I.J. Image, Graphics and Signal Processing, **2022, 5, 44-57**
Published Online on October 8, 2022 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijigsp.2022.05.04

A Review on HEVC Video Forensic Investigation under Compressed Domain

Neetu Singla
Netaji Subhas University of Technology, New Delhi, India
Email: neetu.co19@nsut.ac.in

Sushama Nagpal
Netaji Subhas University of Technology, New Delhi, India
Email: sushma.nagpal@nsut.ac.in

Jyotsna Singh
Netaji Subhas University of Technology, New Delhi, India
Email: jyotsna.singh@nsut.ac.in

Received: 28 January 2022; Accepted: 29 June 2022; Published: 08 October 2022

Abstract: In recent years, video forensic investigation has become a prominent research area, due to the adverse effect of fake videos on networks, people and society. This paper summarizes all the existing methodologies used for forgery detection in H.265/HEVC videos. HEVC video forgery is generally classified into two categories as video quality forgery and video content forgery. The occurrence of various forgeries such as transcoding, fake-bitrate, inter-frame forgery and intra-frame forgery is deeply analyzed based on features extracted from the HEVC compression domain. The major findings of this research are (i) Less focus on transcoding detection, (ii) Non-availability of HEVC forged video dataset (iii) More focus on double compression detection for forgery detection, and (iv) Non-consideration of adaptive-GOP structure. The forgery detection in the video is critically important due to its wide use as the primary source of information in criminal investigations and proving the authenticity of contents. So, the forgery detection accuracy is of major concern at the present time. Although, various forgery detection methods are developed in past but the findings of this review point out the need of developing more effective detection methods with high accuracy.

Index Terms: Digital Forensics, Video Forensics, HEVC Codec, Transcoding, Fake Bitrate, Double Compression

1. Introduction

The growth of visual digital data including, images and videos, has been exponential in recent years. Every day, millions of video files are shared over social media sites such as Facebook, WhatsApp, YouTube, and Twitter [1,2]. These digital videos are frequently being utilized in a wide range of sectors, including entertainment, surveillance, news production, and evidence gathering. With the widespread availability of powerful and simple-to-use video editing tools, digital videos are increasingly vulnerable to forgery, compromising their authenticity and integrity [3,4]. The forged videos have the potential to have major implications in politics, the economy, law enforcement, and other spheres. As a result, video forensics has piqued the interest of many researchers in the field of information security.

Because of the huge amount of video data and the close correlation between adjoining frames, digital videos are usually stored and transmitted in compressed form [5,6]. Video coding techniques have evolved over the previous decade to facilitate improved data compression while ensuring high visual quality [7,8]. Fig. 1 depicts the progression of video codec standards from the 1990s to the present. JCT-VC (Joint Collaborative Team on Video Coding) has created a new generation video coding standard named, High-Efficiency Video Coding (HEVC) [9]. Due to the widespread popularity of HEVC videos, forgers prefer to directly re-encode their lower definition videos into HEVCs, without improving video quality, to appear as HD videos [10]. This increases the popularity and profit gains when published on YouTube or other social media.

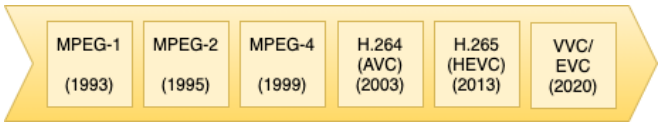

Fig. 1. Evolution of video coding standards.

Video forensic science is a branch of digital forensics that investigates the artefacts left behind by tampering to determine if a digital video is genuine or forged [11]. In the literature, a variety of techniques for video forensic analysis has been proposed. Several surveys have been presented to evaluate video forensic works [12-15]. Authors in [12] examined inter-frame and intra-frame forgery detection techniques on image and video data, but do not discuss transcoding and fake bitrate attacks. Shelke et al. [13] give an in-depth analysis of video forensic techniques, however, work on HEVC-coded videos is only briefly discussed. Few more surveys [14,15], offer an examination of forensic techniques on videos encoded with old coding standards, but these techniques may not be proficient in high-definition and ultra-high-definition video data generated these days. HEVC has a unique quad-tree-like coding structure. In the recent past, many authors have conducted HEVC video forensic investigations by discovering abnormalities in various coding units. To our knowledge, there are no published surveys on HEVC video forensic analysis that cover all potential HEVC video forgeries and focus on the detection features extracted from the compressed domain.

The column graph in Fig. 2 shows the proportion of articles that exploits features extracted specifically from the HEVC compression domain for forgery detection. The rest of the proportion includes generalized forgery detection techniques which are applicable to videos compressed using previous generations of codecs as well as HEVC. The major objectives of the present survey are as follows: (i) To discuss the key design elements and available tools for the HEVC video codec. (ii) Representation of the generalized architecture focusing on the generation of forged videos. (iii) Summarization of different types of video forgeries. (iv) Analysing the research gaps in the existing literature in the field of forgery detection to get future research perspectives.

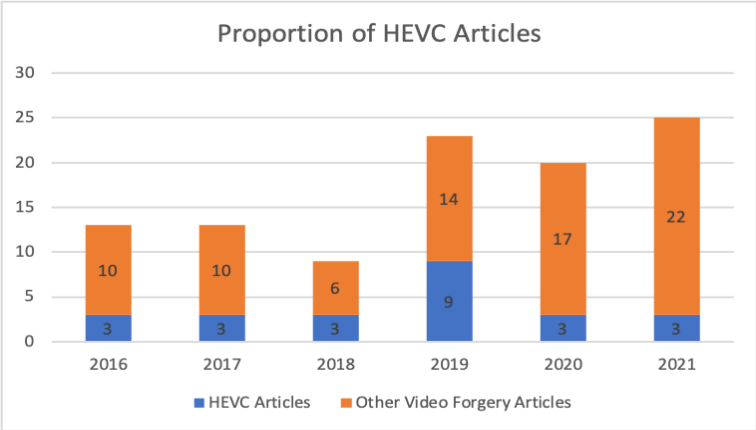

Fig. 2. The proportion of HEVC articles. The blue part represents the proportion of forensic articles that exploits HEVC compression domain features and orange part represents the rest of video forensic articles.

The rest of the paper is structured as follows. In Section 2 an overview of HEVC coding structure, highlighting the key feature domain for forgery detection is presented. Section 3, classifies various types of possible forgeries in HEVC videos. Further, the review of HEVC forgery detection approaches presented by various researchers is discussed in Section 4. Section 5 lists the major findings of this research. Finally, Section 6, concludes the study and provides the future research perspectives.

2. Overview of HEVC Coding Structure

The HEVC standard is designed to accomplish the objective of providing high-quality compression at reduced bitrate, to meet the increasing demand for storing and transferring high-definition video contents. This need is substantially larger on mobile devices and other consumer applications such as digital TV broadcasting [16].

The general architecture of encoding and decoding a sequence of video frames is depicted in Fig. 3. A typical encoder divides each frame into multiple block-shaped units. These units are predicted using intra prediction (by referring to neighbouring regions in the same image) or inter prediction (by referring to neighbouring frames). The first frame of each GOP (Group of pictures) is always intra-predicted, while the rest of the frames are both intra or inter-predicted. The residual information generated by the prediction unit is transformed, quantized and entropy encoded to generate a valid bitstream. This bitstream can be stored and transmitted. The decoder reverses this process by decompressing the bitstream to produce the sequence of frames. In comparison to previous coding standards, efficient partitioning and prediction algorithms have been adopted in the HEVC standard to provide increased compression

efficiency. The following subsections briefly discuss the key design elements of the basic encoding unit. In literature, researchers have performed forensic investigations based on the artefacts found in these coding elements.

