Digital Investigation 9 (2012) 151–159

Contents lists available at SciVerse ScienceDirect

Digital Investigation

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

A MCEA based passive forensics scheme for detecting frame-based video tampering

Qiong Dong, Gaobo Yang*, Ningbo Zhu

School of Information Science & Engineering, Hunan University, Lushan Road, Changsha, 410082 Hunan, China

ARTICLE INFO

Article history:
Received 4 February 2012
Received in revised form 23 May 2012
Accepted 10 July 2012

Keywords:
Digital media forensics
Passive forensics
Motion-compensated edge artifact
Fourier transform
GOP

ABSTRACT

Without the use of digital signature or digital watermark, video passive forensics only utilizes the statistical characteristics of digital video to verify its integrity and authenticity. For frame-based video tampering, it usually suffers from double MPEG compression. In this paper, a motion-compensated edge artifact (MCEA) based passive forensics scheme is proposed for detecting frame-based video manipulation. It exploits the MCEA difference between adjacent P frames, and the decision is made by judging whether there are any spikes in the Fourier transform domain after double MPEG compression. Experimental results show that the proposed approach is effective for frame-based tampering, such as adding/deleting frames and GOP structure change, and can predict the GOP structure of original video.

1. Introduction

With the wide availability of digital video camera and the prevalence of video sharing websites, digital videos are playing important roles in our daily life. Meanwhile, it is becoming much easier to manipulate and tamper digital video without leaving any visual clues with the continuous development of advanced video editing tools (Rocha et al., 2011). As a consequence, various video forgery operations for malicious purposes are more common than ever. There is an urgent need to develop effective forensics techniques for exposing those malicious video manipulations (Chuang et al., 2011). The conventional active methods must embed digital signature or digital watermark into video data in advance to verify its origin or authenticity. Passive video forensics aims at providing tools to support blind investigation because it utilizes only the statistical characteristics of digital video itself. Therefore, passive video forensics does not assume any a-priori knowledge about the original video, which appeals the research efforts in the field of information security.

Digital video can be regarded as an extension of digital image in the time axis. Though there are many works about digital image forensics, the research on digital video forensics is still in its infancy. The reasons are summarized as follows. First, the tampering of digital video is more sophisticated and time-consuming than digital image. Furthermore, due to the large amount of video data, it is usually encoded before storage and transmission. As a result, it is more difficult for video forensics. Second, since digital video has an additional temporal dimension, this brings some forgery operations specific to digital video, such as frame-based tampering. In this paper, we put emphasis on the passive forensics for detecting frame-based tampering.

For an MPEG video, it is usually re-saved in MPEG format after tampering operations. This leads to the so-called double MPEG compression in video forensics. In the literature, there are already several kinds of approaches for detecting double MPEG compression. The most representative algorithm proposed by Wang and Farid (2006) exploits the static and temporal artifacts introduced by double MPEG compression. I frame is viewed as a static image which is similarly

* Corresponding author. Tel.: +86 0731 88821341.
E-mail address: yanggaobo@hnu.edu.cn (G. Yang).

1742-2876/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.diin.2012.07.002

subjected to JPEG encoding, and the double JPEG compression detection algorithm is directly extended to double MPEG compression detection. In the temporal domain, it has been stated that motion compensation errors for P-frames are a function over time exhibiting a periodic pattern after frame deletions and recompression. However, this property can only be exploited with some constrains: The number of deleted frames must be multiple times of frame number in a GOP (Group of Picture), and the GOP structure must be kept during tampering. For the detection of GOP structure change in video tampering, Qin et al. (2010) propose a blind forensics technique based on GOP abnormality. It utilizes the Fourier analysis of motion errors. It is effective for the detection of video splicing. Luo et al. (2008) present a feature curve to reveal the compression history of an MPEG video file with a given GOP structure, and use the temporal patterns of block artifacts as evidence to detect tampering. Su and Zhang (2009) utilizes the motion-compensated edge artifacts (MCEA) for the exposing of digital video forgery. However, it needs a hard threshold factor α to detect frame-deleting forgery. Moreover, at least three P frames must be deleted. This seriously constrains its adaptability in practice.

