J. Vis. Commun. Image R. 35 (2016) 55–66

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Double compression detection based on local motion vector field analysis in static-background videos ☆

Peisong He, Xinghao Jiang *, Tanfeng Sun, Shilin Wang

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China
National Engineering Lab on Information Content Analysis Techniques, GT036001 Shanghai, China

ARTICLE INFO

Article history:
Received 22 September 2015
Accepted 26 November 2015
Available online 11 December 2015

Keywords:
Video forensics
Static-background videos
Double MPEG-4 compression
Background segmentation
Local motion vector field
Periodic analysis
GOP estimation
Bit-rate control mode

ABSTRACT

Videos captured by stationary cameras are widely used in video surveillance and video conference. This kind of video often has static or gradually changed background. By analyzing the properties of static-background videos, this work presents a novel approach to detect double MPEG-4 compression based on local motion vector field analysis in static-background videos. For a given suspicious video, the local motion vector field is used to segment background regions in each frame. According to the segmentation of backgrounds and the motion strength of foregrounds, the modified prediction residual sequence is calculated, which retains robust fingerprints of double compression. After post-processing, the detection and GOP estimation results are obtained by applying the temporal periodic analysis method to the final feature sequence. Experimental results have demonstrated better robustness and efficiency of the proposed method in comparison to several state-of-the-art methods. Besides, the proposed method is more robust to various rate control modes.

1. Introduction

With the development of capturing devices and network transmission techniques, digital video has become an important method for people to satisfy their demands for information and entertainment. In many practical applications, such as video surveillance and video conference, videos captured by static cameras (referred to as static-background videos) have been widely used. To facilitate transmission and storage, these videos are often stored in the lossy compression format. In the past few years, MPEG-4 standard has been adopted in many video surveillance systems and employed for video content sharing over the Internet [1]. However, the popularization of video editing software enables attackers to change the semantics of the original video content easily, which has posed a threat to the authenticity of digital videos. To detect the manipulation of digital video, great efforts have been devoted to active authentication solutions, such as digital watermarking [2]. The main limitation of these methods is that watermark information should be extrinsically embedded in the recording process. Contrary to active approaches, the passive forensics techniques have been proposed only utilizing the traces left during the capturing process and editing operations. Therefore, passive video forensics has attracted more and more research interests in the last decade [3].

In real forensics scenarios, most of tampering processes are conducted on decompressed domain and tampered videos have to undergo the recompression process. The analysis of double compression can be regarded as the first step of video forensics. The pioneering work in video double compression detection has been done by Wang and Farid [4], who used the Discrete Fourier Transform (DFT) of the average prediction residual sequence to expose the trace left by double MPEG compression. In most video coding standards, Group of Pictures (GOP) is used as the encoding unit. The first frame of a GOP is the intra-coded frame (I-frame) which does not depend on other reference frames. The rest frames in a GOP employ motion estimation to reduce temporal redundancy, e.g. predictive frames (P-frames) and bi-directionally predictive frames (B-frames). Therefore, double compression can be classified into two categories according to whether the structure of GOP between the primary and the secondary compression are matched or not.

For double compression with the matched GOP structure (referred to as Type I double compression hereinafter), some successful detection methods have been proposed. The basic idea is that, in this kind of double compressed videos, the re-encoded I-frame has the similar properties of double JPEG compression.

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author at: School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, China.

E-mail addresses: gokeyhps@sjtu.edu.cn (P. He), xhjiang@sjtu.edu.cn (X. Jiang), tfsun@sjtu.edu.cn (T. Sun), wsl@sjtu.edu.cn (S. Wang).

http://dx.doi.org/10.1016/j.jvcir.2015.11.014
1047-3203/

According to different types of cues used in double compression detection with the matched GOP structure, published methods can be categorized into four groups: (i) DCT coefficient distribution based methods [5–7]; (ii) first digit statistics based methods [8,9]; (iii) Markov statistics based methods [10,11]; (iv) methods based on other cues [12–14].

For double compression with different GOP structures (referred as Type II double compression), most of the above-mentioned detection methods fail to achieve satisfactory performance. It is because the GOP structure will shift during the recompression process and some original I-frames are coded as P-frames in the second compression. Since the fixed GOP is usually used, the periodic analysis of the temporal feature sequence is efficient to detect double compression with different GOP structures. According to different types of cues used in double compression detection with different GOP structures, published methods can be categorized into three groups:

