Interframe Video Forgery Detection and Localization using Histogram-Oriented Gradients for Surveillance Videos

LIBA MANOPRIYA J[1]
AROCKIA JANSIRANI P[2]

Department of Computer Science and Engineering,
Manonmaniam Sundaranar University,
Tirunelveli, Tamil Nadu - 627 012, India.
[1]libamanopriya06@gmail.com
[2]jansicse@msuniv.ac.in

Abstract. Surveillance cameras are generally used in real-time scenarios to provide assurance and security. These videos often serve as crucial evidence in court proceedings. Currently, technology is growing rapidly, resulting in the availability of various editing tools, which are essential for checking the integrity and trustworthiness of video content. Forgery detection is most commonly accomplished through pixel-correlation methods that take a long time to calculate since each pixel of a video frame is compared to identify a forgery. So, the statistical value-based histogram approach effectively detected inter-frame forgeries such as frame insertion, deletion, and duplication. This paper proposes a method to detect forged videos using Histograms of Gradients (HOG) with Discrete Wavelet Transform (DWT). The experimental outcome suggests that the proposed method is more accurate than the existing method and gives a 0.92 accuracy score with a faster execution time.

Keywords: Interframe video forgery detection, Discrete Wavelet Transform (DWT), Histogram of Gradients (HOG).

(Received November 12th, 2022 / Accepted June 1st, 2023)

1 Introduction

Today, as a result of the rapid development of digital technology, multimedia data such as images, video, and audio have been increasing. In criminal cases, videos and images which have been captured by surveillance or CCTV cameras play a crucial role since most of the evidence in court trials comes from these sources. As a result, legal courts consider such multimedia data to be important for maintaining integrity and reliability. Furthermore, the availability of software editing tools such as Adobe Photoshop, Adobe After Effects, MovieMaker, and others are useful in manipulating multimedia data [18, 6, 17]. However, to destroy video evidence, these tools misuse original videos and produce forged or morphed content. Inter-frame video forgery generally falls into three categories: frame insertion, frame deletion, and frame duplication [7, 19]. A frame insertion refers to adding a clip or frame to a video from another video, a frame deletion refers to deleting one or more frames in a video, and a frame duplication is said to occur when frames are copied and pasted into another location within the video [2, 3]. The correlation between adjacent frames of an original video is high, whereas the correlation between frames of a forged video is low. It is crucial to detect forgeries by using different approaches when the correlation factor is considered. Video forensic detection methods are commonly categorized into active and passive approaches [23]. The active approach involves pre-embedding information during the acquisition process, such as incorporating watermarks or digital signatures. However, this approach tends to be more ex-

pensive compared to passive methods.Contrary to this, passive video analysis does not embed information in advance but merely observes forgery traces. This paper presents a technique that uses passive detection techniques to identify different types of interframe forgeries in surveillance videos. The following is an overview of the sections in this paper: In Section 2, a comprehensive review of the literature on video forgery detection methods is presented. Section 3 offers a concise description and technical analysis of the proposed system, which aims to identify and locate different types of interframe forgeries. The experimental results are presented and discussed in Section 4. Finally, Section 5 concluded the research work presented in this paper, summarizing the finding and highlighting their implications.

2 Related Work

Over the past decade, a great deal of research has been conducted in the field of digital forensics. Many noteworthy studies have also been conducted regarding the detection of video forgeries, such as the works reported in [2,3,6,9,10,12]. These studies aim to detect common inter-frame forgery detection such as Frame insertion, deletion, and duplication. Nowadays Passive or Blind detection techniques are used to detect interframe forgery videos in a surveillance video.

Bakas et al.,[4] used three modules for identifying forgery frames in a video. In the first module by exploiting Prediction Footprint Variation (PFV), abnormal p-frames are detected in videos. In the second module, outliers are rectified again so that the false positive rate is decreased. These modules also identify the exact location of the forgery. The final module has classified the video according to the type of forgery that occurred. In the case of inserting or deleting a Group of Pictures (GOP), this method is not appropriate. Fadl et al.,[8] attempted to combine the spatiotemporal average in each frame, and a 2D Convolutional Neural Network (CNN) is used to extract useful features. Following that, multiclass support vector machines (SVM) and Gaussian Radial Basis functions (RBF) are used to classify the video. This method is ineffective when more than one forged attack is used to manipulate the video. Fadl et al.,[9] utilized the Histogram of Oriented Gradient (HOG) to find the frame insertion and deletion forgery. Correlation coefficients and Motion Energy Image (MEI) values are used to detect abnormal points in the frame. There are some cases in which a forgery cannot be detected in more than one type. According to Zhao et al.,[33] features are extracted using the Speeded-Up Robust Features (SURF) and similarity is determined by Fast Library Approximate Nearest Neighbors (FLANN) based on the Hue-Saturation-Value (HSV). But it cannot handle situations where a shot is incorrectly obtained.

