SEPTEMBER 2021 M.Sc. in Electrical and Electronics Engineering MUHARREM BALCI

REPUBLIC OF TURKEY

GAZİANTEP UNIVERSITY

GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES

FORGERY DETECTION ON VIDEO

M.Sc. THESIS

IN

ELECTRICAL AND ELECTRONICS ENGINEERING

BY

MUHARREM BALCI

SEPTEMBER 2021

FORGERY DETECTION ON VIDEO

M.Sc. Thesis

in

Electrical and Electronics Engineering

Gaziantep University

Supervisor

Asst. Prof. Dr. Serkan ÖZBAY

by

Muharrem BALCI

September 2021

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Muharrem BALCI

ABSTRACT

FORGERY DETECTION ON VIDEO

BALCI, Muharrem
M.Sc. in Electrical and Electronics Engineering
Supervisor: Asst. Prof. Dr. Serkan ÖZBAY
September 2021
60 Pages

The whole world experiences digitalization at a rapid pace. Recently, smart phones, internet opportunities and digital video recording devices have accelerated digitalization. In social media, video sharing has become incredibly widespread. At the same time, videos are the most reliable and vital evidence both in social media and especially in courts. Videos are trusted more than pictures and any other evidence. However, many powerful and inexpensive video authoring tools for modifying video content are easily available by everyone today. A malicious attacker can destroy evidence in videos by deleting a sequence, adding a new sequence, changing the video stream order, etc. The easiness of these types of tampering has also reduced the trust to the videos. Therefore, videos can be accepted as reliable evidence by testing the originality of the video. In spite of the increasing variety of attacks on video content, the variety of authenticity testing methods increase day by day. In this study, a method is proposed to detect attacks such as deleting a part of the video content, adding a new section or changing the video stream order. Videos are formed by the flow of frame groups, such as pictures, in the time axis. Algorithm which is developed for our method is based on the using of correlation coefficients during the transitions of video frames. The method has been implemented with Matlab and tested on 120 videos. Test results have shown that the proposed method works quickly and successfully.

Keywords: Digital Video Forensics, Inter-Frame Video Forgery, Deletion/Insertion/Shuffling Forgery, Video Forgery Detection, Correlation Coefficient.

ÖZET

VİDEO GÖRÜNTÜLERİNDE SAHTECİLİK TESPİTİ

BALCI, Muharrem
Yüksek Lisans Tezi, Elektrik ve Elektronik Mühendisliği
Danışman: Dr. Öğr. Üyesi Serkan ÖZBAY
Eylül 2021
60 sayfa

Tüm dünya büyük bir hızla dijitalleşme yaşamaktadır. Son zamanlarda akıllı telefonlara ulaşmanın kolaylaşması, internet imkanlarının artması ve dijital kamera ile diğer video kayıt cihazlarına kolayca erişilmesi dijitalleşmeyi daha da hızlandırmıştır. Sosyal medya ilgisinin de artmasıyla video paylaşımları inanılmaz dercede yaygınlaşmıştır. Aynı zamanda videolar, gerek sosyal medya paylaşımlarında, gerek mahkemelerde en güvenilir, hayati bir kanıttır. Videolara resim ve diğer tüm delillerden daha fazla güvenilmektedir. Ancak video kullanımlarının artmasıyla beraber video içeriğinin değiştirilmesine yönelik çok sayıda güçlü ve ucuz video yazma araçları bugün herkes tarafından kolayca kullanılabilmektedir. Kötü niyetli bir saldırgan, videolarda bir diziyi silerek, yeni bir dizi ekleyerek, video akış sıralamasını değiştirerek vb birçok kurcalama yaparak delilleri yok edebilir. Bu tip kurcalamaların kolaylaşması videolara olan güveni de iyice azaltmıştır. Bu yuzden videoların güvenilir bir delil kabul edilebilmesi, video orjinalliğinin test edilmesiyle mümkündür. Video içeriğine yapılan saldırı çeşitliliğine karşılık, orijinallik test yöntemlerinin çeşitleri de gün geçtikçe artmaktadır. Bu çalışmada video içerinin bir kısmının silinmesi, yeni bir bölüm eklenmesi veya video akış sıralamasının değiştirilmesi gibi yapılan saldırıların tespit edilmesi için yöntem önerilmiştir. Videolar resimler gibi çerçeve gruplarının zaman ekseninde akışıyla oluşmaktadır. Yöntemimiz için, geliştirlen algoritma video çerçevelerinin geçişleri esnasında korelasyon katsayılarının kullanımına dayanmaktdır. Yöntem Matlab ile uygulanmış ve 120 video üzerinde test edilmiştir. Test sonuçları, önerilen yöntemin hızlı ve başarılı bir şekilde çalıştığını göstermiştir.

Anahtar Kelimeler: Dijital Video Adli Tıp, Çerçeveler Arası Video Sahteciliği, Silme/Ekleme/Karıştırma Sahtecilikleri, Video Sahtecilik Tespiti, Korelasyon Katsayısı.

''Dedicated to my family''

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Asist. Prof. Dr. Serkan ÖZBAY for his guidance and support throughout the study. I am thankful for his encouragement and motivation.

I would like to express my love and gratitude to my family for their support, always best wishes. Special thanks to my wife for her support and encouragement

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS

Φ	Phi
σ	Sigma

LIST OF ABBREVIATIONS

DCT	Discrete Cosine Transform
DVF	Digital Video Forensics
DNA	Deoxyribo Nücleic Acid
SPN	Sensor Pattern Noise
FRUC	Frame Rate Up–Conversion
MVs	Motion Vectors
GOP	Group Of Pictures
RGB	Red-Green-Blue
SD	Standard Deviation
MPEG	Motion Picture Experts Group
I-Frames	Intra frames
P-Frames	Predictive Frames
B-Frames	Bi-directionally Predictive Frames
SULFA	Survey University Library for Forensic Analysis
PU	Punjab University
VIFFD-A	Video Inter Frame Forgery Detection
REWIND	Reverse Engineering of Audio Visual Content Data
MATLAB	Matrix Laboratory
2D-3D	Two Dimensional-Three Dimensional
F1	F1 score.
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
A	Accuracy
P	Precision
R	Recall
f/sec	Frame/second

PC	Personal Computer
GUI	Graphical User Interfaces
RAM	Random Access Memory
HD	Hard Disk
CPU	Central Processing Unit
GB	Gigabyte
GHz	Giga Hertz
MOV	QuickTime Movie (file extension)
AVI	Audio Video Interleave (file extension)
MP4	MPEG Layer-4 Audio (file extension)

CHAPTER 1

INTRODUCTION

1.1 General View

People tend to believe what they see as natural. Also this fact is valid to both seeing it directly and seeing with second-hand through pictorial representations. Among all the types of pictures, nothing reminds more trust, or has a stronger effect on our intuition, than videos. Looking at a video of an event is like witnessing the real event. This power of representation makes video evidence a particularly prerogative form of evidence. Therefore, videos have a great importance in our life and this is getting more and more common. In today's world where digitalization has become widespread, videos are used for entertainment purposes and also serve in many areas. The visual content of digital videos creates the basis of limitless very important decisions in the fields of journalism, politics, medical diagnosis, insurance claims, criminal trials and defense planning. According to a forensic viewpoint, videos have a high persuasive value. For example digital videos create one of the most inculpatory proof in the court of law. DNA and fingerprints are forms of forensic evidence that require lots of processes due to their structure. However, videos have the quality of an eyewitness to an event and because that is very difficult not to trust what we see with our own eyes often the testimony which is provided by a video is often indisputable. Therefore, it is necessary to be aware of the flaws of videos, in particular their sensitivity to deliberate semantic manipulations, in order to use their evidence effectively and safely.

1.2 Video Forensics Analysis

Today, the use of video data as evidence in any case is inevitable. Even the fact that video data is open to changes and manipulations cannot prevent it from being accepted as evidence. Therefore, in order to rely on the testimony of video evidence, one must first be sure of its reliability. If subjective review fails to provide the desired degree of content accuracy, special forensic measures must be employed. These measures are provided by a dedicated branch of digital forensics. That is known as Digital Video Forensics (DVF). The DVF is mostly used in legal matters to determine the integrity and legitimacy of video evidence. Here, the DVF performs recovery and repair, scientific investigation and review, comparison and detailed assessment, and interpretation for video. Considered most generally, the most basic tasks of the forensic analysis process of video evidence are shown in Figure 1.1

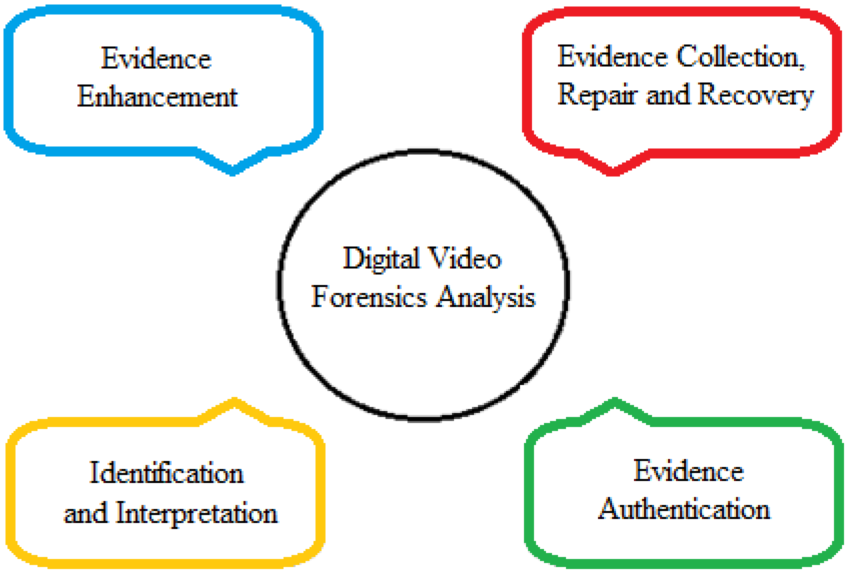


Figure 1.1 Basic tasks of the video forensic analysis process.

1.3 Video Tampering Detection

Video data is protected against tampering by two main methods called active method and passive method. The active method which is used to protect the integrity and authenticate the video data consists of watermarking and digital signatures. Active methods have significant disadvantages. Situations in which these methods can be applied are largely limited. They also reduce video quality and require special hardware [1]. In passive methods, transactions are made without relying on previously embedded information such as digital signatures. That's why more useful and advanced passive techniques are used to detect frauds in video data. In these passive video protection methods, statistical details and tampering traces in the video data are used to recognize the validity of data. Therefore, the whole purpose of video forgery detection is whether the video data has been modified after the recording period or not. All detection methods are generally shown in Figure 1.2.

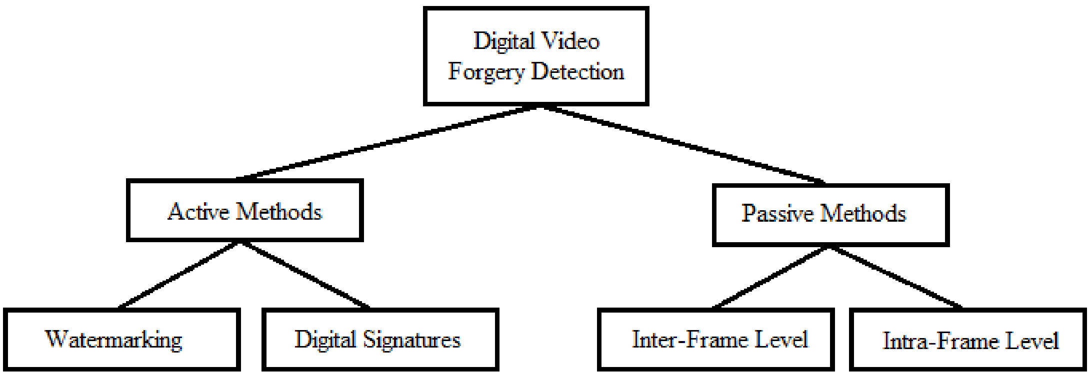


Figure 1.2 General view of digital video forgery detection

The visual contents of digital videos can be modified in a number of ways. Digital video forgeries have been classified into two kinds of tampering in the passive techniques [2], [3]. Intra-frame forgeries are one of them. These tamperings are done in both spatial and spatio-temporal domain. For example, region splicing; it is the merging of different regions in a video into the original video because of changing the content. Again if it is also copy-move, it is copying parts of a frame and pasting it to other locations in the same frame or it is copying specific regions from a series of frames and pasting them into another sequence in the same video. The other one is inter-frame forgeries. Inter-frame forgery is usually used in lookout videos because of its simplicity and imperceptibility. These tamperings are done only in temporal domain. For example, duplication means that a group of frames in the video has been copied and is pasted in another location in the same video. Insertion, taking a frame group from a foreign video is added to the original video. Deletion is the deletion of some frames from the original video and shuffling is changing the order of some squares.

1.4 Objectives of the Study

Digital videos present much of the most possible convincing evidence, both inside and outside the court, and according to this epistemology unique feature of video evidence is grounded in both the realistic quality of its definition and its allegation of mechanical objectivity. By the way, the sensibility of videos to conscious falsifications raises concerns regarding their ability to use as reliable evidence. Among of these reverse influences, the challenge of establishing the accuracy and reliability of video evidence becomes especially significant. This study revolves around of developing a forgery detection framework for digital video authentication and also inter-frame (deletion, insertion and shuffling) forgery is detected with high accuracy and low runtime in large amount of video datasets. The present study generally aims at addressing the following main objectives:

First Objective: Video evidences have issues and challenges related to forensic analysis. These need to be determined.

All research fields present certain challenges and unexplained issues, and the explanation and evaluation of these issues both make it easier the process of their resolution and help us to understand the natural limitations of that specific research field. By identifying the issues and challenges associated with the forensic analysis of video evidence, it is planned to gain an insight into the various requirements of this research area.

Second Objective: Definition of the relationship between the natural properties of test videos and parameters of different detection methods.

Video data is defined by wide variability in terms of its intrinsic attributes and also the features of its visual contents. For this reason, it becomes very important to know that the types of impact the various natural features of the goal video have on the variable factors of the experimental technique designed to authenticate it. Disclosure of such operational dependencies is very important about that both to enhance the current level of understanding of this topic, and to formulate strategies whose knowledge of them maintains objectivity without being affected by the variable characteristics of the given video.

Third Objective: Development of video tampering detection framework that can effectively distinguish malicious forgeries.

Verifying video evidence has a paramount social importance. However, this is a significant challenge. Therefore, developing precise forensic solutions which can be relied upon in critical situations is essential. In addition, real-life judicial situations are volatile and unpredictable. These situations necessitate that the developed solution must be strong. The main purpose of this study is to develop a video authentication framework that meets the mentioned requirements.

Fourth Objective: It is necessary to evaluate the performance and effectiveness of the developed framework by testing its capabilities on different datasets. Comparing the proposed methodology with existing video tampering detection techniques is also required.

An experimental technique should be extensively validated on extensive datasets and compared with similar researches. However, in this way it can be mentioned about an effective applicability and whether it provides any advantage over other techniques in the same field. Therefore, it is very important to evaluate the performance of the proposed solutions in realistic environments and to verify them with numerous tests. These are necessary to compare the forensic capabilities of the proposed technique with existing solutions.

1.5 Thesis Organization

This study is organized as follows:

Chapter 1 has an introductory view on the importance of video, its usage areas, and its place in our lives. It also includes attacks on videos and explanations on the authentication of video content. On the last of the chapter, objectives, and thesis organization are included.

Chapter 2 addresses the various issues and challenges associated with the forensic analysis of video evidence, the types of video tampering, and the types of video forgery detection methods. It also includes an analysis of the existing literature in the field of video tampering detection and some similar topics. This chapter lays the foundation for the successful realization of the study.

Chapter 3 includes the proposed method for inter-frame forgery detection in fully digital videos and the step-by-step explanation of the developed algorithm. In addition, it includes that presentation the performance strength of the proposed method used to see video dataset and video properties. Moreover, the requirements for the proposed algorithm are given in this section.

Chapter 4 covers the implementation of the proposed method and the dataset design which is used for the detection of inter-frame frauds (i.e., frame insertion, frame deletion, frame shuffling, and frame duplication) in digital videos. Examples of the application for the proposed method are included. It also includes performance analysis for all videos and video datasets.

Chapter 5 evaluates the performance of proposed method and experimental results.

Chapter 6 includes the overall conclusion of the thesis and ideas for future work.

CHAPTER 2

LITERATURE REVIEW

The literature review has revealed that there are significant studies on video forgery detection. Researchers have proposed and developed various forensic video methods for the detection of video forgery, generally by using active and passive techniques. This section generally includes types of video forgery, methods of video forgery detection and related studies.

