Tampering Detection in Digital Video - A Review of Temporal Fingerprints Based Techniques

Vaishali Joshi
BVICAM, New Delhi, INDIA
Email Id: vaishali.joshi22@gmail.com

Sanjay Jain
School of Computer Application, ITM University,
Gwalior, INDIA
Email Id: sanjayjainitm@gmail.com

Abstract - With the availability of advanced electronic gadgets like high-quality digital video camera, cell phone etc. there is an explosion of digital videos. When coupled with sophisticated digital video processing techniques there is a good increase in count and quality of tampered video. Video forensic is comparatively a new research domain which comes with many threats as well as many opportunities. This paper presents classification of video authentication techniques and their fundamental features, developed in last few years for tempering detection in digital videos.

Keywords - Tampering detection, video forensic, temporal fingerprints, digital forensic

I. INTRODUCTION

Video forensic is still a young field in forensic science. Forgery creation and detection are complimentary to each other and go hand in hand with each other. The understanding and knowledge of forgery creation promotes sophistication in forgery detection techniques and visa versa.

The automatic and scientific technique for detecting tempered videos has become the biggest challenge for researchers. The research in this domain is ultimately shaping towards finding more generalized solutions and techniques, building standardized data sets, bench marking the techniques, evaluation criteria etc. so as to minimize these forgeries. A perfect video authentication system must carry the properties like vulnerability to malicious alteration, robustness to regular video processing operations, localization and self recovery of altered regions, tolerance against loss of information and computational feasibility. The basic process of video forensic system is to prove if the given video is maliciously altered or not but it might be more significant if the system can tell where and when the modification is made.

Temporal fingerprints based Video forensic techniques make available information regarding digital video contents without relying on exterior descriptors such as metadata tags or extrinsically inserted information such as digital watermarks both fragile and semi-fragile. Alternatively, these techniques make use of fingerprints left in digital video content itself by editing operations or the digital video capturing process [1].

When is video is captured by a video recording device, it records the scene in front of lens of capturing device, frame by frame. Thus a digital video sequence can be considered as a collection of successive frames with temporal dependency, in a 3-D plane. When tampering is carried out on a video sequence, it either attacks on contents of video or on the temporal dependency between the frames. So, based on regional property video sequences, video tampering attacks can be broadly classified in three categories: spatial tampering attacks, temporal tampering attacks and spatio-temporal attacks [2].

II. CLASSIFICATION OF AUTHENTICATION TECHNIQUES

Authors in [3] have classified video authentication techniques in four categories: Watermark based technique, Digital Signature based techniques, machine learning based intelligent techniques and other authentication techniques based on use of inbuilt video features. Among these techniques, digital signature and fragile watermarking techniques are commonly used for video authentication.

If we compare these categories, each of them has some pros and cons. And each one has an application area related to type of video sequences. The main shortcoming of the digital watermarking technique is that a watermark must be embedded at exactly the time of capturing video, which confines this approach to specially designed digital cameras. Furthermore, embedding a watermark may alter video contents and which is not acceptable in legal proceedings.

Video authentication techniques based on digital signatures offer improved results about robustness, since the digital signature remains unaffected when there is a change in pixel values of the frames in a video sequence. But if the location where digital signature is stored is compromised then it is easy to mislead the authentication system. The third category, intelligent techniques investigate the new scope in video authentication. This learning based intelligent tampering detection algorithm for video authentication does not require working out and storage of any key or embedding of secret information in the video data but it requires a large database of doctored and original video sequences to learn the algorithm so that it can classify whether the given video is authentic or not. Major drawback with the intelligent techniques is that, for even a single kind of attack, they need a sufficient large amount of databases of tampered and authentic video

978-9-3805-4416-8/15/$31.00 ©2015 IEEE

sequences to learn. That makes these techniques a little bit slower in comparison to other techniques.

III. BASICS OF DIGITAL VIDEO

Digital videos are stored in compressed form. Video compression techniques are used to reduce and remove redundancy in video sequences so that it can be stored easily on disks and can be effectively transferred on networks. Different compression techniques are available but most of the vendors use standard techniques to facilitate compatibility and interoperability. There are three most popular video compression standards. They are Motion JPEG, MPEG-4 and H.264. H.264/AVC is the most recent and most efficient digital video compression standard.