As described by Han et al.,[10] residual features are determined by calculating spatial energy, temporal energy, as well as signal-to-noise ratio values. This method becomes ineffective when the deleted frames are static scenes. Liu et al.,[13] have addressed a variety of forgery attacks such as frame addition, deletion, and duplication. It is reported in the paper that every frame is transformed into 2D opponent chromaticity space, and then Zernike Opponent Chromaticity Moments (ZOCM) are calculated by detecting abnormal points between frames in response to ZOCMs. Additionally, the coarseness feature of Tamura is used to reduce the false-positive rate in videos. However, this method is more time-consuming. Raahat et al.,[21] have used optical flow and prediction residual techniques. This method was ineffective due to the multiple compressions involved. Ulutas et al.,[25] have calculated the peak signal-to-noise ratio and distance between frames, after which binary features are extracted to evaluate the similarity between features. As a result, it requires a lot of computation time when processing each video frame.

In the frame insertion and deletion, Li et al.,[12] have utilized a method called Mean Structural Similarity Measure (MSSM), which calculates the quotient between the MSSM values of adjacent frames. According to Zhang et al.,[32] the inconsistencies between frames are detected by the quotients of correlation coefficients between local binary patterns (QoCCLBP), and the Tchebyshev inequality is used to determine the abnormality. In any case, deleting five frames will yield poor accuracy. Wang et al.,[28] have used the optical flow method, in which the abnormality is detected based on the Gaussian distribution values, and Grabb's test is used to find the forgery videos exactly. By using the optical flow technique, Chao et al.,[5] have extracted the feature from each frame but the time complexity of this technique is high.

Su et al.,[24] have extracted MPEG video features using Discrete Cosine Transform (DCT) coefficients. However, this method is not suitable for all video encoding types. It is also difficult to detect when a single Group of Pictures(GOP) or multiple GOPs are deleted. Wang et al.,[27] have detected the forgery videos by using the spatial and temporal correlation between adjacent frames. As soon as the frames are split into overlapping sequences, the temporal correlation for each frame is computed to detect frame duplication, then the spatial correlation is calculated to detect spatial manip-

ulation within a video. Although this is an acceptable technique for reducing the block size, it is not applicable to increasing the block size. Fayaaz et.al.,[11] have exploited the Locally Adaptive Discrete Cosine Transform (LADCT) filtering and weighted average technique to calculate the sensor pattern noise in surveillance videos and finally have detected forgeries using cross-correlation analysis. In this case, the forgery is not detected even when the attacker generated noise in the middle of the video. Sitara K. and Mehtre BM. [22] utilized the Velocity Field Intensity (VFI), Variation of Prediction Artifact (VPF), and generalized extreme studentized deviation (ESD) approaches in order to detect the forgery and locate the forged component in a video. One of the drawbacks of this approach is that the VFI and VPF values see a sharp increase whenever there is a quick shift in the camera lens. In this scenario, the video is considered to be forged even if it is in perfect condition because it does not meet the criteria for authenticity.

Qi et al.,[26] identified the inconsistency in the velocity field, the change in the prediction footprint, and the correlation coefficient values, and then estimated the texture coarseness to identify the forged section. In this method, dynamic background videos make it difficult to detect forgery. Rahul et al.,[15] investigated an approach based on the Normalized Multi-Scale One Level Subtraction (NMOLS) and localized the forging point using the Extreme Student Deviate (ESD) test. This method works only when the video is still and more than five frames have been added or taken away. Yu et al.,[31] described two feature extraction steps. First, remove the magnitude prediction residual. Second, count intra-coded P-frame macroblocks and fuse both features to find the deleted frame. This method is not suitable for slow-motion videos with fewer than five frames.