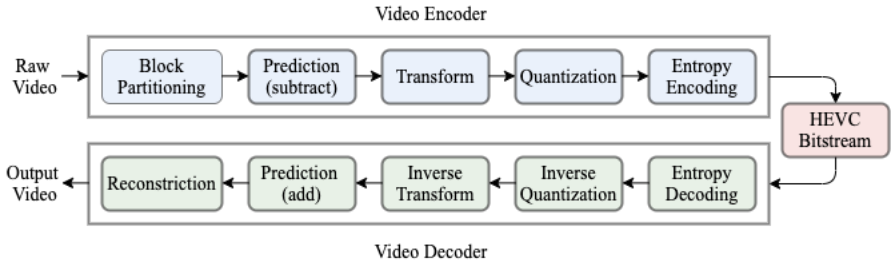

Fig. 3. Block diagram of HEVC encoder-decoder.

A. Coding units (CUs)

Block partitioning divides each frame into equal-sized units called Coding Tree Units (CTUs). Replacing the conventional macroblock structure where the size was fixed to 16x16, CTUs in HEVC are subdivided into flexible CUs to form a quadtree structure. To find the optimal portioning, an efficient Rate-Distortion Cost (RDCost) algorithm is employed [5]. Smoother regions of high-resolution pictures are encoded using larger-sized CUs to provide improved compression efficiency. Whereas smaller-sized CUs are more efficient in dealing with finer details of dense areas resulting in improved picture quality. Fig. 4 shows the partitioning structure of a CTU into CUs and the resulting quad-tree formation for that portion of the video frame.

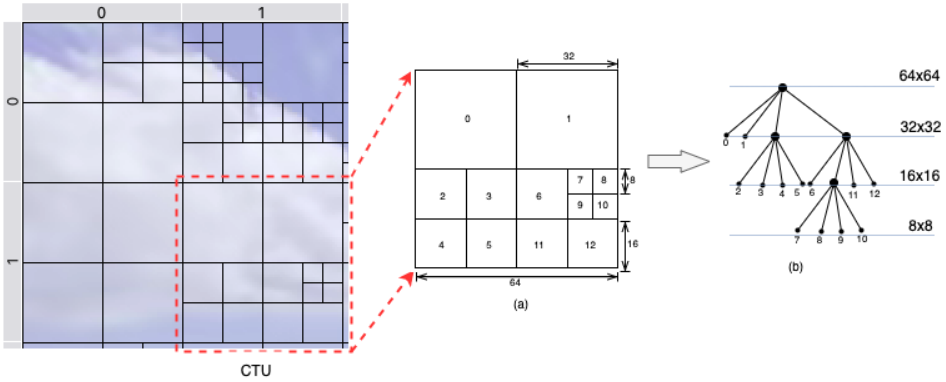

Fig. 4. Partitioning of a CTU into CUs. (a) A CTU with partitioning. (b) Corresponding quadtree representation.

B. Prediction units (PUs)

Image blocks can be predicted in two modes: intra or inter. At the CU level, the decision is made whether to code an image block using intra prediction or inter prediction. Every CU is further split into one or more PUs. HEVC allows prediction block sizes ranging from 64x64 to 4x4. For intra-prediction mode, except for the smallest coding block size, all prediction blocks are of the same size as the coding block. For intra predictions, reference is taken from previously transformed and decoded prediction blocks of the same picture. HEVC offers DC prediction, planar prediction, and 33 different directional predictions to improve intra signal prediction effectiveness. For inter prediction reference is taken from one or two neighbouring pictures. Compared to intra-prediction, HEVC supports symmetric as well as asymmetric shapes for inter-prediction units. To adapt to a variety of texture characteristics, intra prediction supports 5 sizes and inter prediction supports 25 different sizes in HEVC standard.

C. Transform units (TUs)

Block transforms are used to encode the prediction residual. Every CU is recursively subdivided into one or more TUs. A residual TU quadtree structure, rooted at the CU level, is formed. The transform blocks at the leaf of the residual quadtree are processed by transform encoding. For the square transform blocks of sizes 4x4, 8x8, 16x16, and 32x32, integer transforms derived from Discrete Cosine Transform (DCT) are defined. An integer transform generated from a kind of Discrete Sine Transform (DST) is described as an alternate for the 4x4 sized intra-predicted transform blocks. In contrast to prior standards, the HEVC architecture permits a transform block to extend across several prediction blocks for inter-predicted CUs. This exploits the inherent compression efficiency benefits of quadtree-structured partitioning. Fig. 5 depicts the CU partitioning, PU partitioning and TU partitioning for a portion of a video

frame. Table 1 displays the various permitted sizes of the coding blocks, inter and intra prediction blocks and transform blocks.

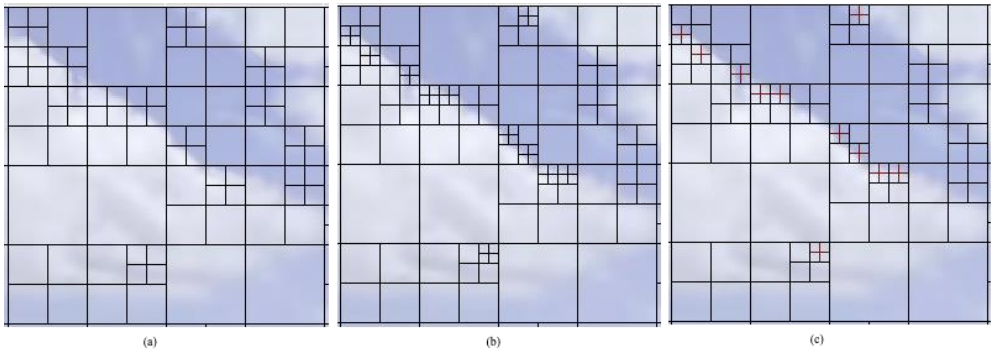

Fig. 5. (a) CU partitioning (b) PU partitioning (c) TU partitioning.

Table 1. Size chart of CU, PU and TU partitioning types in HEVC

CU	PU for I-frames	PU for P-frames				TU
64x64	64x64	64x64	32x32	16x16	8x8	64x64
32x32	32x32	64x32	32x24	16x12	8x4	32x32
16x16	16x16	32x64	24x32	12x16	4x8	16x16
8x8	8x8	64x48	32x16	16x8	4x4	8x8
	4x4	48x64	16x32	8x16		4x4
		64x16	32x8	16x4		
		16x64	8x32	4x16		

D. HEVC Tools

HEVC tools consist of software programs for generating and analysing standard HEVC bitstreams. To perform forensic investigations, researchers need test sequences encoded with the HEVC codec. These sequences can be obtained by encoding and decoding uncompressed YUV sequences with the help of coding software. For the HEVC standard, three open-source implementations are available: HM Reference Model, X265 and FFmpeg. While performing encoding operation the user can set desired configuration parameters such as GOP structure, quantization parameter (QP) value, rate control modes etc.