MPEG-2 video system adopts a hybrid coding structure, which integrates these three classical techniques: prediction coding, transform coding and entropy coding. When coarse quantization is combined with motion compensation prediction, the blocking artifacts propagate from I-frames into subsequent frames and accumulate. This will cause structured high frequency noise. The MCEA involves high frequency noise within those blocks in every P frame. In one GOP, the P frames' MCEAs are non-decreasing. By observation, we found that the frame-based forgery operations, such as adding frames, deleting frames or changing the GOP structure, will make the MCEAs of adjacent P frames larger, and they are shown as a periodic characteristics. In this paper, a MCEA based passive forensics scheme is proposed for frame-based video tampering. It is in fact an improved algorithm on Su's work (Su and Zhang, 2009) for detecting double MPEG compression. The block diagram of the proposed approach is illustrated in Fig. 1. The MCEA difference sequences between adjacent P frames are exploited to judge whether there are any spikes in the Fourier transform domain after double MPEG compression. The main contribution of the proposed approach is that it overcomes the shortcomings of the hard threshold in the Su's work (Su and Zhang, 2009). It can not only detect the frame adding/deleting operations, but also is effective for the forensics of GOP structure change.

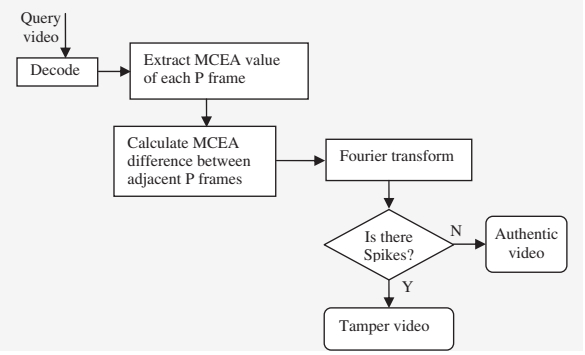


Fig. 1. Block diagram of the proposed approach.

The rest of this paper is organized as follows. In Section 2, the MPEG-2 video codec and re-compression process are briefly introduced. Section 3 discusses the calculation of the P frame's MCEA and its application for forensics. Section 3 presents the proposed video tampering detection algorithm. Experimental results are reported in Section 5, and conclusions and future work are given in Section 6.

2. MPEG-2 codec and double MPEG compression

MPEG videos are compressed by removing both the temporal redundancy and spatial redundancy. In the general MPEG architecture, there are three types of frames in video encoding: I (intra-coded) frame, P (forward predictive coded) frame and B (bi-directionally predictive coded) frame. Let N be the total number of frames in a given GOP structure, and M be the minimum distance between P-frames. For example, the GOP structure shown in Fig. 2 can be represented as ($N = 12$, $M = 3$).

Due to the complexity of bitstream syntax and the high correlations between adjacent frames, it is difficult to manipulate the MPEG video directly in compressed domain. For the video forgery operations, they usually decode the input video stream into frames, and then re-encode the forged video frames into MPEG. Therefore, the falsified video often suffers from the double MPEG compression. In this paper, we investigate two types of frame-based tampering and their re-compression.

2.1. Re-compression after deleting/adding frames

Fig. 3 shows an example of MPEG re-compression after frame deletion. The original video sequence is compressed with a GOP structure of ($N = 15$, $M = 3$). After deleting 6 consecutive frames, the rest frames are re-encoded with the same GOP structure. Apparently, some frames have changed their types because of frame deletion and re-compression. For example, the third P frame (marked in red) in each GOP is in fact I frame of previous GOP in original sequences. The change of frame types and double MPEG-2 compression will lead to some changes in the statistical characteristics of P frame, which can be used as clues for tampering detection. The most common clues used in existing work include block artifacts (Luo et al., 2008), motion errors (Su and Zhang, 2009), motion vectors (Qin et al., 2010), prediction residuals (Xiong et al., 2008) and pattern noises (Wang et al., 2008).

