EXPOSING VIDEO INTER-FRAME FORGERY BASED ON VELOCITY FIELD CONSISTENCY

Yuxing Wu, Xinghao Jiang, Tanfeng Sun and Wan Wang*

School of Electronic Information and Electrical Engineering,
Shanghai Jiao Tong University, Shanghai, China
{wuyuxing, xhjiang, tfsun, wangwan}@sjtu.edu.cn

ABSTRACT

In recent years, video forensics has become an important issue. Video inter-frame forgery detection is a significant branch of forensics. In this paper, a new algorithm based on the consistency of velocity field is proposed to detect video inter-frame forgery (i.e., consecutive frame deletion and consecutive frame duplication). The generalized extreme studentized deviate (ESD) test is applied to identify the forgery types and locate the manipulated positions in forged videos. Experiments show the effectiveness of our algorithm.

Index Terms— Video forensics, Inter-frame forgery detection, Velocity field, Generalized extreme studentized deviate

1. INTRODUCTION

Nowadays, surveillance camera systems have been widely deployed in many circumstances to monitor illegal activities. Surveillance videos have already been regarded as the judicial proofs in the court. However, with the development of advanced video editors, their integrity cannot be guaranteed anymore. Therefore, how to authenticate the surveillance videos has become a significant issue.

So far, many video forensics techniques have been studied [1]. [2]-[4] proposed to detect double compression, [5]-[7] detected video forgery with sensor noise patterns, and [8]-[10] exposed forgery based on the videos' content. In the aspect of inter-frame forgery detection, Wang and Farid [2] first exposed the frame deletion or insertion by prediction error. They discovered that frames moving from one group of picture (GOP) to another will have larger motion estimation errors. However, their method would fail if a complete GOP is deleted. Mondaini et al. [5] proposed to detect frame insertion/duplication by the photoresponse non-uniformity noise (PRNU) fingerprinting technique. Chao et al. [10] proposed to detect frame deletion and insertion through optical flow. They found that inter-frame forgery operations would cause discontinuity in optical flow sequence.

In this paper, we propose a new approach to detect surveillance video inter-frame forgery based on the consistency of velocity field. This method is able to distinguish the tampered video, identify the forgery types (i.e., consecutive frame deletion, consecutive frame duplication) and locate the manipulated positions in forged videos as well. Our algorithm follows three steps. First, obtain velocity field sequence by applying block-based cross correlation. Then, calculate the corresponding relative factor sequence from velocity field sequence. Finally, determine the authenticity, the forgery type and manipulated locations with generalized extreme studentized deviate (ESD) algorithm.

2. VELOCITY FIELD IN VIDEO FORGERY DETECTION

Velocity field is a term induced from Particle Image Velocimetry (PIV) technique [11]. The key point of PIV is to compare adjacent video frames and estimate their displacements caused by time separation. It is considered that any inter-frame operations, like frame deletion and duplication will enlarge the displacements. In this section, we will show how to form the velocity field sequence and illustrate traces left in it after different forgery operations.

2.1. Velocity field sequence estimation

The velocity field computation is done by PIVlab [12]. Its PIV algorithm is set to FFT window deformation with one-pass 16×16 pixel interrogation window and 75% overlap factor. (1) and (2) are the mathematical descriptions of the computation process.

$$R_C(u,v)=\mathcal{F}^{-1}\left(\left[\mathcal{F}\left(I(i,j,t)\right)\mathcal{F}\left(I(i,j,t+1)\right)\right]^*\right) \quad (1)$$

$$\underset{u,v}{arg\,max}\,Re\left\{R_C(u,v)\right\} \quad (2)$$

where $I(i,j,t)$ and $I(i,j,t+1)$ are the interrogation windows at (i,j) location in t and $(t+1)$ frame respectively. $\mathcal{F}$, $\mathcal{F}^{-1}$ are 2-D Fourier transform operator and inverse Fourier transform operator respectively, * is the complex conjugate

* Corresponding Author is Xinghao Jiang

function, $Re\{\cdot\}$ obtains the real part of its parameter. According to these formulas, (u,v) is regarded as the displacement (also called velocity vector) between the two interrogation windows. To express accurately, we denote (u,v) as $[u(i,j,t),v(i,j,t)]$, which indicates the velocity vector at (i,j) location of t frame. Therefore, we can define the velocity field intensity (VFI) as follows:

$$VFI(t)_h = \sum_i \sum_j |u(i,j,t)|;\ VFI(t)_v = \sum_i \sum_j |v(i,j,t)| \quad (3)$$

where $VFI(t)_h$, $VFI(t)_v$ indicate the horizontal and vertical velocity field intensity respectively. We then denote $\{VFI(t)_h | t \in [1, L-1]\}$ and $\{VFI(t)_v | t \in [1, L-1]\}$ as the horizontal and vertical VFI sequence, where L is the number of frames.