A bitstream analyser is a graphical tool that allows users to examine in-depth coding information of an encoded bitstream. The HEVC analyser help video coding professional in analysing and visualizing the statistics of inherent design elements. Some of the popular HEVC analysers are GitlHEVCAnalyzer, openHEVC, Zond265, Elecard StreamEye, Parabola Explorer and CodecVisa [17,18]. Among these GitlHEVCAnalyzer and openHEVC provide open-source implementations, while the rest are commercial analysers.

3. HEVC Video Forgeries

There are various kinds of tampering attacks being applied to HEVC videos, commonly classified as Video Quality Forgery and Video Content Forgery [19]. Former, fake about the quality of videos and later, alter the contents of the video. Fig. 6 shows the classification chart of HEVC video forgeries.

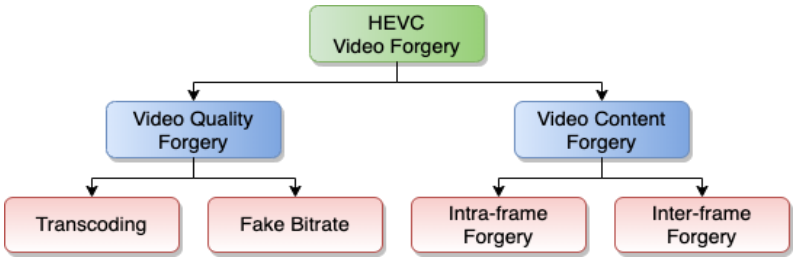

Fig. 6. Classification of HEVC video forgery.

A. Video Quality Forgery

HEVC videos are essentially high-quality videos. In this type of forgery, encoding characteristics of low-quality videos are altered in order to publish them as high-quality HEVC videos. Video quality forgeries are subdivided into Transcoding and Fake Bitrate.

Transcoding Video transcoding is the conversion of digital content from one encoding scheme to another. With the popularity of HEVC, as a symbol of high-quality videos, fraudsters re-encode their originally low-quality MPEG/AVC videos as HEVC videos. In order to gain more financial benefits from advertising and click-through rate, contents are transcoded using newer codecs, without any improvement in the quality. In the entertainment sector, where hundreds of hours of videos are uploaded to YouTube every minute, quality control measures must be extremely efficient. In [10,20,21], the authors presented forensic approaches for determining if it is an authentic HEVC video or it is generated by transcoding an existing MPEG/AVC video sequence.

Fake Bitrate Bitrate is another measure to judge the quality of a digital video. Fraudulent persons tend to increase the bitrate of videos without adding any more information, especially for videos that are shared over the Internet. In this scenario, the ostensibly high bitrate video would have been of inferior visual quality. Many internet bitrate conversion applications that can accomplish this operation are now available for free. These fake upscaled bitrate videos deceive media consumers and potentially waste a lot of storage space and network bandwidth. This paper provides a comparative analysis of the work presented by authors in [19,22,23] for fake bitrate attacks tested for varied bitrates and resolutions.

B. Video Content Forgery

This type of forgery tampers the video either by manipulating the contents inside individual frames or by altering the original sequence of frames [13]. It is divided into two sub-categories: Intra-frame forgery and Inter-frame Forgery.

Intra-frame Forgery A digital video is essentially a collection of still images called frames [24]. A counterfeiter manipulates the contents of a frame in intra-frame forgery. Intra-frame forgeries can be carried out at the pixel level, where specific pixels are altered, or at the object level, where a region containing a specific object is tampered. Typically, the goal of this manipulation is to clone, splice, or inpaint an area or object at a specific location in a frame. Fig. 7 shows an example of intra-frame object forgery, wherein part (a) depicts the cloning operation. A white truck from the authentic sequence of frames is copied and pasted at some other location in the same sequence of frames. In part (b), a person from some other video is spliced into an authentic video to produce a spliced video. An example of an inpainted video is shown in part (c), where a cycle object is removed from the original video by replacing the object region with neighbouring pixel values.

Fig. 7. Example of intra-frame forgery. (a) Cloning forgery; a white truck is copied and pasted, in the same sequence, at some other location (b) Splicing forgery; a lady is spliced in an authentic sequence, from another video (c) In-painting forgery; a cycle object is removed from the authentic sequence.

Inter-frame Forgery Inter-frame forgery is the act of altering a series of frames in a video. A fraudster may purposefully remove, insert, or duplicate a sequence of frames from a genuine video [25]. The objective could be to

destroy evidence of a criminal act or to stage an event at an incorrect moment, in order to slander someone. Surveillance videos are frequently used in criminal case investigations as vital evidence. If such manipulation takes place in surveillance footage, it may go unnoticed by human eyes, jeopardizing justice. Fig. 8 shows an example of video inter-frame forgery. Row (a) depicts an authentic video sequence obtained from a surveillance camera. In row (b), frame deletion forgery is performed by removing a sequence of frames representing the arrival of a person in an office premise. Frame insertion forgery is performed in row (c), wherein, a forged sequence of frames representing the arrival of a person is inserted in the original sequence. The inserted sequence is obtained from the same surveillance camera, recorded on some other day. In row (d), a sequence of frames is copied and pasted in same surveillance footage to depict frame duplicated forged video.

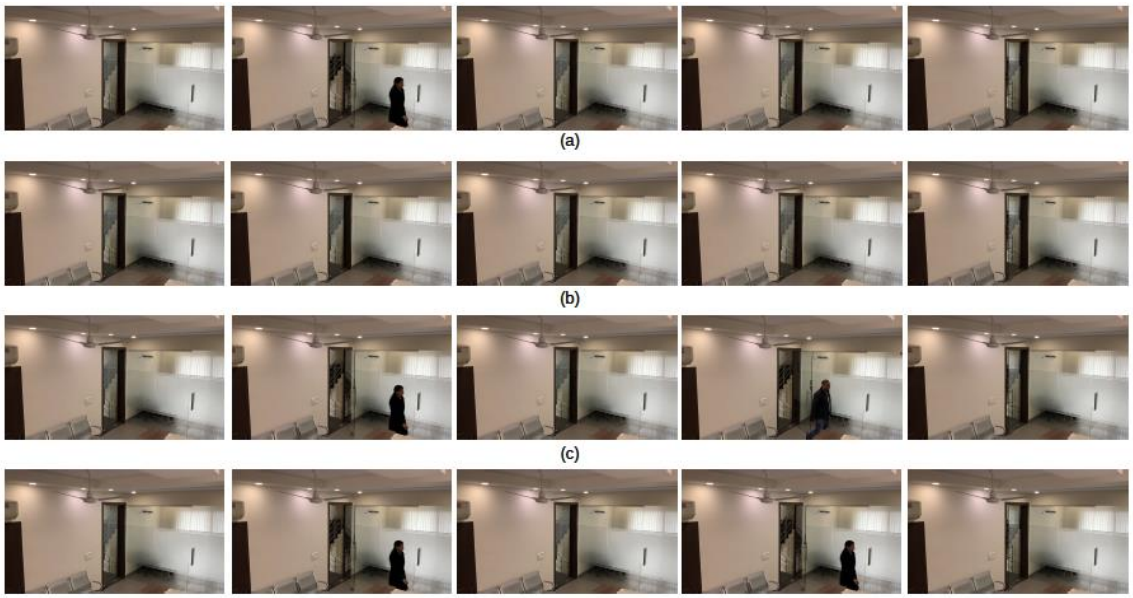

Fig. 8. Example of inter-frame forgery. (a) Authentic video (b) Frame deletion forgery; frames containing a person's arrival are removed. (c) Frame insertion forgery; frames containing a person's arrival are inserted from some other footage (d) Frame duplication forgery; frames containing a person's arrival are duplicated.

