ISSN 2063-5346

AN ARTIFICIAL NEURAL NETWORK-BASED APPROACH FOR DETECTING FRAME DELETION IN VIDEO PROCESSING

J.Liba Manopriya[1], P.ArockiaJansi Rani[2]

Article History: Received: 01.02.2023 Revised: 07.03.2023 Accepted: 10.04.2023

Abstract

In the age of digital multimedia, it is increasingly important to ensure the integrity and authenticity of vast amounts of video data. As technology advances, it becomes easier for individuals with malicious intent to manipulate video footage for fraudulent purposes. One common and relatively simple technique for video forgery is frame deletion, which involves removing frames from a video sequence to alter the content. This paper proposes a novel approach to detect video forgery using a combination of Discrete Cosine Transform (DCT) and Artificial Neural Network (ANN). The proposed method measures the changes in difference between adjacent frames using feature descriptors to identify the breakpoint of frame deletion in video. Experimental results demonstrate that the proposed approach achieves an accuracy of 98% in detecting frame deletion for video forgery.

Keywords-Digital video forensic, Frame deletion, Video Forgery

[1] Research Scholar, Department of Computer Science & Engineering, ManonmaniamSundaranar University,Tirunelveli, Tamil Nadu, India-627012. *Email:libamanopriya06@gmail.com*

[2] Associate Professor, Department of Computer Science & Engineering, ManonmaniamSundaranar University,Tirunelveli, Tamil Nadu, India-627012. Email: jansi_cse@msuniv.ac.in*@gmail.com*

DOI:10.31838/ecb/2023.12.s1-B.375

Introduction

Advancements in technology, particularly smartphones, and the widespread availability of internet access have greatly facilitated access to social media sites. These sites offer a plethora of digital images and videos that can significantly influence a user's perception of events or individuals [1]. Unfortunately, these platforms also enable the spread of false information through manipulated images or videos, misleading the public. To combat such activities, CCTV cameras are commonly installed for security purposes in various locations such as offices, homes, malls, and shops, with their footage frequently serving as crucial evidence in police investigations. However, it is crucial to verify the accuracy of the conveyed information, as criminals may manipulate these videos to conceal their actions, and the sophisticated image and video editing tools readily available make such manipulation challenging to detect. Video forgery, or video tampering, can be achieved through three techniques: temporal, spatial, and spatiotemporal [2]. Active and passive forgery detection is two approaches to identify manipulated videos. Active forgery detection involves embedding a watermark or signature to authenticate, but it necessitates specialized hardware, which may affect video quality. In contrast, passive forgery detection relies on statistical properties to identify forgeries without the original video [3].

This paper proposes a new passive detection technique for identifying interframe forgeries in surveillance videos. The literature review in Section 2 summarizes current video forgery detection methods, identifying research gaps. Section 3 presents a technical analysis of the proposed system to locate and identify frame duplication forgery. The effectiveness of the system is evaluated through experiments discussed in Section 4, and the paper concludes with Section 5.

Literature Review

Over the past decade, extensive research has been conducted in the field of digital forensics. Currently, passive detection techniques are used to detect inter-frame forgeries. The goal of these investigations is to identify attacks involving frame duplication. Fadl et al. [4] attempted to enhance detection by integrating spatiotemporal averages in each frame and extracting relevant features using 2D Convolutional Neural Network (CNN). They then employed multiclass support vector machine (SVM) and Gaussian Radial Basis Functions (RBF) to categorize the video. Meanwhile, Zhao et al. [5] utilized the speeded-up Robust Features (SURF) to extract important information and Fast Library Approximate Nearest Neighbors (FLANN) based on Hue-Saturation-Value (HSV) to determine similarity. However, this method may not work if the shot is improperly obtained. To address this issue, Liu et al. [6] proposed a method in which each frame was transformed into a 2D opponent chromaticity space and abnormal points between frames were detected using Zernike Opponent Chromaticity Moments (ZOCM). They further improved the accuracy and reduced the false-positive rate by utilizing the coarseness feature of the Tamura technique. Nevertheless, this approach is more time-consuming despite its enhanced performance. In their research, Raahat et al. [7] attempted to use optical flow and residual prediction techniques but found them inadequate due to multiple compression stages. In contrast, Ulutas et al. [8] adopted a different approach that involved computing the peak signal-to-noise ratio and distance between frames and extracting binary features to evaluate their similarity. Although this method produced satisfactory results, it was computationally intensive and required a significant amount of time to process each frame. Li et al. [9] utilized the Mean Structural Similarity Measure (MSSM) technique, which involves