2.2. Re-compression with changing GOP structure

Fig. 4 shows an example of re-compression with the GOP structure change. The original sequence is encoded with GOP (12, 3), and then it is re-compressed with GOP (15, 3). It can be seen that some P frames in the re-compressed sequences are derived from I frames in the original GOP, and this derivation is periodic. Therefore, it can be used to detect whether an MPEG file has been double compressed or not. Furthermore, if the original GOP

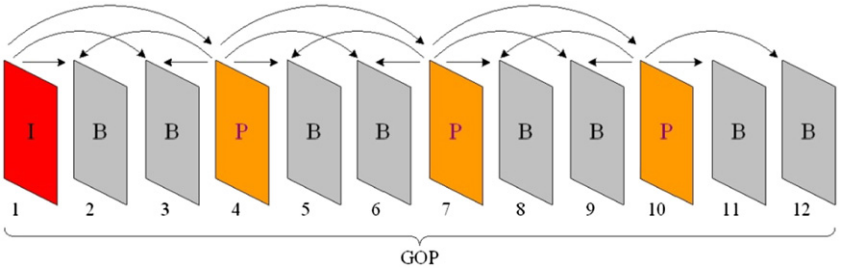


Fig. 2. A GOP structure ($N = 12$, $M = 3$).

structure can be conjectured or calculated from the re-compressed video, it will be desirable for forensic.

3. The calculation of MCEA and its application in forensics

For typical in video codec, when coarse quantization is combined with motion compensation prediction, the blocking artifacts propagate from I-frames into subsequent frames. It causes structured high frequency noise that is no longer located at block boundaries. These kind of motion-compensated edge artifacts (MCEA) are referred to be false edges, and their energies accumulate in each GOP (Leontaris et al., 2007). As a no-reference video quality metrics, MCEA is proposed to measure the impairment of motion compensated edge artifacts. It is an energy-based approach, using both compressed bitstream information and decoded pixels. In this paper, MCEA is introduced into video forensics. As what illustrated in Figs. 3 and 4, when some frames are deleted or the GOP structure is changed during video forgery operations, the additional high frequency noise will increase. In the following, we firstly make a brief introduction on how to compute the MCEA values of P frame.

3.1. The computation of P frame's MCEA

The energy of MCEA is calculated on a block basis. The following terminologies are defined for simplicity.

T: the set of all 8×8 blocks in a frame;
$C_\tau^n(i,j)$: the received DCT coefficient at location (i,j) in an 8×8 block τ of frame n ($\tau \in T$);
$m_\tau^n(i,j)$: the measured DCT coefficient obtained from the reconstructed frame;
$s_\tau^n(i,j)$: a DCT coefficient from a block in the original source frame;
$p_\tau^n(i,j)$: a DCT coefficient from the motion-compensated prediction block.

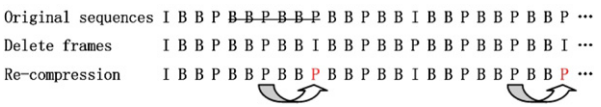


Fig. 3. The MPEG re-compression after deleting 6 frames. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We consider the set N of all AC DCT coefficients. The encoder compresses and transmits the prediction residuals $s_\tau^n(i,j) - p_\tau^n(i,j)$ with available bits, resulting in quantized residuals $c_\tau^n(i,j)$, which help reduce prediction errors. The transmitted residual DCT energy is

$$C_\tau^n = \sum_{(i,j)\in N} \left(c_\tau^n(i,j)\right)^2 \tag{1}$$

and the measured DCT energy is

$$M_\tau^n = \sum_{(i,j)\in N} \left(m_\tau^n(i,j)\right)^2 \tag{2}$$

Using the measured DCT coefficients in the reconstruction process of $m_\tau^n(i,j)$ and the received coefficients $c_\tau^n(i,j)$, the energy P_τ^n in the prediction block can be exactly computed as follows:

$$P_\tau^n = \sum_{(i,j)\in N} \left(m_\tau^n(i,j) - c_\tau^n(i,j)\right)^2 \tag{3}$$

Given the decoded pixels and the compressed bitstream, P_τ^n, M_τ^n and C_τ^n are known. MCEA is designed as a non-reference metric that estimates the percentage of high-frequency energy in M_τ^n, which is not part of the original image content.