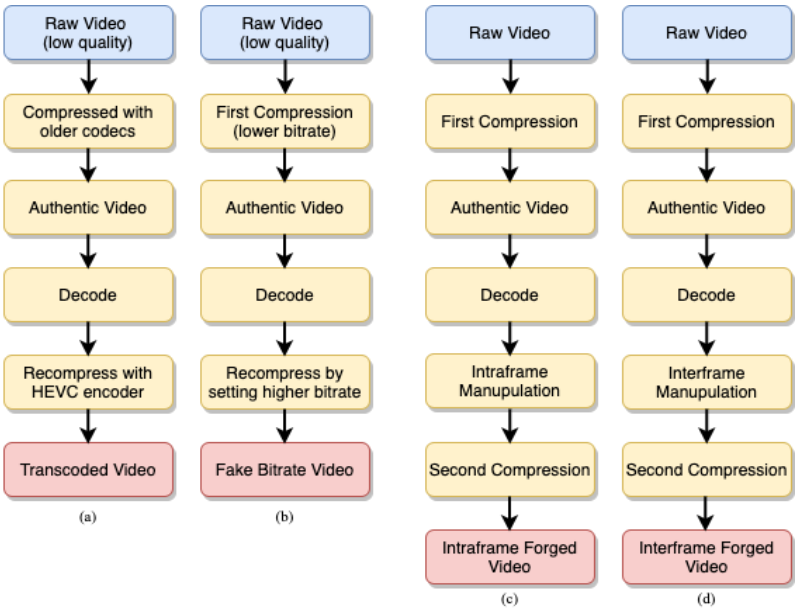

Fig. 9. The general process of generating forged videos. (a) Transcoded (b) Fake Bitrate (c) Intra-frame forged and (d) Inter-frame forged.

Fig. 9 illustrates the process of generating forged videos including, transcoded, fake bitrate, intra-frame forged and inter-frame forged. Typical forgery operations involve decoding the authentic video, performing manipulations and recompressing. Thus, a forgery operation always includes double compression. To detect forgery several double compression detection techniques are proposed in the literature. These techniques extract abnormal features either from decoded frames known as, decompressed domain or extract irregular coding features directly from encoded bitstream

known as compression domain. The approaches proposed in the compression domain require less computation time and demand less cache for processing operations [13]. Thus, for data of high dimension or beyond, analysing compression domain features is more desirable. This paper aims at reviewing HEVC forensic techniques utilizing artefacts that occurred in HEVC coding elements.

4. HEVC Forensic Investigation

This section summarizes the various methods used by different researchers in past for the video forgery detection using the systematic methodology as given in Fig.10. The stepwise description of the methodology adopted for this survey is explained below.

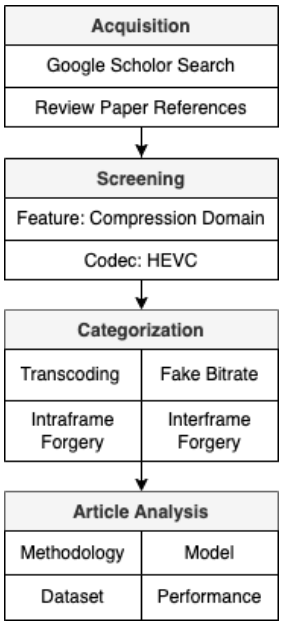

Fig. 10. The methodology adopted for the survey.

1. Data Acquisition: Articles have been collected from google scholar with the help of keywords video forensics, video forgery detection, inter-frame forgery, object forgery, digital forensics etc. A total number of 352 articles have been collected.

2. Article Screening: An elimination approach based on the type of data used for forgery detection has been implemented in order to obtain the relevant articles. The collected research articles contain various articles performing forensic investigations on image data. Such articles are removed and articles with video data are retained. Further, articles investigating compression domain features of the HEVC codec are identified.

3. Categorization: The selected articles are categorized as per the general objective of forgery detection as transcoding detection, fake-bitrate detection, inter-frame forgery detection and intra-frame forgery detection.

4. Article Analysis: Finally, the working concept, model used, datasets, and the performance measures of existing methodologies are examined and analysed, in order to get depth knowledge of this field.

As discussed, any type of forgery operation requires the video sequence to be recompressed, which inevitably leaves artefacts in the fundamental characteristics of coding elements. Such artefacts have been actively investigated by several authors in the past. The Venn diagram in Fig. 11 shows the distributions of articles based on the coding unit utilized for forensic feature extraction. The objective of this section is to provide a comprehensive review of the techniques proposed in the literature for detecting various forgeries in HEVC-coded videos using compression domain features. The detection techniques are grouped into three categories: transcoding detection, fake bitrate detection, and double compression detection. The pie chart in Fig. 12 graphically depicts the proportion of articles published by various researchers among these three categories.

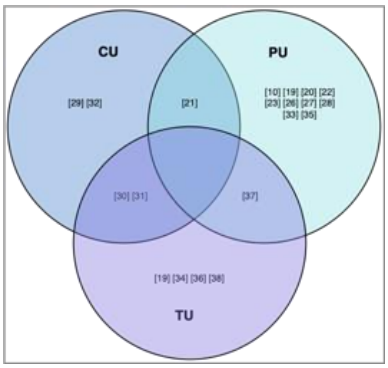

Fig. 11. Distributions of articles among various feature domains.

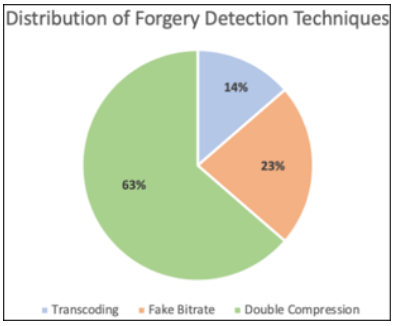

Fig. 12. The proportion of video forensic techniques proposed for various categories.

A. Transcoding detection techniques

Genuine HEVC videos are directly formed from a sequence of raw uncompressed images, whereas transcoded HEVC videos are generated by re-encoding the decoded AVC/MPEG content. Costanzo et al. [10] analysed the frequency of uni-predicted CUs and bi-predicted CU's in all B-frames of genuine HEVC videos and transcoded HEVC videos. The authors observed that in genuine HEVC videos the number of bi-predicted CUs is substantially larger than the number of uni-predicted CUs. The number of uni-predicted CUs, on the other hand, makes up the vast bulk of all units when the HEVC sequence is generated by re-encoding a lower-quality AVC sequence. This approach, however, can only be used if the quantization mode is set to a constant value and QP in re-encoding is less than that first encoding.

Bian et al. [20] explored the differences in prediction unit statistics of I-frame and P-frames for single and double encoded videos. Differences in quantization parameters between the first and second encoding cause the distortion of the optimal encoding scheme thus, to achieve the least cost the partitioning scheme of transcoded video tries different patterns. Authors utilized this observation and came to a conclusion that there exists more fine-grained PUs, i.e., smaller-sized blocks, in transcoded HEVC videos, whereas, the amount of coarse-grained PUs is much less accordingly in transcoded HEVC videos. A 30-D feature vector comprising the statistics of PUs in I and P-frames is utilised for training and testing using an SVM classifier to support this claim. The proposed technique is robust against video enhancement.