calculating the ratio between the MSSM values of consecutive frames. Meanwhile, Zhang et al. [10] detected inconsistencies between frames using the quotients of correlation coefficients between local binary patterns (QoCCLBP) and determined the abnormality by applying the Tchebyshev inequality. Wang et al.[11] used an optical flow approach to identify abnormalities in videos, where the Gaussian distribution values and the Grabb's test were used to verify the video's authenticity. In contrast, Chao et al.[12] utilized a similar optical flow technique to extract features from each frame, but the computational complexity was high, resulting in a time-consuming method.

Singla et al. [13] focused on detecting inter-frame forgery in videos encoded with the High-Efficiency Video Coding (HEVC) standard. The authors used coding features such as intra-prediction, skip prediction, and transform energy to identify outliers in the video frames. These outliers were then further analysed using a spectral similarity metric to compare the similarity between adjacent frames. In order to confirm the presence of forgery and eliminate false positives detected by the feature detection methods, the authors employed additional techniques. Meanwhile, Kumar et al. [14] used the minimum distance score and dual-threshold method to detect forgery videos. Initially, they calculated the minimum distance score using correlation coefficient values. Then, they established lower and upper adaptive threshold values through statistical measures such as the mean, standard deviation, and sigma coefficients. These methods were used to identify and differentiate between authentic and manipulated videos. Shelke et al. [15] proposed a passive algorithm that detects video forgeries by analysing the correlation consistency of entropy-coded frames. The algorithm utilises entropy-based texture features, including two-dimensional distribution entropy (DistrEn2D) and bi-dimensional multiscale entropy (MSE2D), to detect forgeries in videos. Ren et al. [16] developed a technique that detects duplicate frames based on the improved Levenshtein distance, but it cannot distinguish duplicate frames with dynamic backgrounds. On the other hand, priyadharshini et al. [17] utilised Earth Mover's Distance (EMD) to detect forgeries and anomalous points. This was accomplished by dividing the input video into a series of frames and representing each frame with the distribution of pixel intensities in the CIE-Lab colour space and the motion flow between frames. The EMD flow was then analysed to detect any sudden changes. Raskar and Shah [18] attempted to detect suspicious frames using the VFDHSOG (histogram of second-order gradients) technique. They utilised Contrast Limited Adaptive Histogram Equalization (CLAHE) correlation coefficient values to calculate the HSOG value by determining the distance and similarity threshold. However, the success of this method relies heavily on determining the threshold correctly. Oraibi et al. [19] used spatial and temporal information to identify features in videos. Specifically, they employed a three-dimensional Convolutional neural network (3D-CNN) to calculate the difference between each adjacent frame. To detect forgeries, it is necessary to analyse the temporal feature with Long Short Term Memory, and the probability can be calculated based on the result. Shelke et al. [20] utilised two feature extraction techniques, namely Polar Cosine Transform (PCT) and Neighborhood Binary Angular Pattern (NBAP), and employed the GoogleNet architecture to identify inter-frame and intra-frame forgeries.

MaterialsandMethods

A general approach to video forgery detection involves pre-processing the input frame and then converting it into an Artificial Neural Network (ANN) that uses cosine transform for classification. Then ANN subsequently determines the authenticity of the input video frame. A more detailed explanation of the methodology is provided in the following section.

Preprocessing

To preprocess a color video into gray-level frames, it is necessary to transform each color frame of the video into a grayscale representation. This involves merging the red, green, and blue color channels of the image to form a single channel that encodes the brightness or luminance of the image. Additional preprocessing techniques can then be applied to the grayscale frames to improve their quality or extract desired features from the image data. Once these preprocessing steps have been performed, the resulting grayscale frames can be utilized as input data for further analysis.