Any digital video is captured with a lot of redundancy among each the frames of video sequence. MPEG video compression technique makes the most of this redundancy by predicting some frames and encoding residual error between these predicted frames. The MPEG standard is developed to minimize both spatial redundancy within each video frame and temporal redundancy across all video frames. During MPEG video compression, to put a stop to error propagation, the video sequence is divided into fragments, where each fragment is referred to as a group of pictures (GOP). Frame prediction is done within each fragment, but never across fragments, thus preventing decoding errors in one frame from extending throughout the video sequence. Within each group of pictures (GOP), video frames are categorized into three types: intra-frames (I-frames), predicted-frames (P-frames), and bidirectional-frames (B-frames). Each Group of Picture starts with an I-frame followed by a number of P-frames and B-frames. No prediction is carried out when encoding I-frames therefore each I-frame is encoded and decoded independently using compression similar to JPEG compression. A P-frame can reference only preceding I- or P-frames, while a B-frame can reference both preceding and succeeding I- or P-frames. Video compression algorithms such as MPEG-4 and H.264 use inter frame prediction to reduce video data between a series of frames [11].

IV. NEED FOR ADAPTIVE GROUP OF PICTURE (GOP) STRUCTURE

Today digital resources are used in highly networked environment. Therefore, different video coding systems are used to reduce the bit rate required to transfer these digital resources over network. Most of the encoders use fixed group of pictures (GOP) to encode these digital resources mainly digital video sequences. Fixed GOP structures are easy to implement but they prevent capturing and coding devices from adapting to temporal variations in video sequences and so prevent them from improving their coding efficiency. A proper intra-frame is inserted to GOP structure, to avoid prediction error sequence to go beyond GOP and to provide the capabilities of random access on video streaming applications.

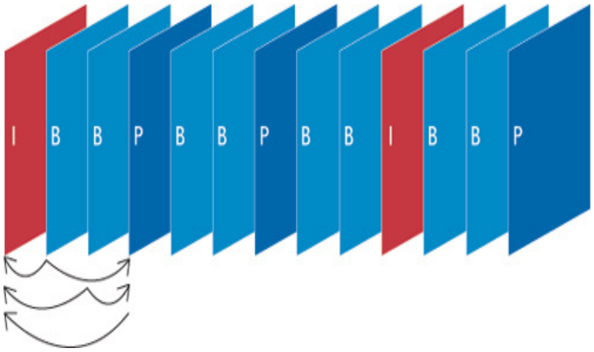

Fig. 1. A typical video sequence with I-, B- and P-frames

Normally if there is no scene change in a video sequence, the inter-coding can achieve better performance than the intra-coding due to motion compensation. Hence, an adaptive group of picture (AGOP) structure can adjust the length of GOP according to video contents, to reduce RD (rate–distortion) cost [14]. The AGS scheme adaptively changes the sizes of GOP structure as per the temporal characteristics of a video sequence to improve the coding efficiency. For Group of Picture (GOP) to be adaptive, temporal characteristics of video sequence need to be found out. There are two possibilities for this – Static GOP, when video sequence is known beforehand to the encoder, and Dynamic GOP when characteristics of video sequence is not known in advance and video sequence needs to be processed by encoder. In general, an I_frame uses more bits than a P_frame. However, if a scene changes too rapidly, using Intra mode is better than using Inter mode. It is clear that it is important to detect scene changes to determine which coding mode is to be used. Figure 2 shows Open and Closed GOP Structure used for Inter-frame prediction depending on temporal characteristics of video sequence. In an open GOP structure P- and B-frame can reference an outside I-or P-fame which belong to other Group of Picture (GOP). While in Close GOP, P- and B-frame can reference I-or P-fame belonging to the same Group of Picture (GOP).

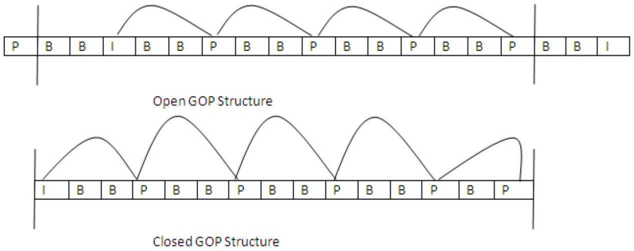
P B B I B B P B B P B B P B B P B B I
Open GOP Structure
I B B P B B P B B P B B P B P
Closed GOP Structure

Fig. 2. Open and Closed GOP Structure used for Inter-frame prediction

V. WORK DONE IN THE FIELD

Authors in [4] described the development of digital watermarking based video authentication algorithm which can deal with two tampering scenarios named cut-and-splice and cut-insert-splice. The algorithm provides watermark survivability even after video has gone through MPEG compression / decompression cycles. But it limits at false authentication of poor quality video and should be further extended for effective ambiguity resolution.

Authors in [5] describe a technique that can determine if a video was resaved after its original recording. But the algorithm cannot determine if it was tampered with or simply re-saved from within video editing software after harmless viewing.