Zhang et al. [21] concatenated the characteristics of CU partition types to the PU partition types. A 38-D feature vector is designed by extracting mean frequencies of CU and PU partition types from I-frames and first P-frames of each GOP. The features are then classified as transcoded or not using an SVM classifier with a polynomial kernel. The technique obtained an average accuracy of 98.98% for high-resolution movies, which is slightly better (0.81%) than the scheme described by [20]. The approach claimed to be resilient to both a shifted GOP structure attack and a frame deletion attack. A comparative study of transcoding detection techniques is given in Table 2.

Table 2. Study of transcoding detection techniques

Ref.	Features	Model	Resolution	Performance	Tools
[10]	Frequency of uni-predicted and bi-predicted PUs	Threshold Based Model	352x288	AUC= 94.95	FFmpeg, Commercial analyser
[20]	Statistics of PU types, 25-D feature from P-Frame and 5-D feature from I-Frame of each GOP	SVM classifier with RBF Kernel	1280x720	Accuracy= 98.17%	HM 12.0, JM 19.0, GitlHEVCAnalyzer
[21]	A 38-D features set of CU and PU partition types of I and first P frame of each GOP	SVM with PolySVC kernel	352x288, 1920x1080	Accuracy= 98.98%	HM, JM

B. Fake Bitrate detection techniques

Rate control modes decide the number of bits to spend on a given frame. HEVC standard supports two rate control modes: CBR and VBR. Constant bitrate mode (CBR), forces the video encoder to always use a certain bitrate for all frames. Variable bitrate mode (VBR) enables the encoder to use more bits for complex and fast-moving content while saving bits from smooth textures. Rate control is a critical step in determining the size-quality trade-off. The higher the number of bits, the better the video quality. Using tools like FFmpeg a forger can re-encode an originally low-quality video with higher bitrates, to give the illusion of being a high-quality video.

To expose fake bitrate forgery Liang et al. [22] obtained PU statistics from the first P frame of each GOP. A 25-D feature vector is retrieved from the histogram of the PU partitions, known as HPP. An average of the HPP features of all GOPs in a video sequence is calculated to obtain the final detection feature. The features are then classified using an SVM classifier with a polynomial kernel to obtain an average accuracy of 87%.

In another study, the authors improved average accuracy by extracting multiple prediction features from both I and P frames [23]. Features extracted from I-frames include a 4-D CU partitioning mode (PMoICU), 5-D PU partitioning mode (PMoIPU) and 6-D intra partitioning modes (PMoIIPM). From P-frames, a 4-D CU partitioning mode (PMoPCU) and 25-D PU (PMoPPU) partitioning mode information is extracted. A systematic analysis of individual and combination of features is performed to detect videos with fake bitrate.

Recently, Yu et al. [19] proposed a 10-D feature vector named, PMF formed from the features extracted from intra and inter-prediction modes. HEVC introduced 35 intra-prediction modes. Out of these, Planar, DC, H0 and V0 are suitable for representing homogenous textures. Fake HD videos with low-quality origin have reduced texture complexity, thus during re-encoding more Planar, DC, H0 and V0 modes would be used for intra-prediction in comparison to videos with high-quality origin. A 4-D feature vector of intra-prediction modes is derived from each I-Frame. Further, a 6-D feature vector is extracted from P-frames and B frames, comprising information of Skip, Merge and AMVP inter prediction modes. An ensemble classifier was employed to achieve state of art results with average accuracy near to 99%.

For detecting fake bitrates authors have conducted experiments on videos with both low-resolution (QCIF and CIF) and high-resolution (720p and 1080p). A comparison chart of average detection accuracies achieved by the above works are presented in Table 3. It has been observed that the higher the difference between the bitrate of first encoding and second encoding, the higher the variation in the prediction mode features thus, the higher the detection accuracy. Table 4 presents a summary of Fake bitrate detection techniques in HEVC videos.

Table 3. Comparison of fake bitrate detection accuracy in QCIF, CIF, 720p, and 1080p videos (in percentage). The difference denotes the difference in bitrate between the first and second encodings

	QCIF			CIF		
Difference	[22]	[23]	[19]	[22]	[23]	[19]
100 Kbps	94.47	94.67	-	92.9	94.63	94.63
200 Kbps	100	100	98.5	96.45	98.2	98.2
300 Kbps	100	100	100	100	100	100
	720p			1080p		
Difference	[22]	[23]	[19]	[22]	[23]	[19]
<10 M	91.7	96.53	-	92.25	-	-
10-20 M	97.91	100	98.96	98.65	96.83	97.3
>20 M	100	100	100	99.15	100	100

Table 4. Study of fake bitrate detection techniques

Ref.	Features	Model	Resolution	Tools
[22]	25-D HPP (Histogram of PU Partition Types) feature set from First P-frame of each GOP	SVM with PolySVC kernel	176 x 144, 352 x 288, 1280x720, 1920x1080	H.265/HEVC reference software
[23]	Coding unit prediction modes for I and P frames; Prediction unit prediction modes for I and P frames, Intra prediction modes of I-picture.	SVM with PolySVC kernel	176 x 144, 352 x 288, 1280x720, 1920x1080	H.265/HEVC reference software
[19]	A 4-D feature set extracted from DC, Planar, H0 and V0 direction intra prediction modes; A 6-D feature set extracted from Skip, Merge and AMVP inter prediction modes.	Ensemble classifier	176 x 144, 352 x 288, 1280x720, 1920x1080	x265, HM16.15

C. Double Compression detection techniques

Based upon the parameter selection for recompression, we analyse double compression detection techniques by categorizing them into Non-aligned GOP, Same quantization parameter and Different quantization parameters for first and second encoding.

Double Compression: Non-Aligned GOP Inter-frame forgeries such as frame deletion, frame insertion or frame duplication involves alteration in the temporal sequence of the original video. This changes the original GOP structure causing some I-frames to be re-encoded as P-frames (I-P frames) as shown in Fig. 13 and some P frames to be re-encoded as P frames (P-P frames). When compared to neighbouring P-P frames, relocated I-P frames possess abnormal characteristics in basic coding elements, such as variation in statistics of CU, PU and TU; disproportion of intra, inter and skip units; and disparity in Intra modes has been observed [26-31].

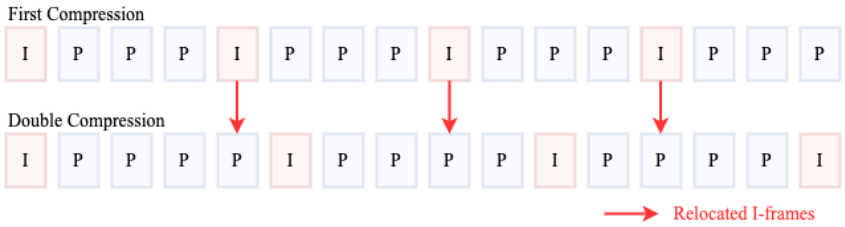

Fig. 13. Relocated I-frames in double compression

Authors in [26] counted the number of intra, inter and skip PUs and designed a feature vector named, SN-PUPM, by taking absolute difference with adjacent three frames. To reduce noise effects, twice averaging filters were introduced. Sequences with abnormal peak values at a periodic interval, indicating an I-P frames change, were classified as double compressed using the SVM classifier. However, when the adaptive GOP algorithm is employed, the proposed sequence-based approach fails because the periodic nature of the recompression traces would be destroyed.