Yao et al.,[30] made use of a method known as frame interpolation, in which Adaptive Overlapped Block Motion Compensation (AOBMC), and global and local residual characteristics were utilized for the purpose of determining whether or not a frame had been deleted. On the other hand, there is a significant amount of time complexity. Wei et al.,[29] used a multiscale standardized mutual information procedure to find frame deletion and duplication. But this does not work for a video that has more than one forgery on it. According to Aghamaleki et al.,[1] the DCT coefficient and quantization residual values can be used to determine spatio-temporal information. The fused values are used to determine if a video frame has been inserted or removed. However, this strategy is not applicable to videos in a dynamic environment. Ren et al.,[16] developed a technique for detecting duplicate frames using the improved Levenshtein distance; however, this experiment cannot be used to identify duplicate frames with dynamic backgrounds. Priyadharshini et al.,[20] used Earth Mover's Distance (EMD) to detect forgeries and abnormal points. This method does not work when frames are inserted or deleted at the video's beginning or end.

3 Proposed Inter-frame Video Forgery Detection Methodology

A general inter-frame forgery involves frame insertion, frame deletion, and frame duplication. After preprocessing, an input video frame is transformed into wavelet coefficients. Then, the HoG-based feature extraction technique is used to extract significant features from the approximation coefficients of the wavelet-transformed output. Finally, the features extracted are given to a two-dimensional Convolutional Neural Network (2D CNN) for classification. At this stage, the input video frame is detected if it is forged or not. Then, the forged video is subject to localization in order to detect frame insertion, deletion, and duplication. The detailed methodology of the proposed work is described below:

3.1 Preprocessing

In preprocessing, video frames are extracted and resized. Subsequently, the frames are converted into gray levels using the equation:

$$F(x,y) = 0.299xR(x,y)+0.587xG(x,y)+0.114xB(x,y) \quad (1)$$

Here, R(x, y), G(x, y), and B(x, y) represent the Red, Green, and Blue channel intensities of the image, and F(x,y) represent the gray level intensity of the image.

3.2 Mathematical transformation under frequency domain

As part of this procedure, DWT is utilized to extract spatial and temporal information. The application of DWT on images gives four bands namely LL, LH, HL, and HH bands. DWT supports multilevel decomposition and the LL band should be used to construct the next levels of decompression since it has the maximum amount of information about the frame. As opposed to dividing the image into small blocks, this method focuses on the picture as a whole and can find the signal with great precision. Therefore, the preprocessed input

is transformed using DWT. The transformation process converts the spatial details into respective transform coefficients. The LL subband after the final level of decomposition is used in the interframe detection process.

3.3 Feature Extraction

A histogram of oriented gradients (HOG) is a feature descriptor technique that extracts gradients from an image to find the edge and local shape information. To reduce the influence of light and noise, the HOG descriptor performs color and gamma normalization on the images. In this F represents the image and (i, j) represents the row and column of an image. By using the following equations, the horizontal (G_x) and vertical (G_y) gradient is calculated.

$$G_x = F(i, j+1), -F(i, j-1) \tag{2}$$

$$G_y = F(i-1, j) - F(i+1, j) \tag{3}$$

Based on the Horizontal and vertical gradient values, the Magnitude and direction of the gradient values are calculated.

$$Magitude\mu(i,j) = \sqrt{G_x^2 + G_y^2} \tag{4}$$

$$Directionangle\theta = tan^{-1}(\frac{G_y}{G_x}) \tag{5}$$

In this gradient, a range of 0° - 360° is provided in each of the nine orientation bins. Weights are derived from gradient amplitudes in each direction. In order to construct the feature vector of each block, the image is divided into 16 x 16 pixel blocks and the histograms of the cells are concatenated. L2 normalization is applied to feature vectors to reduce the effects of local illumination variations and visual angle changes. To generate the picture's HOG feature vector, all of the block feature vectors are concatenated into one vector. For a 128 x 64 sliding detection window, obtain a 1D-dimensional HOG feature vector, which can be described by 7 x 15 blocks.

3.4 Classification

Convolutional Neural Networks (CNNs) are utilized for the analysis of two-dimensional images. These networks consist of several layers, including convolutional, pooling, and fully connected layers. The architecture of 2D-CNN typically comprises three convolutional layers (conv2d, conv2d_1, conv2d_2), followed by three max pooling layers (max_pooling2d, max_pooling2d_1,max_pooling2d_2), fully connected layers. Each convolutional layer employs 3 x 3 kernels and incorporates 16, 64 and 128 filters, respectively.