1. *Prediction Residual Based Methods*: Wang and Farid [4] used the periodicity of the average prediction residual sequence to detect double compression with shifted GOP. In [15], the spectrum of the average prediction residual sequence was utilized to reveal the trace left by double compression. However, methods in [4,15] both required the human identification of tampering fingerprint. Stamm et al. [16] developed an automatic double compression detection technique based on [4] using peak detection and the corresponding anti-forensics method.
2. *Variation of Macroblock Prediction Types Based Methods*: Vázquez-Padín et al. [17] proposed a technique based on the variation of macroblock prediction types in the re-encoded P-frames. It has robust detection ability with various settings of compression parameters. Besides, this scheme is able to estimate the first GOP size. Based on the method in [17], Gironi et al. [18] proposed an inter-frame tampering detection scheme and Labartino et al. [19] proposed an intra-frame tampering detection scheme.
3. *Method Based on Other Cues*: Luo et al. [20] considered this issue by analyzing block artifact of recompression videos after removing different frames for a given sequence. However, the final decision requires human identification. Su et al. [21] proposed two potential forensics methods which are able to counter the proposed anti-forensics method in H.264 videos. The first method is to employ the deblocking filter of H.264/AVC and the second method is related to the selection of QP. He et al. [22] applied an adaptive post-filtering technique to extract block artifact measurement for double compression detection in MPEG-4 videos.

Compared with Type I double compression detection, Type II double compression detection is more complicated, since the original GOP size is unknown to analysts and the inter-coding process is more complex. Existing studies about Type II double compression detection focus on different cues in both compression domain [4,15–17] and decompression domain [20–22], but few of them consider the relationship between the compression artifact and video content. In static-background videos, background regions usually have strong temporal redundancy. For this kind of videos, the tampering operation is more easily to be conducted without any noticeable visual distortion, since the background regions between adjacent frames are nearly consistent. However, due to lack of consideration about different properties between backgrounds and foregrounds, conventional Type II double compression detection methods are not robust and efficient to deal with static-background videos.

To solve this problem, we proposed a novel Type II double MPEG-4 compression detection method based on local motion vector field analysis in static-background videos. Different from conventional double compression detection methods, the proposed method considers the varied motion strength of foregrounds and the temporal consistence of backgrounds to extract more robust prediction-residual-based fingerprints. For a given video, the local motion vector field analysis is applied to roughly segment backgrounds and foregrounds in each frame. According to the macroblock-wise segmentation and strengths of motion vectors in each macroblock, the prediction residuals in background regions and slowly moving foregrounds are retained while the disturbance of fast moving foregrounds is reduced. The post-processing operation is applied to further enhance the robustness of the modified prediction residual sequence against various coding parameter settings. Experimental results demonstrate that the proposed method outperforms several state-of-the-art methods in double compression detection and GOP estimation in static-background videos.

The rest of this paper is organized as follows. Section 2 provides the analysis of double compression in MPEG-4 static-background videos. Section 3 presents the proposed double compression detection scheme. In Section 4, the performance of the proposed method is evaluated in comparison to the state-of-the-art alternatives. Section 5 draws the conclusion.

2. Analysis of double compression in MPEG-4 static-background videos

In this section, the process of Type II double compression is first modeled and several properties about Type II double compression in static-background videos are theoretically investigated. Then, the temporal model for prediction residual in Type II double compression is introduced. Finally, an illustrative example is presented. For convenience, if not specified, double compression refers to Type II double MPEG-4 compression in static-background videos hereinafter.

2.1. Modeling the double compression process

We assume that $\mathbf{F} = \{F_t, t = 0, 1, \ldots, T-1\}$ is the uncompressed image sequence with the resolution $M \times N$, where T denotes the total number of frames. To simplify, only I-frames and P-frames are considered.

2.1.1. The procedure of the single compression

As show in Fig. 1, F_{t-1} is inter-coded while its following uncompressed image F_t is intra-coded in single compression. They belong

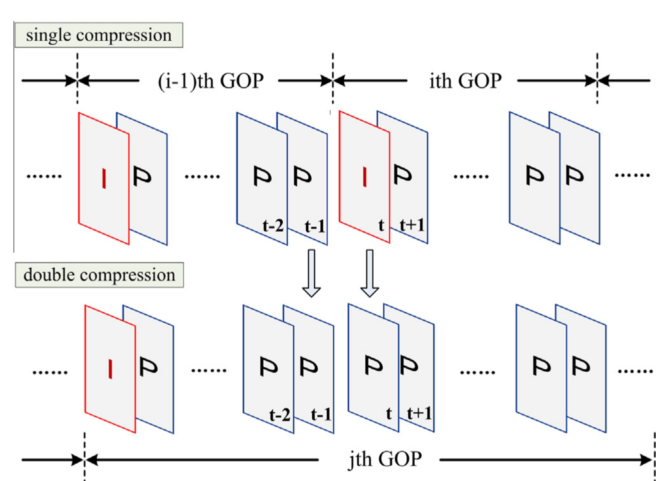

Fig. 1. The process of double compression with different GOP structures.