2.1. Kinds of the Video Forgeries

The contents of digital videos can be changed by many different methods. The scheme of all known types of video forgery is shown in Figure 2.1

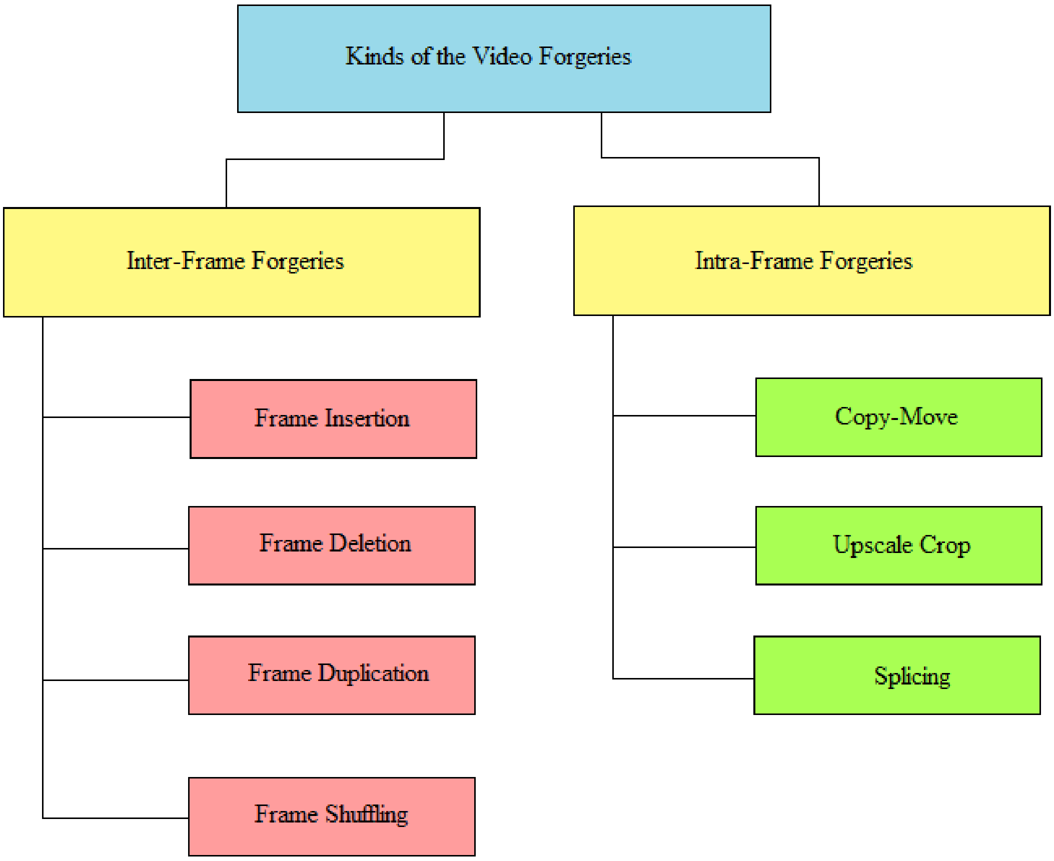


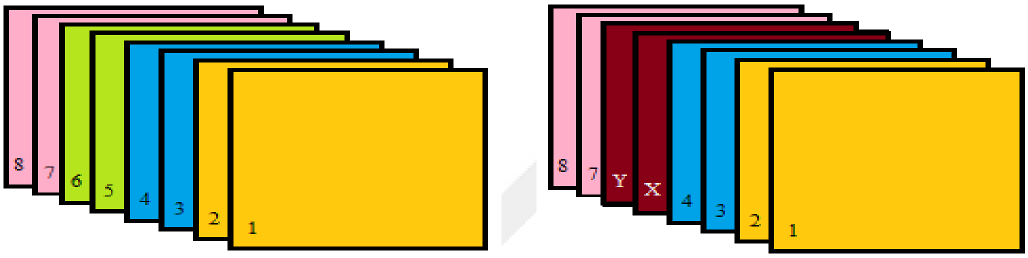
Figure 2.1 Classification of Video Forgeries

2.1.1 Inter-Frame Video Forgeries

Inter-frame forgeries are forgeries that are changed the sequence of frames in a video. These changes occur only in the temporal domain. There are types of forgery, such as frame insertion, frame deletion, frame duplication, frame shuffling. In a video with long or complex stage elements, seeing comfortably this kind of tampering is not easy for a person.

2.1.1.1 Frame Insertion

This type of forgery involves adding a group of frames from a different video source to a different time sequence of the video which is illustrated in Figure 2.2.



(a) Original Frame Sequence (b) Frame-Insertion Sequence

Figure 2.2 (a) Representation of the original frame sequence, (b) Representative illustration of frame-insertion forgery, where frames x and y have been inserted into the video from outside.

As an example, in Figure 2.3, after the 108th frame of the original video, a total of 104 frames were added and forgery video was produced. The added 104 frames were taken from a different video source.

Original frames 107-112 are shown in Figure 2.3 (a) and in the Figure 2.3 (b), the beginning of forgery frames is shown against the same frames (the first forgery frame is 109). In Figure 2.3 (c), the original frames between 209 and 214 are shown. In Figure 2.3 (d), the end of forgery frames is shown against the same frames (the last forgery frame is 212).

(a) Original Frames (107-112)

(b) Forged Frames (107-112)

(c) Original Frames (209-214)

(d) Forged Frames (209-214)

Figure 2.3 Frame examples for frame-insert forgery by using two different video sources [4]

The location of the forgery-original frames of the same test video is schematized in Figure 2.4.

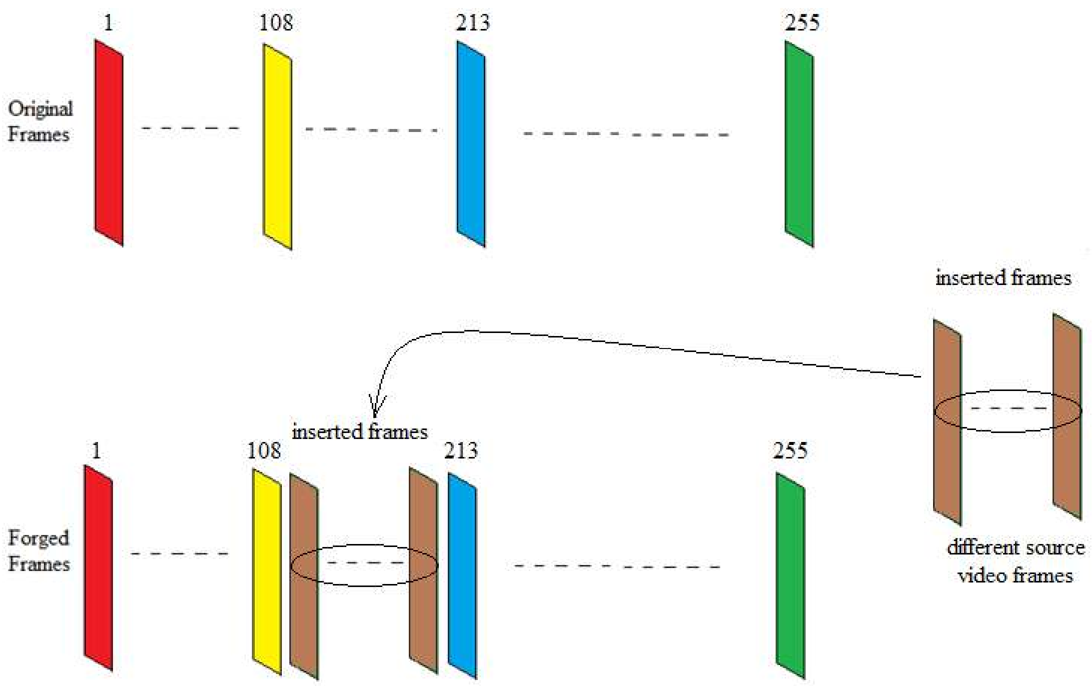
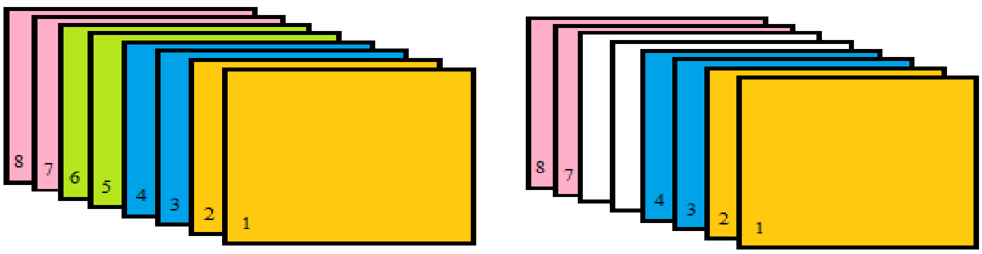


Figure 2.4 Schematized form of the Figure 2.3

2.1.1.2 Frame Deletion

Frame deletion is a form of forgery that involves removing a group of frames from a video sequence which is shown in Figure 2.5.



(a) Original Frame Sequence (b) After Frame Deletion

Figure 2.5 (a) Representation of the original frame sequence, (b) Representative illustration of frame-removal forgery, where frames 5 and 6 (denoted in white) have been removed from the video.

As shown in the Figure 2.6 (a), after the 106th frame of the original video, a total of 81 frames were deleted and forgery video was produced as in the Figure 2.6 (b) while 106th frame is the same for forgery and original video, 107th frame is different. 187th frame in the original video corresponds to 107th frame in the tampered video.

(a) Original Frames (104-109 and 183-188)

b) Forged Frames (104-109)

Figure 2.6 Frame examples for frame-deletion forgery [5]

The location of the forgery-original frames of the same test video is schematized in Figure 2.7.

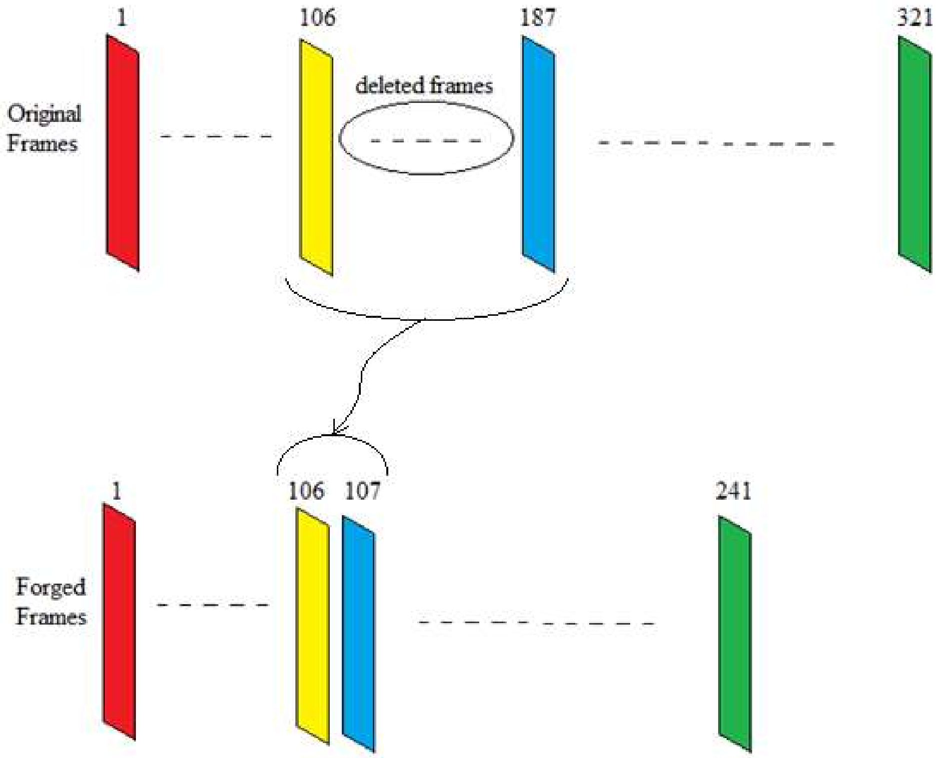
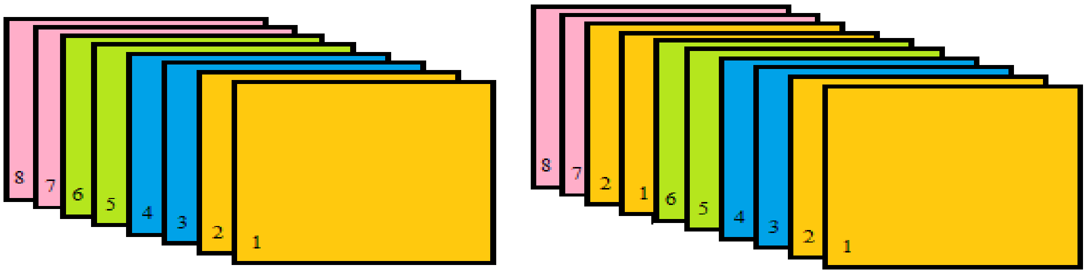


Figure 2.7 Schematized form of the Figure 2.6

2.1.1.3 Frame Duplication

Frame duplication is a type of forgery that occurs when a group of frames is copied from within a video sequence to another group at another temporal domain as illustrated in Figure 2.8.



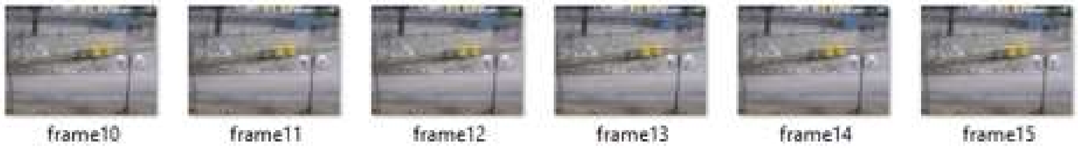
(a) Original Frame Sequence (b) After Frame Duplication

Figure 2.8 (a) Representation of the original frame sequence, (b) Representative illustration of frame-duplication forgery

In Figure 2.8 (b), frames 1 and 2 (denoted in orange) have been pasted repeat after frame 6. Any frame has not been deleted while pasting frames 1 and 2.

As shown in Figure 2.9 (c), total 50 frames of the original video between 361-411 were copied as in Figure 2.9 (b) and a forgery video was produced by pasting it again between 1-50. In Figure 2.9, a few frame sequences are shown as an example of this process.

Pasted 50 frames between 1-50 differ from frame-insertion forgery because they are copied from a different time domain of the same video. For frame-insertion forgery, added frames should be taken from a different video source.



(a) Original Frames (10-15 frames from 1-50)

(b) Forged Frames (10-15 frames from 1-50)

(c) Original Frames (370-375 frames from 361-411)

(d) Forged Video (370-375 frames from 361-411)

Figure 2.9 Frame examples for frame-duplication forgery [6]

The location of the forgery-original frames of the same test video is schematized in Figure 2.10.

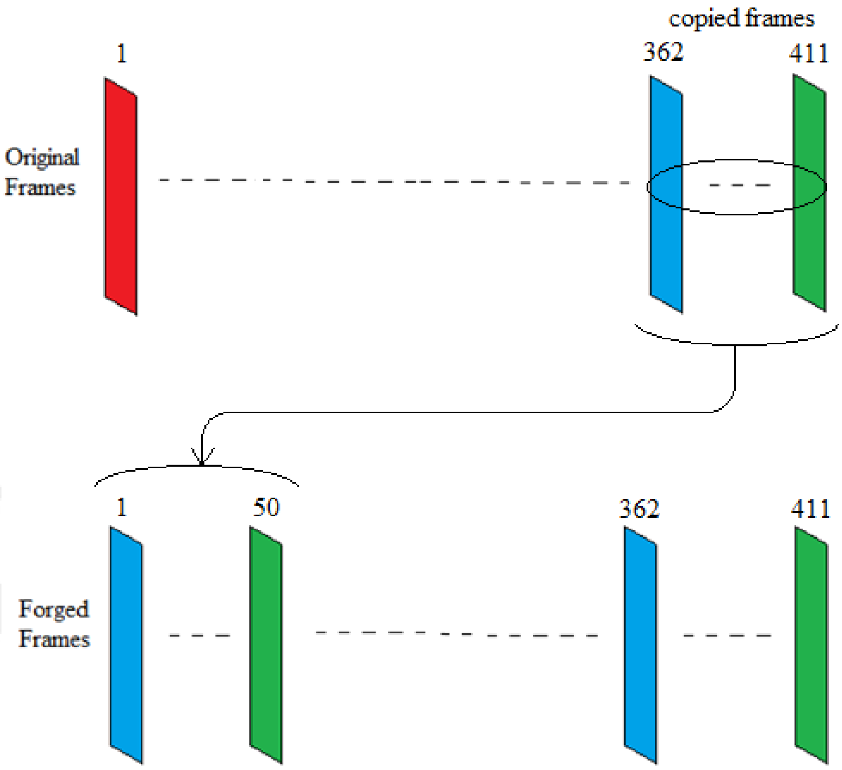
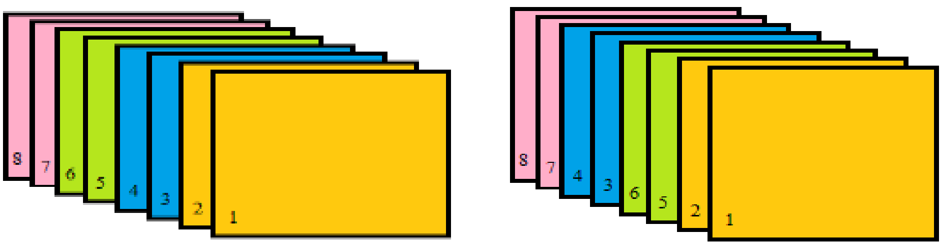


Figure 2.10 Schematized form of the Figure 2.9

2.1.1.4 Frame Shuffling

Frame shuffling is forgery that occurs by moving or swapping a few frames in a video sequence and is represented in Figure 2.11.



(a) Original Frame Sequence (b) After Frame Shuffling

Figure 2.11 (a) Representation of the original frame sequence, (b) Representative illustration of frame-shuffling forgery

In the Figure 2.11 (b), frames 5 and 6 have been changed with frames 3 and 4 at another location in the video.