The LeakyRelu function is the activation function for all layers except the fully connected layer. The activation layer formula of LeakyRelu is as follows:

$$f(x) = \begin{cases} 0.01x, & x < 0 \\ x & x >= 0 \end{cases} \tag{6}$$

In many cases, Rectified Linear units (ReLu) activations are used since they do not suffer from backpropagation errors. Meanwhile, one or more neurons become inactive or worth zero, leading to the Dying ReLu Problem. To overcome this issue, the LeakyReLu activation function is used, which results in a more efficient and effective model. Finally, flatten layers are used for converting two-dimensional values into one-dimensional data and provide 204 feature vectors. In addition, the binary cross-entropy is applied to the loss function and optimization using the Adam function. The softmax activation function is used as the final step in the classification process to detect the various types of forgeries. Table 1 includes information about all layers in 2D-CNN architecture.

Table 1: Detailed information about 2D-CNN Architecture

$Layer(type)$	$OutputShape$	$Param\#$
$conv2d(Conv2D)$	(128,64,16)	160
$leaky\_re\_lu$ $(LeakyReLU)$	128,64,16	0
$max\_pooling2d$ $(MaxPooling2D)$	64,32,16	0
$conv2d\_1(Conv2D)$	64,32,64	9280
$leaky\_re\_lu\_1$ $(LeakyReLU)$	64,32,64	0
$max\_pooling2d\_1$ $(MaxPooling2D)$	32,16,64	0
$conv2d\_2(Conv2D)$	32,16,128	73856
$leaky\_re\_lu\_2$ $(LeakyReLU)$	32,16,128	0
$max\_pooling2d\_2$ $(MaxPooling2D)$	16,8,128	0
$flatten(Flatten)$	(16384)	0
$dense(Dense)$	(50)	819250
$dense\_1(Dense)$	(4)	204

Proposed Algorithm for Inter-frame Forgery detection This section presents the detailed procedure for inter-frame forgery detection using the proposed method. The proposed system is capable of detecting insertion, deletion, and duplication attacks.

Algorithm 1- Procedure for Discrete Wavelet Transform with Histogram Oriented Gradients

$I_{img} \rightarrow InputImage$

Perform 2D DWT for I_{img}, extract LL approximate image of an input image

[LL LH, HL, HH] = 2dwt(I_{img})

Perform Histogram Oriented Gradients for LL image

1. Image Resizing

R_{img} = Convert image into 68 x 128

2. Calculating gradients for R_{img}

$G_x \rightarrow$ Calculate x direction gradient

$G_y \rightarrow$ Calculate y direction gradient

3. Compute the Magnitude and Orientation

$\sqrt{(G_x)^2 + (G_y)^2}$

$G_a \rightarrow$ atan$(\frac{(G_y)^2}{G_x)^2}$

4. Compute Histogram of Gradients in 8x8 cells

In this stage, the image is divided into 8x8 cells. For each of these cells, a gradient histogram is computer. This process generates a histogram matrix of size 9 x 1 for each individual cell.

5. Normalize gradients in 16x16 cell

The next step involves combining the 8x8 cells into 16x16 blocks. This combination results in a histogram value matrix of size 36 1. To normalize this matrix, each of these values is divided by the square root of the sum of squares of all the values

$V_m \rightarrow [c1, c2, ...c36]$

Perform root of the sum of squares

$R_s \rightarrow \sqrt{(c1)^2 + (c2)^2 + (c3)^2, ...(c36)^2}$

Normalize the vector

$N^{vec} \rightarrow [\frac{c1}{R_s} + \frac{c2}{R_s} + \frac{c3}{R_s} + ...\frac{c36}{R_s}]$

6. Calculate the Histogram of Oriented Gradients feature vector

The 36 x 1 vector are concatenated into one giant vector to calculate the final feature vector for the full image patch.

7. Classify the model using two-dimensional Convolutional Neural Network(2D CNN)

4 Experimental results and discussion

The proposed work has been developed on the Spyder (Anaconda3) platform, using the Intel Core i7 with 8 GB RAM and a 2.60GHz processor. The next section discusses the dataset description, followed by the experimental results, and a comparison with other existing state-of-the-art methods.

4.1 Dataset Description

VIFFD [14] dataset is used as a benchmark to evaluate the performance of the proposed work. A set of four data subsets are created based on frame insertion, deletion, duplication, and a combination of all forgery types, namely InsertionDB, DeletionDB, DuplicationDB, and MixedDB. Each video sequence is neither longer than 30 seconds nor smaller than 10 seconds. Forged frames typically have a total length between 100 and 140 inches. Figures 1 to 4 show examples of forged videos taken from each dataset.

4.2 Performance Metrics

The following performance metrics are employed to evaluate the detection performance: Accuracy, Precision, Recall, and F1-Score.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10}$$

Precision+Recall Where True Positives (TP) are the number of forged videos having classified as forged, the True Negatives (TN) are the number of genuine videos having classified as genuine, the False Positives (FP) are the number of forged videos having classified as genuine and the False Negative (FN) are the number of genuine videos having classified as forged. Performance of the proposed system used 60% as the training set and 40% as a testing set.

