Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/cose

Computers & Security

A multi-channel approach through fusion of audio for detecting video inter-frame forgery

Tianqiang Huang[a,b,c], Xueli Zhang[a,b,c,*], Wei Huang[a,b,c], Lingpeng Lin[a,b,c], Weifeng Su[d]

[a] College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350007, China
[b] Fujian Digital Institute of Big Data Security Technology, Fuzhou 350007, China
[c] Fujian Provincial Engineering Research Center of Big Data Analysis and Application, Fuzhou 350007, China
[d] BNU-HKBU United International College, Zhuhai 519000, China

ARTICLE INFO

Article history:
Received 31 October 2017
Revised 18 April 2018
Accepted 23 April 2018
Available online 27 April 2018

Keywords:
Video forensics
Frame insertion and deletion forgeries
Copy-move forgery
Wavelet singularity analysis
Perceptual hashing
Quaternion discrete cosine transform (QDCT)

ABSTRACT

The forgery operation of digital video in the temporal domain is often accompanied by the synchronization of the audio channel operation. In this paper, we proposed a fusion of audio forensics detection methods for video inter-frame forgery. First, the audio channel of the video is extracted, and discrete wavelet packet decomposition and analysis of singularity points of audio signals are used to locate the forged singularity points. Next, features of each frame of the video are extracted with the perceptual hash and used to calculate the similarity between consecutive frames, to locate the forgery position in the video frame sequence. We fused the results of the audio channel and the video frame sequence channel. The QDCT feature is used to further fine detect the suspected forgery location. Our method can position replication source locations for copy-move forgery. Experiments show that our method has higher accuracy and better performance in comparison with similar methods, especially on the delete forgery operation.

1. Introduction

The authenticity and integrity of digital video has long been a hot topic in the field of information security. With the popularity of digital imaging devices and the growing power of multimedia editing software, people can easily modify images or videos. It can be determined if a whether video has been tampered with using active and passive methods. Passive forensics technology has a wider range of applications, because of its universality and convenience.

Recently, passive detection techniques for video tamper detection were studied by Wang and Farid (2009), Stamm et al. (2012), Feng et al. (2014), and Gironi et al. (2014) for MPEG and H.264 video frame tamper detection. These algorithms are limited by the encoding format used by the video. HUANG et al. (2011) proposed a method using gray values to depict the video content, and determine if the correlation continuity of the inter-frame is changed. With the purpose of improving the efficiency, the Chebyshev inequality was used twice to adaptively set the threshold and detect outliers. YIN et al. (2014) proposed that the video sequence is represented by a four-dimensional tensor. The video tensor is decomposed using the Tucker decomposition, where each frame is extracted using the dimensions of the time-dimension factor matrix to represent the content. The relevance is then calculated to

[*] Corresponding author at: College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350007, China.
E-mail address: sherryiszxl@outlook.com (X. Zhang).

https://doi.org/10.1016/j.cose.2018.04.013
0167-4048/© 2018 Elsevier Ltd. All rights reserved.

determine a frame insertion or deletion. Chao et al. (2012) proposed video inter-frame forgery detection based on optical flow consistency. The algorithm is based on the premise that the optical flow between the original video sequences is consistent, and the tampering of the video sequence will disturb this inconsistency. Zhang et al. (2015) proposed a method based on the consistency of quotients of mean of structural similarity (MSSIM) to detect video-frame insertion and deletion, According to the phenomena, MSSIM between every two adjacent frames will mutate, and the Chebyshev inequality is used to detect outliers.

The above algorithm has a good detection effect on most videos that have been tampered with by inter-frame forgery. However, if the detection of video image sequences changes violently, the detection methods only based on the similarity between adjacent frames will cause a large number of false detections. Furthermore, if the deletion does not affect the continuous video clips for video sequences with static backgrounds, the detection algorithm for adjacent inter-frame similarity is prone to missed detections. The above tampering operation is difficult to detect on the sequence of video frames, but will cause the audio signal to be obviously discontinuous. Therefore, we propose fusion audio forgery detection to detect video tampering.

