Machine Learning Approach to Detect Tampering in H.264 Video

Remya R. S, Anupama Pradeep

Abstract*— Now a days there are plenty of software's available to access and edit digital videos. Therefore video tampering detection is crucial for legal, medical and surveillance applications. Digital videos are considered as more reliable source of evidence than still images. The abundance of compressed video forms a potential thread of evidence in court rooms. In case of artifacts and possibility of fraud videos court usually calls forensic investigators for examining the problem of authenticating multimedia content. An automated objective assessment of digital video helps to increase the accuracy of videos. Existing schemes are based on MPEG codec. This paper proposes a novel technique to detect tampering in H.264 videos by using neural network. This paper identifies video tampering by using a feature called sequence of average residual of P-frames (SARP). Then time and frequency domain features of sequence of average residual of P-frames are calculated. The detection system is trained with these features. Then the detection system is applied to the video sequence under examination. This method identifies video tampering by differences in time domain and frequency domain features of tampered video from original video. By using machine learning approach, it classifies type of tampering such as insertion, deletion and copy-move. PNN is used for training. The proposed method is applicable for different codec.

***Index Terms*—video tampering detection, SARP, Time domain feature, Frequency domain feature, Training.**

I. INTRODUCTION

In the recent years, a great deal of forensics researches in video comes forward as a discipline aiming at investigating the history of digital videos. Video forensics technique have been developed to carry out different kind of tasks such as source identification, traces of forgery detection and compression history of multimedia contents. The detection of tampering and validation of a legal property of multimedia content is difficult since the original owner is unknown. Multimedia contents can be no longer considered as an "proof of evidence" in court rooms since their source and authenticity is not trusted. Basic goal is to understand very first step is the history of content, namely source identification. Source identification is a method in which Camera that is used to take the video is identified. In this phase video is declared as authentic if the camera identified in this method is matched with one that is provided as evidence. The identification of the type of video tampering such as inserting, deleting or duplicating frames is done under doctoring analysis. Compression analysis checks whether the video is doubly compressed or even multiple times.

Revised Version Manuscript Received on June 10, 2015.

Asst. Prof. Remya R. S, Department of Computer Science, College of Engineering, Karunagapally, Kollam, India.

Anupama Pradeep, PG Scholar, Department of Computer Science, College of Engineering, Karunagpally, India.

First encoding occurs at the time of acquisition and the second occurs after tampering. Double compression can be regarded as an evidence of tampering since the genuine video undergoes only single compression. Forensic researches also studies truthfulness of digital videos that helps to identify inter-frame forgeries and intra-frame forgeries. Inter-frame forgery, where the forger made manipulations on entire number of frames (deletion, insertion or copy-move)of (group of) entire frames. Intra-frame forgery, where the forger changes the content of single frames (e.g. insertion of an object). H.264 is a new video compression standard which is expected to become the video standard of choice in coming years. It is also known as MPEG4-AVC. H.264/MPEG4-AVC is established as a joint standard project between ITU-T's Video Coding Experts group and the ISO/IEC Moving Picture Experts Group in December 2001. H.264 is the name used by ITU-T, while ISO/IEC has named it as MPEG4-AVC. Without compromising the image quality an H.264 encoder can reduce the size of a multimedia file by more than 80% compared with the Motion JPEG format and as much as 50%more than with MPEG-4. H.264/AVC has seven profiles, each targeting a specific class of applications. Surveillance cameras and video encoders mostly uses a profile called Baseline profile. It have core compression capabilities, error resilience, e.g. for video conferencing, mobile video. In main profile, it have high level of compression and quality, e.g. for broadcasting. In extended profile they added features for efficient streaming. H.264 has 11 levels or degree of capability to limit performance, bandwidth and memory requirements. Each level defines bit rate and encoding rate. The higher the resolution, the higher the level is required. Depending on the profile, H.264 encoder uses different types of frames such as I-frame, P-frame and B-frames, may be used by an encoder. A variety of methods can be used to reduce video data, both within the image frame and series of frames. Within the image frame data can be reduced simply by reducing the redundant data. In a sequence of frames multimedia content can be reduces by such methods as difference coding. Typical blocky artifacts can be seen in highly compressed videos such motion JPEG and MPEG standards other than H.264. H.264 can reduce blocky artifacts by using an in-loop deblocking filter. H.264 has been applied in various areas such as high definition DVD (e.g.blu-ray), high definition TV, online multimedia content (e.g. YouTube) and third generation mobile telephony. H.264 is expected to replace other standards and methods in use today. H.264 is broadly available in network cameras and video encoders. Thus most of the surveillance cameras are based on H.264 format.