The max-sample technique is used to exclude those frames with extremely low VFI. The low VFI is probably caused by the similarity of two neighbor frames in data, which was introduced by camera coding error. Every three frames are sampled into one frame with the maximum VFI. And the sample process starts at the position where the number of the remaining frames can be divided by 3. Then we have the new VFI sequences as follows:

$$\{SVFI(t)_h | t \in [1,T]\}, \{SVFI(t)_v | t \in [1,T]\}$$

where $SVFI(t)$ denotes VFI sequence after max-sampling, $\lfloor\ \rfloor$ represents round down function and $T = \lfloor (L-1)/3 \rfloor$.

The consistency of the VFI sequences in both directions will be destroyed if the video is manipulated by some inter-frame forgery operations. Therefore, the relative factors RF_h and RF_v are defined to reveal these changes.

$$RF_h(t) = \frac{SVFI(t-1)_h + SVFI(t+1)_h}{SVFI(t-1)_h \times SVFI(t+1)_h} \times SVFI(t)_h \quad (4)$$

$$RF_v(t) = \frac{SVFI(t-1)_v + SVFI(t+1)_v}{SVFI(t-1)_v \times SVFI(t+1)_v} \times SVFI(t)_v \quad (5)$$

In the relative factor sequences $\{RF(t)_h | t \in [2,T-1]\}$ and $\{RF(t)_v | t \in [2,T-1]\}$, the discontinuity peaks introduced by the forgery operations will be obviously highlighted.

2.2. Traces in relative factor sequence

In this paper, two types of forgeries, consecutive frame deletion and consecutive frame duplication are considered. Different forgery operations will introduce different numbers of discontinuous peaks in the relative factor sequence. Fig. 1 shows the corresponding relative factor sequences of a given video before and after manipulation.

2.2.1. Original video

These videos are directly from surveillance cameras without any modifications. And there is no discontinuous peak in the relative factor sequence in this type of video.

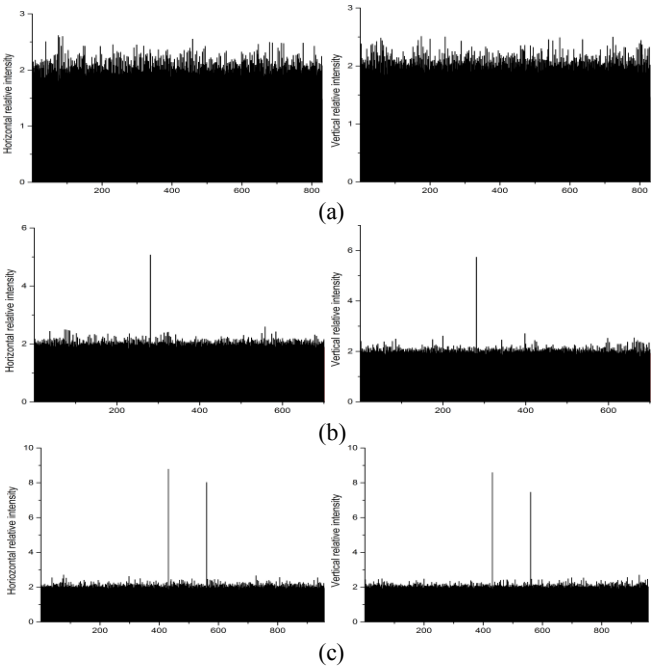


Fig. 1. The horizontal (left) and vertical (right) relative factor sequences. (a) original video; (b) frame deletion video; (c) frame duplication video.

(a) (b) (c) (d)

Fig. 2. Four representative frames of a video. Over 600 consecutive frames have been deleted between (a) and (d) to cover a suspicious man walking out of the elevator. There will be no visual differences before and after forgery process.