For a given block τ in frame n, let $\sigma(\tau)$ be the set of (up to) four on-grid blocks in frame $n-1$ that are used to predict block τ, and $\omega(\beta)$ be the weighting coefficient for block β, $\beta \in \sigma(\tau)$. Then, the energy E_τ^n can be estimated as follows.

$$E_\tau^n = \sum_{\beta\in\sigma(\tau)} \omega(\beta) \sum_{(i,j)\in N} \left(m_\beta^{n-1}(i,j)\right)^2 \tag{4}$$

It approximates the energy content of the source block τ in current frame. The recursive per-block MCEA energy metric is obtained as follows:

$$\mu_\tau^n = \left(P_\tau^n - E_\tau^n\right) - C_\tau^n + \sum_{\beta\in\sigma(\tau)} \omega(\beta)\mu_\beta^{n-1} \tag{5}$$

For the whole frame n, its MCEA is the average of μ_τ^n for all the blocks. That is,

$$MCEA = \frac{\sum_{\tau\in T} \mu_\tau^n}{M^n} \tag{6}$$

Fig. 4. The original GOP (12, 3) re-encode to GOP (15, 3).

where $M^n = \sum_{\tau \in T} M_\tau^n$ is the measured energy content of the entire frame n. The P frames' MCEA can be calculated by the Formula (6). It increases with the distance d from the most recent I-frame in one GOP.

3.2. The application of MCEA into video forensics

The clues for video forensics include intrinsic fingerprints left by in-camera operations and extrinsic artifacts left by the signal processing process of video tampering. Blocking artifacts, ringing artifacts and blurriness are the most common artifacts for digital forensics. In this paper, we investigate the quality degradations of P frame via MCEA caused by double compression. The adopted no-reference metric (MCEA) does not require any information about the original video, thus may allowing an efficient and passive scheme for frame-based video tampering.

For the frame deletion operation, it leads to the decrease of temporal distance for adjacent frames in the re-encoded video sequence. As a result, the motion compensated errors will be bigger. In Fig. 3, we observe that the third P frame of a GOP in re-encoded video sequence comes from I frame in the original GOP. Therefore, the MCEA distribution of P-frames in one GOP is directly influenced because the coding types change between I frame and P frame. This provides useful clues for tamper detection. In the Su's work (Su and Zhang, 2009), an impact factor α as Equation (7) is defined to represent the change of P-frame's MCEA distribution in a GOP. $MCEA^i$, $MCEA^{i+1}$, $MCEA^{i+2}$ denote the MCEA values of the first, second and third P-frame in a GOP, respectively.

$$\alpha = \frac{|MCEA^{i+2} - MCEA^{i+1}|}{|MCEA^{i+1} - MCEA^{i}|} \tag{7}$$

We have repeated the experiments in (Su and Zhang, 2009) to test the effectiveness of impact factor α. Two test video sequences, *Carphone* and *Hall* are selected for experiments (Testing samples). The original YUV sequences are encoded into MPEG video with GOP (15, 3), and then the α in each GOP is computed. After deleting 6 frames, the factor α in each GOP is also calculated in the re-encoded MPEG video. The experimental results are illustrated in Fig. 5. In the first row of (a) and (b), the left is the MCEA values of each P frames in the original MPEG video, the right is the correspondent impact factor α in each GOP. After deleting 6 frames, the factor α in each GOP is also calculated in the re-encoded MPEG video. Similarly, in the second row of (a) and (b), the left is the MCEA values of each P frames in tampered MPEG video, the right is the correspondent impact factor α in each GOP. Apparently, the factor α might exceed the range (0.5, 2) for true video, and the factor α of tampered video might falls in this scope as well. Therefore, it is difficult judge whether the video frames have been tampered simply by a factor α. Furthermore, the calculation of α needs at least three P frames in one GOP. It will fail for those video with only 2 frames in a GOP, such as becomes ineffective in only two P frames, such as GOP (9, 3).

4. The proposed forensics approach for frame tampering detection

4.1. The FFT of MCEA difference

For one GOP (N, M), the number of P frames N_p is $N/M - 1$. According to the Formula (6), the MCEA values of all the P frames can be calculated, and a difference sequence ΔM is defined as the difference of MCEA between adjacent P frames.

