Noon
121.	Uni Road 03	23.33	1920×1080	700	30	Samsung Galaxy A51	Noon
122.	Uni Road 04	13.33	1920×1080	400	30	Samsung Galaxy A51	Noon
123.	UTurn 01	23.33	1920×1080	700	30	Samsung Galaxy A51	Noon
124.	UTurn 02	17	1920×1080	510	30	Samsung Galaxy A51	Noon
125.	UTurn 03	9.367	1920×1080	281	30	Samsung Galaxy A51	Noon
126.	Water 01	10.7	1920×1080	321	30	Samsung Galaxy A51	Morning
127.	Water 02	11.67	1920×1080	350	30	Samsung Galaxy A51	Morning

128.	Water 06	11.1	1920×1080	333	30	Samsung Galaxy A51	Morning
129.	Water Glass	7.761	640×480	193	24.9968	Wifi Robotic Camera V380	Evening
130.	Waterfall 02	47	1920×1080	1410	30	Samsung Galaxy A51	Evening
131.	Waterfall 03	30	1920×1080	900	30	Samsung Galaxy A51	Evening
132.	Whiteboard	6.7627	1920×1080	168	25	Nikkon D5200	Night
133.	White Car	10.32	1920×1080	258	25	Nikkon D5200	Evening

Table A2 Video Tampered in Spatial Domain

No.	Video Name	Duration	Total Frames	Camera Model	Type of Video	Resolution	Frame Rate	No. of Frame Tamper	Tampered Area	Tampering Percentage(%)
1	Oil Tanker	09.1733	228	Nikkon D5200	Copy-Move	1920×1080	25.00	156-194	720×324	11.25
2	Light Bulb	08.8747	264	Redmi Note 4x	Copy-Move	1280×720	29.97	1-15	104×128	01.44
3	Name Plate	07.7227	229	Oppo F11	Copy-Move	1920×1080	29.97	1-20	132×276	01.76
5	School Painting	10.2400	255	Nikkon D5200	Copy-Move	1920×1080	25	All Frames	255×255	03.14
4	No-Plate	08.7467	260	Nikkon L29	Splicing	1280×720	30.00	30-45	226×124	03.04
6	Standing Person	08.2133	245	Nikkon L29	Splicing	1280×720	29.9976	61-70	560×250	15.19
7	CPU	06.0160	74	Redmi Note 4x	Splicing	1280×720	12.37	38-74	254×152	04.19
8	CCTV Still Bike(M)	06.1227	91	HIKVISION Turbo HD Outer	Splicing	1280×720	14.80	1-182	268×254	07.39
9	Plants	8.28	207	Nikkon D5200	Splicing	1920×1080	25	All Frames	207×207	02.07
10	Whiteboard	6.7627	168	Nikkon D5200	Splicing	1920×1080	25.00	100-150	168×168	01.36

References

[1] Marcoux, T., Agarwal, N., Erol, R., Obadimu, A., and Hussain, M.N., "*Analyzing cyber influence campaigns on YouTube using YouTubeTracker,*" Big Data and Social Media Analytics: Trending Applications, pp. 101-111, 2021. DOI: https://doi.org/10.1007/978-3-030-67044-3_5.

[2] Saddique, M., Asghar, K., Bajwa, U.I., Hussain, M., and Habib, Z., "*Spatial video forgery detection and localization using texture analysis of consecutive frames,*" Advances in Electrical Computer Engineering, vol. **19**(3), pp. 97-108, 2019.

[3] Richao, C.G., Yang Ningbo, Zhu, "*Detection of object-based manipulation by the statistical features of object contour,*" Forensic science international, vol. **236**, pp. 164-169, 2014. DOI: https://doi.org/10.1016/j.forsciint.2013.12.022.