2.2.2. Consecutive frame deletion video

These videos are tampered by deleting consecutive frames. After the forgery process, two originally unrelated frames have become neighbors, generating a salient increase in the VFI sequence. Therefore, one discontinuous peak would be observed in the relative factor sequence.

2.2.3. Consecutive frame duplication video

These videos are modified by duplicating consecutive frames from one time point to another. Hence, two discontinuous peaks would be observed.

Again, note that we only consider the videos recorded by static surveillance cameras. That is to say, only the inter-frame forgery operations will introduce the obvious discontinuity in the relative factor sequence. In addition, we focus on detecting videos with meaningful forgeries, which mean no visual differences will be perceived before and after forgery process. Fig. 2 demonstrates an example of meaningful consecutive frames deletion forgery.

3. VIDEO FORGERY IDENTIFICATION

The discontinuous peaks in the relative factor sequence are regarded as the evidence of video forgery. The generalized ESD test has been applied to extract the peaks and identify the forgery types. The detail of the identification algorithm is described in this section.

3.1. Generalized ESD test

We find that the probability distribution of the relative factor sequence follows an approximate normal distribution. Hence, Generalized ESD test [13] is able to be employed in our identification algorithm.

There are two important parameters in the test, the upper bound number of outliers r and significance level α. First compute R_1 from

$$R_i = \max_i \left| x_i - \bar{x} \right| / s \tag{6}$$

where $\bar{x}$ and s denote the mean and standard deviation of the n samples respectively. Remove the observation that maximizes $\left| x_i - \bar{x} \right| / s$ and then re-compute the above statistic with $n-1$ observations. Repeat this process until $R_1, R_2, ..., R_i$ have all been computed. Finally pick the corresponding r critical values λ_i at the chosen confidence level α. The number of outliers is determined by finding the largest i such that $R_i > \lambda_i$.

In order to determine the exact number of peaks in the relative factor sequence, we have fine-tuned the critical values λ_i by multiplying a coefficient η, and the new definition is as follows:

$$\lambda_i^{'} = \eta \times \frac{t_{(p,n-i-1)} \times (n-i)}{\sqrt{\left(n-i-1+t^2_{(p,n-i-1)}\right) \times (n-i+1)}} \tag{7}$$

$$p = 1 - \frac{\alpha}{2 \times (n-i+1)} \tag{8}$$

where $t_{(p,n-i-1)}$ is the pth percentile of a t distribution with $(n-i+1)$ degrees of freedom.

Some fake peaks might be found in the relative factor sequence. Comparing with real forgery peaks, these peaks are with relatively low intensities, which were probably introduced by camera noise or video encoding. The fake peaks would be determined as outliers with the original λ_i, while the fine tuning, which slightly raises the critical values is helpful to refuse these fake peaks and pick out the forgery peaks accurately as well.

3.2. Identification algorithm

According to the description in section 2.2, there are at most two discontinuity peaks in the relative factor sequence, hence we set the upper bound number of outliers $r = 2$. Moreover, the generalized ESD test is carried out on both horizontal and vertical relative factor sequences, which helps to improve the identification accuracy. Let N_h and N_v denote the detected number of the discontinuity peaks in horizontal and vertical sequence, respectively. The flowchart of the identification algorithm is given in Fig. 3.

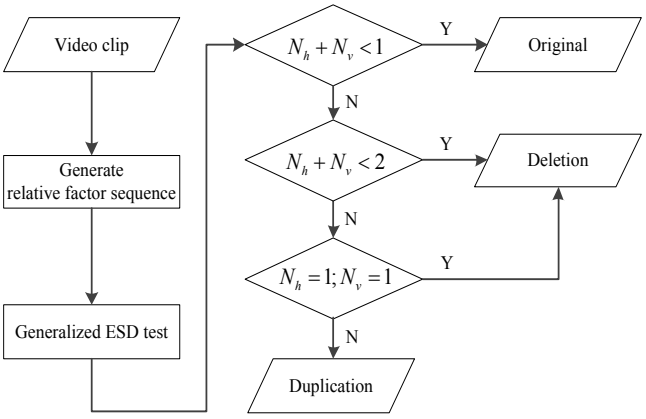
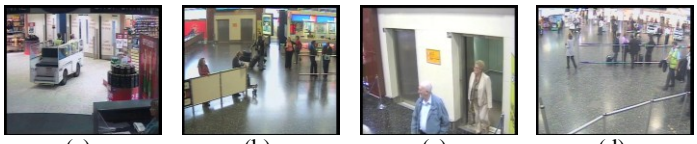