$$\Delta M = MCEA^i - MCEA^{i+1}, \quad i \in (1, N_p - 1) \tag{8}$$

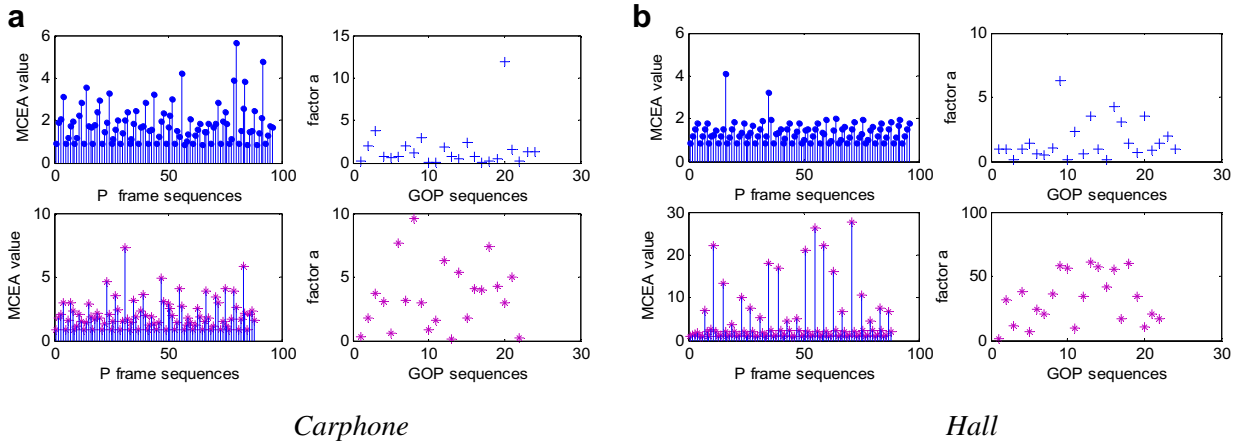


Fig. 5. MCEA value sequences of adjacent P frames and factor in each GOP.

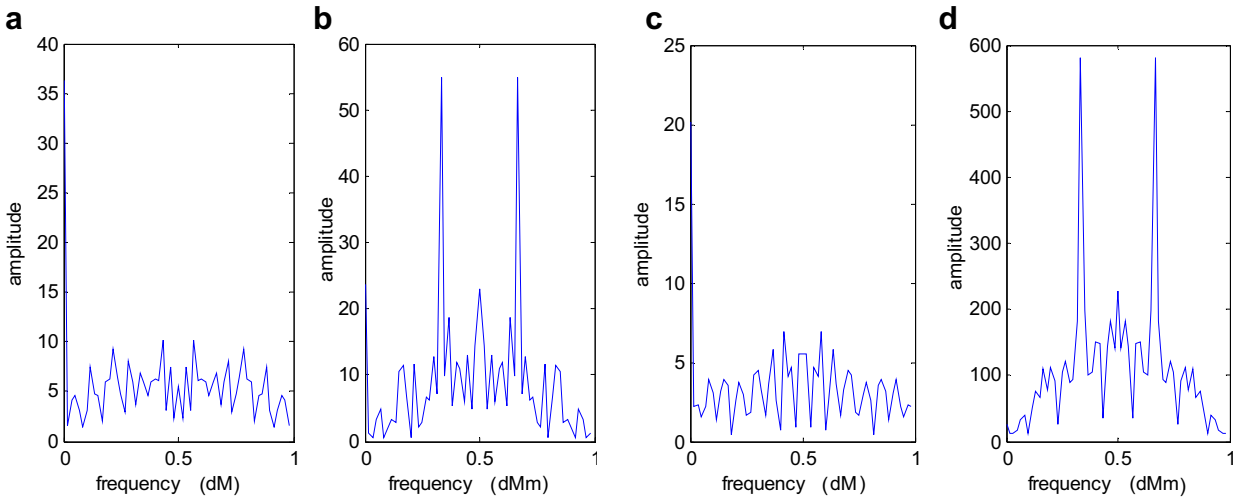


Fig. 6. Fourier transform spectrum: (a) true video, *Carphone* (b) after deleting 6 frames, *Carphone*; (c) true video, *Hall* (d) after deleting 6 frames, *Hall*.