Therefore video tampering detection based on H.264 videos is a crucial issue. Existing schemes in video tampering is based MPEG videos. Less work is done based on H.264 standard. This paper focuses on H.264 standard. One of the main contributions of tampering detection was proposed by Wang and Farid[1]and focuses on MPEG standard. This paper give an idea about double compressed video and its presence can be used as an evidence of tampering. A doubly compressed digital video sequence introduces specific static and temporal perturbations. Double compression leads to double quantization. Quantization is a point wise operation. First, extract I-frames from the group of frames. Then Discrete Fourier Transform (DCT) of I-frames is computed for detect double quantization. An I-frame is compressed twice with different compression quantities, the DCT coefficients are subjected to two levels of quantization. Plot the histogram of singly quantized and doubly quantized images. Note the periodic artifacts in the histogram of doubly quantized images and thus results in tampering. But H.264 does not work with these methods. An attempt to forgery detection by using noise characteristics was proposed by Kobayakshi [3]. This approach detects forgery from a video sequence traced from static scene by using its noise characteristics. Basic idea of this method is to use noise inconsistencies between genuine video and altered video. Photon shot noise[4] is used as a clue for tampering. Photon shot noise [5] results from the quantum nature of photons and it follows a Poisson distribution where the variance of photons is equal to mean of photons. This relation is formulated as noise level function. By evaluating each pixel in this way, per pixel forgery can detect for a given video. Altered regions captured from digital video camera taken under different situations can be classified when the noise characteristics of the region are inconsistent with rest of the video sequence. Dynamic videos are not supported in this method. Another approach for detecting forgery in real time was proposed by Evan[6]. This approach is put into practice in real time detection of camera tampering and was developed for network surveillance and security applications. Some examples for camera tampering are hiding the camera lens with hands, spray painting on lens, turn on camera and point towards different directions. When a live video is received by the program, it is stored in two buffers. First buffer is named as short term buffer which stores video frames that are less than 10-50 seconds old. The second buffer is long term buffer which stores video frames until they are 2 minutes old. Both of these buffers are First In First Out structures. After the time limit frames from short term buffer goes to long term buffer after a specified time. Each and every time a new frame is pushed into long term pool, the short term and long term pool must be compared by using 3 image dissimilarity features. This method gives high detection rate while at the same time using a very small number of false alarms. Wang and Farid [7] propose another method to detect frame duplication in MPEG videos. Partition the full length video sequences into short overlapping sub-sequences. Similarly in the temporal and spatial correlation is used as evidence of duplications. Correlation coefficient is used to measure of similarity. The spatial and temporal correlation matrices of sub-sequences are used to detect duplicated frames in a full length video. Correlation matrix for all temporal overlapping sequences is computed. Any two sub-sequences with in a correlation above a threshold (close to 1) is measured as a candidate for duplication. The spatial correlation matrices of these candidates of short sub- sequences are compared. If the correlation coefficient of all pairs of these matrix is above a specified threshold ,then the sub-sequence are considered to be spatially and temporally correlated and duplicated. But it is difficult to detect duplication in small regions. Shiang and Lin[8] proposes an approach to detecting frame duplication based on temporal and spatial analysis. Each sub-sequence is used as a query clip. A block based correlation algorithm is developed to spatial correlation of corresponding frame between query clip and the candidate one. Stamm[9] et al proposes an approach to detect forgeries in video that is undetectable by digital forensic technique. This technique overcomes the shortcoming of Wang and Farid[1] method. Now a days, it is easy to alter multimedia content. There are so many digital forensic techniques have been developed to authenticate multimedia content. Likewise there are various anti-forensic operations are applied to this video to make forgery undetectable. Consider e(n) be the prediction error sequence of p-frames. Let $e_1(n)$ be the prediction error sequence of unaltered video is set as null hypothesis and $e_2(n)$ be the prediction error sequence of altered video. In hypothesis testing inequality of prediction error sequence of both videos shows the presence of an additional term. Hence the questioned video is tampered otherwise original. P-frame prediction sequence is obtained by median filtering. Gironi[10] proposes an iterative method to detect frame deletion and insertion of whole frames in a video and propose a detecting system that is able to locate the point where tampering(insertion or deletion) occurs. Here Variation of Prediction Footprint(VPF) is used as a tool for detecting whether the video has been encoded twice. VPF is evaluated as a product of slopes. This method supports different types of codec including H.264. The efficiency of this method is high. Tamer[11] examines the authenticity of video and suggests a machine learning approach to frame deletion. Several number of features are extracted from the frames of video sequences. The features are based on prediction residuals, percentage of intracoded macro blocks, quantization scales and reconstruction quality. The mean and standard deviation of these features are computed. In machine learning, the typical system is trained with unaltered and forged videos. These feature vectors are used for normalization. The spectral regression is used to reduce the dimensionality of features. The system is trained with features of both videos. The features of altered vectors are entirely different from original video. Then the model is applied to the video under question. These features can be applied for both VBR and CBR. Wang and Huang[12] propose an approach to sequence matching based on variance of color correlation. Color correlation is defined as the arrangement of red, green and blue. These color components are arranged in order of intensity. Measure the percentage of pixels belonging to their corresponding color correlation and obtain 6 normalized real values. Resulting 6 numbers are truncated and first 5 numbers are stored in a binary form. Fugui[13]