As shown in Figure 2.12 (a) and in Figure 2.12 (c), frames 75-273 from the original video have been replayed by reversing as in Figure 2.12 (b). In this way, shuffling-forgery video was obtained. The Figure 2.12 (b) shows the beginning of shuffle (the first tampered frame is 75) and Figure 2.12 (d) indicates the end of shuffle. (The last tampered frame is 273). As a result, the order of the original frames has been changed as follows, (.…73, 74, 273, 272, 271…77, 76, 75, 274, 275, 276…)

(a) Original Frames (73-78 frames from 75-273)

(b) Forged Frames (73-78 frames from 1-50)

(c) Original Frames (271-276 frames)

(d) Forged Frames (271-276 frames from 1-50)

Figure 2.12 Frame examples for frame-shuffling forgery [5]

The location of the forgery-original frames of the same test video is schematized in Figure 2.13.

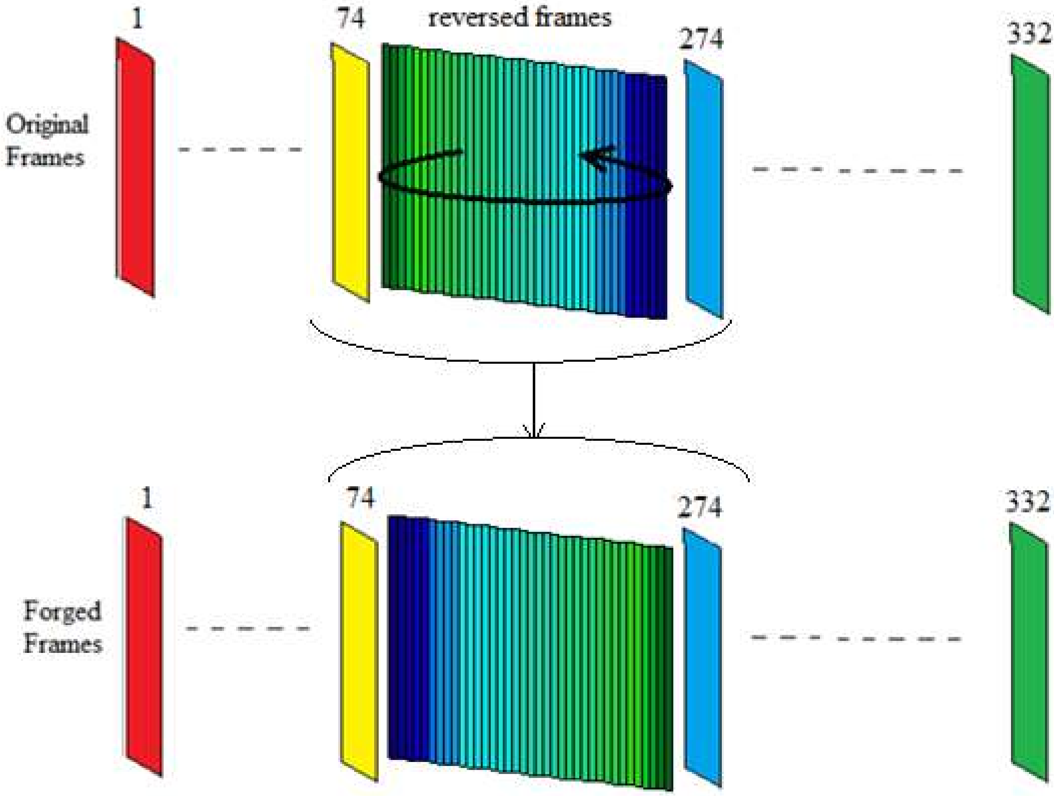


Figure 2.13 Schematized form of the Figure 2.12

Inter-frame forgeries are inherently one of the most uncompromising but effective techniques of tampering. For example, by deleting some frames from a video, all signs can be easily removed from a video in a real event.

Likewise, by pasting a set of frames from one location to another temporal location, an event can easily be made liking as to occur during a period of time that does not normally occur.

2.1.2 Intra–Frame Forgeries

Intra-frame forgeries are fraud in a specific region of the video scene. They involve the direct modification of information which are presented by a video scene, and are relatively difficult to generate it than inter-frame forgeries. There are types of forgery, such as copy-move, splicing and upscale–crop forgeries.

2.1.2.1 Copy–Move Forgery

Copy –move tampering involves removing only an area from video scene or copying into only an area of video scene. This type of forgery is occasionally referred to as partial tampering, as only a small area of the frame is manipulated while the rest of the frame remains untouched.

Figure 2.14 (a) represents sample frame from original videos, while (b) represents corresponding frame from tampered version of this video. In Figure 2.14 (b), the fixed object on the right side of the moving person is deleted.

(a) (b)

Figure 2.14 Examples of copy-move tampering [7]

2.1.2.2 Upscale–Crop Forgery

Upscale–crop tampering requires cropping video frames to eliminate scene content from the endpoints of said frames, and then expanding the resolution of the cropped frames to match the rest of the frames. Frame enlargement is mainly done through the process of interpolation, which creates new pixels with the help of the values of the existing pixels of the frame.

Figure 2.15 (a) and (c) represent original video frames, which are cropped and resized to eliminate objects at the extremities of the frames, thereby producing the tampered frames (b) and (d), respectively. People in the first frames of the videos in Figure 2.15 (a) and (c) have been deleted.

(a) (b)

(c) (d)

Figure 2.15 Examples of upscale–crop forgery [5]

2.1.2.3 Splicing

Splicing tampering is done by attaching an object / region to a video from a different video.

(a) Original Frame (b) Splicing-Forgery Frame

Figure 2.16 Examples of splicing forgery [8]

Figure 2.16 (a) represents sample frame from original video, while (b) represents corresponding frame from tampered version of this video. Another yellow-marked car has been added to the parking place in Figure 2.16 (b).

2.2 Video Forgery Detection Techniques

The various forensic features which are used by current tamper detection techniques are categorized in Figure 2.17.

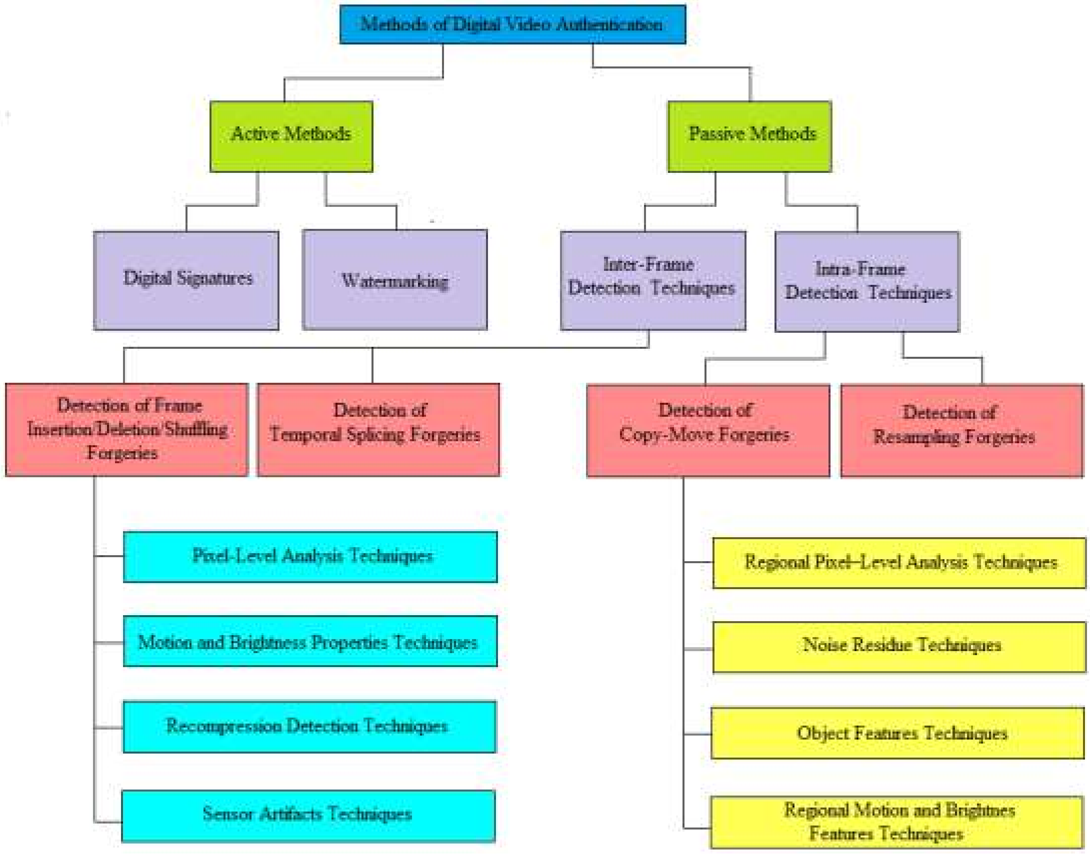


Figure 2.17 Categorization of forensic features and analyses used for the detection of different kinds of video forgeries.

There are two basic techniques for video forgery detection. These are active techniques and passive techniques [9].

2.2.1 Active Methods of Video Authentication

Active methods of forgery detection include two techniques. These two techniques are digital signatures and digital watermarking. Active techniques are required by authentic content ownership and copyright violations. The basic application of watermarking and signatures is copyright protection. They can be used for fingerprint, forgery detection, error concealment, etc. There are several disadvantages on the active technique as it requires a signature or watermark to be embedded during the obtaining period when video is recording or an individual person to embed it after obtaining period at the time of sending. Because of the need of distinctive hardware, this limits the application of active technique. Another problem which has an influence on the sturdiness of watermarks and signatures are factors like compression, scaling, noise, etc [10].

2.2.1.1 Digital Signatures

Diffie and Hellman devised the digital signature in 1976. A digital signature depends on confidential data known to the signer [11]. Content of video data matches the data contained in the digital signature. So this cannot be forged. The sender firstly removes the key from the original video, and then the data encrypted with a private key that give signature [12]. The receiver can use incoming public key and then the receiver decrypt the signature to authenticate the received video [12]. The signature is stored elsewhere than the media. And the signature is stored separately by user in defined area [11].

2.2.1.2 Watermarking

In watermarking, there is data which is used as authentication. These are embedded with multimedia data. Various watermarking is recommended to prevent illegal copying. The watermarking methods work on uncompressed and compressed information [13]. These techniques can be worked in spatial or frequency domain.

2.2.2 Passive Methods of Video Forgery Detection

Passive forgery detection techniques are noted as a developing route in video safety. Passive techniques are very different from active techniques. These techniques work in without the limitation and they do not need any first information about the video contents. These techniques use some basic assumptions like as features which are stable and some natural in original videos. If a video is forged these features are changed. Passive techniques eject these features from a video. And then they analyze them for many forgery detection methods [14].

2.2.2.1 Inter–Frame Forgery Detection Techniques

Inter–frame forgeries involve forgery with the editing of frames in a video. Creating these types of forgeries is not very complicated, but if created reasonably they can be quite difficult to detect.

2.2.2.1.1 Detection of Frame–Insertion/Deletion/Shuffling

Such manipulations are an easy way to cause an event to occur during a time period in which it would not normally occur. All that is required is to place or copy the respective frame sets to another temporal location in the video. Likewise, by removing some frames, all evidence can be comfortably removed from the video that an event has occurred. These techniques generally include sensor artifacts, recompression, motion and brightness features and pixel-level analysis sub-techniques.

Pixel-Level Analysis Techniques: Pixel–correlation analysis is one of the most effective methods for the detection of frame–tampering forgeries. Frame–duplication is repetition of a set of frames within the video at different temporal location. This means that the tampered video contains the same set of frames at the least two different locations. For this reason, inter–frame correlation sequences are obtained for adjacent frames in a target video. Although videos that include same sets of frames have abnormally high correlation corresponding to the repeated sets of frames, original videos have normal correlation forms. Since the illusion of sustained motion requires adjacent frames to be some similar, some little correlation is all the time exists among adjacent video frames. However, if there are frame–duplicated videos, correlation or the inter-frame similarity is still abnormally high. This content usually prefers for the detection of frame–duplication forgeries.

Motion and Brightness Properties Techniques: Recompression artifacts provide valuable information to the rendering history of the target video. However, their entity as alone is not an adequate pointer of content error. Because completely statutory digital content encountered in everyday life has often been subjected to more than one post-production compression over its existence whereas sensor artifacts and recompression base very seriously on forensic evidence provided by recompression artifacts. This is why often they could not be succeeded to individuate tampered videos from harmlessly recompressed ones. Evidence-based Inter-frame tampering detection methods other than recompression often use one of the key features such as Motion Features, Prediction Residual, Brightness Features, Optical Flow, and Velocity Field Features.

Recompression Detection Techniques: The category of inter-frame forgery detection techniques relates to the events involved in performing a typical inter-frame forgery to change the content of a particular video. First, each frame is extracted from the target video and forged is performed. This tampered frame sequence is then combined to create a new video. Lastly, a certain amount of compression is done and the re-encoding process is performed. This has a simple conclusion that recompression is an unwanted but unavoidable consequence of tampering with a compressed video. Therefore, the existence of recompression artifacts is accepted to be an evidence of post-production content tampering. All recompression detection techniques are used one of the four basic features such as Discrete Cosine Transform (DCT) Coefficients, Block Artifacts, Prediction Features, and Markov Features.

Sensor Artifacts Techniques: Video recorders have specific detectable marks or artifacts in all the content they record. Such marks are mostly used for source identification. These are also used in the domain of tamper detection. Sensor artifacts based tamper detection techniques use forensic features as Read Out Noise, Sensor Pattern Noise (SPN), and Photon Shot Noise.

2.2.2.1.2 Detection of Temporal Splicing

Temporary splicing is defined as the process of combining frames of multiple source videos to create a new video. If the frame rates of the source videos do not match, the merging is performed before they are merged with the help of a process known as Frame Rate Up–Conversion (FRUC) [15].

2.2.2.2 Intra–Frame Forgery Detection Techniques

Intra-frame detection includes techniques to detect tampering on only a certain region or an object on the video screen.

2.2.2.2.1 Detection of Copy–Move/Paste Forgeries

Copy-move/paste forgery involves adding or removing objects from video frames. Detecting techniques of this type of forgery look for structures that arise in the underlying statistical properties of a frame containing identical regions. These techniques generally include pixel-level analysis, noise features, object features and motion and brightness sub-techniques.

Regional Pixel–Level Analysis Techniques: The techniques belonging to this category actually deal with similarities or correlations among certain regions of adjacent frames or specific regions of the same frame. In other words, these techniques focus on regional level, not between frames (spatial domain or spatio-temporal domain). Video frames often have some natural similarity between nearby regions due to spatial redundancy. However, if this similarity outruns a certain predetermined threshold, it means that forgery can occur. An alternative way to detect copy-move/paste forgery is to identify regions of the frame with abnormal pixel correlations by comparing different regions of the frame with other regions of the frame. Generally, such techniques identify frame regions that showing abnormally associated pixels. For this, a machine learning model is trained.

Noise Residue Techniques: Noise residue techniques are different techniques used to determine the same regions within video frames. This technique investigates correlations between residual noise components in video frames. From the noise-removed type, the given frame is subtracted and the noise residue is calculated. This technique is based on the fact that the correlation between the noise residues in the copy-move/paste regions has a considerable deviation from the standards. By detecting this deviation, the tampered region is detected.

Object Features Techniques: Copy-move/paste forgery can be detected by detecting artifacts caused by removing a region from the frame. Ghost shadow artifacts, which appear when a moving object is removed from a series of frames, are one of the most evident artifacts of region removal. These artifacts are detected by first dividing the video frames into a background and a foreground. Any inconsistency between the two means forgery.

Regional Motion and Brightness Features Techniques: Object removal forgeries are detected by observing contradictions in motion information in consecutive video frames. Motion Vectors (MVs) are calculated for consecutive frames of the video and moving objects are detected by comparing the sizes of the MVs. Then, the monotony of the MVs in the foreground and background regions of each frame is focused. If there is insufficient monotony in any foreground region, it indicates the presence of object removal forgery. This technique is used to detect the removal of moving objects from videos with a fixed background.

2.2.2.2.2 Detection of Resampling

Upscale−cropping involves trimming the ends of a series of video frames and enhancing them to maintain resolution consistent with the rest of the video frames. In the case of insertion, regions of different video frames are combined to create new tampering frames. Both types of tampering are characterized by a common process known as resampling. When cropping, the cropped frames should match the resolution of the remaining frames of the video. For this, it is necessary to enlarge it to match and do a resampling process. Likewise, to create reasonable and unobtrusive added frames, it becomes important to resize the copied and pasted object. Sometimes it is necessary to maintain consistency both within the tampered frames and within the video as a whole. This resizing is a pixel rearrangement. This process essentially involves resampling.

2.3 Related Work

Video forgery detection techniques are divided into two categories and these consist of active approaches and passive approaches [9]. Active techniques require data such as watermarks and signatures to be embedded in advance. But this is not required for passive techniques. Passive approaches are one of the detection digital video tampering methods. These approaches are the most important techniques.

Table 2.1 summarizes similar studies on passive approaches. The classifier headings include author, methods of work, number of test video, frame dimension, accuracy and performance evaluation.

Table 2.1 Summary of previous works for video forgery detection

Author's	Methods of Work	Test/Frame Dimension	Accuracy	Performance Evaluation
Wang, W., [16], (2007)	Correlation coefficient in duplicate frames, removing people or objects from a video	2 videos 480×720	84.2%	The method is for automatic image frame to detect frame and region copy-move
Su, Y., et al. [17], (2009)	With respect to the motion-compensated edge build is the recommended frame deletion	5 videos Unknown	Unknown	Proposed method has shown a successful performance against distinct artifacts like pixel blocks in a frame.