Jiang et al. [27] proposed an enhanced approach for dealing with PU prediction mode statistics. Instead of classifying each sequence authors worked on GOP-based statistics. To generate PU sequences for each GOP unit, the ratio of I-PUs to S-PUs is first calculated. A three-tap median filter was used on PU sequences to reduce the impact of different video contents. A 6-D feature set was formed by applying several low-order statistics, including, maximum, minimum, average and variance. These GOP-based features were then classified using a multi-layer perceptron (MLP). In order to render double compression decision score fusion practice was used. The approach works well for adaptive GOP but when the GOP size of the re-encoding operation is multiple of first encoding, the suggested technique fails.

Further, Wu et al. [28] combined the count of inter PUs with optical flow residuals to make a fusion feature. Their technique improved the detection of relocated I-P frames by 4.59% in comparison to the SNPUPM feature [26]. Hong et al. [32] developed a method for detecting frame deletion forgeries in HEVC encoded videos. They extracted three coding features, namely, intra CU area, residual energy of transform unit and statistics of TUs from the P-frame of each GOP. For a GOP of size N, 3 (N-1) sized feature vectors were generated. To check the authenticity of videos, these features are classified using LDA with k-NN and MLP classifiers. The proposed features are robust for detecting flawless frame deletion and forgery performed in static scenes. However, the technique can work only for fixed GOP sizes.

Recently, deep neural networks have been successfully applied in many areas of computer vision such as object recognition, action recognition, plant disease detection and image forgery detection. These models are capable of automatically learning features from visual data. The authors in [30,31] presented approaches by using a convolutional neural network. Authors in [30] proposed an attention-based two-stream ResNet (ATResNet) network integrated with LSTM model. To identify relocated I-frames two types of feature maps are extracted from compressed bitstream: CSM (CU size map) and CPM (CU prediction mode map). Spatial variations in generated maps are learned by a pre-trained ResNet network where an attention model was employed to optimize network weights and increase detection accuracy. The coding maps extracted from the input video are divided into short clips and LSTM was employed to capture temporal variations. The proposed model is efficient in learning spatio-temporal features, that outperform other forgery detection models.

He et al. [31] examined that low order statistics applied on the distribution of CU types, including intra, inter and skip CUs, can efficiently predict relocated I-frames under increased bitrate. Whereas these features are insufficient to detect relocated I-frames encoded with decreased bitrate. In the first stage, the authors proposed an efficient and compact 50-D feature vector consisting of information related to CU prediction modes, block sizes, and motion vectors. Using these features, an SVM classifier with RBF kernel was able to classify relocated I-frames with increased bitrate correctly. In the second stage, to discriminate between relocated I-frames compressed with decreased bitrate and single compressed frames a Convolutional neural network with dense connections (S-DenseNet) is built. This model learns spatio-temporal representations from CU-type information of three continuous frames and generates output probabilities at the softmax layer. The proposed CNN model obtains an accuracy of 88.51%. Table 5 comparatively studies HEVC double compression detection techniques under Non-aligned GOP.

Table 5. Study of double compression detection techniques under non-aligned GOP.

Ref.	Features	Feature Domain	Model	Resolution	Performance	Tools
[26]	Sequence of Number of Prediction Unit with Intra, Inter and Skip Prediction Modes (SN-PUPM)	Sequence-Based	SVM classifier with RBF Kernal	Not Specified	AUC = 0.9226	Encoder X265 Decoder HM16.15
[27]	A 6-D GOP-based feature set using average, variation and max of S-PU's and I-PU's.	GOP-Based	MLP Classifier with sigmoid function and RPROP optimizer	1920x1080, 1280x720	Accuracy = 94.31%	x265
[28]	Optical flow residuals (OFRB-NIPU); Prediction mode feature proposed by [26]	Sequence-Based	SVM classifier with RBF Kernal	1920x1080, 1280x720	AUC = 0.9707	Encoder X265 Decoder HM16.15
[29]	Depth map of CU and TU partitions and grouping them in GoP units.	GOP-Based	3-D Convolutional Neural Network	1920x1080, 1280x720	Precision=0.9865 Recall = 0.9869 F1-score = 0.9867	HM, Low Delay P mode, Random Access mode
[30]	CU Size Map (CSM); CU Prediction mode Map (CPM)	Frame-Based	Attention-based Ressidual Nework (ATResNet) and LSTM	1920x1080, 1280x720	Accuracy = 94.07% TPR = 92.96% TNR = 95.18%	x265
[31]	Distribution of block size and prediction modes of CU types and Kullback–Leibler divergence between adjacent frames.	Frame-Based	SVM classifier with RBF Kernal; Shallow CNN with Dense connections (S-DenseNet)	1920x1080, 1280x720	Accuracy = 88.51%	x265
[32]	Intra CU Area; Total TU Residual Energy; and Number of TUs extracted from P-Frames.	GOP-Based	LDA and k-NN, MultiLayer Perceptron (MLP)	832 x 480	Accuracy LDA+k-NN=0.823 MLP =0.883	HM16.7 low-delay P mode

Double Compression: Aligned GOP, Same QP Authors in [33-35] studied the recompression artefacts under aligned GOP structure and the same quantization parameter setting for first and second encoding. Jia et al. [33] proposed an approach utilizing the count of 4x4 PUs in each I-frame and computed standard deviation between the single compressed and double compressed versions of the video. Experiments were conducted for QP (22, 24, 26, 28, 32) and an appropriate threshold was selected for detection results. An accuracy of 82.22% was achieved. However, the testing dataset comprised low-resolution QCIF videos.

Elrowayati et al. [34] concatenated the intra-prediction modes with the DST coefficients of each 4x4 PU block extracted from I-frames. A feature named Average Number of Changing PBs (ANCPB) was formulated. ANCPBs are able to recognize the double compression under the same QP, however, the distinctive power of the feature followed a downward trend with the quality of the video.

A novel double compression detection approach under the same QP, based on PU modes in I-frames is presented by Jiang et al. [35]. The proposed feature vector, named Intra Prediction Unit Prediction Mode (IPUPM) composes PU statistics of block sizes: 4x4, 8x8 and 16x16; and intra prediction PU mode of each block with values varying between 0 to 34. The author claimed that by fusing different PU features accuracy of detection has been improved. Moreover, unlike the work in [33] experiments were carried out on HD sequences, resulting in more reliable outcomes. A comparative summary of proposed techniques is given in Table 6.