proposes an approach to detect copy-move forgery in a video based on structural similarity. A new algorithm for structural similarity is used. The range of value of similarity between duplicated frames is higher than that between normal inter frames. For an original video sequence based on continuity of the content in the video, both first and last frame of the video sequence are highly similar to adjacent frames. But duplicated video sequences have no continuity. So the value of similarity between them will be relatively low. In the next section II we explain overall system design and our experimental results are illustrated in section III. The conclusions are presented in section IV.

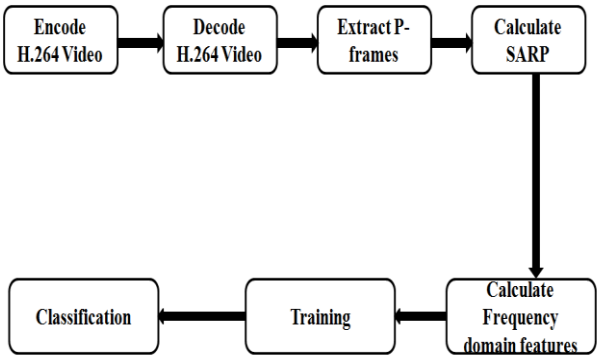

Figure 1: Major Steps in Forgery Detection

II. SYSTEM DESIGN

The major steps in the sytem is shown in figure 1 . Here we analyse the effects of tampering in H.264 video. First encode the H.264 video to extract P-frames from the sequence[14]. During encoding ,I –frame is used as a reference for P-frames .First P-frame is used as reference for rest of P-frames. Further steps are as follows

A. Computation of SARP

Sequence of average residual (SARP) of P-frames from the video sequence is computed. In one GOP P-frames are strongly correlated because they refer to intial I-frame directly or indirectly. The SARP of the video is computed by the equation [11] ,

$$r(n)=\frac{1}{N}\sum_{(i,j)} r_n'(i,j) \quad (1)$$

where N is the number of pixels in one frame and $r_n'(i,j)$ be the residual of n^{th} p-frame at pixel location (i,j). Let matrix Y_k be the k^{th} frame and Y_t' be the t^{th} reconstructed frame. As per [14] ,

$$r_k = Y_k - C(Y_t') \quad (2)$$

Motion vector is obtained by subtracting orginal frame from the reference background. Matrix r_k is the residual of k^{th} frame, C is the motion compensation operator and Y_t' serves as the reference frame of Y_k. Thus we get the following equation for the decoding purpose.

$$Y_k' = F(r_k' + C(Y_t')) \quad (3)$$

This equation is used for decoding process Y_k is the k^{th} reconstructed frame. F is deblocking operator. For simplicity, F is not consider from now. r_k' be the k^{th} decoded residual. The compression noise of k^{th} frame id defined as

$$n_k = Y_k' - Y_k \quad (4)$$

Thus residual energy can be expressed as

$$r_k' = Y_k - C(Y_t') + n_k - C(n_t') \quad (5)$$

As per the equations of [11] and [14] .

