A Forgery Video Detection Algorithm for Resolution Promotion Manipulations Using Frequency Spectrum Analysis

Saihua Song
University of Shanghai for Science and Technology
Shanghai, China
(086)15800968773
sshusst@163.com

Heng Yao*
University of Shanghai for Science and Technology
Shanghai, China
(086)13661488750
hyao@usst.edu.cn

Chuan Qin
University of Shanghai for Science and Technology
Shanghai
China
qin@usst.edu.cn

Ying Tian
University of Shanghai for Science and Technology
Shanghai
China
tianying@usst.edu.cn

ABSTRACT
Resolution is one of the important factors for video visual quality. With the development of the video editing software, one video sequence can be easily enlarged and inserted into another video sequence to generate a forgery video. The detection of video resolution modifications has become an important issue for digital video authentication. In this paper, we propose a passive detection algorithm based on the zero-crossings of the high-frequency component of the key frames in frequency domain. Firstly, the key frames of video are extracted before the second-order differential is executed. Then, the zero crossings of differential are calculated to generate a one-dimensional binary sequence. Finally, the frequency analysis is applied to expose the original resolution of the video for detection. Based on this method, we also develop a forgery detection method for sequence inserting. The experimental results demonstrate that the proposed method can accurately estimate the video original resolution, as well as detect the forgery traces of synthetic videos with maliciously inserted irrelevant video sequences.

CCS Concepts
• **Mathematics of computing→Mathematical analysis**

Keywords
Digital video forensics; Video amplification; Spectrum analysis; Zero-crossing

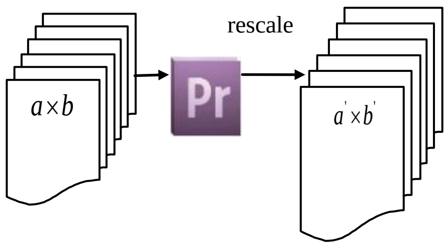
1. INTRODUCTION
With the development of multimedia processing techniques, many video processing softwares, such as Adobe Premiere, can easily modify video content without leaving any obvious trace. The video tampering authentication has become a significant research topic. For most video forensic circumstances, the original video is damaged due to some malicious tampering and cannot be restored. In order to expose the authenticity of the video, the researchers carried out a lot of works and proposed the amount of methods. A video bit rate tampering detection algorithm was proposed by Bian et al. [1] based on Benford's Law combined with Mean Structural Similarity, which had a 16-dimensional feature vector. The periodic spectrum of excessive prediction residuals in the H.264/AVC codec was proposed by Su et al. [2]. In their work, the energy of residuals with the deblocking filter were evaluated to reveal the video manipulation, and the mechanism of rate control was also utilized to check the quantization parameters. An inter-frame tampering detection algorithm based on the characteristic peak of P-frame prediction errors in time domain and frequency domain was proposed by Kang et al [3]. The forensic and anti-forensic techniques for video frame deletion is proposed by Stamm et. al. [4], they investigated a game theory framework for analyzing the interplay between forensics and anti-forensics. A Double MPEG-4 compression detection method based on local motion vector field analysis in static-background videos was proposed by He et. al. [5]. An estimation method for double compression was proposed by Bestagini et. al. [6] to exploit the coding-based footprints to identify both the codec and the size of the group of pictures (GOPs). In the pioneering work, the algorithm based on the periodicity of the average prediction residual sequence is proposed by Wang and Farid [7] to detect double compression with shifted GOP. In addition, a forensic method to detect the forgery of objects in ballistic motion based on physical inconsistencies was proposed by Conotter et al. [8].

As we know, there is no existing literatures to detect the traces of video amplifications. However, for image resampling forensics, there are many existing related works, the method of second-order differential frequency analysis, proposed by Gallagher [9], was one of representative work among them. Birajdar et al. [10] made improvements on [9] to enhance the accuracy of the interpolation factor by using zero-crossing concept. Motivated by [9] and [10], in this paper, we propose a passive detection algorithm for resolution promoted videos based on the zero-crossings of the

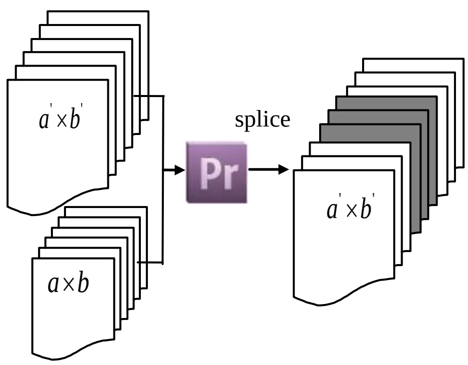
ICNCC'16, December 17-21, 2016, Kyoto, Japan
© 2016 ACM. ISBN 978-1-4503-4793-8/16/12…$15.00
DOI: http://dx.doi.org/10.1145/3033288.3033304

high-frequency component of the key frames in frequency domain. This method can also be extended to locate the partial inserted frames by detecting the traces of resolution promotions. Two types of tampering forms, consisting of the overall resolution promotion and frame-insertion of partial amplified sequences, are shown in Fig. 1.