Table 6. Study of double compression detection techniques under same QP

Ref.	Features	Model	Resolution	Accuracy	Tools
[33]	Standard deviation in number of 4×4 PUs of I-frames (SDoPU)	Threshold Based Model	176x144	82.23%	HM10.0
[34]	Quantized residual DST coefficients and Intra prediction modes of 4x4 PUs in I- frames.	Threshold Based Model	Not specified	83.24%	HM10.0
[35]	Statistics of 4x4,8x8,16x16 PUs and Intra prediction modes in I-frames	SVM classifier with RBF Kernal	1920x1080, 1280x720	97.55%	FFmpeg

Double Compression: Aligned GOP, Different QP For the aligned GOP case, authors in [36-38] presented their studies considering the different QPs for the two compressions. Huang et al. [36] analysed that the quantization errors in recompressed videos cause variation in DCT coefficients. The authors extracted a 136-D feature set consisting of four co-occurrence matrices of neighbouring DCT coefficients and several higher-order statistics, to prove this fact. The proposed algorithm performs well when the first quantization parameter is higher than the second quantization parameter, however, performance is poor when the first quantization parameter is lower or equal to the second quantization parameter.

Li et al. [37] combined the distribution of various TU sizes, specifically its mean, variance, kurtosis and skewness with the DCT coefficients. Authors developed a reduced feature set of size 17-D and achieved comparable detection accuracy to [37] with less computation cost requirements.

Fang et al. [38] proposed a novel methodology to study the influence of DCT coefficients based on transform unit partitioning. The authors extracted a set of three features including, the proportional distribution of DCT coefficients in different types of Tus (PDDC), the autocorrelation property of the DCT coefficients in 4x4 Tus (APDC), and the proportion distribution of the four types of TUs (PDTU). For experiments, only I-frames were considered. During feature selection it has been observed that for TUs of size 16x16 the distribution pattern of DCT coefficients in single and double compressed videos is not distinguishable, thus authors have skipped considering them in feature selection. Table 7 presents a summary of discussed techniques.

Table 7. Study of double compression detection techniques under different QP.

Ref.	Features	Model	Resolution	Accuracy	Tools
[36]	A 136-D feature set formed from co-occurrence matrix of distribution of DCT coefficients	SVM classifier with RBF Kernel	176 ×144, 416 ×240, 832 ×480, 1280×720, 1920×1080, 2560×1600	82.26%	HM12.0
[37]	Distributions of Discrete Cosine Transform (DCT) coefficients and Transform Unit (TU)	SVM classifier with RBF Kernel	176 ×144, 416 ×240, 832 ×480, 1280×720, 1920×1080, 2560×1600	82.11%	HM12.0.
[38]	Proportional distribution of types of TUs (PDTU)	SVM classifier with RBF Kernel	1920×1080	91.10%	x265; HM16.15

5. Discussion and Findings

The major findings derived from the literature concerning to the HEVC video forgery detection methods under the compression domain are summarized below:

- Finding 1: To conduct experiments, researchers have prepared their own forged dataset using uncompressed YUV sequences acquired from the Derf library[1]. However, due to the non-availability of a single repository of forged HEVC videos, a robust comparison of proposed techniques is not possible. Moreover, the library doesn't have videos comprising fast motion. The prime focus of video forgery is surveillance footage; however, the library lacks the collection of surveillance videos for investigation. A centralized repository of forged videos encoded with new generation codecs, such as HEVC, containing a diversity of contents is required.
- Finding 2: Fig. 2 depict that only (20-22)% of the available literature related to the video forgery detection under the compression domain uses HEVC coding features.
- Finding 3: The size of the CU, PU and TU partitioning type also plays a major role during the forgery detection. Fig. 10 show that studies related to prediction unit features have been exploited the most for investigations and no research article has utilized features from all three domains together. Moreover, transform unit features are never studied by articles detecting transcoding and fake bitrate forgeries.
- Finding 4: The distribution of the forgery detection techniques as provided in Fig. 12 point out the very less attention of the researchers on transcoding detection. Transcoding is a major menace for social media users in the present time. Further, it is observed that the main focus of the researchers is on double compression detection, the reason being researchers working for inter-frame and intra-frame forgery detections have anticipated recompression detection.
- Finding 5: This research highlights the presence of four kinds of forgeries such as transcoding, fake-bitrate, inter-frame forgery and intra-frame forgery in HEVC-coded videos. The various features, models, tools and performance measures used by the different authors in past for all types of forgeries are summarized in Table 2-6. To investigate video content forgeries, techniques based on the detection of double compression traces have been utilized. However, because a video can be recompressed while being transferred from one network to another, exclusive inter-frame and intra-frame forgery detection techniques for high-definition data are required.
- Finding 6: The performances of a vast majority of the methodologies discussed in this survey depend on the use of fixed GOP structures, (assumes that the test video under investigation consists of GOPs with a fixed number of frames). However, HEVC compression standards use adaptive GOP structures, and the lengths of these GOPs can be up to 250 frames. Consequently, such techniques may fail entirely for realistic videos. Furthermore, algorithms that exploit relocated I-frames are also unsuitable for this encoding standard. As a result, approaches that take into account adaptive GOP must be developed.

[1] http://media.xiph.org/

6. Conclusion

This paper provides an in-depth analysis of different HEVC video forgery detection techniques operated under the compression domain environment. Such analysis is done based on dataset, resolution, features, methods, type of forgery, performance measures along with their limitations. The present study reveals that most of the video forgery detection techniques addressed a single type of forgery detection instead of multiple forgeries which may be of great interest for future research in this field. Although various forgery detection methods have been developed in past, the findings given in Section 5 raise the need of developing more effective methods for the same. Further, the present survey will give deep insights to the researchers in the field of video forgery detection.

References

[1] A. K. Jain, S. R. Sahoo and J. Kaubiyal, "Online social networks security and privacy: comprehensive review and analysis," *Complex & Intelligent Systems*, pp. 1-21, 2021.

[2] K. K. Sindhu and B. B. Meshram, "Digital Forensic Investigation Tools and Procedures," *International Journal of Computer Network and Information Security*, vol. 4, pp. 39-48, 2012.

[3] P. Johnston and E. Elyan, "A Review of Digital Video Tampering: From Simple Editing to Full Synthesis," *Digital Investigation*, vol. 29, March 2019.

[4] F. F. Chamasemani and L. S. Affendey, "Systematic Review and Classification on Video Surveillance Systems," *International Journal of Information Technology and Computer Science*, vol. 5, pp. 87-102, 2013.

[5] G. J. Sullivan and T. Wiegand, "Rate-distortion optimization for video compression," *IEEE Signal Processing Magazine*, vol. 15, pp. 74-90, 1998.

[6] S. P. Jaiswal and S. V. Dhavale, "Video Forensics in Temporal Domain using Machine Learning Techniques," *International Journal of Computer Network and Information Security*, vol. 5, pp. 58-67, 2013.

[7] X. H. Nguyen, Y. Hu, M. A. Amin, K. G. Hayat, V. T. Le and D. T. Truong, "Three-dimensional Region Forgery Detection and Localization in Videos," *International Journal of Image, Graphics and Signal Processing*, vol. 11, pp. 1-13, 2019.

[8] S. Subbarayappa and K. Rao, "Overview and Extensions of the High Efficiency Video Coding (HEVC) and Beyond (Versatile Video Coding)," December 2019.

[9] G. J. Sullivan, J.-R. Ohm, W.-J. Han and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, pp. 1649-1668, 2012.

[10] A. Costanzo and M. Barni, "Detection of double AVC/HEVC encoding," in *2016 24th European Signal Processing Conference (EUSIPCO)*, 2016.

[11] O. Al-Sanjary, A. Ahmed and G. Sulong, "Development of a Video Tampering Dataset for Forensic Investigation," *Forensic Science International*, vol. 266, July 2016.