B. Computation of Time domain feature

Consider T be the number of frames in one GOP and G be the number of GOPs. Therefore The periodicity of SARP after deletion of some frames is also T. Consider the i^{th} GOP, $\varphi(i)$ be the average residual of P-frame whose position is larges amoung the T P-frames. The position vector [14] of SARP is defined as ,

$$V(i) = \begin{cases} T & if\ \varphi(i)\ isa\ multiple\ of\ T \\ \varphi(i)\ mod T & if\ \varphi(i)\ is\ not\ multiple\ of\ T \end{cases} \quad (6)$$

The value of i ranges from 1,2……G,1≤V(i)≤T.Thus the position vectors and largest value of position vectors is obtained from Sequence of Average Residual of P-frames shown in Figure 2

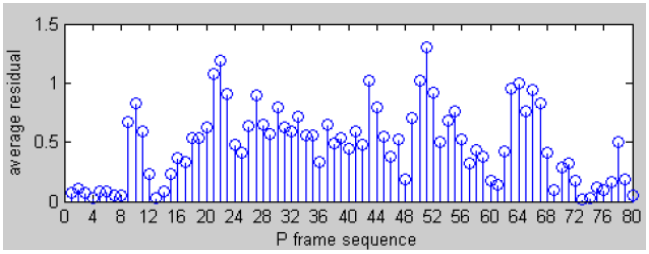

Fig. 2: SARP of P- frames

Next , compute the mean and variance of S(j). S(j) be the times that value of j occurs in V(i), μ be the mean value of S and σ^2 be the variance of S. Therefore we get following equations as per[11] and [14],

$$\sigma^2 = \frac{\sum_{j=1}^{T}(S(j)-\mu)^2}{T-1} \quad (7)$$

$$\mu = \frac{\sum_{j=1}^{T} S(j)}{T} = \frac{G}{T} \quad (8)$$

$$\sigma_{max}^2 = \frac{(G-G/T)^2+(T-1)(G/T)^2}{T-1} \quad (9)$$

The sum of all elements in S is G. so variance achieves its largest value when one element in S is G and all other elements are zero. Normalized variance is to define the Time domain feature ratio Q_t Time domain feature[14] is expressed as follows,

$$Q_t = \frac{\sigma^2}{\sigma_{max}^2} = \frac{T\sum_{j=1}^{T}(S(j)-G/T)^2}{(T-1)G^2} \quad (10)$$

We know that S is relatievely scattered for tampered videos and it results larger variance. Therefore Q_t is relatievely small for original videos. The value of Q_t varies and it depends upon type of tamering.

C. Computation of Frquency domain feature

Discrete Time Fourier Transform is used to transform the SARP into frequency domain. The DTFT of SARP [14] is denoted as $R(e^{j\omega})$. The periodicity of $|R(e^{j\omega})|$ is 2π and symmetric to the vertical axis. $|R(e^{j\omega})|$ attains highest value when ω is zero. If r(n) is strictly periodic ,

$$R\left(e^{j(\omega+2\pi/T)}\right) = R(e^{j\omega}) \quad (11)$$

There is also a large value when ω is a multiple of $2\pi/T$. Thus r(n) shows periodicity, but it is not strictly periodic for tampered videos. We describe h_{min} as per[14],

$$h_{min} = |R(e^{j2m\pi/T})|_{\min} \quad (12)$$

where $m \in K, K$ is the set of all integers in (0,T/2]. Thus h_{min} is small for original videos. Sum of all values of SARP is relatively stable after deleting some frames. Therefore $|R(e^{j\omega})|_{max}$ is relatively stable. But, actual experiments shows that $|R(e^{j\omega})|_{max}$ changes little when some frames are tampered. The frequency domain feature [14] ratio is described as ,

$$Q_f = \frac{h_{min}}{|R(e^{j\omega})|_{max}} \quad (13)$$

Q_f is likely to be smaller for the original video.