Feature Extraction

The Discrete Cosine Transform (DCT) is a widely used mathematical technique in video processing. The DCT coefficients of a video frame are sensitive to changes in the frame content, and deleting a frame can cause a noticeable change in the DCT coefficients of the surrounding frames. By comparing the DCT coefficients of adjacent frames, it is possible to detect a sudden change or discontinuity that may indicate a deleted frame. Here are the detailed steps involved in performing a DCT:

i. The first step in DCT is to select the input data. This data can be two-dimensional signal that is represented as a sequence of samples, such as an image.
ii. Divide the input data into non-overlapping blocks of the desired size. The block size is typically chosen to be a power of two, such as 8x8, 16x16, or 32x32.
iii. Multiply each block of data by a window function, such as a Hamming window, to reduce the spectral leakage and improve the accuracy of the DCT coefficients.
iv. Apply the DCT algorithm to each block of data to compute the corresponding DCT coefficients. The DCT algorithm transforms the input data from the time domain to the frequency domain by expressing it as a linear combination of cosine functions of increasing frequency.
v. Quantize the DCT coefficients by dividing them by a quantization factor and rounding them to the nearest integer. The quantization factor controls the amount of compression and the quality of the reconstructed signal.
vi. Encode the quantized coefficients using an entropy coding scheme, such as Huffman coding, to reduce the number of bits required to represent the data. The encoded data can be transmitted or stored for later use.
vii. To reconstruct the original data, reverse the process by applying the inverse DCT algorithm to the encoded and quantized data. The inverse DCT algorithm expresses the frequency domain coefficients as a linear combination of the original data points in the time domain.

Classification

In the context of frame deletion in videos, an Artificial Neural Network (ANN) can be trained to classify video frames as either deleted or not deleted. The ANN can be trained using a set of input features, such as DCT coefficients or other relevant video characteristics. The ANN consists of a series of interconnected processing nodes, or neurons, organized in layers. Each neuron receives inputs from other neurons or from the input features, and applies a non-linear transformation to the inputs to produce an output. The outputs of the neurons in one layer are then used as inputs to the neurons in the next layer, until the final output layer produces a classification decision. During training, the ANN is presented with a set of labeled examples, each consisting of a set of input features and the corresponding desired output (i.e., deleted or not deleted). The ANN adjusts the strength of the connections between its neurons, or

weights, in order to minimize the error between its predictions and the desired outputs. Once the ANN is trained, it can be used for inference, where it receives the DCT coefficients of a video frame as input and produces a classification decision as output (i.e., deleted or not deleted).

Experimental results and discussion

The system used for the experimentation has an Intel Core i7 processor,, 8GB of RAM, and a clock speed of 2.60GHz, and the Spyder(Anaconda3) platform was employed. Details about the dataset are provided in the following section, and the experimental results, as well as a comparison to other state-of-the-art techniques, are presented.

Dataset Description

The proposed work's performance is evaluated using the VIFFD[21] dataset as a benchmark. A collection of videos was produced by deleting frames, with the number of manipulated frames being 20, 15 and 5. Each video sequence has a length between 10 and 30 seconds, and manipulated frames typically span a total length of 100 to 14o frames.

Performance Measure

The following performance metrics are used to assess the detection performance: Accuracy, precision, recall, Root-Mean-Square Error (RMSE) and Matthews Correlation Coefficient (MCC).

Accuracy:

It is the ratio of correctly classified instances to the total number of instances in the dataset.

$$Accuracy = \frac{Number\ of\ correctly\ classified\ instances}{Total\ number\ of\ instances}$$

Precision:

It measures the proportion of true positive predictions out of all positive prediction made by a classifier.

$$Precision = \frac{True\ Positive}{(True\ Positive + False\ Positive)}$$

Recall:

It is a measure of how well the method avoids false negatives and correctly identifies negative results.

$$Recall = \frac{True\ Positive}{(True\ positive + False\ Negative\ )}$$

Root-Mean-Squared Error (RMSE):

It is a measure of the deviation between the predicted and actual values. It measures the accuracy of a model's predictions by comparing the predicted values with the actual values.

$$RMSE = \sqrt{\frac{Sum(Predicted\ values - Actual\ values)^2}{n}}$$

Matthews Correlation Coefficient (MCC):

It is a measure of the quality of binary classification models. It produces a value between -1 and +1, where 1 represents a perfect prediction, 0 represents a random prediction, and -1 represents a perfectly inverse prediction.

$$MCC = \frac{(TP \times TN - FP \times FN)}{\sqrt{((TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN))}}$$

Where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives and FN is the number of false negatives.