[12] H. Kaur and N. Jindal, "Image and Video Forensics: A Critical Survey," *Wireless Personal Communications*, vol. 112, pp. 1281-1302, 2020.

[13] N. Shelke and S. Singh Kasana, "A comprehensive survey on passive techniques for digital video forgery detection," *Multimedia Tools and Applications*, vol. 80, pp. 1-64, February 2021.

[14] R. Singh and N. Aggarwal, "Video content authentication techniques: a comprehensive survey," *Multimedia Systems*, vol. 24, February 2017.

[15] S. Bourouis, R. Alroobaea, A. Alharbi, M. Andejany and S. Rubaiee, "Recent Advances in Digital Multimedia Tampering Detection for Forensics Analysis," *Symmetry*, vol. 12, October 2020.

[16] I.-K. Kim, J. Min, T. Lee and W.-J. Han, "Block Partitioning Structure in the HEVC Standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, December 2012.

[17] Gitlhevcanalyzer: Gitl hevc/h.265 analyzer based on qt. custom filters supported.

[18] S. Jain, A. Fell and A. S. Motra, "A framework for video coding analyzer," in *2016 IEEE Annual India Conference (INDICON)*, 2016.

[19] L. Xiaoyun, Z. Li, Y. Yang, Z. Zhang and Y. U. Zhang, "Detection of Double Compression for HEVC Videos with Fake Bitrate," *IEEE Access*, vol. PP, pp. 1-1, September 2018.

[20] S. Bian, H. Li, T. Gu and A. C. Kot, "Exposing Video Compression History by Detecting Transcoded HEVC Videos from AVC Coding," *Symmetry*, vol. 11, p. 67, 2019.

[21] Z. Zhang, C. Liu, Z. Li, L. Yu and H. Yan, "Detection of Transcoding from H.264/AVC to HEVC Based on CU and PU Partition Types," *Symmetry*, vol. 11, 2019.

[22] X. Liang, Z. Li, Y. Yang, Z. Zhang and Y. Zhang, "Detection of Double Compression for HEVC Videos With Fake Bitrate," *IEEE Access*, vol. 6, pp. 53243-53253, 2018.

[23] X. Liang, Z. Li, Z. Li and Z. Zhang, "Fake Bitrate Detection of HEVC Videos Based on Prediction Process," *Symmetry*, vol. 11, 2019.

[24] S. Jia, Z. Xu, H. Wang, C. Feng and T. Wang, "Coarse-to-Fine Copy-Move Forgery Detection for Video Forensics," *IEEE Access*, vol. PP, pp. 1-1, March 2018.

[25] S. M. Fadl, Q. Han and Q. Li, "Inter-frame forgery detection based on differential energy of residue," *IET Image Process.*, vol. 13, pp. 522-528, 2019.

[26] Q. Xu, T. Sun, X. Jiang and Y. Dong, "HEVC Double Compression Detection Based on SN-PUPM Feature," 2017.

[27] X. Jiang, P. He, T. Sun and R. Wang, "Detection of Double Compressed HEVC Videos Using GOP-Based PU Type Statistics," *IEEE Access*, vol. 7, pp. 95364-95375, 2019.

[28] Q. Wu, T. Sun, X. Jiang, K. Xu, Q. Xu and P. He, "HEVC Double Compression Detection with Non-Aligned GOP Structures Based on a Fusion Feature with Optical Flow and Prediction Units," in *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2019.
[29] J. H. Hong, Y. Yoonmo and T. O. Byung, "Detection of Frame Deletion Using Convolutional Neural Network," *Journal of Broadcast Engineering*, vol. 23, p. 886–895, November 2018.
[30] P. He, H. Li, H. Wang, S. Wang, X. Jiang and R. Zhang, "Frame-Wise Detection of Double HEVC Compression by Learning Deep Spatio-Temporal Representations in Compression Domain," *IEEE Transactions on Multimedia*, vol. 23, pp. 3179-3192, 2021.
[31] P. He, H. Wang, R. Zhang and Y. Li, "A Two-Stage Cascaded Detection Scheme for Double HEVC Compression Based on Temporal Inconsistency," *Security and Communication Networks*, vol. 2021, p. 14, November 2021.
[32] J. H. Hong, Y. Yang and B. T. Oh, "Detection of frame deletion in HEVC-Coded video in the compressed domain," *Digit. Investig.*, vol. 30, pp. 23-31, 2019.
[33] J. Rui-Shi, L. Zhao-Hong, Z. Zhen-Zhen and L. Dong-Dong, "Double hevc compression detection with the same qps based on the pu numbers," *ITM Web Conf.*, 7:02010, 2016.
[34] A. A. Elrowayati, M. F. L. Abdullah, A. A. Manaf and A. S. Alfagi, "Tampering detection of double-compression with the same quantization parameter in HEVC video streams," in *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, 2017.
[35] X. Jiang, Q. Xu, T. Sun, B. Li and P. He, "Detection of HEVC Double Compression With the Same Coding Parameters Based on Analysis of Intra Coding Quality Degradation Process," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 250-263, 2020.
[36] M. Huang, R. Wang, J. Xu, D. Xu and Q. Li, "Detection of Double Compression for HEVC Videos Based on the Co-occurrence Matrix of DCT Coefficients," in *IWDW*, 2015.
[37] Q. Li, R. Wang and D. Xu, "Detection of double compression in HEVC videos based on TU size and quantised DCT coefficients," *IET Inf. Secur.*, vol. 13, pp. 1-6, 2019.
[38] Q. Fang, X. Jiang, T. Sun, Q. Xu and K. Xu, "Detection of HEVC Double Compression with Different Quantization Parameters Based on Property of DCT Coefficients and TUs," *2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1-6, 2019.

Authors' Profiles

Neetu Singla is currently pursuing her PhD from the Department of Computer Science and Engineering, Netaji Subhas University of Technology, New Delhi. She has received her MTech in Computer Engineering from Maharishi Markandeshwar University and BTech in Information Technology from Kurukshetra University. She has 12 years of experience in teaching. Her research interests include computer vision, video analytics, and video forensics.

Sushama Nagpal is currently working as Professor in the Division of Computer Engineering at NSUT, New Delhi. She has almost 25 years of experience in teaching and has actively engaged herself in research. Her areas of interests include Software Quality Measurement, Data Warehouse, Data Mining/Machine Learning, Social Network Analysis and Recommender Systems. She has published various research papers in reputed international journals and conferences. She has reviewed number of research articles for reputed international journals and acted as Member, Technical Program Committee for international conferences.

Jyotsna Singh is working as a Professor in the Department of Electronics and Communication Engineering at Netaji Subhas University of Technology, New Delhi. She has been working with the University for more than 20 years. She is a senior member of IEEE and a life member of IETE. She has also been on the technical program committees of various international conferences such as INDICON, SPIN, ICACCI, CCAIS etc. She has published papers in more than 50 Journals and conferences. Her research interests include Signal Processing, Multimedia Security, Image and Audio Compression.

How to cite this paper: Neetu Singla, Sushama Nagpal, Jyotsna Singh, " A Review on HEVC Video Forensic Investigation under Compressed Domain", International Journal of Image, Graphics and Signal Processing(IJIGSP), Vol.14, No.5, pp. 44-57, 2022. DOI:10.5815/ijigsp.2022.05.04