Table 2.1 (Continuous)

Author's	Methods of Work	Test/Frame Dimension	Accuracy	Performance Evaluation
Zhang, J., et al. [18], (2009)	While the active object is removed from the video with inpainting, ghost shadow is formed during the process. The method is based on this artifact. The gray level were used.	3 videos 720x480	Unknown	Accuracy was not mentioned for method. Ghost shadow is efficient for detecting. Compared to other researching, a very small data set was used.
Chetty, G., et al [19]. (2010)	Extraction of pixel sub-block noise residue features between three distinct types of correlation	Internet streamed movies Unknown	92%	More results are needed for the method. Also very accurate segmentation results were obtained, effectively removing video tampering detection.
Bestagini, P., et al. [20], (2013)	The tampered were detected using the footprint residual. Two properties are calculated between adjacent frames.	20 videos 320x240	87%	The important periods of the proposed techniques are feature extraction.
Wang, Q., et al [21], (2014)	Correlation coefficients gray, normalization, quantization distinguishing features	5 videos 256×256	98.79%	The method contains a small data set. It is not very efficient in terms of frame dimension and amount of test video in classifying frame insertion and deletion fraud.

Wang et al. [21] presented a method between consecutive video frames based on the assumption that the correlation coefficients of the gray values lie. Correlation coefficients of gray values between the following video sequences were normalized and quantified to detect inter-frame forgery. For this, a small dataset (five videos) was used and accuracy was found to be 98.79%. Also Wang et al. [16], in 2007, only tested with 2 videos and got only 84.2% accuracy.

Moreover, Chetty et al. [19] proposed a method of video tampering detection based on multi-mode combination and conversion of several intra-frame and inter-frame pixel sub-blocks from video sequences. The quantization residue features situation for the all experiments is like noise residue features. And copy-move tamper scene revealed this. However, the method focused on frame-level forgery and therefore could not find regional level falsification and localization problem.

Su et al [17] used a tampering detection method for frame-deletion with motion compensated edge artifact. Motion-compensated edge artifact was used to define the changes of correlation among adjacent frames.

Similarly, Bestagini et al. [20] carried out analysis of the footprints for video frames with the help of a detection algorithm that enables a forensic researcher to define video forgeries and place them in the domain of spatio-temporal. 120 original frame sequences were analyzed. They had had with the resolution of 320x240 pixels and with 20 videos. The results had an accuracy of 87%.

CHAPTER 3

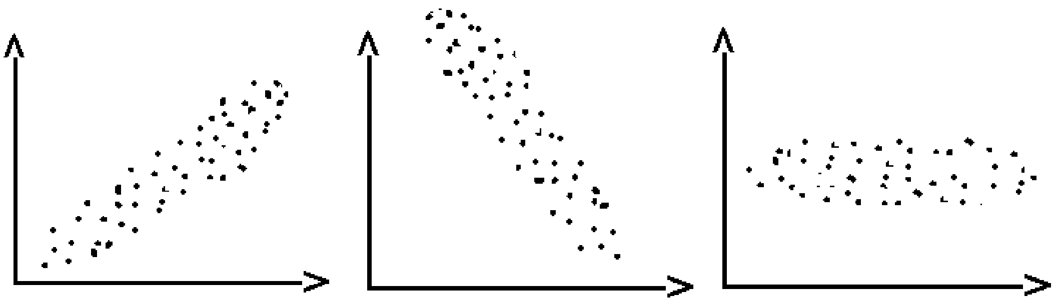
METHODOLOGY AND REQUIREMENTS

3.1 Proposed Method

In this section, algorithms, methods, systems, materials and basic functions of the planned system are discussed for the generally proposed approaches and methods which are focused on inter-frame video forgery detection, such as frame insertion, deletion, and shuffling. A method based on correlation coefficient difference and 3sigma-rule has been proposed to classify original videos and inter-frame tampering. Because of correlation coefficient difference, the Group Of Pictures (GOP) of video, gray GOP levels of the video and the pixel values of each frame were used. The software and hardware features which are required for testing and using the proposed method are shown in Table 3.1.

3.1.1 Correlation Coefficient

Correlation shows the linear relationship between two or more variables [22]. It is the amount of relationship between two variables, called binary or simple correlation calculated by correlation techniques. Correlation coefficient equation is used to find how powerful a relationship is between at least two data. The equations proceed a value between -1 and 1. '1' defines a powerful positive relationship. '-1' defines a powerful negative relationship. '0' defines no relationship at all which are all shown in Figure 3.1.



(a) (b) (c)

Figure 3.1 Correlation illustration for -1, 0 and +1, where (a) Positive Correlation, (b) Negative Correlation, and (c) No Correlation

Correlation coefficient calculations are used in many fields and with different variations [22]. In the proposed method, it is checked whether the video content is normally continuous, so Equation (3.1) is used to define the correlation coefficient between the two frames.

$$r_k = \frac{\sum_i \sum_j (F_{ra_k}(i,j) - \overline{F_{ra_k}})(F_{ra_{k+1}}(i,j) - \overline{F_{ra_{k+1}}})}{\sqrt{(\sum_i \sum_j (F_{ra_k}(i,j) - \overline{F_{ra_k}})^2)(\sum_i \sum_j (F_{ra_{k+1}}(i,j) - \overline{F_{ra_{k+1}}})^2)}} \tag{3.1}$$

where k and k+1 are frame numbers (k=1, 2, ..., n-1), r_k is the correlation coefficient between frames k and k + 1. $F_{ra_k}(i,j)$ represents the gray value of the pixel. For frame k, i is row and j is column. $\overline{F_{ra_k}}$ is the average value of the $F_{ra_k}(i,j)$. Accordingly, from a continuous frame sequence of which the length is n, a correlation sequence of n −1 is obtained with Equation (3.1).

While the content correlation between adjacent frames is high for the original video, the content correlation between distant frames in the temporal domain is relatively low [23]. In the proposed method, it should be checked whether the video content is normally continuous. At each frame consistency of correlation coefficient (Equation (3.1) together with gray value) is used to detect it. Because of using consistency of correlation coefficient, video frames must be extracted. On the other hand, if a video has a large number of frames, each frame needs high storage at Red-Green-Blue level (RGB level). Therefore, gray-level frames are used to prevent slowdown in using consistency of correlation coefficient.

3.1.2 3σ-Rule (3sigma-Rule)

The 3sigma-Rule is a simple and widely used for outlier detection. 3sigma is the name of some statistical hypothesis tests whose test statistics are normalized or known as normalized. Researchers and statisticians use this calculation to identify outliers in data and adjust their findings accordingly [24]. An investigation is considered as an outlier if its least squares residual overlaps three times its Standard Deviation (SD) [25]. The Equation (3.2) formula is used as the basic for the 3sigma theory.

The Equation (3.2) formula is used as the basic for the 3sigma theory. X is normally $N(a, \sigma^2)$ distributed random variable. For any k>0,

$$\mathrm{P}\{\, |\mathrm{X} - \mathrm{a}| < k\sigma \,\} = 2\Phi(k) - 1 \tag{3.2}$$

where $\Phi(k)$ is the distribution function of the standard normal law, σ is standard deviation (SD).

In particular, for k=3 it follows that

$$P\{a - 3\sigma < X < a + 3\sigma\} = 0.099730 \tag{3.3}$$

The Equation (3.3) means that the values of X can differ from its expectation a by amount extreme 3σ on the average in not more than 3 times. This situation is occasionally used by a tester in particular problems of probability theory and mathematical statistics, by assuming that the event $\{|X-a|>3\sigma\}$ is actually impossible and, as a result, the event $\{|X-a|<3\sigma\}$ is actually certain [26].

3.1.3 GOP Structure

The differences between adjacent frames symbolize the movements of the objects recorded by the camera [27]. It is a Motion Picture Experts Group (MPEG) structure and a display format, and not usually edited after encoding [28]. MPEG has better compression ratios and lower bitrates. In the MPEG compression, each frame of the video stream is compressed individually. The compressed frames are categorized in three ways [27].

Intra frames (I-Frames): I-frames are the highest quality frames and use the most amount of bits within a video stream. At the same time, they have the lowest compression. In these, the spatial redundancy within a single frame is removed and compression is achieved. I-frames are also called key frames [29].

Predictive Frames (P-Frames): P-Frames predicts the movement relative to previous frames [29]. The frame is compressed by using the information from previous frames. So, information from previous frames is necessary. Shown in Figure 3.2, P-frames have much better compression than I-frames and these use less bits within a video stream [30].

Bi-directionally Predictive Frames (B-Frames): B-frames depend on with motion compression like P-frames, but B-frames are bi-directional because they associated with the frames that came before and after them as shown in Figure 3.2. Therefore, B-frames offer the highest compression level and have the fewest bits within a video stream [29] [31].

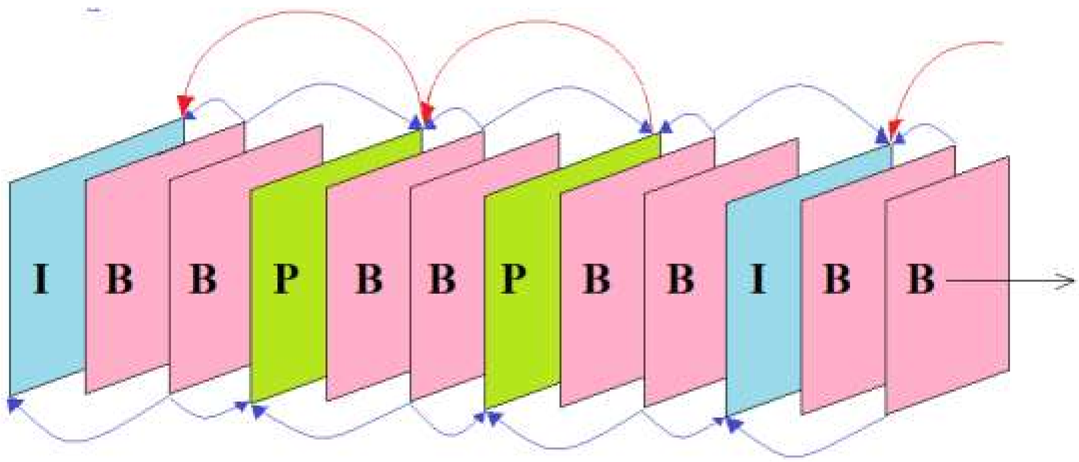


Figure 3.2 Group of Structure (GOP)

Digital videos are made up of frames, and these frame groups are called Group of Picture (GOP). The GOP structure consists of I, B and P frames. At the same time, a GOP is the distance between two key frames, measured in the number of frames, or the amount of time between key frames. The beginning of a GOP starts with an I-frame and ends with the next I-frame. This is called the length of the GOP and is defined by N. In the GOP, the distance between two P-frames or I-P frames is denoted by M. The GOP structure and sample representation for different N and M values are given in Figure 3.3 and Figure 3.4.

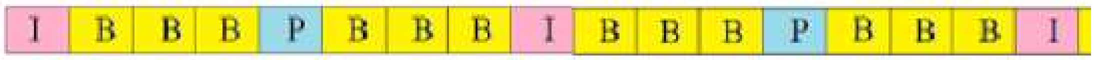


Figure 3.3 GOP Structure for N=8 and M=4

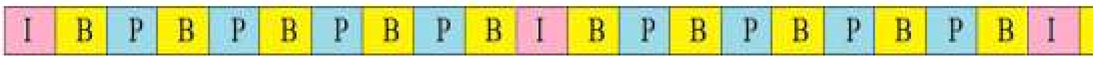


Figure 3.4 GOP Structure for N=10 and M=2

If the video in Figure 3.4 is tampered with and re-recorded in MPEG format, the given frame order is broken and so it can be detected as forgery. This frame tampering leaves both spatial and temporal traces to the frame in the resulting forgery video stream. The post-tamper video stream is shown in Figure 3.5.

The frame sequence in Figure 3.5 (a) represents the original MPEG video stream. The six frames marked in red in Figure 3.5 (b) were forged by deleting the frames. The frame row in Figure 3.5 (c) shows the forgery frame sequence obtained by recompressing after forgery. Traces of static and temporal artifacts play an important role in forgery detection in this compressed forgery MPEG video.

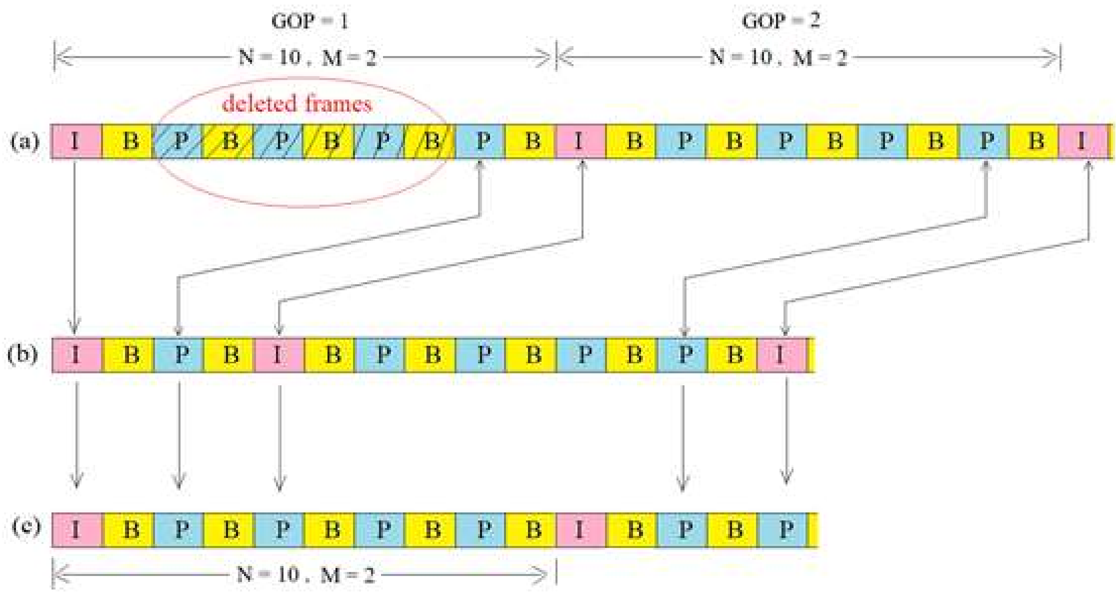


Figure 3.5 Change of GOP Structure with Tampering

3.1.4 Proposed Algorithm

In videos, inter-frame tampering is a frame-level tampering. Also, because the videos consist of a group of pictures (GOP), frames are extracted from the video. Thus, the frames are processed and the video is analyzed. Short videos can take a long time to process because that videos have a large amount of frames (such as a video of about 10 seconds has more than 300 frames on average). Therefore, mostly short-length videos are preferred for video analysis (about 10-second videos). At the same time, all frames (GOP) are turned gray level in order to speed up video processing in inter-frame analysis (gray GOP level). Correlation coefficient difference values are calculated by using all pixel values in a gray level frame. This is repeated throughout the video for every two adjacent frames. Proposed algorithm steps are shown in Figure 3.6.

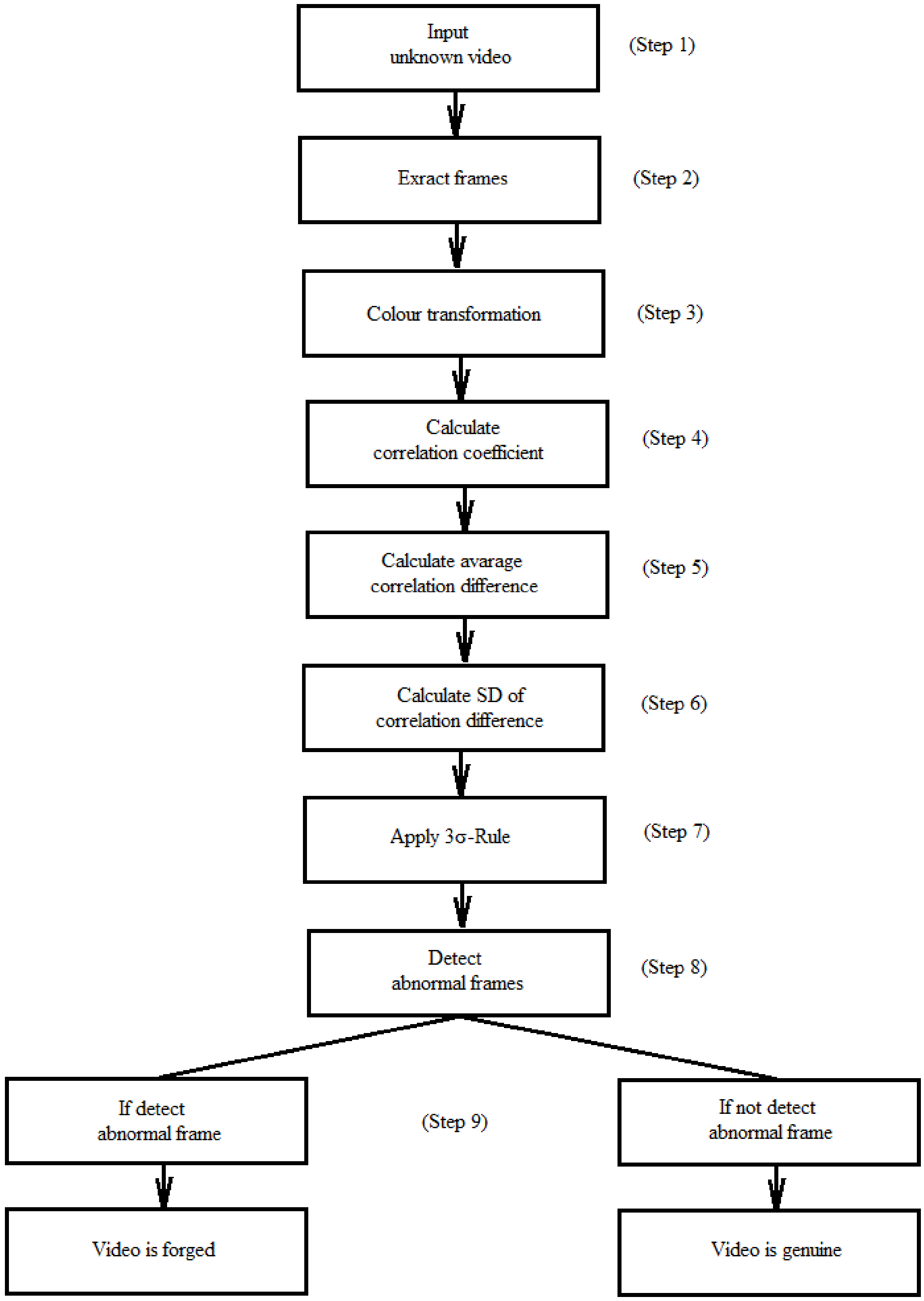


Figure 3.6 Algorithm of the Proposed Method

Step 1: A dataset folder is created. Random videos are placed in the folder (forged and original videos). For the test, any of the videos is selected and read.