The rest of this paper is organized as follows. Section 2 discusses an amplified video resolution detection algorithm. Section 3 introduces the location detection algorithm for inserted frames whose resolution is amplified. Section 4 and Section 5 present experimental results and conclusions, respectively.



(a) Resolution promotion



(b) Frame insertion of partial amplified sequence

Figure 1. Two types of tampered videos

2. ESTIMATION METHOD OF VIDEO AMPLIFICATION FACTOR

In this section, the detection method is proposed to identify whether there is an amplification operation by seeking peak magnitudes of key frames in the spectrum of the video, and furtherly estimate the amplification factor. The flow diagram of factor estimation method for video amplification is shown in Fig 2. First, the key frames from to-be-identified video are extracted before converting them into YCbCr space. Then, the second differences of Y component are computed and averaged weightedly along the vertical direction. Next, the binary sequence indicated by the zero crossings of the averaged second difference sequence are generated and transformed to frequency domain by Fast Fourier Transform (FFT), and finally, the spectrum analysis is executed to derive the amplification factor. More details are represented in the remaining of this section step by step.

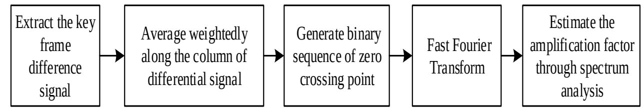


Figure 2. The flow chart of factor.

The Group of Picture(GOP) is regarded as a compression coding unit in the video standard. The GOP contains I, P and B frames. The first frame, i.e. I-frame, is an intra-coded frame and does not refer to any other coding frames, so it is a key frame in the video codec process. P-frame is a forward predicted frame which records the difference between the current frame and the key frame or P-frame. B-frame is a bi-directional prediction frame which records the difference between the current frame and the front and rear frames. The length of GOP is fixed in most video coding and the video stream can be extracted key frames at regular intervals. The key frames in video coding are sampled, quantized and coded in the form of a whole image. Thus, I-frame is more suitable for detecting the resampling trace of video than P-frame and B-frame. Similar to the principle of image amplification, the resampled pixels in the amplified video are derived by using the interpolation technique according to the surrounding pixels.

Due to the interpolation manipulation on the key frames of to-be-amplified video, the residuals between the interpolated pixels and their corresponding predictions are relatively smaller than the residuals of the original pixels and their predictions. To exploit this features, following [9], the second-order differential operation is carried out along the row of each key frame. Suppose the pixel in input key frame is denoted by $p(i, j)$, where $0 \le i < R$ and $0 < j < C-1$. R and C are the number of rows and columns in the key frame. The second-order differential signal of each row is calculated as

$$s(i, j) = 2p(i, j) - p(i, j+1) - p(i, j-1) \tag{1}$$

After the second-order differential manipulated on each pixel p(i, j), a difference matrix with the size of $R \times C$-2 is generated and denoted by S. Since the resampled pixels of each row of each key frame in amplified video are undergone the same interpolation operation. The magnitudes of the column in **S** are averaged together to form a one-dimensional signal $v_p(j)$ as

$$v_p(j) = \frac{1}{R} \sum_{i=0}^{R} s(i, j) \tag{2}$$

According to [10], in order to overcome the deficiency of common second-order differential-based resampling detection method, which cannot estimate the amplification factor of 2, the concept of zero crossing of second-order differential signal is introduced. The zero crossing is the intersection of the gray axis of the image and the extremum of second derivative. A binary zero crossing sequence, denoted by p, is generated by comparing the current and next averaged signal with 0 as

$$p(n) = \begin{cases} 1 & v_p(j) > 0 \; and \; v_p(i+1) \le 0 \\ 1 & v_p(j) < 0 \; and \; v_p(j+1) > 0 \\ 0 & otherwise \end{cases} \tag{3}$$