D. Machine Learning System

Learning system is trained with the features which extracted from video sequences.The features are time and frequency domian thresholds. Let τ_t and τ_f be the time and frequency domain thresholds. The thresholds changes with type of tampering(insertion,deletion and copy-move). An H.264 video is considered as an original video if $Q_t < \tau_t$ and $Q_f < \tau_f$. If either of these conditions is violated ,then the H.264 video is consider as a tampered video. Probabilistic Neural Network (PNN) is used for classification. PNN have 3 layers. Input layer, target layer and output layer. PNN is used here because it reduces the time during training rather than other machine learning techniques. The detection system is trained with these features. Then the detection system is applied to the video under examination. By using this approach, it classifies type of tampering such as deletion ,insertion and copy-move.

III. EXPERIMENTAL RESULTS

This section analyses the working of proposed method in discriminating between original video and tampered video. In our experiments videos in AVI[14] format is used. x264 encoder is used to encode the video. Quality parameter of h.264 video is 27. P-frames is extracted from the video. For all AVI sequences are encoded to generate original H.264 videos, which are then decoded back into pixel domain. For experimental purpose, each decoded videos are tampered. Tampering such as deletion, insertion and copy-move . Table 1 shows time and frequency domain features of original and deleted videos. (τ_t, τ_f) pair of original videos is set as (0.2740 , 0.0275). If the features of a video , (Q_t , Q_f) is less than or equal to τ_t and τ_f is come under original category. For a deleted video (τ_t, τ_f) pair will be (0.2740,0.4000). If Q_t is greater than or equal to τ_t otherwise Q_f is less than τ_f is considered as a deleted video.

Table 1: Features of original and deleted videos

Index	Original videos	Deleted Videos
Vid 1	(0.2740, 0.0270)	(0.2820, 0.0265)
Vid 2	(0.2634, 0.0225)	(0.2940, 0.3255)
Vid 3	(0.2560, 0.0183)	(0.2846, 0.2375)
Vid 4	(0.2740, 0.0243)	(0.2740, 0.3395)
Vid 5	(0.2740, 0.0214)	(0.2740, 0.2275)

Table 2 shows time and frequency domain features of original and inserted videos. For an inserted video the Q_t is less than or equal to τ_t otherwise Q_f is less than τ_f . (τ_t, τ_f) pair for an inserted video is (2.1690,0.00036). Table 3 shows time and frequency domain features of original and copied videos. For an copied video the (Q_t , Q_f) pair is less than or equal to (τ_t, τ_f) pair, ie (0.4786, 0.5000). The value of thresholds depends upon the type of tampering. The original videos and tampered videos are encoded and decoded again. Next step is to extract SARP in the decoding process. SARP is shown in Fig 2. Probabilistic Neural Network is used for training. Then the time-domain feature ratio and frequency –domain feature ratios are obtained from the SARP.

Table 2: Features of original and inserted videos

Index	Original videos	Inserted Videos
Vid 1	(0.2740, 0.0270)	(2.1690,0.00036)
Vid 2	(0.2634, 0.0225)	(1.2680,0.00027)
Vid 3	(0.2560, 0.0183)	(1.1567,0.00019)
Vid 4	(0.2740, 0.0243)	(1.1699,0.00011)
Vid 5	(0.2740, 0.0214)	(1.1568,0.00009)

Fig 3 shows the Time and frequency domain ratios (Q_t and Q_f) of video sequences. X-axis shows Time domain feature and Y-axis shows frequency domain features. Black, Red, Blue and Cyan represents original, deleted ,inserted and copied video respectively. From the Fig 3 , we can conclude that tampered videos can classify by thresholds adaptively. For a video sequence to be tested, we get a (Q_t , Q_f) pair. The value of (τ_t, τ_f) pair varies depends upon type of tampering. If the value of (Q_t , Q_f) pair varies from specified value of (τ_t, τ_f) pair, then the video is considered as a tampered video.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 3: Features of original and copied videos

Index	Original videos	Copied Videos
Vid 1	(0.2740, 0.0270)	(0.4786,0.0088)
Vid 2	(0.2634, 0.0225)	(0.4534,0.4084)
Vid 3	(0.2560, 0.0183)	(0.3297,0.5000)
Vid 4	(0.2740, 0.0243)	(0.2319,0.4453)
Vid 5	(0.2740, 0.0214)	(0.1986,0.4084)