[4] D'Amiano, L., Cozzolino, D., Poggi, G., and Verdoliva, L., "*A patchmatch-based dense-field algorithm for video copy–move detection and localization,*" IEEE Transactions on Circuits and Systems for Video Technology, vol. **29**(3), pp. 669-682, 2018.

[5] Singh, R.D. and Aggarwal, N., "*Video content authentication techniques: a comprehensive survey,*" Multimedia Systems, vol. **24**(2), pp. 211-240, 2018. DOI: https://doi.org/10.1007/s00530-017-0538-9.

[6] Kaur, S. and Kushwaha, A.K.S. "*A Comparative study of various Video Tampering detection methods*". In Proceedings of *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*. 2018. IEEE.

[7] Saddique, M., Asghar, K., Bajwa, U.I., Hussain, M., and Habib, Z., "*Spatial Video Forgery Detection and Localization using Texture Analysis of Consecutive Frames,*" Advances in Electrical and Computer Engineering, vol. **19**(3), pp. 97-108, 2019. DOI: 10.4316/AECE.2019.03012.

[8] Sitara, K. and Mehtre, B., "*Digital video tampering detection: An overview of passive techniques,*" Digital Investigation, vol. **18**, pp. 8-22, 2016. DOI: https://doi.org/10.1016/j.diin.2016.06.003.

[9] Nabi, S.T., Kumar, M., Singh, P., Aggarwal, N., and Kumar, K., "*A comprehensive survey of image and video forgery techniques: variants, challenges, and future directions,*" Multimedia Systems, vol. **28**(3), pp. 939-992, 2022. DOI: https://doi.org/10.1007/s00530-021-00873-8.

[10] Sitara, K. and Mehtre, B.M., "*Digital video tampering detection: An overview of passive techniques,*" Digital Investigation, vol. **18**, pp. 8-22, 2016.

[11] Jia, S., Xu, Z., Wang, H., Feng, C., and Wang, T., "*Coarse-to-fine copy-move forgery detection for video forensics,*" IEEE Access, vol. **6**, pp. 25323-25335, 2018.

[12] Akhtar, N., Saddique, M., Asghar, K., Bajwa, U.I., Hussain, M., and Habib, Z., "*Digital Video Tampering Detection and Localization: Review, Representations, Challenges and Algorithm,*" Mathematics, vol. **10**(2), pp. 168, 2022. DOI: https://doi.org/10.3390/math10020168.

[13] Saddique, M., Asghar, K., Mehmood, T., Muhammad Hussain, and Habib, Z., "*Robust Video Content Authentication using Video Binary Pattern and Extreme Learning Machine,*" International Journal of Advanced Computer Science and Applications, vol. **10 (8)**(8), pp. 264-269, 2019.

[14] Feng, C., Xu, Z., Jia, S., Zhang, W., and Xu, Y., "*Motion-adaptive frame deletion detection for digital video forensics,*" IEEE Transactions on Circuits and Systems for Video Technology, vol. **27**(12), pp. 2543-2554, 2016.

[15] Verdoliva, D.C.G.P.L., "*Extracting camera-based fingerprints for video forensics,*" Proc. of CVPRW, 2019.

[16] Hong, J.H., Yang, Y., and Oh, B.T., "*Detection of frame deletion in HEVC-Coded video in the compressed domain,*" Digital Investigation, vol. **30**, pp. 23-31, 2019.

[17] Stamm, M.C., Lin, W.S., and Liu, K.R., "*Temporal forensics and anti-forensics for motion compensated video,*" IEEE Transactions on Information Forensics and Security, vol. **7**(4), pp. 1315-1329, 2012. DOI: https://doi.org/10.1109/TIFS.2012.2205568.

[18] Zhao, D.-N., Wang, R.-K., and Lu, Z.-M., "*Inter-frame passive-blind forgery detection for video shot based on similarity analysis,*" Multimedia Tools and Applications, vol. **77**(19), pp. 25389-25408, 2018.