Once the video resolution is promoted, the second-order differential signal of each key frame will exhibit periodicity and can be detected by spectrum analysis. The obtained binary sequence p is transferred to its transform domain, denoted by P, through discrete Fast Fourier Transform (FFT) as

$$P(k)=\sum_{n=0}^{C-3} p(n)e^{-j\frac{2\pi}{C-2}nk}, k=0,...,C-3 \tag{4}$$

The peak point in frequency spectrum indicates the most probable periodicity in second-differential signal, which can be manually or automatic selected. In order to achieve a batch execution, we propose an adaptive threshold method for automatically seeking of the peak frequency. Because the information of the video screen is mainly concentrated in the low-frequency, the spectrum of the low-frequency part is ignored in the search of the maximum spectrum. The maximum amplitude of the frequency spectrum, denoted by $P_{\max}$, is identified as peak point once it satisfies

$$P_{\max} - \mu \geq n\sigma \tag{5}$$

where μ and σ are the mean and standard deviation of the normalized frequency amplitude, respectively, and n is the preset threshold parameters which is set at 1.35 in our experiment. Denote the corresponding frequency of peak point $P_{\max}$ as f, and the relationship between f and the interpolation factor $\tilde{N}$ is

$$f = \begin{cases} 1 - 1/\tilde{N} & 1 < \tilde{N} \leq 2 \\ 1/\tilde{N} & \tilde{N} > 2 \end{cases} \tag{6}$$

It should be noted that each f corresponds to two possible amplification factors, and if there is no any obvious peak point to be detected in our seeking algorithm, the video for detection will be regarded as an authentic video.

3. LOCATION DETECTION ALGORITHM OF INSERTED FRAMES

If frames with lower resolution are intended to insert into another load video with larger resolution, it's inevitable to adjust the resolution of the frames in accordance with that of load video. The detection method described in Section 2 can be extended to locate the amplified frames through the similar detection method of spectrum analysis. Fig. 3 illustrates the flow chart of the detection algorithm to locate the amplified partial sequences. First, the peak points in the frequency spectrums of each key frame is sought according to the method introduced in Section II. As can be seen in Fig. 3, the GOPs for detection are from NO. 1 to NO. M , and the peak points can be detected in the frequency spectrums from the GOP number of m_1 to m_2. Thus, the key-frame of GOP NO. m_1 and NO. m_2 are treated as temporal start and end frame of inserted frames. Since the inserted start position probably appears in the P-frame or B-frame of NO. m1-1 GOP, i.e. the preceding GOP of NO. m_1. Similarly, the end position maybe in the P-frame or B-frame of NO. m_2 GOP as well. Thus, to further seek the more precise inserted position, we traverse all frames in NO. m1-1 and NO. m_2 GOP to seek the first and last peak points in each spectrum of each frame by using the method introduced in Section II. Suppose T_0 and T_d be our detected start and end position in the overall video, and they can be determined by

$$\begin{cases} T_0 = \dfrac{(m_1 - 2) \times N_{GOP} + N_0}{v_{rate}} \\ T_d = \dfrac{(m_2 - 1) \times N_{GOP} + N_d}{v_{rate}} \end{cases} \tag{7}$$

where N_0 and N_d are the frame number that appears peak frequency in NO. m_1-1 and NO. m_2 GOP, respectively. N_{GOP} is the length of GOP and v_{rate} is the frame rate of the video.

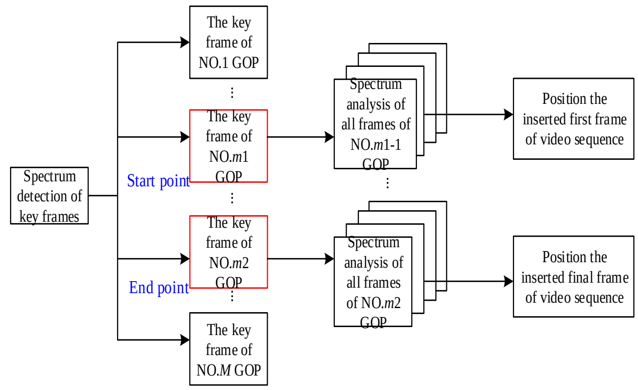


Figure 3. The flow chart of location detection algorithm of inserted frames.