Step 2: Frames are extracted from the video and frames of video is converted to group of pictures (GOP). Because all videos consist of images.

Step 3: GOP is converted into gray level picture frames. Each frame needs high storage at Red-Green-Blue level (RGB level). Therefore, gray-level frames are used to prevent slowdown in using consistency of correlation coefficient. The amount of computation needs to be reduced. While RGB color image is (256×256×256), gray image is (256).

Step 4: The values of the inter-frame correlation coefficient between adjacent frames of the gray values of GOP are kept and they are calculated. Equation (3.1) is used for this calculation. Between adjacent two frames the time range is short and the correlation is high whereas between the remote frames the time range is long and the correlation is low.

Step 5: The average of the correlation coefficient difference between adjacent frames is calculated. In between adjacent frames for inter-frame the correlation of the video should be continuous [32]. Equation (3.4) is used for the correlation coefficient difference.

$$\Delta r_k = \begin{cases} |r_k - r_{k+1}|, & k \geq 2 \\ 0, & k = 1 \end{cases} \tag{3.4}$$

where $\Delta r_k = \{\Delta r_1, \Delta r_2, \Delta r_3, \dots, \Delta r_{n-4}, \Delta r_{n-3}, \Delta r_{n-2}, \Delta r_{n-1}\}$ represents the variability of the content of a total video [14]. The variability of the original video should be consistent or else, the consistency is destroyed by the tampered frames [32].

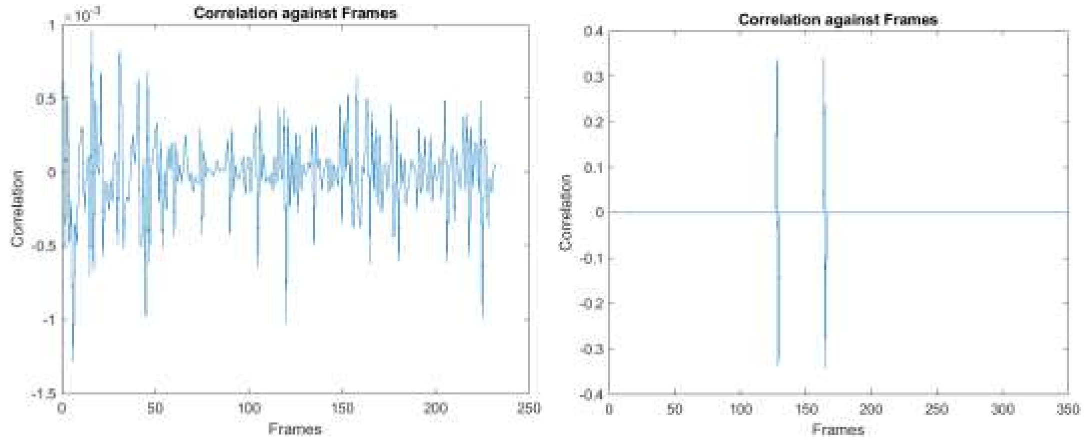
Step 6: The Standard Deviation (SD) of the correlation coefficient difference is calculated. SD is calculated to be able to apply 3sigma and see more accurate results of correlation coefficient.

Step 7: For the extreme values of the distribution, the 3sigma rule is applied. Outliers in the data are identified and findings are adjusted accordingly. This is considered as an outlier if its least squares residual overlaps three times its Standard Deviation (SD) [25].

Step 8: Abnormal points are detected between two adjacent frames for all GOPs. The correlation value differs between adjacent frames and distant frames. Frame deletion/insertion/shuffling is expected to change the correlation value in the opposite direction of what was predicted. In the temporal domain, this change point creates an abnormal point for inter-frame level.

Step 9: If the calculated abnormal point is a positive return result, the video may be forgery because the temporal correlation between the two frame sets is forged, otherwise it is original.

An example is given in Figure 3.7 for the graph that appears on the result screen:



(a) Original Video (b) Forgery Video

Figure 3.7 Example of correlation coefficient where, original video has 242 frames, forgery video has 352 frames

Correlation values of the original video are in the range of ($1x10^{-3}$)-($-1.5x10^{-3}$), while the correlation values of the forgery video are in the range of (0.4)-(-0.4). (The graphic of the original has been enlarged approximately 500 times compared to the graphic of forgery.)

3.2 Dataset

Four different benchmark datasets are used in this study. The first dataset is from SULFA [5]. The dataset contains a total of 46 videos. It contains 26 tampered videos under different inter-frame forgery and 20 original videos. The second dataset is from PU [8]. The dataset contains a total of 10 videos. 4 of them are forgery, while 6 of them are original videos. The third dataset is from VIFFD [33]. The dataset contains a total of 34 videos. It contains 14 tampered videos under different inter-frame forgery and 20 original videos. The fourth dataset is from REWIND [6]. The dataset contains a total of 6 videos. 2 of them are forgery, while 4 of them are original videos. The final dataset is created from various datasets. There are a total of 24 videos in this dataset, of which 10 are original videos and 14 are forgery videos.

All datasets have original videos, duplicated forgery videos, shuffling forgery videos, deletion forgery videos, and insertion forgery videos. The detailed classification of these datasets and the detailed features of the videos in them are discussed in Chapter 4 (In Dataset Design and Video Features Section).

3.3 Requirements

The software and hardware features required for testing and using the proposed method are shown in the subtitles and tables. At the same time, Matrix Laboratory (MATLAB), one of the programming languages, is used to perform the tests and implement the proposed algorithm.

3.3.1 MATLAB Programming

MATLAB was created using the initial parts of the words from Matrix Laboratory. MATLAB is very rich in terms of features and also has a very wide usage field. Some of these features and usage fields are listed below:

- ✓ MATLAB performs data visualizations, data analyzes and numerical calculations with a simple and practical environment [34].

- ✓ MATLAB has large built-in archives that includes definitions of its main functions, exemplary code, related demos, and general help pages.

- ✓ MATLAB, whose fundamental data type is matrix, treats all data as some kinds of matrix. For example, it accepts single values as 1×1 matrices.

- ✓ It has perfect graphics abilities and a high-level programming language for matrix-type operations.

- ✓ MATLAB algorithm development environment has many services. Some of them are a command-line interface, 2D-3D plotting functions, numerical and string processing features, and the capability to create Graphical User Interfaces (GUI).

- ✓ Thanks to the programming language, commands are interpreted without the need for compilation, which shortens the programming time.

- ✓ Furthermore, MATLAB is preferred for solving many numerical problems because the instruction statements are very similar to those used in mathematics and engineering. Because that MATLAB is easier to use than C, FORTRAN and other languages.

- ✓ MATLAB includes powerful practicability and necessary tools developed for many different fields. It has widespread usage in many fields from neural networks to finance analysis and signal processing such as control system design and analysis, image and video processing.

It has been used actively and effectively in this study due to its video processing capabilities. The fact that the videos consist of Groups of Pictures (GOP) and have a large number of tools used in both video and image processing has been particularly preferred when compared to different programs. In a similar way, images are represented as strings of real values. They consist of grayscale or color data items [35, 36]. MATLAB generally uses a two-dimensional array to manipulate images. Each element of an array matches to a pixel value of the image. Videos can be thought of as an extension of images in time or perspective. Every frame of a video is a static image.

Videos are stored by adding a dimension to the image array. This dimension is represented by time and display information [37].

3.3.2 Hardware and Software Requirements

Personal Computer (PC) is used for the experiments and the implementation of the algorithm steps. Software and hardware features such as Random Access Memory (RAM), Hard Disk (HD) and Central Processing Unit (CPU) of the Personal Computer and version information of installed the MATLAB program on the computer are shown in Table 3.1.

Table 3.1 Hardware and Software Features Used for the Proposed Method

System Requirements	Using Hardware	Personal Computer (PC)	DELL Inc. Inspiron N4050
		Central Processing Unit (CPU)	Intel(R) Core (TM) i5-2410M CPU @ 2.30GHz (4CPUs), ~2.3GHz
		Random Access Memory (RAM)	4.00 GB
		Hard Disk (HD)	232GB
		Video Card	Intel(R) HD Graphs 3000
	Using Software	Operation System (OS)	Windows 10 Education PC
		Coding	MATLAB R2020a Program

CHAPTER 4

DESIGN AND IMPLEMENTATION OF THE PROPOSED METHOD

4.1 Dataset Design and Video Features

SULFA dataset, PU dataset, VIFFD-A dataset, REWIND dataset and also dataset consisting of different various datasets are used for inter-frame forgery video detection. A large video dataset is created by taking various numbers of videos from each of these different datasets. Detailed information about the videos in the dataset consisting of 60 tampering videos and 60 original videos is shown in Table 4.1 and Table 4.2. All videos from these datasets were tested to demonstrate the accuracy and performance analyses of the proposed method.

Table 4.1 Video dataset information

Dataset	Total Video	Forged Video	Original Video	Types of Forgery
SULFA [5]	46	26	20	Deletion
				Insertion
				Shuffling
PU [8]	10	4	6	Insertion
				Shuffling
VIFFD-A [33]	34	14	20	Deletion
				Insertion
				Copy-move
REWIND [6]	6	2	4	Insertion
Various [4]	24	14	10	Deletion
				Insertion
				Shuffling
Total Dataset	120	60	60	

Table 4.2 Video information

Dataset Videos	Frame Rate	Resolution	Format	Length
SULFA Videos	29-30 f/sec	320x240	avi / mp4	8-15 sec.
PU Videos	28-30 f/sec	640x360	avi / mp4	12-16 sec.
VIFFD-A Videos	25-30 f/sec	720x404	mov / avi / mp4	7-15 sec.
REWIND Videos	29.97 f/sec	320x240	avi / mp4	9-10 sec.
Various Videos	25-30 f/sec	320x240	mov / avi / mp4	7-18 sec.

4.2 Implementation of Method

Dataset videos were run with the codes developed in the MATLAB R2020a version program. The algorithm steps of the proposed method are applied and some results are exemplified in sub-titles at below. Applications focus on detecting frame-deletion forgery, frame-duplication forgery and frame-shuffling forgery. Detailed information on the videos tested for each type of forgery detection is given in the tables of the sub-titles at below.

4.2.1 Implementation of Frame-Deletion Forgery Detection

Table 4.3 contains information about the original and forged video used for frame-deletion forgery detection. At the same time, in the forged video, the location and amount of the frame sequence in which the attack was made are shown.

Table 4.3 Information of example videos from VIFFD-A Dataset [33]

Features	**Original Video**	**Forged Video**
Name of Video	salman exercise1_1	delete75-100 salman exercise1_1
Number of Frame	206	181
Frame Rate	25.00 f/sec	25.00 f/sec
Resolution	720x404	720x404
Format	mp4	mp4
Length	8 sec.	7 sec.
Types of Forgery	x	deletion
Deleted Frames	x	76th – 100th (total 25 frames)

GOP is created by extracting all frames of the tested video. It has been already discussed in Chapter 3 that all videos consist of pictures. Therefore, firstly RGB-level images are obtained for the video as shown in Figure 4.1.

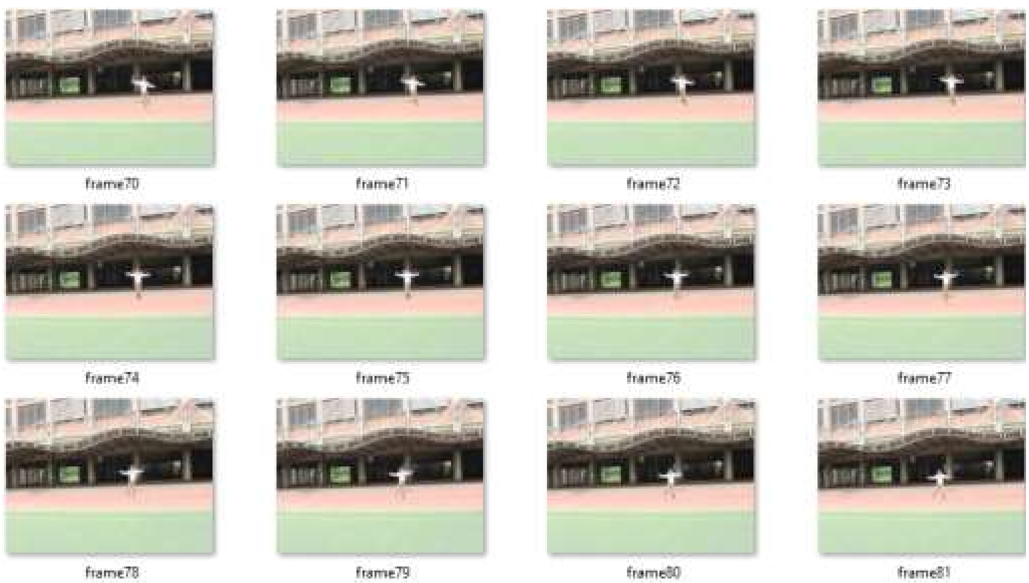


Figure 4.1 RGB frames of original video

The high amount of frames in the processing of videos creates some disadvantages in terms of both storage space and processing speed. Therefore, all frames are converted to gray-GOP level, stored and processed. Figure 4.2 shows the gray forms of the original frames shown in Figure 4.1.

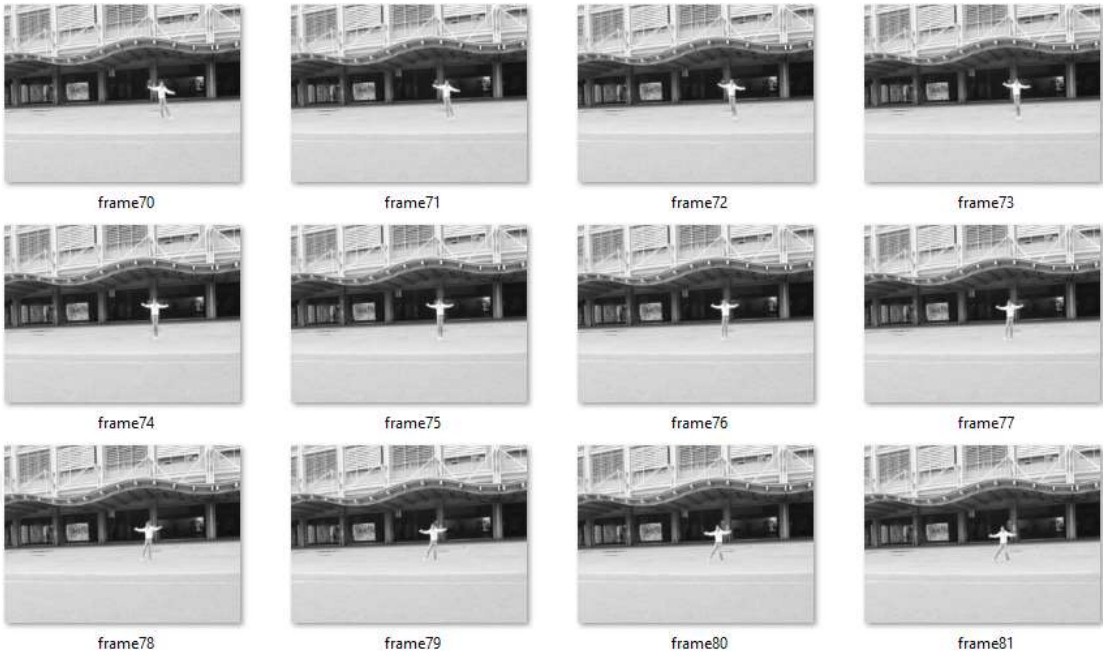


Figure 4.2 Gray frames of original video

The process of extracting frames from the video are similarly done for the forged video. Representation of a group RGB frame is shown in Figure 4.3.