For all the GOP groups in a sequence, their difference sequence dM can be computed. The change of dM is relatively steady. However, the double recompression after frame manipulation or changing the GOP structure brings greater motion compensation errors of P frame. Since the adjacent P frames might come from different GOP in the original sequence, some kinds of transitions will occur in their ΔM. Moreover, they periodically occur in the whole sequence. There will appear apparent peak spikes in the spectrum after discrete Fourier transform. Fig. 6 gives the experimental results of the Fourier transform spectrum for real video sequences and the forged sequence after frame deletion according to Fig. 3. Apparently, the Fourier transform spectrum is relatively smooth, yet there are obvious spikes in the tampered video. The spikes in the FFT spectrum are important and useful clues for the forensics of frame-based manipulation. When there are spikes, further judgment is to determine the tampering operation is frame-deleting or changing the GOP structure. For the re-compression after deleting frames, there is only a periodic pattern and the main peak is $1/(N_p - 1)$. For the re-compression with GOP structure change, there are several cycles that present multi-spikes in spectrum. Therefore, it can be used as a clue to predict the original GOP structure before tampering. It is discussed in next section.

4.2. Calculate the original GOP structure

For the GOP structure change in the video tampering process, the positions of P frames between adjacent GOP will change. Consequently, there occur multiple spikes in the Fourier spectrum. Without loss of generality, we suppose that the original GOP ($N1$, $M1$) is changed to GOP ($N2$, $M2$) in the double MPEG compression. In Fig. 4, an example of the GOP structure change is illustrated. In order to explain the pattern of re-compression more clearly, the B frames in Fig. 4 can be omitted as Fig. 7. Then, the deduction of original GOP structure can be simplified as follows.

The GOP lengths are 12 and 15 before and after re-compression, respectively. That is, $N1 = 12$ and $N2 = 15$.

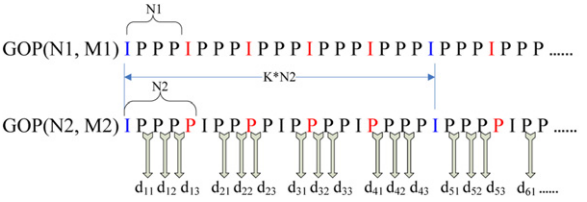


Fig. 7. The simplified GOP structure change during re-compression. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

According to the principle of least common multiple, it is easy to find the value of K as follows:

$$K = [N1, N2]/N2 \tag{9}$$

$[N1, N2]$ is the least common multiples of $N1$ and $N2$. Let $\{d_{11}, d_{12}, d_{13}, d_{21}, d_{22}, d_{23}, \ldots\ldots\}$ represent the MCEA difference sequence. Due to those P frames (marked in red) in the re-compressed sequence, the stationary difference sequences appear a transition. For example, d_{13} is one kind of difference transition, which again appears in d_{53}, and this transition periodically occurs in the whole sequence. The period T_0 is equal to 3^*K in the Fig. 7, that is $K^*(N_p - 1)$. There will appear apparent peak spikes in the FFT

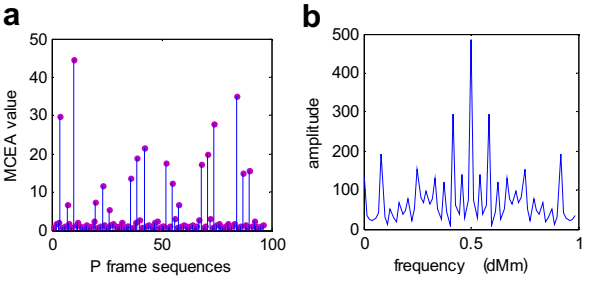


Fig. 8. The MCEA difference and its FFT spectrum of original sequences (12, 3) re-compress to GOP (15, 3).