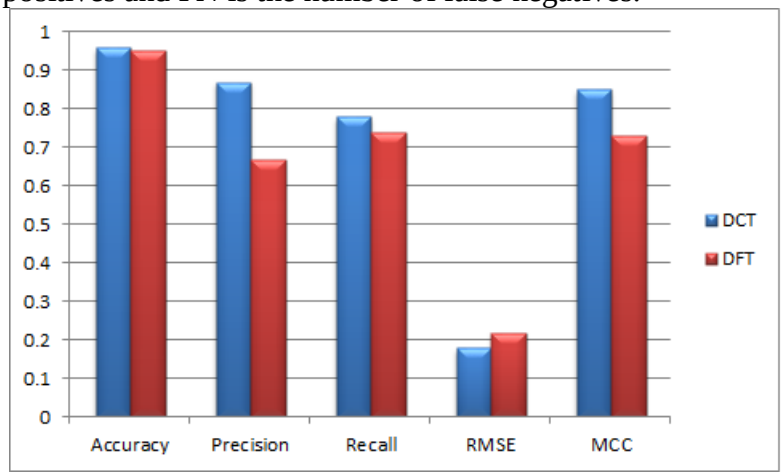

Figure 1: Performance analysis of various transform domains based on frame deletion

Figure 1 presents an analysis of various transform domains based on frame deletion. The results show that the accuracy of DCT is 0.96, indicating a high level of accuracy in classifying data, while the accuracy of DFT is slightly lower at 0.95. In terms of precision, DCT outperforms DFT with a precision score of 0.87 compared to DFT's score of 0.67. On the other hand, DFT has a higher recall score of 0.74 compared to DCT's score of 0.78. The RMSE of DCT is also lower at 0.18 than DFT's score of 0.22, indicating a smaller deviation between the predicted and actual values. Additionally, DCT has a higher PSNR score of 0.15 compared to DFT's score of 0.13, indicating a better quality of the reconstructed signal. Overall, the table suggests that DCT has a higher level of accuracy and precision, lower RMSE, and higher PSNR than DFT, while DFT performs better in terms ofrecall.

Performance analysis of various block size in frame deletion

DCT (Discrete Cosine Transform) is a family of mathematical transforms that are widely used in signal processing, data compression, and image and video processing. There are several types of DCT, denoted as DCT-I, DCT-II, DCT-III, and DCT-IV that differ in their mathematical formulations and properties. DCT-I is also known as the Fejér-Konyves transform. It is an odd function and is used for type 1 symmetric boundary conditions. The DCT-I has real transform coefficients and the output sequence is purely real. DCT-II is the most commonly used type of DCT. It is also known as the "standard" DCT and is used for type 2 even boundary conditions. The DCT-II is even-symmetric, which means that its transformcoefficients are symmetric around the center of the input signal. The output sequence is real and is represented using N/2 complex values, where N is the length of the input sequence. DCT-III is used for type 3 odd boundary conditions. The DCT-III is an odd-symmetric transform, which means that its transform coefficients are antisymmetric around the center of the input signal. The output sequence is also real and is represented using N/2 complex values. DCT-IV is used for type 4 boundary conditions. The DCT-IV is neither odd nor even symmetric, which means that it has both odd and even transforms coefficients. The output

sequence is real and is represented using N/2 complex values.

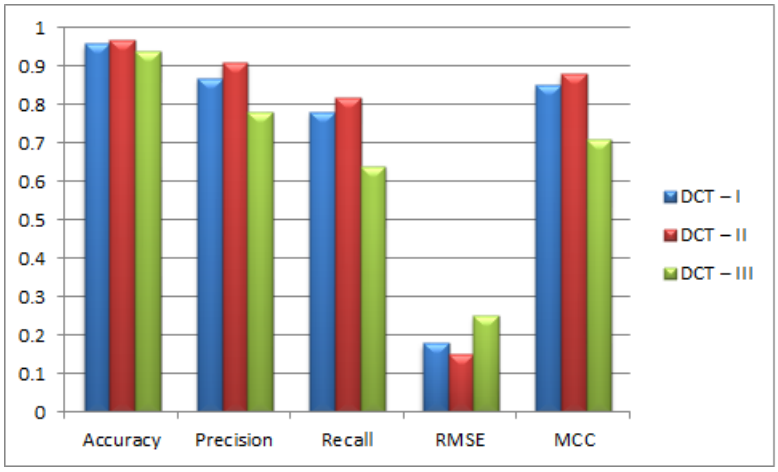

Figure 2:Performance analysis of various coefficient types

According to the Figure 2, DCT II outperforms the other two methods in terms of accuracy, precision, recall and MCC, and also has the lowest RMSE value, indicating the best overall performance. DCT III has the lowest accuracy, precision, and recall, and the highest RMSE value, indicating the poorest performance among the three methods.