Performance analysis of the proposed work

The following section examines the performance of the proposed method in detecting inter-frame forgeries using accuracy, precision, recall, and the f1-score.

To evaluate the performance of the proposed work, attacks such as frame insertion, deletion, duplication, and mixed attacks are considered. Table 2 shows the discrete wavelet transform applied to the input video frame. A given image is transformed into LL, LH, HL, and HH sub-bands. According to our analysis, the DWT LL band provides better F1 scores.

As shown in Table 3, different inter-frame forgery detection results are summarized based on Histogram oriented gradients. A histogram-oriented gradient is applied to the input video frame for feature extraction. Compared to other attacks, duplication attack does not provide a good result in the analysis. Furthermore, frame insertion produces an f1-score value of 50%. Table 4 summarizes the results of inter-frame forgery attacks using the proposed work. Analysis of 5 to 20 forged frames is conducted under each attack. In this work, an input video frame is given to DWT then the DWT LL band is given to Histogram Oriented Gradients. The final step is the classification process using the 2D-CNN. The analysis shows that the accuracy of detecting forged video remains the same for different

Figure 1: Sample forged videos from InsertionDB

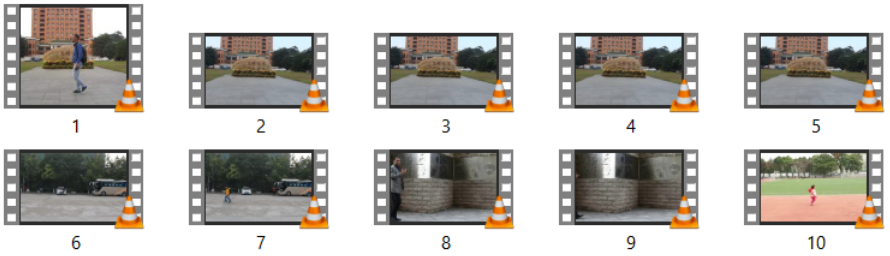

Figure 2: Sample forged videos from DeletionDB

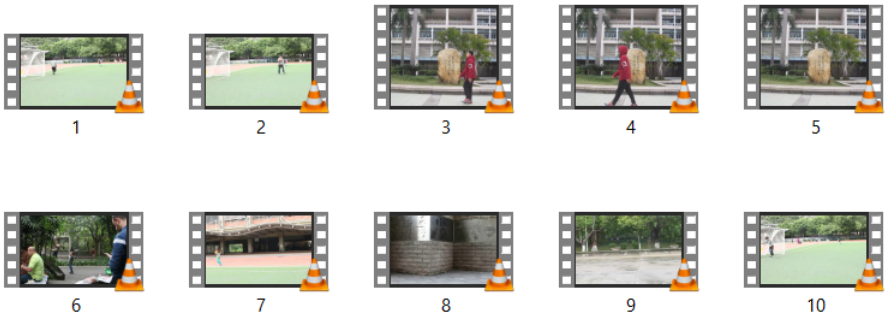

Figure 3: Sample forged videos from DuplicationDB

Figure 4: Sample forged videos from MixedDB

Table 2: Performance analysis of DWT band result based on different inter-frame attacks

Type of attack	*Feature Extraction Method*	*Accuracy*	*Precision*	*Recall*	*F1_Score*
Insertion	**DWT- LL**	**0.98**	**0.62**	**0.37**	**0.46**
	DWT-LH	0.97	1.00	0.03	0.06
	DWT-HH	0.97	0.52	0.26	0.35
	DWT-HL	0.97	0.58	0.29	0.39
Deletion	**DWT-LL**	**0.98**	**0.89**	**0.69**	**0.78**
	DWT-LH	0.98	0.92	0.31	0.46
	DWT-HH	0.97	0.50	0.09	0.15
	DWT-HL	0.97	0.91	0.27	0.42
Deletion	**DWT-LL**	**0.99**	**1.00**	**0.76**	**0.86**
	DWT-LH	0.98	1.00	0.46	0.63
	DWT-HH	0.97	1.00	0.46	0.63
	DWT-HL	0.97	1.00	0.46	0.63
Mixed	**DWT-LL**	**0.98**	**0.63**	**0.47**	**0.54**
	DWT-LH	0.96	0.49	0.276	0.35
	DWT-HH	0.96	0.49	0.27	0.35
	DWT-HL	0.97	0.99	0.32	0.48