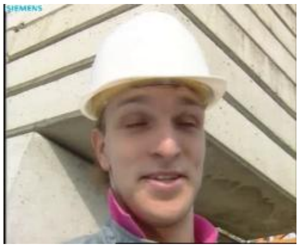
4. EXPERIMENTAL RESULTS

4.1 The Experimental Results on Amplified Video

In the experiment, test video *foreman* was downloaded from the database of Xiph.org Video Test Media [11] and, was amplified with 10 different amplification factors from 1.2 to 2.4 with the aid of the video editing software of Adobe Premiere. The detection method was tested in the platform of OpenCV3.0 with VS2013 libraries. Fig. 4 presents one example of the test results of the spectrum analysis of key frames with the comparison of the original video *foreman* and its corresponding double amplified video. Fig. 4 (a) shows a key frame of the original video and Fig. 4 (d) shows its corresponding frequency spectrum by using the method proposed in Section 2. The key frame amplified with a factor of 2 is presented in Fig. 4 (b) and its spectrum is shown in Fig. 4 (e). As can be observed in Fig. 4 (e), there is a significant spike corresponding to the normalized frequency of 0.5 in Fig. 4 (e). This phenomenon has revealed the video is amplified with the factor of 2. Similarly, the spectrum of another amplified key frame shown in Fig 4 (c) appears a similar peak as shown in Fig. 4 (f). The experimental results demonstrated that each key frame of the video resolution with amplification manipulations showed significant periodicity in high frequency components. The detection performance for each interpolation factors is listed in Table 1. For each factor, we compute the detection accuracy by averaging all deviations between the sought peak frequencies of all key frames and the actual peak frequency for each factor. The experimental results show that the detection accuracy rate of different factors is around 99.0%.

Table 1. Detection accuracy of different interpolation factor.

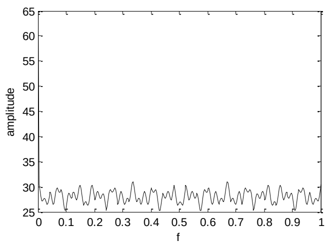
Scaling factor	1.2	1.3	1.4	1.5	1.6	1.7	1.8	2.0	2.2	2.4
Peak frequency	0.16	0.24	0.28	0.34	0.38	0.41	0.45	0.50	0.46	0.41
Accuracy（%）	99.2	98.4	99.2	99.0	98.2	99.7	99.0	100	98.8	98.4



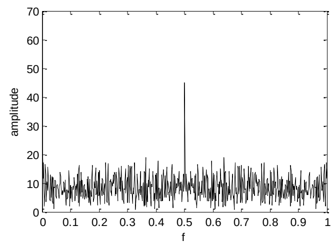
(a) The key frame of the first GOP

(b) Amplified key frame of the first GOP (factor: 2.0)

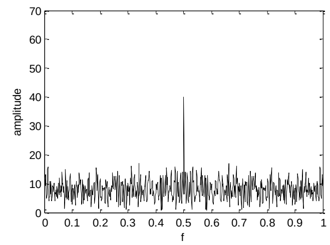
(c) Amplified key frame of the tenth GOP (factor: 2.0)



(d) The frequency spectrum of the key frame of the first GOP



(e) The frequency spectrum of the amplified key frame of the first GOP



(f) The frequency spectrum of the amplified key frame of the tenth GOP

Figure 4. The comparison of the frequency spectrum between original key frame and amplified video key frames.

4.2 The Experimental Results of the Identifications of the Partial Inserted Frames

We generated 20 synthetic videos by inserting some partial frames into another video sequence as shown in Fig. 1(b). Due to the resolution inconsistency between each sequence, we amplified the inserted frames in accordance with the size of the load sequence. The frame rate of to-be-detected video, the length of GOP and other related parameters are supposed to be read out. In our experiment, the frame rate of video is 25 f/s and the length of GOP is 12. In Fig. 5, for example, a resolution of 352 × 288 video sequence of *akiyo* that is amplified with a factor of 2.2 is inserted into the resolution of 774 × 634 of the original *foreman* video. The middle two video frames are the start frame and the last frame of inserted frames in the synthetic video and the remaining two video frames are the original frames at the beginning and the end time of the synthetic video. According to our detection method, the start and end positions of inserted frames can be accurately located in the tampered video. Through the amount of the experiments, the adaptive peak detection threshold is suggested to set at 1.35σ and by this setting, the system can accurately locate the time nodes of the inserted frames. Fig. 6(a) and Fig. 6(b) present the automatically detected peak point of the start and final frame in the synthetic video shown in Fig. 5, respectively. Observed from both spectrums, the normalized frequencies corresponding to the peak points are both at around 0.46. According to Equation (6), the interpolation factor can be estimated by $\tilde{N} = 2.2$.