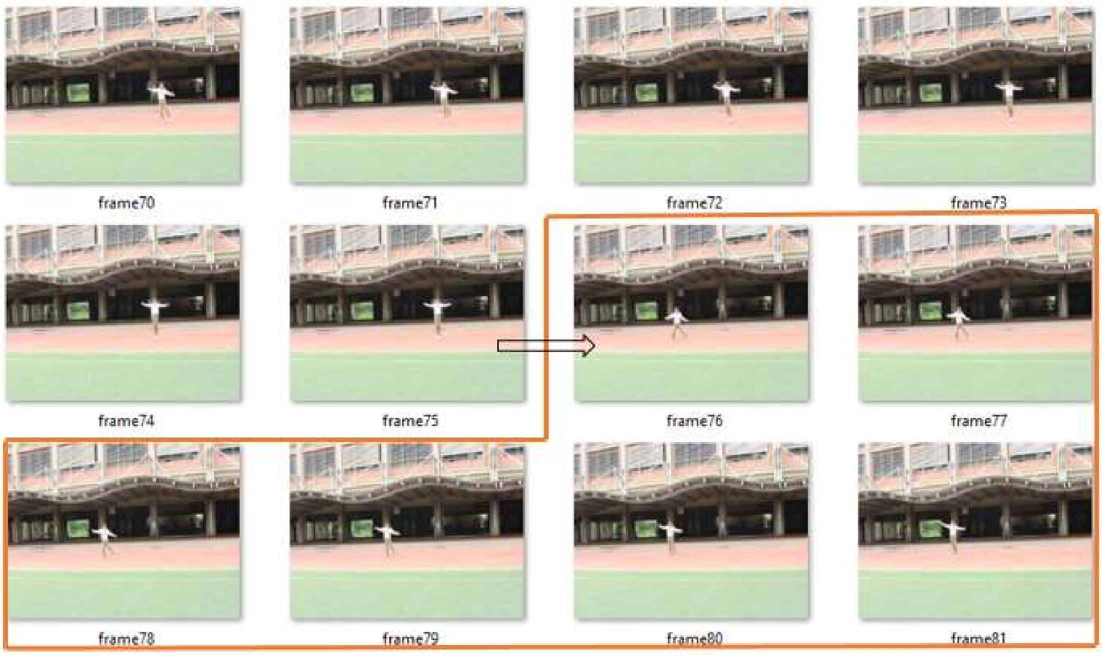


Figure 4.3 RGB frames of forged video

Again for the same purpose, the conversion of RGB frames to gray frames is applied which is illustrated in Figure 4.1.

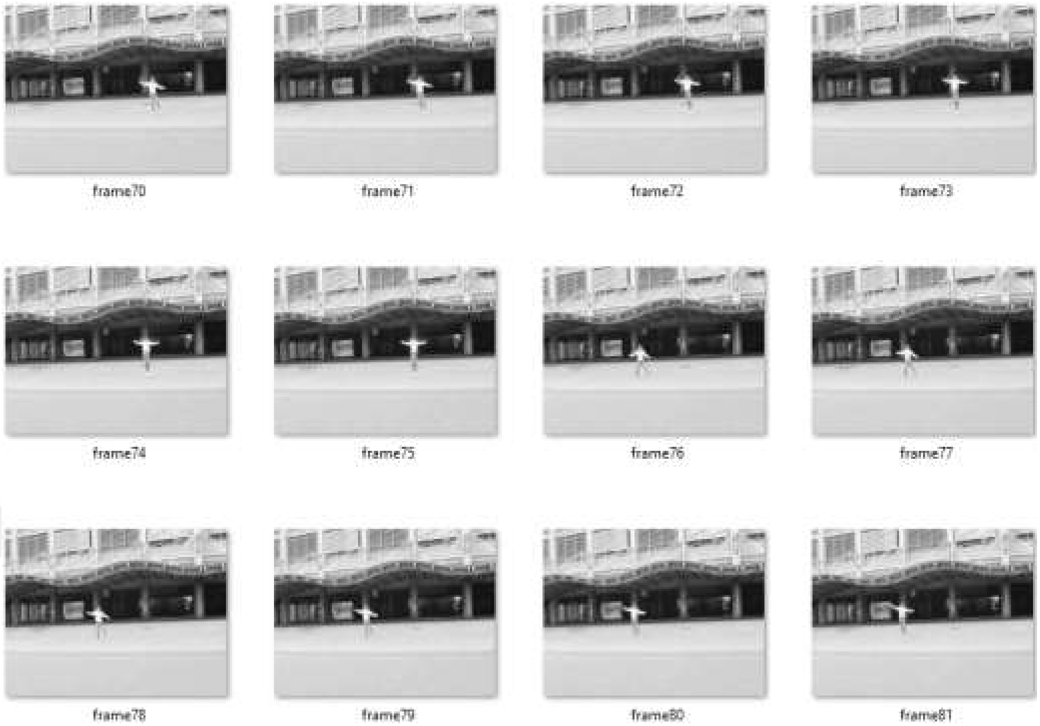


Figure 4.4 Gray frames of forged video

Total of 25 frames between 76th-100th frames are deleted from the original video as shown in Figure 4.5. 101st frame in the original video matched 76th frame in the forged video. After deletion processing, a forged video is produced by re-compressing.

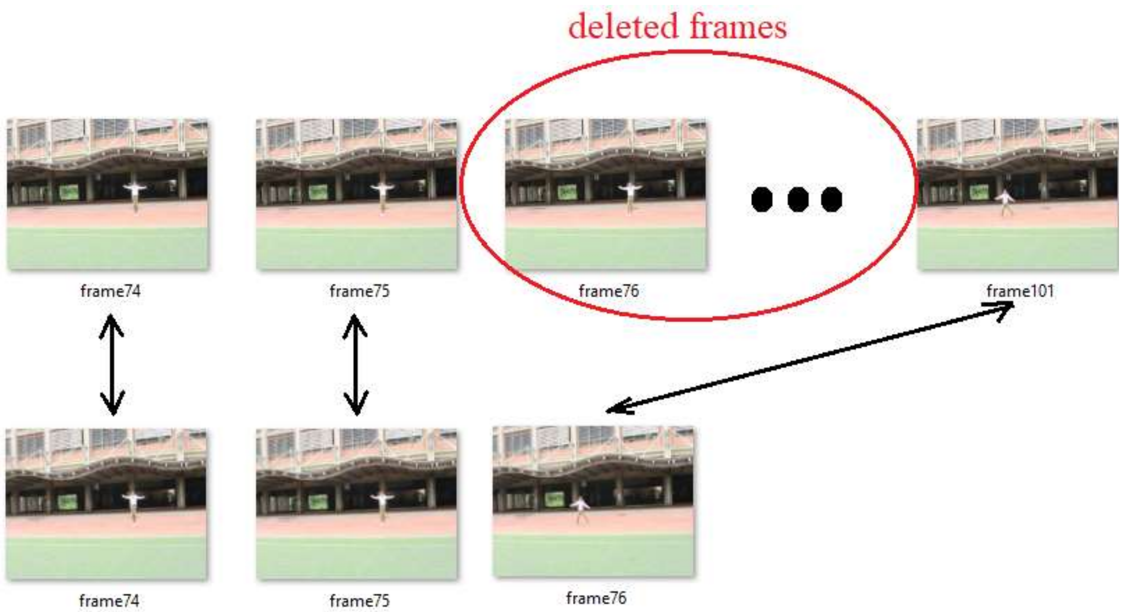
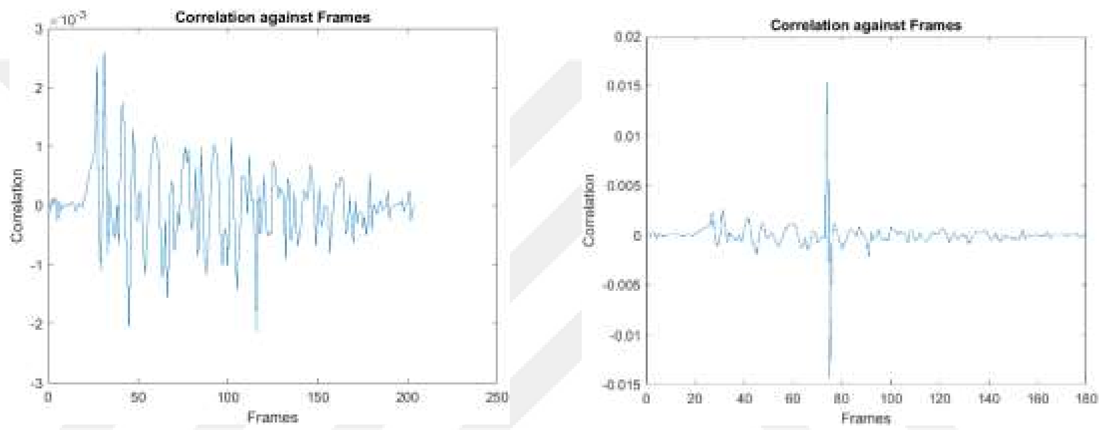


Figure 4.5 Location of frame deletion forgery

In Figure 4.6 (a), it is observed that the graph of correlation coefficients for original video is very smooth. (The graph of original video is approximately 6.5 times larger than the graph of forged video. If the graph of original video were displayed in the same range as the graph of forged video, it would be expected to observe the graph of original as close to a straight line.) The inter-frame tampering in the forged video started after 76th frame as shown in Figure 4.6 (b). In the graph obtained by calculating the deviation in the correlation coefficient difference of the adjacent frames, a great change is observed after the 76th frame on forged video.



(a) (b)

Figure 4.6 (a) and (b) Representations of the correlation coefficient for the original video and the forged video, respectively.

In inter-frame deletion forgery, there is only at one point change in the video stream. So the correlation coefficient graph peaks at one point in the inter-frame forgery. Since the pixel values of the frame will change with an object entering the scene in the video stream, the correlation coefficient graphic shows a sudden change and this change fades over time as shown in Figure 4.6 (a). In the original video, when person entered the scene, the graphic line increased because of this and then disappeared over time with the deviation in the correlation coefficient difference.

4.2.2 Implementation of Frame-Shuffling Forgery Detection

A video sample from the SULFA dataset has been tampered with at the inter-frame level. Forged video has shuffling tampering. Table 4.4 contains information about the original and forged video used for frame-shuffling forgery detection. At the same time, the location and amount of the frame sequence which the attack was made are shown in the forged video.

Table 4.4 Information of example videos from SULFA Dataset [5]

Features	**Original Video**	**Forged Video**
Name of Video	can_220_flap(2)	can_220_flap(2) forged
Number of Frame	332	332
Frame Rate	28.97 f/sec	30.00 f/sec
Resolution	320x240	320x240
Format	mov	mp4
Length	11	11
Types of Forgery	x	shuffling
Shuffled Frames	x	75th – 273th (total 99 frames)

The frames in Figure 4.7 are frames of the original video. For frame shuffling detection, firstly the original video is examined. For this, firstly, the frames are extracted from the video and recorded at RGB level.

Figure 4.7 RGB frames of original video

Then the frames of the same original video were converted to gray level and stored which is illustrated in Figure 4.8.

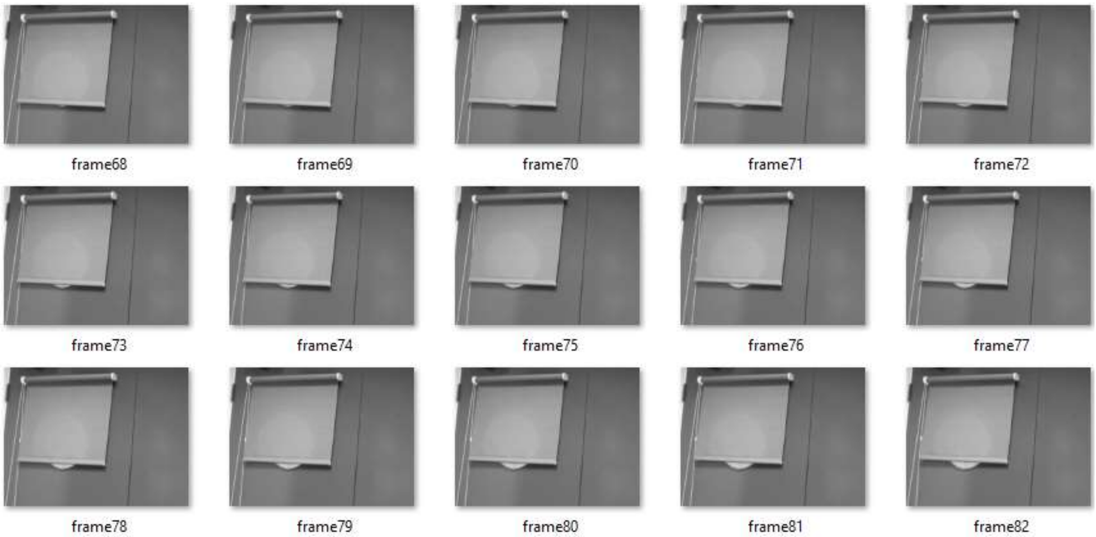


Figure 4.8 Gray frames of original video

The process of extracting frames from the video is similarly done for the forged video as shown in Figure 4.9. It is seen that the tampering started at the 75th frame.

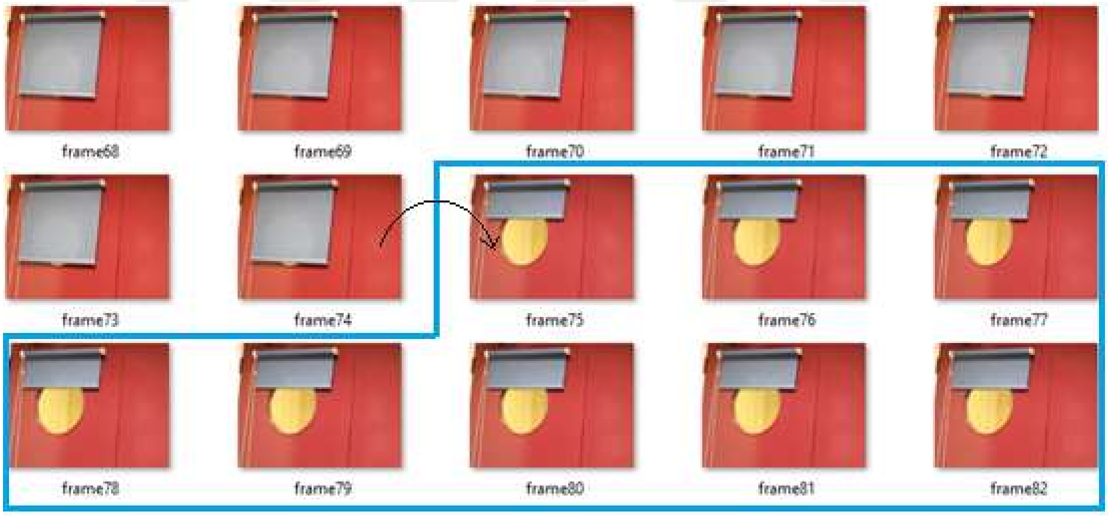


Figure 4.9 RGB frames of forged video

All frames are converted to gray level and Figure 4.10 shows the location where the forgery started and adjacent frames.

Figure 4.10 Gray frames of forged video

It is seen that there is a difference in the 75th and 273rd frames in Figure 4.11, compared to the original video. The 75th frame shows the frames where the forgery started and the 273rd frame shows the frames where the forgery ended. A total of 99 frames between these two frames were copied, reversed and pasted again. A forged video was created by recompression.

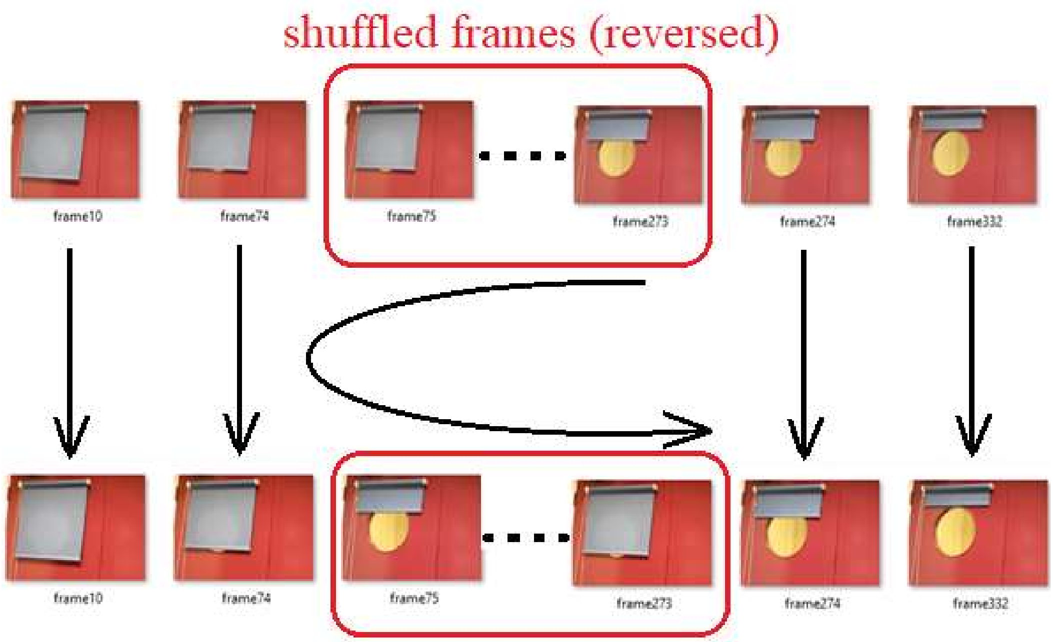
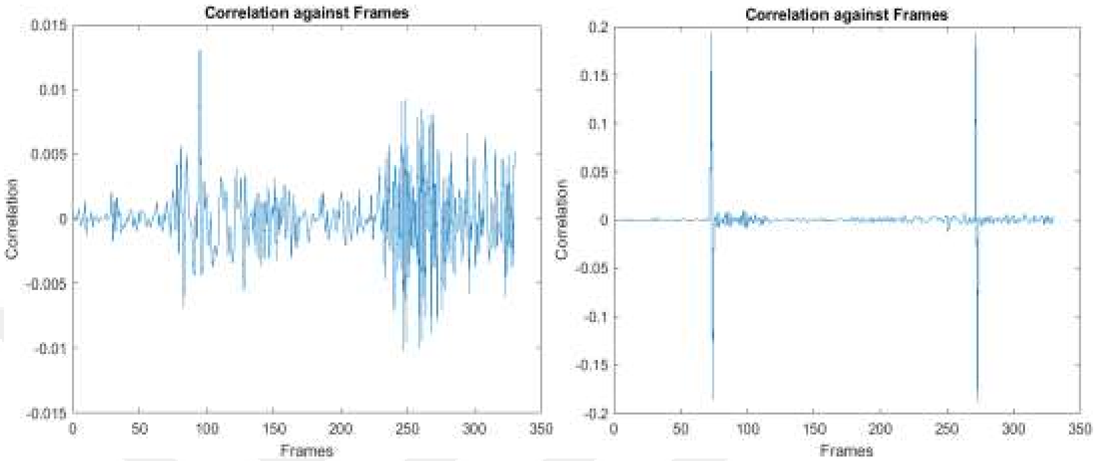


Figure 4.11 Location of frame shuffling forgery

The correlation coefficient results have a smooth graphic for the original video as shown in Figure 4.12 (a). But in the inter-frame deletion forgery, the graphic of the forged video peaked as shown in Figure 4.12 (b) and these peaks occurred at two different locations as expected. This confirms that it is shuffling forgery for inter-frame level.