Fig. 3 Flowchart of the identification algorithm

Finally, the tampered location range is determined based on the relative factor sequence generation process in section 2. The upper bound of the range is $R_u = (P+1) \times 3 + \mathrm{mod}(L_{VFI}, 3)$, where P is the location of the detected peaks in relative factor sequence, "mod" is modulo operation, and L_{VFI} is the length of the corresponding VFI sequence. Hence, the tampered range is $[R_u - 2, R_u]$.



(a) (b) (c) (d)

Fig. 4. Four source videos with different scenes. (a) scene 1; (b) scene 2; (c) scene 3; (d) scene 4.

4. EXPERIMENTS

4.1. Video database

To the best of our knowledge, there is no open database for detecting video inter-frame forgery. Therefore, we have invited some volunteers to build one. Our four different scenes source videos (see in Fig. 4) are downloaded from TRECVID surveillance event detection evaluation [14]. Each source video split out 10 video clips. Each clip contains about 3000 frames with 720×576 resolution. Then the 40 video clips were delicately tampered to generate 40 frame deletion videos and 40 frame duplication videos (defined in section 2.2). Hence, there are totally 120 video clips in our final inter-frame forgery detection database. Note that all the tampered video clips were MPEG-2 re-coded with the same coding standard and parameters as the source videos.

4.2. Results and analysis

The configurations of identification algorithm are as follows. The upper bound number of outliers is $r = 2$, the significance level is $\alpha = 0.05$ and the coefficient of the critical values is $\eta = 1.5$.

4.2.1. Detection accuracy under random deletion
This experiment is to test the sensitivity of our algorithm by computing the detection accuracies when frames were randomly deleted. Table I shows the detection accuracies for randomly deleting 1 frame, 3 consecutive frames and 5 consecutive frames from original videos. The result illustrates that our algorithm could have good accuracy when detecting frame deletion forgery with a few frames removed.

4.2.2. Identification accuracy under meaningful forgery
The confusion matrices for the four scenes of video clips and the overall accuracy are given in Table II. The relatively low accuracies for frame duplication identification are due to the large intensity gaps between their two detected peaks, all the incorrect identification videos are identified as frame deletion forgery. However, the result demonstrates the effectiveness of our algorithm with overall 90.0%, 85.0% and 80.0% accuracies for identifying original video, frame deletion video and frame duplication video. If we only consider whether a video is tampered or not, the overall identification accuracy for the tampered videos is 96.3%, with 10% false positives. We did not do comparison experiments because no papers were found on identifying consecutive frame deletion and duplication forgeries.

4.2.3. Location accuracy under meaningful forgery
The location is considered to be incorrectly identified if one of the detected peaks in both horizontal and vertical VFI sequences is not in the expected range described in section 3.2. The location accuracies for correctly identified forged videos are given in Table III. All the locations of detected peaks in forged videos are correctly identified due to the statistics-based generalized ESD algorithm.

4.3. Robustness against compression

The robustness against lossy compression is tested in this experiment. Each video clips was re-compressed by ffmpeg software with different Qscales (a parameter to control video quality). The identification results with Qscale=1 (lossless compression), 2, 3 are shown in Table IV. When re-compressing with Qscale=2, the bit rate has averagely decreased by 3%, so the accuracy is the same with the result of Qscale=2. While when re-compressing with Qscale=3, the bit rate dropped a lot (with 30%), the duplication identification accuracy have slightly decreased. The reason is that the intensity gap between the two detected peaks enlarged after re-compression, which makes it easy to be identified as frame deletion forgery. Anyway, the accuracies report the robustness of our algorithm to some degree of compression.

5. CONCLUSION

We have proposed a new algorithm to detect video inter-frame forgery. This method is based on the consistency of velocity field. With consecutive frame deletion and frame duplication forgery operations, some discontinuity peaks can be observed in VFI sequence. And the generalized ESD test is applied to extract the peaks and identify the forgery type. Experiments show the effectiveness of our algorithm.