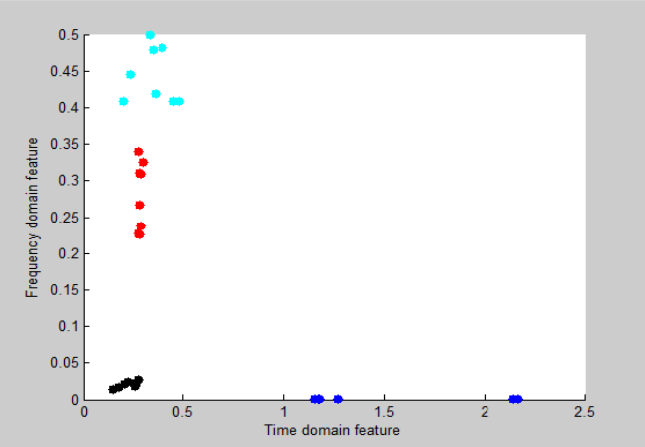

Fig. 3: Original and tampered videos

IV. CONCLUSION

In this study, we developed a method for video tampering detection in H.264 videos. The Key features of this method are: the robustness to use different codecs, and the possibility of distinguishing frame insertion, frame deletion and copy-move. The key feature used in this paper is Sequence of Average Residual of P-frames (SARP). Then time and frequency domain of SARP is computed. The learning system is trained by Probabilistic Neural Network. The tampered videos and original videos are separated by thresholds. Video sequences are tested in our experiments and outcome shows our system is fairly effective for H.264 videos.

REFERENCES

1. W.Wang and H.Farid “Exposing Digital Forgeries in video by detecting double MPEG compression, ACM“MM & Sec’06, Geneva, Switzerland, September 26-27, 2006.
2. W.Wang and H.Farid” Exposing Digital Forgeries in video by detecting double Quantization” , ACM, MM & Sec’09, Geneva, Switzerland, September 6-7,2009
3. Michihiro Kobayashi, Takahiro Okabe, and Yoichi Sato ,”DetectingVideo Forgeries based on Noise characteristics” , Springer-Verlag Berlin, pp.306-317, 2009
4. H. Phelippeau, H. Talbot, M. Akil, H. Phelippeau and S. Bara. "Shot noise Adaptive bilateral Filtering. “ University Paris -Est, Laboratoire A2SI Group ESIEE,2010.
5. Fulu Li, James Barabas, Ankit Mohan and Ramesh Raskar ,”Analysis of Errors due to Photon Noise and Quantization Process with Multiple Images “IEEE, 2010
6. Evan Ribnick, Stefan Atev, Osama Masoud, Nikolaos Papanikolopoulos, And Richard Voyles , ” Rea l - Time Detection of Camera Tampering”, IEEE Computer Society ,2006.
7. W. Wang and H. Farid , “ Exposing Digital Forgeries in video by detecting Duplication ”, ACM, MM&Sec’07, September 20–21, 2007
8. Guo - Shiang Lin, and Jie - Fan Chang,” Detection of Frame Duplication Forgery in videos based on Spatial and Temporal analysis”, IJPRAI,2012
9. Matthew C. Stamm , W. Sabrina Lin and K. J. Ray Liu, " Temporal Forensics and Anti – Forensics for Motion Compensated Video",2006
10. A. Gironi M. Fontaniy , T. Bianchi A. Piva and M. Barnix, "A Video Forensic Technique For Detecting Frame Deletion And Insertion"FET programme , 2012
11. Tamer Shanableh,"Detection of frame deletion for digital videoforensics” Elsevier, Digital investigation,2013
12. Yanqiang Lei, Weiqi Luo, Yuangen Wang and Jiwu Huang,” Video sequence matching based on the invariance of color correlation.”, ACM Digital Library,2012
13. Fugui Li, Tianqiang Huang ,” Video Copy - Move Forgery Detectionand Localization Based on Structural Similarity ” , Springer Proceedings of the 3rd International Conference on Multimedia Technology (ICMT), 2013
14. Hongmei Liu, Songtao Li, and Shan Bian, “Detecting frame deletion in H.264 video”, ISPEC , LNCS 8434, pp. 262–270, 2014
15. Gary J. Sullivan, Jens - Rainer Ohm, Woo - Jin Han and Thomas Wiegand,” Overview of the High Efficiency Video Coding (HEVC) Standard”, IEEE , VOL. 22, NO. 12, December 2012.