(a) (b)

Figure 4.12 (a) and (b) Representations of the correlation coefficient for the original video and the forged video, respectively.

4.2.3 Implementation of Frame-Duplication Forgery Detection

Table 4.5 contains information about the original and forged video used for frame-duplication forgery detection. At the same time, in the forged video, the location and amount of the frame sequence which the attack was made are shown.

Table 4.5 Information of examples videos from SULFA Dataset [5]

Features	**Original Video**	**Forged Video**
Name of Video	fuji_2800_man (2)	fuji_2800_man (2)-110-160-20-forged-compressed
Number of Frame	330	380
Frame Rate	30.00 f/sec	30.00
Resolution	320x240	320x240
Format	avi	mp4
Length	11	12
Types of Forgery	x	Duplication
Duplicated Frames	x	21th – 70th (total 50 frames)

The frames of the original video to be tested are extracted from the video in RGB level. Some of the 330 frames of the whole video are shown in Figure 4.13.

Figure 4.13 RGB frames of original video

All RGB frames are converted to gray-GOP level because of obtaining advantages on both storage space and processing speed as shown in Figure 4.14.

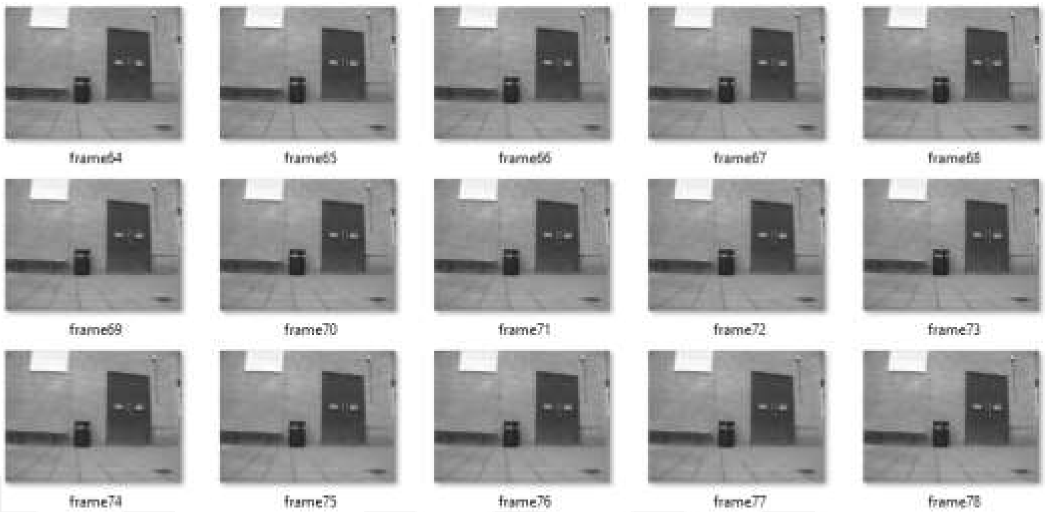


Figure 4.14 Gray frames of original video

Frame extraction from video is done in a similar way for forged video. The critical location where the forgery occurred is shown in Figure 4.15. After the 70th frame, the man in the video suddenly disappears.

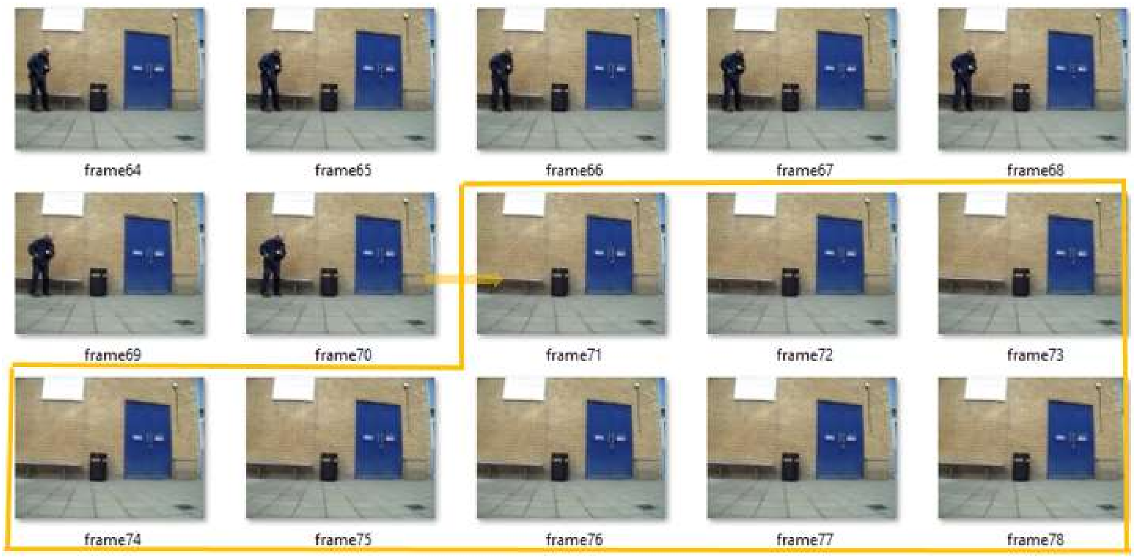


Figure 4.15 RGB frames of forged video

For easy calculation of correlation coefficient, gray level is used instead of RGB level, and therefore RGB frames are converted to gray frames which is illustrated in Figure 4.16.

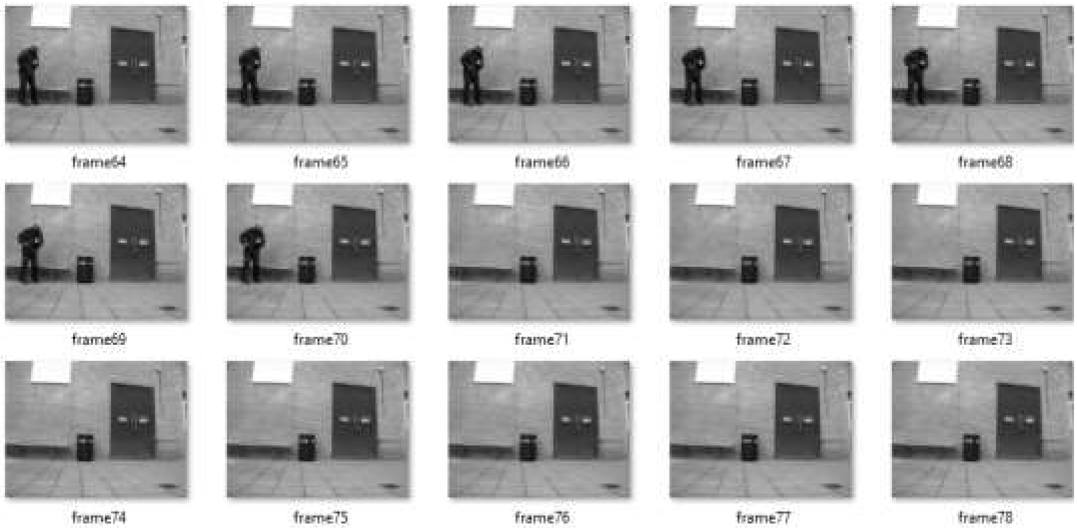
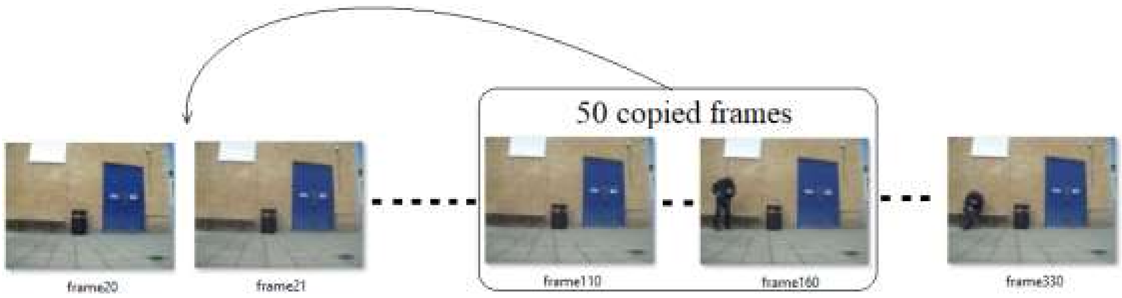
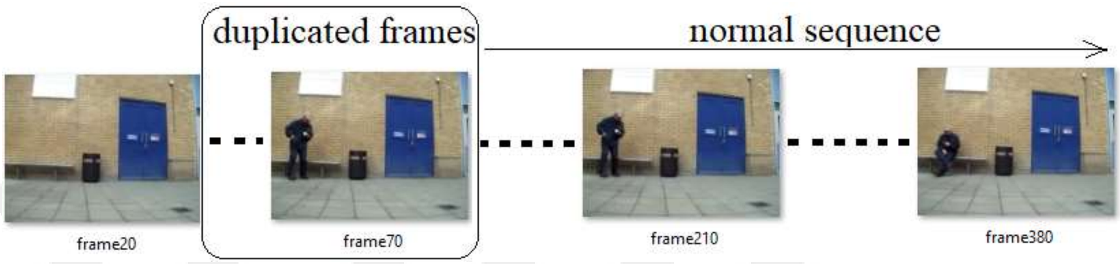


Figure 4.16 Gray frames of forged video

A total of 50 frames between 110th-160th frames are copied from the original video as shown in Figure 4.17 (a). After the 20th frame in the original video, the copied 50 frames are pasted and compressed again, resulting in a forged video as shown in Figure 4.17 (b). Although the man in the original video first appears on stage at 112nd frame, he is twice seen firstly in 22nd frame and secondly in 162nd frame-50 frames ahead- in the forged video. That is, in the forged video, 50 frames are increased and after the 20th frame, each frame number is shifted forward by 50.



(a)

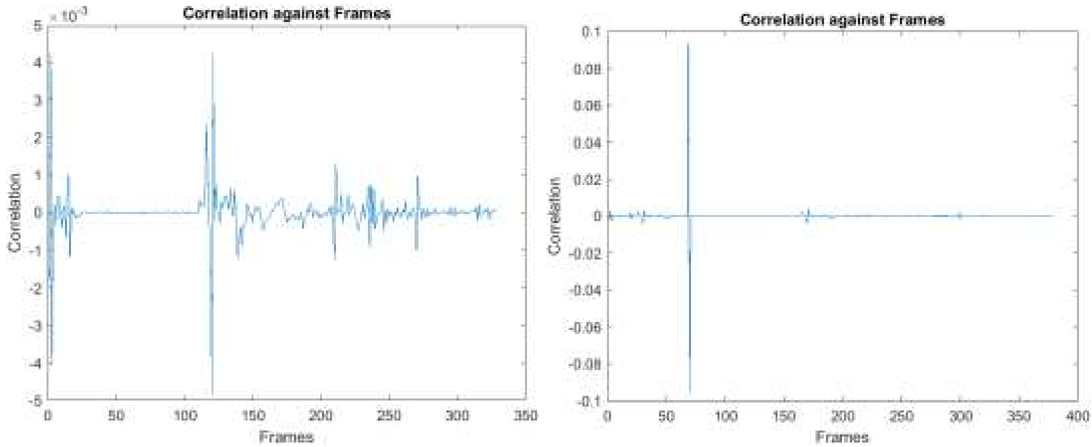


(b)

Figure 4.17 Location of frame duplication forgery

If the pixel values change with a sudden change in the frame level in the video stream, this change is also observed in the correlation coefficient value and the graphic. In Figure 4.18 (a), when the man enters the scene at approximately 112nd frame, this change is immediately visible in the original video's correlation coefficient graph.

Likewise, when the 50 frames copied in the forged video are pasted after the 20th frame, the man is first seen at the 22nd frame. But, the reason why there is no significant change in the correlation coefficient of the forged video at this point is because the correlation coefficient graph of the forged video is reduced 20 times compared to that of the original video. In Figure 4.18 (b), the most obvious change is seen in the 70th frame. The reason for this is that the forged frames end at frame 70 and the man suddenly disappears from the scene.



(a) (b)

Figure 4.18 (a) and (b) Representations of the correlation coefficient for the original video and the forged video, respectively.

CHAPTER 5

EXPERIMENTAL RESULTS AND DISCUSSION

5.1 Performance Analysis

5.1.1 Evaluation Metrics

Evaluation criteria used for the performance of the proposed method and the developed algorithm are shown in Equations (5.1)-(5.4).

$$A = \frac{TP + TN}{TP + FP + FN + TN}\ 100\% \tag{5.1}$$

$$\mathrm{P} = \frac{TP}{TP + FP} \tag{5.2}$$

$$\mathrm{R} = \frac{TP}{TP + FN} \tag{5.3}$$

$$\mathrm{F\,1} = \frac{2 \times R \times P}{R + P} \tag{5.4}$$

where A is accuracy, P is precision, R is recall and F1 is F1 score. At the same time, TP is the true positive number such that the forged video is detected as forged, TN is the true negative number such that the original video is detected as original, FP is the number of false positive such that the original video is detected as forged and FN is the number of false negatives such that forged videos is detected as the original.

5.1.2 Experimental Results

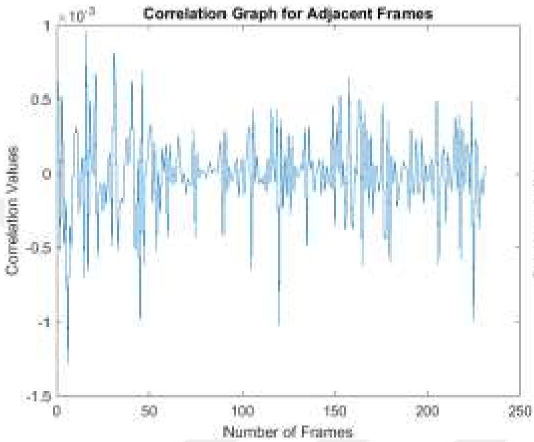
The tests/results are carried out using the five datasets discussed under the dataset design section of this study. All properties of 8 videos randomly selected from these datasets as forgery and original in pairs are shown in Table 5.1. The selected videos are examples of 4 types of tampering focused on the inter-frame level. These binary videos are examples of frame insertion, frame deletion, frame shuffling and frame duplication types at the interframe forgery level.

The correlation results obtained with the application of the proposed algorithm are analyzed in Figure 5.1.

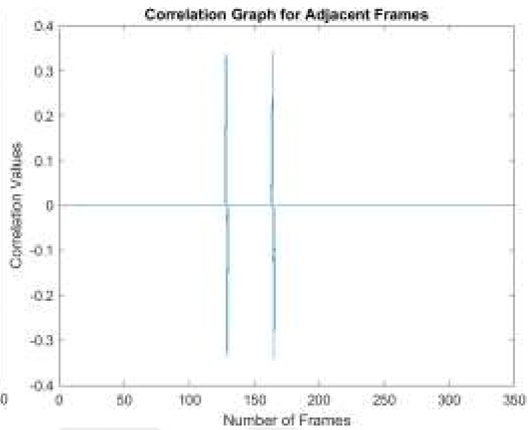
Table 5.1 Range and change of correlation coefficients according to video content.