ACKNOWLEDGMENT

The work of this paper is sponsored by the National Natural Science Foundation of China (No.61272439, 61272249), the Specialized Research Fund for the Doctoral Program of Higher Education (No.20120073110053), and the Fund of State Key Laboratory of Software Engineering, Wuhan University (No.SKLSE2012-09-12). It was also under the Project of International Cooperation and Exchanges supported by Shanghai Committee of Science and Technology (No. 12510708500).

Table I Detection accuracies for random deletion.

Deleted frame number	1	3	5
Accuracy	40%	65%	80%

Table III Location accuracies under meaningful forgery.

Forgery type	deletion	duplication
Accuracy	100%(34/34)[a]	100%(32/32)

[a] (n/m) indicates n of m locations are correctly identified.

Table IV Detection accuracies under different Qscales (%).

Qscale	1	2	3
original	90.0	90.0	90.0
deletion	85.0	85.0	85.0
duplication	80.0	80.0	62.5

Table II Confusion matrix for each scene and their overall accuracy (%). – denotes value 0.

Video	Scene 1			Scene 2			Scene 3			Scene 4			Overall		
Forgery type	ori[a]	del[b]	dup[c]	ori	del	dup	ori	del	dup	ori	del	dup	ori	del	dup
ori	80.0	10.0	10.0	90.0	10.0	-	100	-	-	90.0	10.0	-	90.0	7.5	2.5
del	20.0	70.0	10.0	-	90.0	10.0	10.0	90.0	-	-	90.0	10.0	7.5	85.0	7.5
dup	-	30.0	70.0	-	20.0	80.0	-	30.0	70.0	-	-	100	-	20.0	80.0

[a] Original video type. [b] Frames deletion video type. [c] Frames duplication video type.

REFERENCES

[1] Milani S., Fontani M., Bestagini P. et al., "An overview on video forensics," APSIPA Transactions on Signal and Information Processing, 2012.

[2] Wang W. H. and Farid H., "Exposing digital forgeries in video by detecting double MPEG compression," In proceedings of the 8th ACM workshop on Multimedia and Security, pp.37-47, 2006.

[3] Chen W. and Shi Y. Q., "Detection of double MEPG video compression using first digits statistics," Digital Watermarking, Springer Berlin Heidelberg, pp.16-30, 2009.

[4] Wang W. H. and Farid H., "Exposing digital forgeries in video by detecting double quantization," In proceedings of the 11th ACM Workshop on Multimedia and Security, pp.39-47, 2009.

[5] Mondaini N., Caldelli R., Piva A. et al., "Detection of malevolent changes in digital video for forensic applications," In proceedings of the society of photo-optical instrumentation engineers, vol.6505, pp.T5050-T5050, 2007.

[6] Hsu C.C., Hung T.Y., Lin C.W. et al., "Video forgery detection using correlation of noise residue," IEEE 10th Workshop on Multimedia Signal Processing, pp.170-174, 2008.

[7] Kobayashi M., Okabe T., Sato Y., "Detecting Forgery From Static-Scene Video Based on Inconsistency in Noise Level Functions, " IEEE Transactions on Information Forensics and Security, vol.5, pp.883-892, 2010.

[8] Wang W. H. and Farid H., "Exposing Digital Forgeries in Video by Detecting Duplication," In proceedings of the 9th ACM Workshop on Multimedia and Security, pp.35-42, 2007.

[9] Conotter V., O'Brien J.F., Farid H., "Exposing Digital Forgeries in Ballistic Motion," IEEE Transactions on Information and Security, vol.7, pp.283-296, 2012.

[10] Chao J., Jiang X. H. and Sun T. F., "A novel video inter-frame forgery model detection scheme based on optical flow consistency," International Workshop on Digital Forensics and Watermaking, pp.267-281, 2012.

[11] Grant I., "Particle image velocimetry: A review," In proceedings of the Institution of Mechanical Engineers, vol. 211, pp.55-76, 1997.

[12] Thielicke W., Stamhuis E. J., "PIVlab- Time-Resolved Digital Particle Image Velocimetry Tool for MATLAB version 1.32," 2010.

[13] Iglewicz B., Hoaglin D.C., "How to Detect and Handle Outliers," Milwaukee (Wisconsin): ASQC Quality Press, vol. 16, 1993.

[14] TREC Video Retrieval Evaluation, http://trecvid.nist.gov/