[19] Fadl, S., Megahed, A., Han, Q., and Qiong, L., "*Frame duplication and shuffling forgery detection technique in surveillance videos based on temporal average and gray level co-occurrence matrix,*" Multimedia Tools and Applications, vol. **79**, 2020.

[20] Asghar, K., Sun, X., Rosin, P.L., Saddique, M., Hussain, M., and Habib, Z., "*Edge–texture feature-based image forgery detection with cross-dataset evaluation,*" Machine Vision and Applications, vol. **30**(7-8), pp. 1243-1262, 2019.

[21] Dong, J., Wang, W., and Tan, T. "*CASIA image tampering detection evaluation database*". In Proceedings of *2013 IEEE China summit and international conference on signal and information processing*. 2013. IEEE.

[22] Ng, T.T., Hsu, J., and Chang, S.F., "*Columbia image splicing detection evaluation dataset,*" DVMM lab. Columbia Univ CalPhotos Digit Libr, 2009.

[23] Novozamsky, A., Mahdian, B., and Saic, S. "*IMD2020: A large-scale annotated dataset tailored for detecting manipulated images*". In Proceedings of *the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*. 2020.

[24] Jia, S., Huang, M., Zhou, Z., Ju, Y., Cai, J., and Lyu, S. "*AutoSplice: A Text-prompt Manipulated Image Dataset for Media Forensics*". In Proceedings of *the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

[25] Shullani, D., Al Shaya, O., Iuliani, M., Fontani, M., and Piva, A. "*A dataset for forensic analysis of videos in the wild*". In Proceedings of *International Tyrrhenian*

Workshop on Digital Communication. 2017. Springer, Palermo, Italy, September 18-20, 2017.

[26] Shullani, D., Fontani, M., Iuliani, M., Shaya, O.A., and Piva, A., "*Vision: a video and image dataset for source identification,*" EURASIP Journal on Information Security, vol. **2017**(1), pp. 1-16, 2017. DOI: https://doi.org/10.1186/s13635-017-0067-2.

[27] Akbari, Y., Al-Maadeed, S., Al-Máadeed, N., Al-Ali, A., Khelifi, F., and Lawgaly, A., "*A new forensic video database for source smartphone identification: Description and analysis,*" IEEE Access, vol. **10**, pp. 20080-20091, 2022.

[28] Li, Y., Yang, X., Sun, P., Qi, H., and Lyu, S. "*Celeb-df: A large-scale challenging dataset for deepfake forensics*". In Proceedings of *the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

[29] Zhou, T., Wang, W., Liang, Z., and Shen, J. "*Face forensics in the wild*". In Proceedings of *the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

[30] Jiang, L., Li, R., Wu, W., Qian, C., and Loy, C.C. "*Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection*". In Proceedings of *the IEEE/CVF conference on computer vision and pattern recognition*. 2020.

[31] He, Y., Gan, B., Chen, S., Zhou, Y., Yin, G., Song, L., Sheng, L., Shao, J., and Liu, Z. "*Forgerynet: A versatile benchmark for comprehensive forgery analysis*". In Proceedings of *the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

[32] Mittal, T., Sinha, R., Swaminathan, V., Collomosse, J., and Manocha, D. "*Video manipulations beyond faces: A dataset with human-machine analysis*". In Proceedings of *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2023.

[33] Akbari, Y., Al Maadeed, S., Elharrouss, O., Khelifi, F., and Lawgaly, A., "*A New Dataset for forged smartphone videos detection: description and analysis,*" IEEE Access, vol. **11**, pp. 70387-70395, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3267743.

[34] Singla, N., Singh, J., Nagpal, S., and Tokas, B., "*HEVC based tampered video database development for forensic investigation,*" Multimedia Tools and Applications, vol. **82**(17), pp. 25493-25526, 2023. DOI: https://doi.org/10.1007/s11042-022-14303-y.