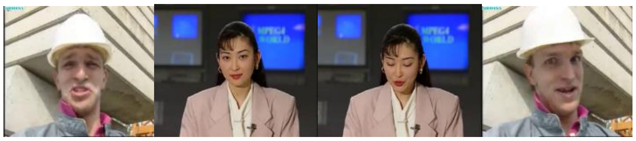
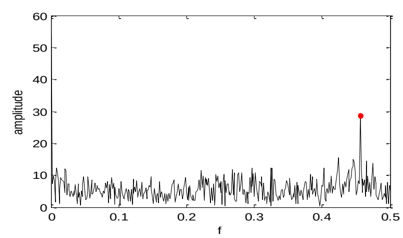
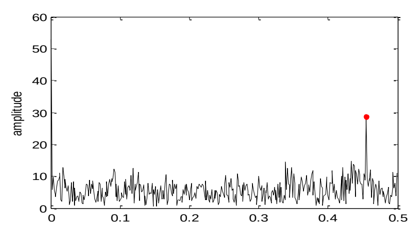


Figure 5. The original frames and inserted frames of synthetic video.



(a) The frequency spectrum of the first frame in the inserted video sequence



(b) The frequency spectrum of the final frame in the inserted video sequence

Figure 6. The peak points detection in spectrum analysis of the first and final inserted frames.

5. CONCLUSIONS

Motivated by the existing image resampling detection method, in this paper, we propose a method to identify the manipulations of video resolution promotion, and estimate the amplification factors by a frequency spectrum analysis. A binary second-order difference sequence can be generated by zero-crossing method and, its frequency spectrum will exhibit obvious periodicity if the frame for detection has been amplified. Besides, by using this factor estimator of video amplification, we also put forward a synthetic video detection method to locate the inserted forgery frames, which has been amplified before they are inserted into another sequence. Experimental results demonstrate that the proposed method can accurately estimate the amplification factor of the video and locate the start and the end frames of the inserted partial sequence.

6. REFERENCES

[1] Bian, S. Luo, W. Q., and Huang, J. W. 2014. Exposing fake bit rate videos and estimating original bit rates. *IEEE Transactions on Circuits and Systems for Video Technology*. 24,12 (Dec. 2014), 2144-2154.

[2] Su, P. C., Suei, P. L., and Chang, M. K. et al. 2015. Forensic and anti-forensic techniques for video shot editing in H.264/AVC. *Journal of Visual Communication & Image Representation*. 29,1 (Feb. 2015), 103-113.

[3] Kang, X., Liu, J., and Liu, H. et al. 2015. Forensics and counter anti-forensics of video inter-frame forgery. *Multimedia Tools and Applications*. 7,8 (Jun. 2015), 1-21.

[4] Stamm, M. C., Lin, W. S., and Liu, K. J. R. 2012. Temporal forensics and anti-forensics for motion compensated video. *IEEE Transactions on Information Forensics and Security*. 7,4 (Aug. 2012), 1315-1329.

[5] He, P., Jiang, X., and Sun, T. et al. 2016. Double compression detection based on local motion vector field analysis in static-background videos. *Journal of Visual Communication & Image Representation*. 35,1 (Nov. 2016), 55-66.

[6] Bestagini, P., Milani, S., and Tagliasacchi, M. et al. 2016. Codec and GOP identification in double compressed videos. *IEEE Transactions on Image Processing*. 25,5 (May. 2016), 2298-2310.

[7] Wang, W. and Farid, H. Exposing digital forgeries in video by detecting double MPEG compression, *In Proceedings of the 8th Workshop on Multimedia and Security*, (The Geneva, The Switzerland, Sep 26 - 27, 2006). MM&Sec '06. ACM, New York, NY, 37-47.

[8] Conotter, V., O'Brien, J. F., and Farid, H. 2012. Exposing Digital Forgeries in Ballistic Motion. *IEEE Transactions on Information Forensics and Security*. 7,1 (Feb. 2012), 283-296.

[9] Gallagher, A. C. 2005. Detection of linear and cubic interpolation in JPEG compressed images. *The Second Canadian Conference on Computer and Robot Vision*. (The Victoria, BC, The Canadian, May 09 - 11, 2005), 65-72.

[10] Birajdar, G. K., and Mankar, V. H. 2013. Blind authentication of resampled images and rescaling factor estimation. *International Conference on Cloud & Ubiquitous Computing & Emerging Technologies*. (The Pune, The India, Nov 15 - 16, 2013), 112-116.

[11] YUV Video Sequences, in, <http://trace.eas.asu.edu/yuv/>, accessed 2013.