There is another method for identifying double quantization in digital video which tells how double MPEG compression can introduce statistical artifacts that can be identified and calculated. This measured artifact can be used further to detect tampering in digital video. The technique works effectively when the ratio between earlier and later quantization scale is greater than 1.7 but accuracy decreases with decrease in this ration. Also count of false positive is small and spatially localized. The scope of this technique is limited to that it is effective only when second compression quality id greater than the previous one [6].

The authors in [7] describe two types of forensic schemes – active and passive. In active schemes tampered area can be pulled out using pre-inserted watermarks or keys used for digital signatures. On the other hand, passive schemes use intrinsic fingerprints to dig out tampered region. When a real world object is recorded using a digital camera, the information about image is processed by the imaging pipeline before final digital image is produced. Every component of this imaging pipeline may put down some intrinsic fingerprint hints in input image. These intrinsic fingerprints are then used to detect tampering. This technique attains high precision rate with small false positive results which are mainly due to improper illumination which may be too high or too low. But as the noise residue is content dependent, this technique may not be efficient for videos with dynamic scenes.

Authors in [8] used a geometric technique to detect forged videos. According to authors trajectories of objects in any video scene can be substantially influenced only by gravity. Any camera motion can be reliably estimated by calculating centre of mass of the object and background elements. Algorithm models the 3-D ballistic motion of the object in free flight and the 2-D projection of object trajectory in the image plane of a camera. Any deviation from this model gives indication of forgery. The main advantage of using geometric technique is that modeling and estimation of geometry is less sensitive to resolution of video. The work can be further extended to estimate the motion of cars, airplanes and other moving objects by finding type of acceleration experienced by these objects.

According to the authors of [9], there are two categories of video forgery detection algorithms. One is tampering detection algorithms that can detect a global manipulation. The other is tampering localization algorithms that aims to detect exactly where and when a video is doctored both in spatial as well as temporal domain. The algorithm proposed by them is purely unsupervised and is able to detect where a spatio-temporal region of sequence was replace by either a still image repeated many times or a region taken from an external sequences. Future work of them is to study the possibility of developing any anti-forensic techniques that can reduce the detector accuracy.

Authors in [12] explored different problems and challenges in the field of digital video forensic. Authors believed that the digital video forensic research domain has various threats and opportunities associated with it. The fundamental problems associated with the field are as follows:

- Whether the video sequence is naturally captured or manipulated later is a big challenge in order to identify the originality and authenticity of video sequence.
- Digital images in the frames of video sequence can be manipulated in many ways by applying simple transformations like translation, scaling, rotation, shear etc. Or by compensation and suppression operations like contrast adjustment, blurring filtering etc. In addition to these operations more complex operations like cropping blending are also possible. These operations may leave visually untraceable artifacts in an image. So it is very difficult to localize the forged region of video sequence.
- Now-a-days, in a highly networked environment it is very hard to track the origin and flow of digital resources over the network.
- Differentiation of forgery and genuine modification is becoming more and more complex with every passing day.
- Benchmarking, building standardized data sets and identifying criteria for evaluation of available video authentication techniques is a big challenge.

In their paper [13], authors proposed an intelligent video authentication technique based on machine learning. Authors took the method proposed in [5] as foundation and used MPEG double quantization technique to detect temporal fingerprints left in video sequence after addition or deletion of frame. Prediction error sequence (PES) is used as an effective training parameter for Support vector machine (SVM) and ensemble based classifier. The trained SVM can be applied to a large number of unknown video clips for forgery detection and shows that a complex task like digital video authentication can be performed without any human involvement.

Related to Adaptive Group of Pictures structure (AGS) authors in [15] described the process of video editing in three basic steps: decoding the input, actual video editing and finally re-encoding of the edited video. According to them, the techniques developed so far are capable of localizing the tampered region but they cannot acquire the knowledge about origin and history of video if video goes under several encoding-decoding cycles. Author in [7] proposed a method to detect double encoding with different sizes of Group of Pictures (GOPs). The main drawback of this approach is rapid the drop in performance as the strength of last compression increases. Motivated by these shortcoming of above stated technique, Authors in [15], proposed a method to detect double encoding along with the estimation of size of GOP in the previous encoding. The advantage of the technique is that the Variation of Prediction footprint is apparent only in intra-coded P-Frame which makes GOP size estimation possible. GOP size estimation can act as a catalyst for further forensic analysis. Their work can be further extended to MPEG-4 and H.264/AVC encoding standards that introduce B-frames in the GOP structure.