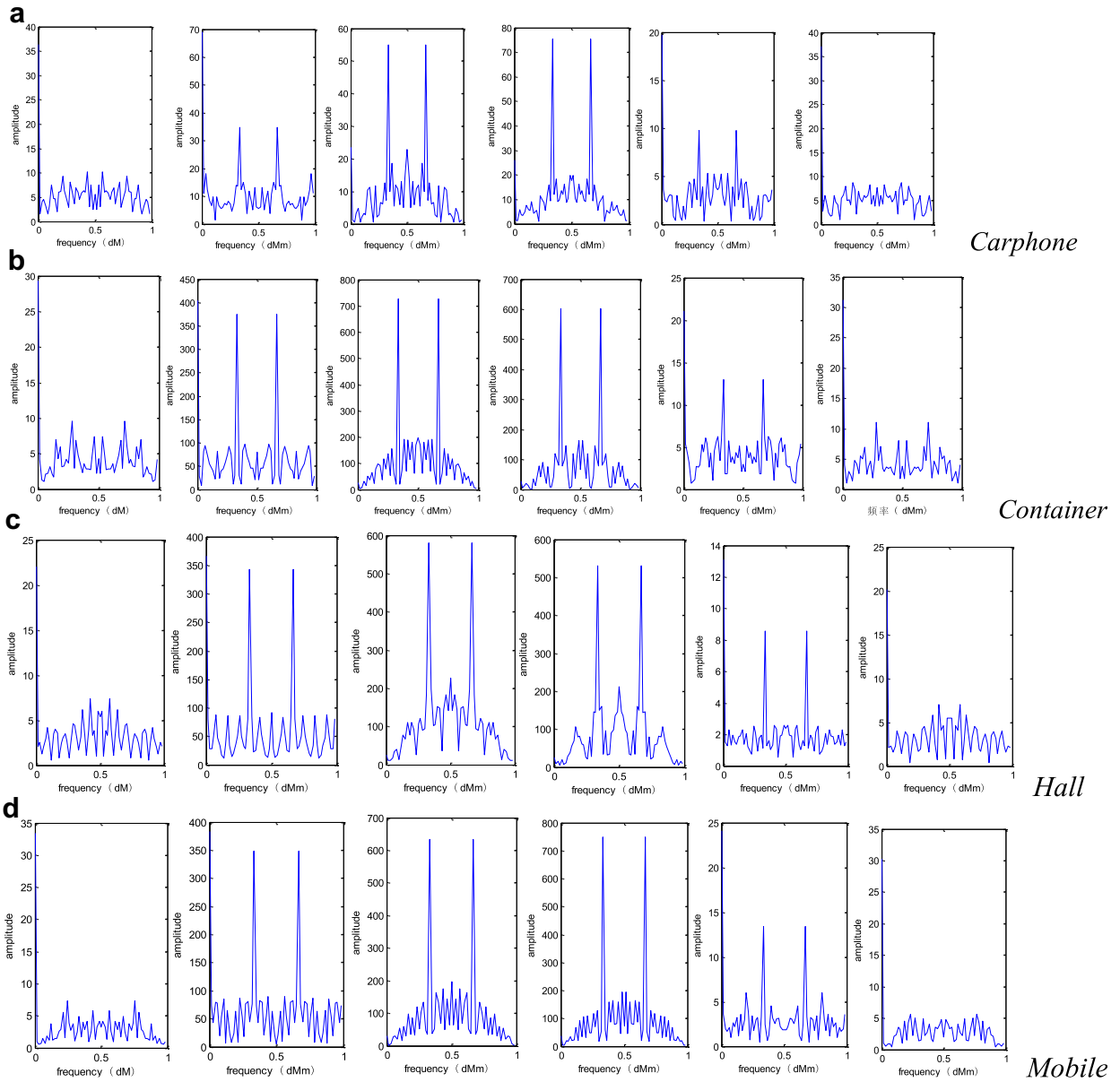


Fig. 9. The FFT magnitude of MCEA difference sequence for the original sequence and frame-deleted video.

Table 1
The main frequency of different GOP structure re-compression and the period T_0.

Re-compression	Main frequency	T_0	[$N1$, $N2$]
GOP(12,3)-GOP(15,3)	0.0833, 0.4167, 0.500, 0.5833, 0.9167	12	60
GOP(9,3)-GOP(15,3)	0.2167, 0.4500, 0.5500, 0.7833	9	45
GOP(15,3)-GOP(9,3)	0.2000, 0.4000, 0.6000, 0.8000	5	45
GOP(12,3)-GOP(9,3)	0.2500, 0. 5000, 0.7500	4	36

spectrum, and its corresponding frequency can be expressed as Formula (10):

$$f_0 = \frac{1}{T_0} = \frac{N2}{[N1, N2]} * \frac{1}{N_p - 1} \tag{10}$$

In all cases, the other spikes appear in the multiplication times of f_0, shown in Fig. 8. The left figure represents the MCEA values of each P frames in re-compression sequences, the right figure represents its FFT spectrum of MCEA difference, and it is showed that $f_0 = 1/12 = 0.0833$. Therefore, the source GOP structure can be deduced by finding the greatest common divisor of all the spikes' frequencies. The other case of changing GOP structure re-compression can be conducted in a similar way.