Table 3: Histogram Oriented Analysis result on different inter-frame attacks

Type attack	*Accuracy*	*Precision*	*Recall*	*F1_Score*
Frame Insertion	0.98	0.50	0.50	0.50
Frame Deletion	0.98	0.99	0.80	0.88
Frame Duplication	0.96	0.30	0.23	0.26
Mixed (combination of all forgery)	0.97	0.86	0.51	0.64

types of attacks, such as frame insertion, frame deletion, frame duplication, and a combination thereof, since true positive and true negative affect the accuracy measure significantly. The precision and recall values are very low for less number of forged frames, say 5. This is due to the fact that false positives and false negatives increase with the number of forged frames. The proposed work guarantees only 50% forgery detection if the number of forged frames is less than five. Since F1-score is the weighted average of precision and recall, the F1-score measure is used in the rest of the research work in evaluating the performance of the proposed work. It shall be acted that F1-score is above 80% for all types of attacks with more than 20 forged frames. For forged frames between 15 to 20, the proposed work gives a 75% forgery detection score whereas, the proposed work finds it difficult to detect forgery detection if the number of forged frames is less than 5 in a video for the insertion, deletion, and duplication attacks. But, the proposed work performs better for mixed attacks which are a combination of insertion, deletion, and duplication. It gives an 80% forgery detection score for forged frames number and it is also capable of detecting forged frames if the number of forged frames is less than five.

Performance comparisons of the proposed works and existing techniques are shown in Table 5. In comparison with other algorithms such as HoG, LBP, DCT, and the proposed algorithm provides high accuracy and F1-Score. In table 5, InsertionDS shows that the proposed method gives a high precision rate, and the F1-score is 0.84, but [32] gives a better recall value. In addition, the InsertionDS videos show that [11] does not perform well. In DeletionDS, the proposed technique yields the best precision and recall values, so the F1-score is 0.91. In DuplicationDS, the proposed method gives better precision and recall values compared to other methods, with an F1-Score of 0.85. With MixedDS, all forgeries are combined, such as frame insertion, frame deletion, and frame duplication. The proposed method yields good precision and recall values. Compared to our manipulated videos, [32] did not provide significant results. It can be observed from Tables 4 and 5 that the proposed work performs better than the existing works for forgery detection. Results indicate that the proposed method is extremely effective for detecting interframe forgeries in surveillance videos.

5 Conclusion and Future work

To detect and localize forgeries, this paper proposes a technique that addresses frame insertion, deletion, and duplication forgeries. The proposed work for forgery

Table 4: Performance analysis of different attacks with the proposed work

$Type\ of\ attack$	$Attacks\ per\ Frame$	$PerformanceMeasure$			
		$Accuracy$	$Precision$	$Recall$	$F1_Score$
Insertion	20 Frames	0.99	0.91	0.78	0.84
	15 Frames	0.98	0.79	0.53	0.63
	5 Frames	0.95	0.68	0.45	0.54
Deletion	20 Frames	0.99	1.00	0.83	0.91
	15 Frames	0.98	1.00	0.72	0.83
	5 Frames	0.94	1.00	0.42	0.59
Duplication	20 Frames	0.99	1.00	0.75	0.85
	15 Frames	0.99	0.70	0.50	0.58
	5 Frames	0.94	0.	0.42	0.59
Duplication	20 Frames	0.97	10.86	0.75	0.80
	15 Frames	0.95	0.76	0.66	0.70
	5 Frames	0.94	0.63	0.55	0.58

Table 5: Performance comparison of the proposed work for forgery detection with existing works

$Dataset$	$Forgery\ detection\ methods$	$Accuracy$	$Precision$	$Recall$	$F1\_Score$
Insertion DS	Zhan et.al.,[32]	0.98	0.64	0.86	0.67
	Fayaaz et. al.,[11]	0.98	0.57	0.50	0.53
	Fadl et. al.,[9]	0.98	0.58	0.70	0.64
	Ours	**0.99**	**0.91**	**0.78**	**0.84**
Deletion DS	Zhan et. al.,[32]	0.99	1.00	0.50	0.67
	Fayaaz et. al.,[11]	0.99	0.81	0.90	0.86
	Fadl et. al.,[9]	0.98	0.80	0.57	0.67
	Ours	**0.99**	**1.00**	**0.83**	**0.91**
Duplication DS Zhan et. al.,[32]	0.99	1.00	0.57	0.73	
	Fayaaz et. al.,[11]	0.97	0.17	0.14	0.15
	Fadl et. al.,[9]	0.99	0.89	0.75	0.81
	Ours	**0.99**	**1.00**	**0.75**	**0.85**
Mixed DS	Zhan et. al.,[32]	0.95	0.54	0.48	0.51
	Fayaaz et. al.,[11]	0.95	0.56	0.49	0.52
	Fadl et. al.,[9]	0.97	0.93	0.91	0.92
	Ours	**0.97**	**0.86**	**0.75**	**0.80**