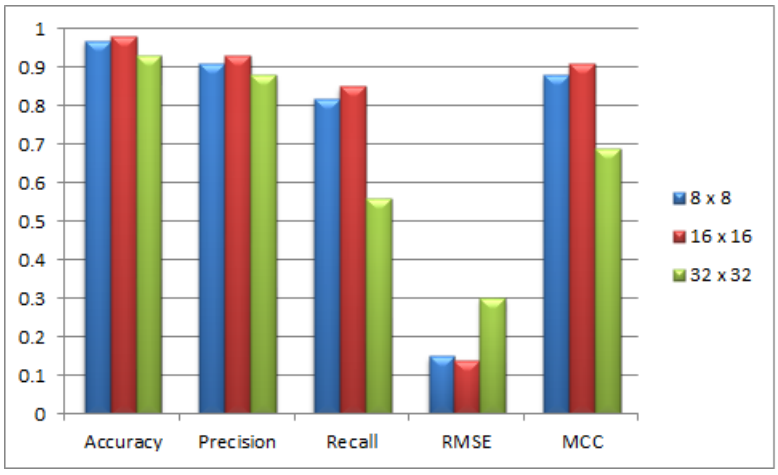

Figure 3: Performance analysis of various quantization factors

Figure 3 provides evaluation metrics for three different quantization factors, where each factor is represented by the size of the extracted features in pixels. Based on the observation, 16 × 16 quantization factor has the highest accuracy and MCC among the three methods, while the 8 × 8 has the lowest RMSE. The 16 × 16 method also has the highest precision and recall indicating that it performs well in correctly identifying positive samples. However, the 32 × 32 method has the lowest precision and recall indicating that it is not as effective in correctly identifying positive samples compared to the other methods. Overall, it suggests that the 16 × 16 method performs the best among the three methods based on the evaluated metrics.

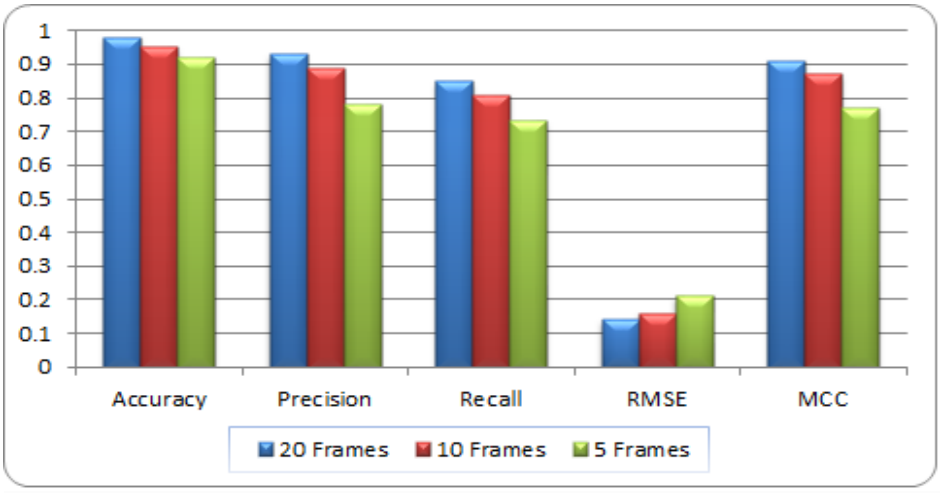

Figure 4:Performance analysis of frame deletion with the proposed method

Figure 4 presents evaluation metrics for three different feature extraction methods, where each method represents the number of video frames used for video forgery detection. Based on the observation, 20 frames have the highest accuracy, precision, recall and MCC among the three methods. It also has the lowest RMSE indicating that it performs well in correctly predicting the video class. The 10 frames method has lower evaluation metrics compared to the 20 frames but still performs reasonably well with an accuracy and MCC. The 5 frames has the lowest performance, suggests that it may not capture enough information for effective video classification.

Conclusion

This paper presents a new algorithm for detecting frame deletion forgery in MPEG video sequences. The proposed technique involves analysing the digital video using Discrete Cosine Transform and extracting the relevant information from the transformed values, which are then subjected to an Artificial Neural Network (ANN) for classification. The VIFFD dataset was used to evaluate the effectiveness of the proposed method against an attack, considering 5, 15 and 20 frames respectively. The results showed that forgery with 5 and 15 frames did not produce higher accuracy than other methods, with an F1-Score of 0.98. The comprehensive performance evaluation demonstrated that the proposed system works effectively. In future work, the proposed solution could be modified and expanded to identify the exact location of deleted frames rather than just detecting their presence**.**

References

[1]Sitara K and Mehtre B. Digital video tampering detection: an overview of passive techniques. *Digital Investigation*,18(Supplement C):8–22,2016.