VI. LIMITATION OF THE STUDY

Most of the papers studied in are specific to a particular type of input and applicable to corresponding field of research only. The research method used in one scenario may not be applied to other similar scenario. Furthermore, another limitation of the study is lack of graphical and statistical data. It would be impractical to add statistical assumptions to the studied articles and which may the direction of actual research. So, the study could not be summarized in tabular form based on statistics.

VII. CONCLUSION AND FUTURE SCOPE

In all above specified tampering detection algorithm a fixed GOP structure is used during both decompression and re-compression. A method should be developed which will identify a few simple properties of the temporal fingerprint not identified in fixed size GOPs. These properties are then used to construct a model of the temporal fingerprints in variable size GOPs, which then can be used to generate a target P-frame prediction error sequence that contains the temporal fingerprint. The new forensic technique should be able to detect the P-frame prediction error sequence of the anti-forensically recompressed video by finding out the target prediction error sequence. Additionally, the new technique should be more generalized and applicable to wide range of video sequences. Furthermore, the possibilities of developing anti-forensic counterparts should also be explored.

REFERENCES

[1] Matthew C. Stamm, W. Sabrina Lin, K. J. Ray Liu, “Temporal Forensics and Anti-Forensics for Motion Compensated Video”, IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 7, NO. 4, AUGUST 2012

[2] Peng Yin, Hong heather Yu, Classification of Video Tampering Methods and Countermeasures using Digital Watermarking Proc. SPIE Vol. 4518, p. 239-246, Multimedia Systems and Applications IV

[3] Saurabh Upadhyay, Sanjay Kumar Singh-“Video Authentication- An Overview”, International Journal of Computer Science & Engineering Survey (IJCSES) Vol.2, No.4, November 2011.

[4] Bijan G. Mobasseri, Michael J.Sieffert, Richard J.Simard, Content Authentication and tamper detection in digital video, Proc. IEEE International conference on Image Processing, Vancouver, September 10-13, 2000.

[5] W. Wang, H. Farid, Exposing Digital Forgeries in Video by Detecting Double MPEG Compression, MM&Sec '06 Proceedings of the 8th workshop on Multimedia and security, pp. 37 - 47, ISBN:1-59593-493-6, doi: 10.1145/1161366.1161375

[6] Weihong Wang, Student Member, IEEE, and Hany Farid, Member, IEEE , Exposing Digital Forgeries in Interlaced and De-Interlaced Video

[7] C.-C. Hsu, T.-Y. Hung, C.-W. Lin, and C.-T. Hsu, Video forgery detection using correlation of noise residue. In IEEE International Workshop on Multimedia Signal Processing, Cairns, Australia, 2008.

[8] Valentina Conotter, James F. O’Brien, Hany Farid, “Exposing Digital Forgeries in Ballistic Motion”, IEEE Transaction on Information Forensics and Security, Vol. 7, No. 1, February 2012.

[9] Paolo Bestagini, Simone Milani, Marco Tagliasacchi, Stefano Tubaro, “Local tampering detection in video sequences”, MMSP’13, Sept. 30- Oct. 2, 2013, Pula (Sardinia), Italy. 978-1-4799-0125-8/13/$31.00 ©2013 IEEE.

[10] Figure 1. Is taken from http://www.axis.com/products/video/about_networkvideo/compression.htm

[11] Matthew C. Stamm and K. J. Ray Liu, “ANTI-FORENSICS FOR FRAME DELETION/ADDITION IN MPEG VIDEO”, Dept. of Electrical and Computer Engineering, University of Maryland, College Park

[12] Shrishail Math, R.C.Tripathi, “Digital Forgeries: Problems and Challenges”, International Journal of Computer Applications (0975 – 8887) Volume 5– No.12, August 2010

[13] Sunil Jaiswal, Sunita Dhjavale, “Video Forensics in Temporal Domain using Machine Learning Techniques”, I.J. Computer Network and Information Security, 2013, 9, 58-67. Published online July 2013 in MECS

[14] Shih-Chang Hsia, Szu-Hong Wang, “High-performance adaptive group-of-picture rate control for H.264/AVC”, Received: 22 May 2009 / Revised: 16 November 2009 / Accepted: 17 November 2009 / Published online: 19 December 2009 © Springer-Verlag London Limited 2009

[15] D. V´azquez-Pad´ın, M. Fontani, T. Bianchi, P. Comesa˜na, A. Piva, M. Barni, “Detection of video double encoding with GOP size estimation”, *WIFS‘2012, December, 2-5, 2012, Tenerife, Spain. 978-1-4244-9080-6/10/$26.00* c *2012 IEEE.*

[16] W. Luo, M. Wu, and J. Huang, “MPEG recompression detection based on block artifacts,” in SPIE Conference on Security, Forensics, Steganography, and Watermarking of Multimedia Contents, 2008.