detection involves analyzing a digital video using Discrete Wavelet Transform and extracting the LL band information. The transformed coefficients are subject to the Histogram Oriented Gradient (HOG) method for feature extraction. The VIFFD dataset was utilized in order to conduct an analysis of how well the proposed technique performs. The DWT LL band provides the highest level of accuracy across all attack types. The HoG method took into account an image as a whole, as well as the application of gradient and magnitude values, to produce the best possible outcome. The suggested approach is applied in each attack, with consideration given to 5, 15, and 20 frames, respectively. According to the findings of the research, forgery with 5 and 15 frames does not produce a superior accuracy compared to other methods. Insertion attack identified with an F1-Score of 0.84, deletion attack detected with an F1-Score of 0.91, duplication attack detected with an F1-Score of 0.85, and combination of all forgery detected with an F1-Score of 0.80. A comprehensive performance evaluation has been done, which demonstrates that the system is working effectively.

Nowadays, deep learning is advancing exponentially in which CNN is used for classification purposes. Furthermore, CNN produces better results for training and testing data. Further research will concentrate on inter-frame video forgery including frame shuffling. By way of definition, frame shuffling means deleting frames after they have been inserted. For that reason, future work will address the above problem and rectify it with a new approach.

References

[1] Abbasi Aghamaleki, J. and Behrad, A. Malicious inter-frame video tampering detection in mpeg videos using time and spatial domain analysis of

quantization effects. *Multimedia Tools and Applications*, 76:20691–20717, 2017.

[2] Affonso, E. T., Nunes, R. D., Rosa, R. L., Pivaro, G. F., and Rodriguez, D. Z. Speech quality assessment in wireless voip communication using deep belief network. *IEEE Access*, 6:77022–77032, 2018.

[3] Affonso, E. T., Rosa, R. L., and Rodriguez, D. Z. Speech quality assessment over lossy transmission channels using deep belief networks. *IEEE Signal Processing Letters*, 25(1):70–74, 2017.

[4] Bakas, J., Naskar, R., and Bakshi, S. Detection and localization of inter-frame forgeries in videos based on macroblock variation and motion vector analysis. *Computers & Electrical Engineering*, 89:106929, 2021.

[5] Chao, J., Jiang, X., and Sun, T. A novel video inter-frame forgery model detection scheme based on optical flow consistency. In *The International Workshop on Digital Forensics and Watermarking 2012: 11th International Workshop, IWDW 2012, Shanghai, China, October 31–November 3, 2012, Revised Selected Papers*, pages 267–281. Springer, 2013.

[6] Dantas Nunes, R., Lopes Rosa, R., and Zegarra Rodríguez, D. Performance improvement of a non-intrusive voice quality metric in lossy networks. *IET Communications*, 13(20):3401–3408, 2019.

[7] de Almeida, F. L., Rosa, R. L., and Rodriguez, D. Z. Voice quality assessment in communication services using deep learning. In *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, pages 1–6. IEEE, 2018.

[8] Fadl, S., Han, Q., and Li, Q. Cnn spatiotemporal features and fusion for surveillance video forgery detection. *Signal Processing: Image Communication*, 90:116066, 2021.

[9] Fadl, S., Han, Q., and Qiong, L. Exposing video inter-frame forgery via histogram of oriented gradients and motion energy image. *Multidimensional Systems and Signal Processing*, 31:1365–1384, 2020.

[10] Fadl, S. M., Han, Q., and Li, Q. Inter-frame forgery detection based on differential energy of residue. *IET Image Processing*, 13(3):522–528, 2019.

[11] Fayyaz, M. A., Anjum, A., Ziauddin, S., Khan, A., and Sarfaraz, A. An improved surveillance video forgery detection technique using sensor pattern noise and correlation of noise residues. *Multimedia Tools and Applications*, 79:5767–5788, 2020.

[12] Li, Z., Zhang, Z., Guo, S., and Wang, J. Video inter-frame forgery identification based on the consistency of quotient of mssim. *Security and Communication Networks*, 9(17):4548–4556, 2016.