Audio is an important part of video. To ensure consistency of the tampered video content, the audio channel will suffer the same tampering operation when people tamper the video. However, most existing detection algorithms are only based on the video frame sequence, ignoring the importance of audio. If only detecting audio channel forgery, some audio noise may affect the test results, resulting in misjudgment. We detected both the audio channel and video sequences, which can more accurately locate the tampering position. Fusing audio channels for video forgery detection can improve the detection accuracy and reduce the leakage detection. Compared to the current video detection algorithm, the proposed method not only has better performance in detection forgery, but can also determine the tampering form and locate the position of copy-move area.

In this study, discrete wavelet packet decomposition and analysis of singularity points of audio signals is used to detect audio channel forgeries (Chen et al., 2016). The location of the audio frame and the video frame sequence is combined with the location of the video frame sequence using the perceptual hash location. For the suspected tampered position, the image quaternion discrete cosine transform (QDCT) coefficient is used to further refine detection, which includes the color information. In addition, we proposed a method to determine the tamper form of the video and locate the source replication area of copy-move forgery.

The innovative works of the dissertation are listed as follows: (a). We first put forward the innovative algorithm which fused the audio and video to detect inter-frame forgery forensics, and has greatly improved the detection accuracy. (b). The algorithm solved the problem of high false detection rate for deletion tampering forensics. (c). An innovational method for estimating the forgery operation form of tampering position is proposed.

The remainder of this paper is organized as follows. Section *Methods* introduces the main structure of our algorithm. Among it, in subsection *Audio forensics detection*, we introduce a method of audio channel tampering forensics. In subsection *Perceptual hash*, we present a detecting method of video sequence channel by perceptual hash. Subsection *Fuse the results of the audio channel and the video frame sequence channel* describes the method of merging the audio channel and video channel forensics results in the time domain. Subsection *The fine detection* introduces a method of reducing the missed detection by using the QDCT feature to further finely detect the suspected forgery points that fails to be accurately determined after the audio-video channel fusion. In subsection *Estimate forgery form*, we propose a method to estimate the forgery form. In Section *Experimental results and discussion*, we describe our comprehensive experiments and provide an analysis of the results. Section *Conclusion* concludes the paper.

2. Methods

2.1. *Audio forensics detection*

There is a strong correlation between the sample point and its neighbors in natural digital audio without tampering. However, the artificial tampering of natural digital audio (such as cutting, inserting, replacing or splicing) will result in the proximity of sampling points in the time domain. The correlation will decrease or disappear, thereby leading to abnormal mutation points known as singular points (Chen et al., 2016).

Step 1: For the audio signal extracted in the video, the wavelet packet is decomposed into 30 sub-bands by performing four levels of wavelet packet decomposition. We reconstructed these sub-bands and denote them as $Rsub_x$, $x = \{1, 2, \cdots, 30\}$, so that the length of each is the same as the original signal.

Step 2: Compute the absolute value of $Rsub_x$, then calculate the mean, denoted as $Msub_x$. The local peak corresponding to the singularity of the tampering is much larger than the mean value of the sub-bands' absolute value. We set $Ratsub_x = \frac{Max_1}{Msub_x}$, where Max_1 is the local peak of the sub-band $Rsub_x$. We design a parameter Ts that is a threshold. If $Ratsub_x$ is greater than the set threshold value Ts, then Max_1 is a singular point. To further reduce the amount of computation, when $Ratsub_x \geq 3 \times Ts$, $Ts = \frac{Ratsub_x}{2}$.

Step 3: The digital audio signal has some inherent singularity, but is generally a cluster. However, the tamper-generated singularity is an isolated one. We can determine if Max_1 is a forged singularity point, by determining whether Max_1 remains in the form of the group, that is, there is no other singular point group around it. The local peak Max_1 is the origin. If there is no other peak in the $[-c2, c1 - c2] \cup [c1, c2]$ coordinate range, as shown in Fig. 1, we consider Max_1 as a forged singularity point.