[2] Vinay Kumar, Abhishek Singh, Vineetkansal and Manish Gaur, "A Comprehensive Analysis On Video Forgery Detection Techniques", *International Conference on Innovative Computing & Communications*, pp. 1-4, 2020.

[3] Bourouis S, Alroobaea R, Alharbi AM, Andejany M and Rubaiee S, "Recent Advances In Digital Multimedia Tampering Detection For Forensics Analysis", *Symmetry, Vol.*12(11), pp. 1811, 2020.

[4] Fadl, Sondos., Han, Qi and Li, Qiong. CNN spatiotemporal features and fusion for surveillance video forgery detection. *SignalProcessing Image Communication*, 90, 2021.

[5] Zhao, D., Wang, N., R. K., and Lu, Z. M. Inter-frame passive-blind forgery detection for video shot based on similarity analysis. *Multimedia Tools and Applications*, 2018.

[6] Liu, Y and Huang, T. Exposing video inter-frame forgery by Zernike opponent chromaticity moments and coarseness analysis. *Multimedia System*, 23(2):223–238, 2017.

[7] RaahatDevender Singh and Naveen Aggarwal, Optical Flow and Prediction Residual Based Hybrid Forensic System for Inter-Frame Tampering Detection,*Journal of Circuits, Systems and Computers*,Vol. 26, Nov.7,pp. 1750107(1-37),2017.

[8] Ulutas, Guzin.,BesteUstubioglu., Mustafa Ulutas and VasifNabiyev. Frame duplication/mirroring detection method with binary features, *IET Image Processing*,11(5),333-342,2017.

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

[10] Zhang Z., Hou J and Ma Q. 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.

[11] Wang W., Xinghao Jiang., Shilin Wang., Meng Wan and Tanfeng Sun. Identifying Video Forgery Process Using Optical Flow.*12th International Workshop on Digital-Forensics and Watermarking*, Vol 8389, pp. 244-257,2013.

[12] Chao J., Jiang X., Sun T. A novel video inter-frame forgery model detection scheme based on optical flow consistency. *In: Proceedings of the 11th international conference on digital forensics and watermaking*, pp 267–281,2013.

[13]Singla N, Nagpal S. & Singh J, "A two-stage forgery detection and localisation framework based on feature classification and similarity metric", Multimedia Systems, 2023.

[14] Kumar V and Gaur M, "Multiple forgery detection in video using inter-frame correlation distance with dual-threshold", *Multimedia Tools and Applications*, 2022.

[15] Shelke NA, Kasana S.S, "Multiple forgeries identification in digital video based on correlation consistency between entropy coded frames", Multimedia Systems, Vol.**28**, pp.267–280,2022.

[16] HongeRen.,WalidAtwa., Haosu Zhang., Shafiq Muhammad and Mahmoud Emam. Frame Duplication Detection and Localization algorithm Based on the Improved Levenshtein Distance. *Scientific Programming*,5595850(10), 2021.

[17] PriyadharsiniSelvaraj., Karuppiah and Muneeswaran. Inter-frame forgery detection and localisation in videos using earth mover's distance metric. *IET Image Processing*,14(16),pp.4168-4177,2020

[18] PunamRaskar and Sanjeevani Shah, "VFDHSOG: Copy-Move Video Forgery Detection Using Histogram Of Second-Order Gradients", Wireless Personal Communications, 2022.

[19] Oraibi, M. R, Radhi, A. M," Enhancement Digital Forensic Approach for Inter-Frame Video Forgery Detection Using a Deep Learning Technique", Iraqi Journal of Science, Vol.63(6), pp. 2686–2701,2022.

[20] Shelke, N.A and Kasana, S.S, Multiple forgery detection and localization technique for digital video using PCT and NBAP. Multimedia Tools and Applications (2021).

[21] nguyen, xuanhau and Hu, Yongjian. VIFFD - A dataset for detecting video inter-frame forgeries. Mendeley Data, V6,doi: 10.17632/r3ss3v53sj.6,2020.