Videos	Number of Frames	Length (sec)	Correlation Range	Location of SD-Correlation	Situation of Video	Figure (5.1)
Video-1	234	7	$-1.5\text{-}1x10^{-3}$	x	original	(a)
Video-2	349	11	-0.4-0.4	130th and165th	insertion	(b)
Video-3	206	8	$-3\text{-}3x10^{-3}$	x	original	(c)
Video-4	181	7	-0.025-0.02	75st	deletion	(d)
Video-5	332	11	$-1.5\text{-}1.5x10^{-3}$	x	original	(e)
Video-6	332	11	-0.2-0.2	75th and 274th	shuffling	(f)
Video-7	326	9	$-4\text{-}4x10^{-4}$	x	original	(g)
Video-8	326	10	-0.08-0.08	30th and 70th	duplication	(h)

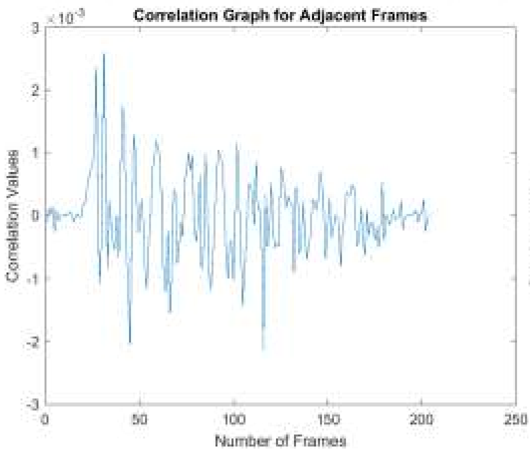
It contains graphs of correlation coefficients of different original videos and different types of forged videos in Figure 5.1. While the correlation coefficient range for the original video has very small values, for forged videos this value is consistently very large. Also, in forged videos, it is observed that there is a peak point once for deletion while it occurs twice for insertion, duplication, and shuffling videos. This is because when there is a change in deletion, there is a change in the beginning and the end of the tampered frames in the other types.



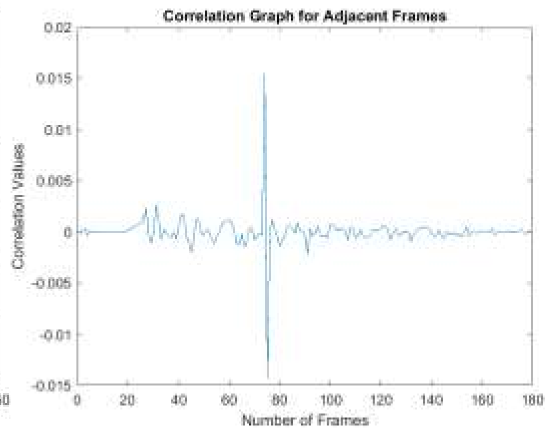
(a) Original video



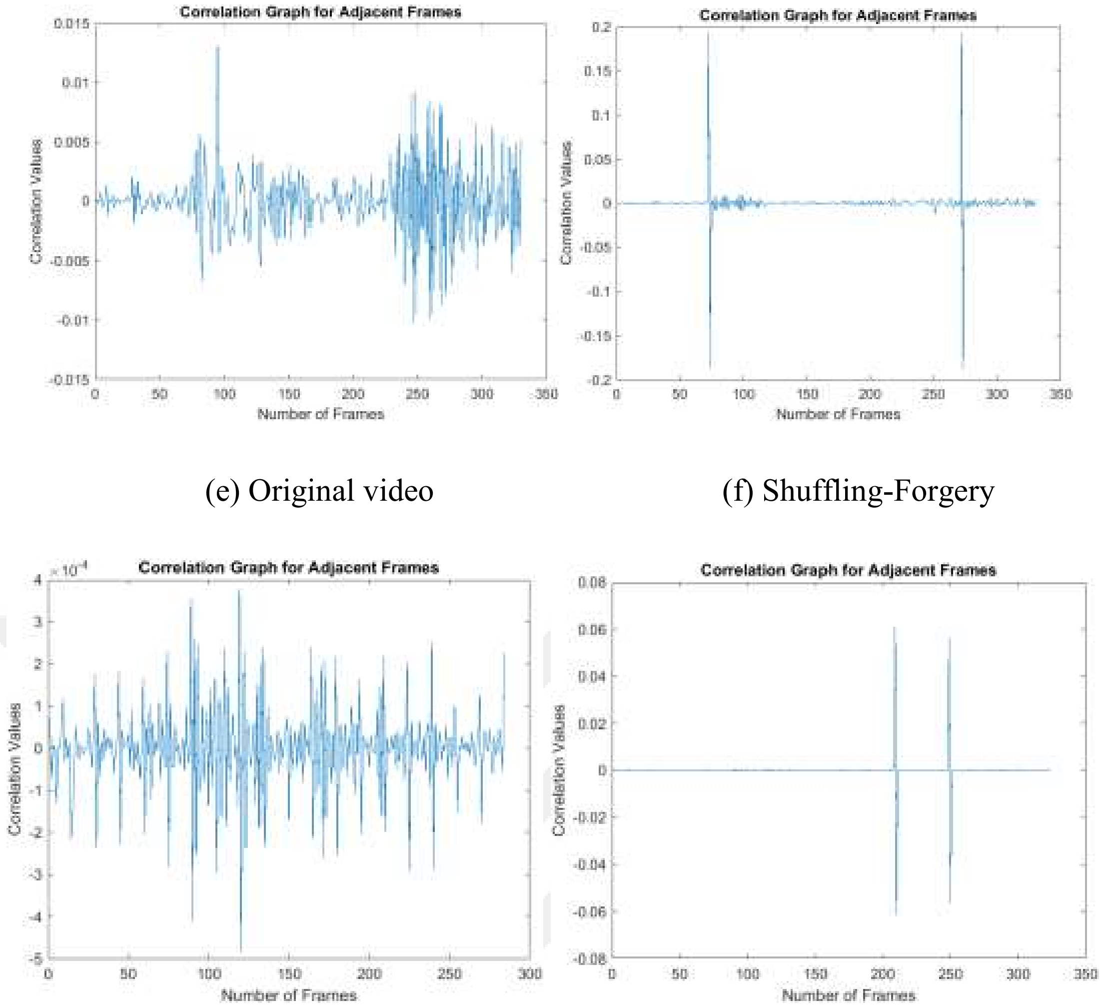
(b) Insertion-Forgery



(c) Original video



(d) Deletion-Forgery



(e) Original video (f) Shuffling-Forgery

(g) Original video (h) Duplication-Forgery

Figure 5.1 Graphs of correlation coefficients.

5.1.3 Detection Performance

The effectiveness and efficiency of the proposed method and algorithm for detecting video forgeries were evaluated. The proposed method for video forgery detection was developed by using the correlation coefficient differences. The effectiveness of the method was examined by using more than 5 types of datasets. In addition, the experimental results obtained with 120 different videos are shown in detail in Table 5.2.

Table 5.2 Detection performance of the proposed method

Dataset	Forged Videos		Original Videos		Accuracy	Precision	Recall	F1 Score
	TP	FN	TN	FP				
SULFA [5]	26		20		100%	1.00	1.00	1.00
	26	0	20	0				
PU [8]	4		6		100%	1.00	1.00	1.00
	4	0	6	0				
VIFFD- A [33]	14		20		97.06%	1.00	0.93	0.96
	13	1	20	0				
Rewind [6]	2		4		100%	1.00	1.00	1.00
	2	0	4	0				
Various [4]	14		10		95.83%	1.00	0.93	0.96
	13	1	10	0				
Total Dataset	60		60		98.33%	1.00	0.96	0.98
	58	2	60	0				

The table contains videos that are compatible with forgery at the inter-frame level. Therefore, only videos compatible with inter-frame forgery were taken from each dataset. From Table 5.2, it is observed that the proposed method has an accuracy of 98.33% which shows that the developed method is very effective in detecting inter-frame video forgery.

5.2 Discussion

In order to show the efficiency of the method and how effective it is, at least 5 types of datasets are used, not dependent on only one dataset video. In addition, a total of 120 videos were tested in these datasets and their results are given in Table 5.2**.** This table contains the experimental results of the proposed algorithm that detects whether a video is forged or not. All of the videos from each dataset focus on the types of forgery at the inter-frame level (frame deletion forgery, frame insertion forgery, frame shuffling forgery, frame duplication forgery).

It has been observed that the accuracy values obtained according to the test results of all datasets vary between 95.83% and 100%. In addition, the total test result of all videos was found to be 98.33%. In other words, when all 60 original videos were tested with the proposed algorithm, they were identified as the original video. Similarly, when the other 60 forged videos were tested, only 2 of them were wrongly evaluated and 118 of the 120 videos were correctly identified.

It has been seen that the accuracy values obtained both in the results obtained separately of the datasets and as a result of the total test have a higher accuracy value than the studies mentioned in Table 2.1.

Moreover, according to the studies described in the same section, the amount of used video is quite high, which shows how powerful and effective the proposed method is. In addition, previous similar studies did not emphasize the point of increasing the speed of the tests.

However, in our method, the conversion of RGB level pictures to gray level pictures is achieved, thus gaining both speed and storage space.

CHAPTER 6

CONCLUSION AND FUTURE WORK

With the developing technology and rapid digitalization, the use of digital materials is becoming widespread as access to technology becomes easier. The center of this situation is the use of digital videos. Another factor in the spread of digital videos is the trust in them. However, changing video content has also become increasingly common. The proposed method is to detect inter-frame forgery, which is the most common of these tamperings. In other words, the focus of the method is to detect whether a video has been tampered with in the area of inter-frame forgery. In the method, frames of the video, gray level transformations and correlation coefficient for each frame are used. By using the gray level of each frame, acceleration is achieved by reducing the need for storage area, unlike RGB. In addition, for the precision of detection, 3sigma was integrated into the correlation coefficient, increasing the success value and achieving a performance of over 98%.

In future studies, some new distinguishing features may be found to identify forgery and it can be determined what types of forgery at the inter-frame level can be. Also, the dataset can be enriched with different videos together with the videos used.

REFERENCES

[1] Chen, W., Shi, Y. Q. (2008, November). Detection of double MPEG compression based on first digit statistics. In *International Workshop on Digital Watermarking* (pp. 16-30). Springer, Berlin, Heidelberg.

[2] Singh, R. D., Aggarwal, N. (2018). Video content authentication techniques: a comprehensive survey. *Multimedia Systems*, **24(2),** 211-240.

[3] Sitara, K., Mehtre, B. M. (2016). Digital video tampering detection: An overview of passive techniques. *Digital Investigation*, **18**, 8-22.

[4] Other references. Various Dataset; IVYLAB Dataset, LASIESTA Dataset etc. Intl.: http://www.gti.ssr.upm.es/data/lasiesta_database.html, 13.05.2019.

[5] SULFA Dataset; Surrey University Library for Forensic Analysis. Intl.: http://sulfa.cs.surrey.ac.uk/, 28.07.2019.

[6] Bestagini, P., Milani, S., Tagliasacchi, M., Tubaro, S. (2013, September). Local tampering detection in video sequences. In *2013 IEEE 15th international workshop on multimedia signal processing (MMSP)* (pp. 488-493). IEEE.

[7] Patwardhan, K. A. Video Inpainting. Intl.: http://kedarpatwardhan.org/Research/VideoInpainting.html, 14.02.2020.

[8] Panjab University. Panjab University Dataset (PU Dataset); Intl.: https://1drv.ms/f/s!Aj8xYEfdOdJ-i0xnQ5YSlUpZl-HrT, 05.10.2019.

[9] Richao, C., Gaobo, Y., Ningbo, Z. (2014). Detection of object-based manipulation by the statistical features of object contour. *Forensic science international*, **236**, 164-169.

[10] Wahab, A. W. A., Bagiwa, M. A., Idris, M. Y. I., Khan, S., Razak, Z., Ariffin, M. R. K. (2014, November). Passive video forgery detection techniques: a survey. In *2014 10th International Conference on Information Assurance and Security* (pp. 29-34). IEEE.

[11] Lin, C. Y. (2001). *Watermarking and digital signature techniques for multimedia authentication and copyright protection*. Columbia University.

[12] Tang, Y. L., Hung, C. J. (2005). Recoverable Authentication of Wavelet-Transformed Images. *ICGST International Journal on Graphics, Vision and Image Processing*, **11**, 61-66.

[13] Hussein, J., Mohammed, A. (2009). Robust video watermarking using multi-band wavelet transform. *arXiv preprint arXiv:0912.1826*.

[14] Kingra, S., Aggarwal, N., Singh, R. D. (2016). Video Inter-frame forgery detection: A survey. *Indian Journal of Science and Technology*, **9(44)**.

[15] Thaipanich, T., Wu, P. H., Kuo, C. C. J. (2009, January). Robust video frame rate up-conversion (FRUC) techniques. In *2009 Digest of Technical Papers International Conference on Consumer Electronics* (pp. 1-2). IEEE.

[16] Wang, W., Farid, H. (2007, September). Exposing digital forgeries in video by detecting duplication. In *Proceedings of the 9th workshop on Multimedia & security* (pp. 35-42).

[17] Su, Y., Zhang, J., Liu, J. (2009, December). Exposing digital video forgery by detecting motion-compensated edge artifact. In *2009 International Conference on Computational Intelligence and Software Engineering* (pp. 1-4). IEEE.

[18] Zhang, J., Su, Y., Zhang, M. (2009, October). Exposing digital video forgery by ghost shadow artifact. In *Proceedings of the First ACM workshop on Multimedia in forensics* (pp. 49-54).

[19] Chetty, G., Biswas, M., Singh, R. (2010, September). Digital video tamper detection based on multimodal fusion of residue features. In *2010 Fourth International Conference on Network and System Security* (pp. 606-613). IEEE.

[20] Bestagini, P., Milani, S., Tagliasacchi, M., Tubaro, S. (2013, September). Local tampering detection in video sequences. In *2013 IEEE 15th international workshop on multimedia signal processing (MMSP)* (pp. 488-493). IEEE.

[21] Wang, Q., Li, Z., Zhang, Z., Ma, Q. (2014). Video Inter-Frame Forgery Identification Based on Consistency of Correlation Coefficients of Gray Values. *Journal of Computer and Communications*, **2(04)**, 51.

[22] Correlation Coefficient. Simple Definition, Formula, Easy Steps. Intl.: https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/#definition, 28.09.2019.

[23] Huang, TQ., Chen, Z. W., Su, L. C., Zheng, Z., Yuan, X. J. (2011) Digital Video Forgeries Detection Based on Content Continuity. *Journal of Nanjing University (Natural Science),* **47**, 493-503.

[24] Businessemt. (2021). Intl.: https://tr.businessemt.com/39-how-5214886-calculate-sigmal-66186, 10.07.2021.

[25] Hekimoglu, S., Koch, K. R. (2000). "How can reliability of the test for outliers be measured?" *Allgemeine Vermessungsnachrichten*, **7**, 247–253.

[26] Smirnov, N. V., Dunin-Barkovskiĭ, I. V. (1969). *Mathematische Statistik in der Technik: Kurzer Lehrgang* (Vol. 9). Deutscher Verlag der Wissenschaften.

[27] MultiCam 6.7 Direct Show Reference. GOP Structure. Intl.: https://documentation.euresys.com/Products/MultiCam/MultiCam_6_16/Content/MultiCam_6_7_HTML_Documentation/26808.htm, 14.02.2020.

[28] AfterDawn. GOP Settings. Intl.: https://www.afterdawn.com/guides/archive/tmpgenc_xpress_4-mpeg-_and_mpeg-1_video_settings_page_4.cfm, 23.06.2019.

[29] Wang, W., Farid, H. (2006). "Exposing digital forgeries in video by detecting double MPEG compression", *MM&Sec.*

[30] AWS Media Blog. Back to basics GOPs explained. Intl.: https://aws.amazon.com/tr/blogs/media/part-1-back-to-basics-gops-explained/, 26.08.2019.

[31] Meng, H.H. (2013). "Digital Video Forensics Detecting MPEG-2 Video Tampering through Motion Errors", *Technical Report RHUL–MA–2013–5*.

[32] Huang, TQ., Chen, Z.W., Su, L. C., Zheng, Z., Yuan, X. J. (2011). Digital Video Forgeries Detection Based on Content Continuity. *Journal of Nanjing University (Natural Science),* **47**, 493-503.

[33] Nguyen, X. H., Hu, Y. (2020). "VIFFD - A dataset for detecting video inter-frame forgeries", *Mendeley Data,* V**6**, doi: 10.17632/r3ss3v53sj.6.

[34] Gilat, A. (2016). *MATLAB: an introduction with applications,* 6th edn, Wiley-IEEE Press.

[35] Gonzalez, R. C., Woods, R. E., Eddins, S. L. (2009). *Digital image processing using MATLAB,* 2nd edn, Gatesmark Publishing.

[36] Solomon, C., Breckon, T. (2011). *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons Press.

[37] Marques, O. (2011). *Practical image and video processing using MATLAB*. John Wiley & Sons Press.

CV

KİŞİSEL BİLGİLER

Adı Soyadı : MUHARREM BALCI

EĞİTİM BİLGİLERİ

Yüksek Lisans : (2019-2021), Gaziantep Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik- Elektronik Mühendisliği, YL (Tezli)

Lisans : (2006-2011), Gaziantep Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği

Lise : (2001-2005), Mustafa Rakım Anadolu Lisesi (ORDU)

AKADEMİK/DENEYİM BİLGİLERİ

Gümüşhane Üniversitesi : Öğretim Görevlisi, Gümüşhane MYO, (2014-...)

Gaziantep Üniversitesi : Sözleşmeli Öğretim Görevlisi, Gaziantep Teknik Bilimler Meslek Yüksekokulu (2011- 2013)

İdari Görevler : Bölüm Başkan Yardımcılığı (2016-2018)
Bölüm Başkanlığı (2018- 2021)
Uzaktan Eğitim Bölüm Koordinatörlüğü (2020- …)
İş Sağlığı ve Güvenliği Eğitim Kurulu Üyesi ve Eğitimci

BİLDİRİLER

Balcı Muharrem, Özbay Serkan. Inter-Frame Digital Video Forgery Detection Based on Correlation Difference and 3Sigma

Yayın Yeri: II. International Science and Innovation Congress (INSI 2021), İngilizce, 25.07.2021