[35] Sharma, H. and Kanwal, N., "*Video interframe forgery detection: Classification, technique & new dataset,*" Journal of Computer Security, vol. **29**(5), pp. 531-550, 2021. DOI: https://doi.org/10.3233/JCS-200105.

[36] Panchal, H.D. and Shah, H.B., "*Video tampering dataset development in temporal domain for video forgery authentication,*" Multimedia Tools and Applications, vol. **79**(33), pp. 24553-24577, 2020. DOI: https://doi.org/10.1007/s11042-020-09205-w.

[37] Ulutas, G., Ustubioglu, B., Ulutas, M., and Nabiyev, V.V., "*Frame duplication detection based on bow model,*" Multimedia Systems, vol. **24**(5), pp. 549-567, 2018. DOI: https://doi.org/10.1007/s00530-017-0581-6.

[38] Le, T.T., Almansa, A., Gousseau, Y., and Masnou, S. "*Motion-consistent video inpainting*". In Proceedings of *2017 IEEE international conference on image processing (ICIP)*. 2017. IEEE.

[39] Al-Sanjary, O.I., Ahmed, A.A., and Sulong, G., "*Development of a video tampering dataset for forensic investigation,*" Forensic science international, vol. **266**, pp. 565-572, 2016.

[40] Su, L., Huang, T., and Yang, J., "*A video forgery detection algorithm based on compressive sensing,*" Multimedia Tools and Applications, vol. **74**(17), pp. 6641-6656, 2015. DOI: https://doi.org/10.1007/s11042-014-1915-4.

[41] Chen, S., Tan, S., Li, B., and Huang, J., "*Automatic detection of object-based forgery in advanced video,*" IEEE Transactions on Circuits and Systems for Video Technology, vol. **26**(11), pp. 2138-2151, 2015. DOI: https://doi.org/10.1109/TCSVT.2015.2473436.

[42] Bestagini, P., Milani, S., Tagliasacchi, M., and Tubaro, S. "*Local tampering detection in video sequences*". In Proceedings of *2013 IEEE 15th international workshop on multimedia signal processing (MMSP)*. 2013. IEEE.

[43] Qadir, G., Yahaya, S., and Ho, A.T., "*Surrey university library for forensic analysis (SULFA) of video content,*" 2012.

[44] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "*FaceForensics: A Large-scale Video Dataset for Forgery Detection in Human Faces,*" arXiv preprint arXiv:1803.09179, 2018.

[45] Ardizzone, E. and Mazzola, G. "*A tool to support the creation of datasets of tampered videos*". In Proceedings of *International Conference on Image Analysis and Processing*. 2015. Springer, Genova, Italy.

[46] Liao, S.Y. and Huang, T.Q. "*Video copy-move forgery detection and localization based on Tamura texture features*". In Proceedings of *6th International Congress on Image and Signal Processing (CISP)*. 2013. IEEE, Hangzhou, China, 16-18 Dec. 2013.

[47] Nguyen, X.H., Hu, Y., Amin, M.A., Khan, G.H., and Truong, D.-T., "*Detecting video inter-frame forgeries based on convolutional neural network model,*" International Journal of Image, Graphics and Signal Processing, vol. **10**(3), pp. 1-12, 2020.

[48] D'Avino, D., Cozzolino, D., Poggi, G., and Verdoliva, L., "*Autoencoder with recurrent neural networks for video forgery detection,*" Electronic Imaging, vol. **2017**(7), pp. 92-99, 2017.

[49] Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M.J.a.p.a., "*Faceforensics: A large-scale video dataset for forgery detection in human faces,*" 2018.

[50] Shullani, D., Fontani, M., Iuliani, M., Al Shaya, O., and Piva, A., "*VISION: a video and image dataset for source identification,*" EURASIP Journal on Information Security, vol. **2017**(1), pp. 1-16, 2017.

[51] Newson, A., Almansa, A., Fradet, M., Gousseau, Y., and Pérez, P., "*Video inpainting of complex scenes,*" Siam journal on imaging sciences, vol. **7**(4), pp. 1993-2019, 2014.