[13] Liu, Y. and Huang, T. Exposing video inter-frame forgery by zernike opponent chromaticity moments and coarseness analysis. *Multimedia Systems*, 23:223–238, 2017.

[14] Nguyen, X. and Hu, Y. Viffdâa dataset for detecting video inter-frame forgeries. *Mendeley Data*, 5, 2020.

[15] Parmani, R., Butala, S., Khanvilkar, A., Pawar, S., and Pulgam, N. Inter frame video forgery detection using normalized multi scale one level subtraction. In *Proceedings of International Conference on Communication and Information Processing (ICCIP)*, 2019.

[16] Ren, H., Atwa, W., Zhang, H., Muhammad, S., and Emam, M. Frame duplication forgery detection and localization algorithm based on the improved levenshtein distance. *Scientific Programming*, 2021:1–10, 2021.

[17] Rodriguez, D. Z. and Bressan, G. Video quality assessments on digital tv and video streaming services using objective metrics. *IEEE Latin America Transactions*, 10(1):1184–1189, 2012.

[18] Rodriguez, D. Z. and Junior, L. C. B. Determining a non-intrusive voice quality model using machine learning and signal analysis in time. *INFOCOMP Journal of Computer Science*, 18(2), 2019.

[19] Rodríguez, D. Z., Rosa, R. L., Almeida, F. L., Mittag, G., and Möller, S. Speech quality assessment in wireless communications with mimo systems using a parametric model. *IEEE Access*, 7:35719–35730, 2019.

[20] Selvaraj, P. and Karuppiah, M. Inter-frame forgery detection and localisation in videos using earth mover's distance metric. *IET Image Processing*, 14(16):4168–4177, 2020.

[21] Singh, R. D. and Aggarwal, N. Optical flow and prediction residual based hybrid forensic system for inter-frame tampering detection. *Journal of Circuits, Systems and Computers*, 26(07):1750107, 2017.

[22] Sitara, K. and Mehtre, B. A comprehensive approach for exposing inter-frame video forgeries. In *2017 IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, pages 73–78. IEEE, 2017.

[23] Sitara, K. and Mehtre, B. M. Digital video tampering detection: An overview of passive techniques. *Digital Investigation*, 18:8–22, 2016.

[24] Su, Y., Nie, W., and Zhang, C. A frame tampering detection algorithm for mpeg videos. In *2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference*, volume 2, pages 461–464. IEEE, 2011.

[25] Ulutas, G., Ustubioglu, B., Ulutas, M., and Nabiyev, V. Frame duplication/mirroring detection method with binary features. *IET Image Processing*, 11(5):333–342, 2017.

[26] Wang, Q., Li, Z., Zhang, Z., and Ma, Q. Video inter-frame forgery identification based on consistency of correlation coefficients of gray values. *Journal of Computer and Communications*, 2(04):51, 2014.

[27] Wang, W. and Farid, H. Exposing digital forgeries in video by detecting duplication. In *Proceedings of the 9th workshop on Multimedia & security*, pages 35–42, 2007.

[28] Wang, W., Jiang, X., Wang, S., Wan, M., and Sun, T. Identifying video forgery process using optical flow. In *Digital-Forensics and Watermarking: 12th International Workshop, IWDW 2013, Auckland, New Zealand, October 1-4, 2013. Revised Selected Papers 12*, pages 244–257. Springer, 2014.

[29] Wei, W., Fan, X., Song, H., and Wang, H. Video tamper detection based on multi-scale mutual information. *Multimedia Tools and Applications*, 78:27109–27126, 2019.

[30] Yao, H., Ni, R., and Zhao, Y. An approach to detect video frame deletion under anti-forensics. *Journal of Real-Time Image Processing*, 16:751–764, 2019.

[31] Yu, L., Wang, H., Han, Q., Niu, X., Yiu, S.-M., Fang, J., and Wang, Z. Exposing frame deletion by detecting abrupt changes in video streams. *Neurocomputing*, 205:84–91, 2016.

[32] Zhang, Z., Hou, J., Ma, Q., and Li, Z. Efficient video frame insertion and deletion detection based on inconsistency of correlations between local binary pattern coded frames. *Security and Communication networks*, 8(2):311–320, 2015.

[33] Zhao, D.-N., Wang, R.-K., and Lu, Z.-M. Interframe passive-blind forgery detection for video shot based on similarity analysis. *Multimedia Tools and Applications*, 77:25389–25408, 2018.