[52] Sadddique, M., Asghar, K., Mehmood, T., Hussain, M., and Habib, Z., "*Robust video content authentication using video binary pattern and extreme learning machine,*" Int. J. Adv. Comput. Sci. Appl., vol. **10**(8), pp. 264-269, 2019.

[53] Liao, S.-Y. and Huang, T.-Q. "*Video copy-move forgery detection and localization based on Tamura texture features*". In Proceedings of *2013 6th international congress on image and signal processing (CISP)*. 2013. IEEE.

[54] Atrey, P.K., Yan, W.-Q., and Kankanhalli, M.S., "*A scalable signature scheme for video authentication,*" Multimedia Tools and Applications, vol. **34**(1), pp. 107-135, 2007.

[55] Subramanyam, A.V. and Emmanuel, S. "*Video forgery detection using HOG features and compression properties*". In Proceedings of *2012 IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*. 2012. IEEE.

[56] Lin, G.-S. and Chang, J.-F., "*Detection of frame duplication forgery in videos based on spatial and temporal analysis,*" International Journal of Pattern Recognition and Artificial Intelligence, vol. **26**(07), pp. 1250017, 2012.

[57] Su, L., Huang, T., Yang, J.J.M.T., and Applications, "*A video forgery detection algorithm based on compressive sensing,*" vol. **74**(17), pp. 6641-6656, 2015.

[58] Javed, A.R., Jalil, Z., Zehra, W., Gadekallu, T.R., Suh, D.Y., and Piran, M.J., "*A comprehensive survey on digital video forensics: Taxonomy, challenges, and future directions,*" Engineering Applications of Artificial Intelligence, vol. **106**, pp. 104456, 2021. DOI: https://doi.org/10.1016/j.engappai.2021.104456.

[59] Johnston, P. and Elyan, E., "*A review of digital video tampering: from simple editing to full synthesis,*" Digital Investigation, vol. **29**, pp. 67-81, 2019. DOI: https://doi.org/10.1016/j.diin.2019.03.006.

[60] Hore, A. and Ziou, D. "*Image quality metrics: PSNR vs. SSIM*". In Proceedings of *2010 20th international conference on pattern recognition*. 2010. IEEE.

[61] Wang, Z. and Li, Q., "*Information content weighting for perceptual image quality assessment,*" IEEE Transactions on image processing, vol. **20**(5), pp. 1185-1198, 2010. DOI: https://doi.org/10.1109/TIP.2010.2092435.

[62] Zhang, L., Zhang, L., Mou, X., and Zhang, D. "*A comprehensive evaluation of full reference image quality assessment algorithms*". In Proceedings of *2012 19th IEEE International Conference on Image Processing*. 2012. IEEE, Orlando, FL, USA.

[63] Li, C. and Bovik, A.C., "*Content-partitioned structural similarity index for image quality assessment,*" Signal Processing: Image Communication, vol. **25**(7), pp. 517-526, 2010. DOI: https://doi.org/10.1016/j.image.2010.03.004.

[64] Machado, P. and Cardoso, A. "*Computing aesthetics*". In Proceedings of *Brazilian symposium on artificial intelligence*. 1998. Springer.

[65] Fadl, S., Han, Q., and Li, Q., "*CNN spatiotemporal features and fusion for surveillance video forgery detection,*" Signal Processing: Image Communication, vol. **90**, pp. 116066, 2021. DOI: https://doi.org/10.1016/j.image.2020.116066.

[66] Akhtar, N., Hussain, M., and Habib, Z., "*DEEP-STA: Deep Learning-Based Detection and Localization of Various Types of Inter-Frame Video Tampering Using Spatiotemporal Analysis,*" Mathematics, vol. **12**(12), pp. 1778, 2024. DOI: https://doi.org/10.3390/math12121778.