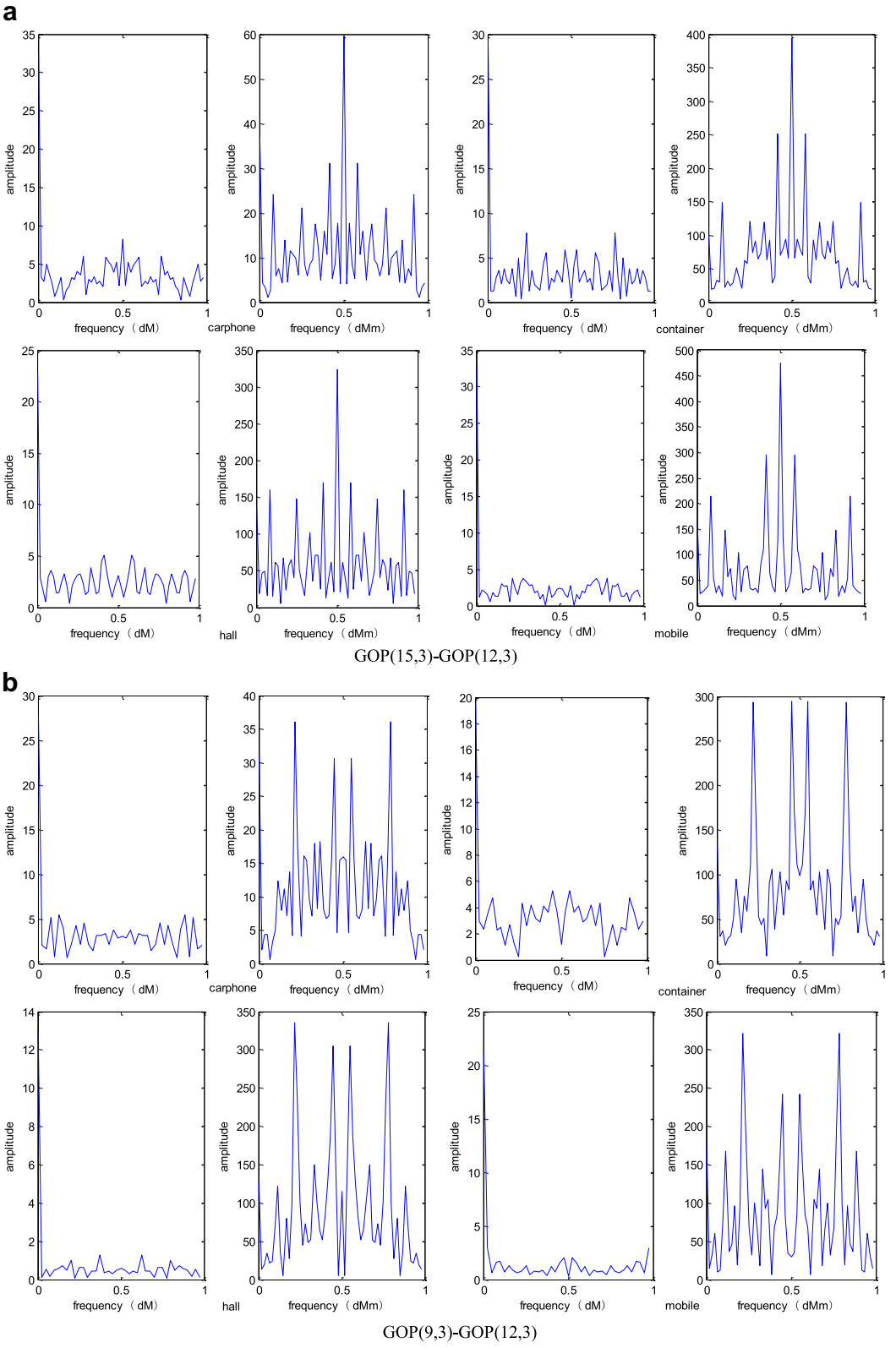
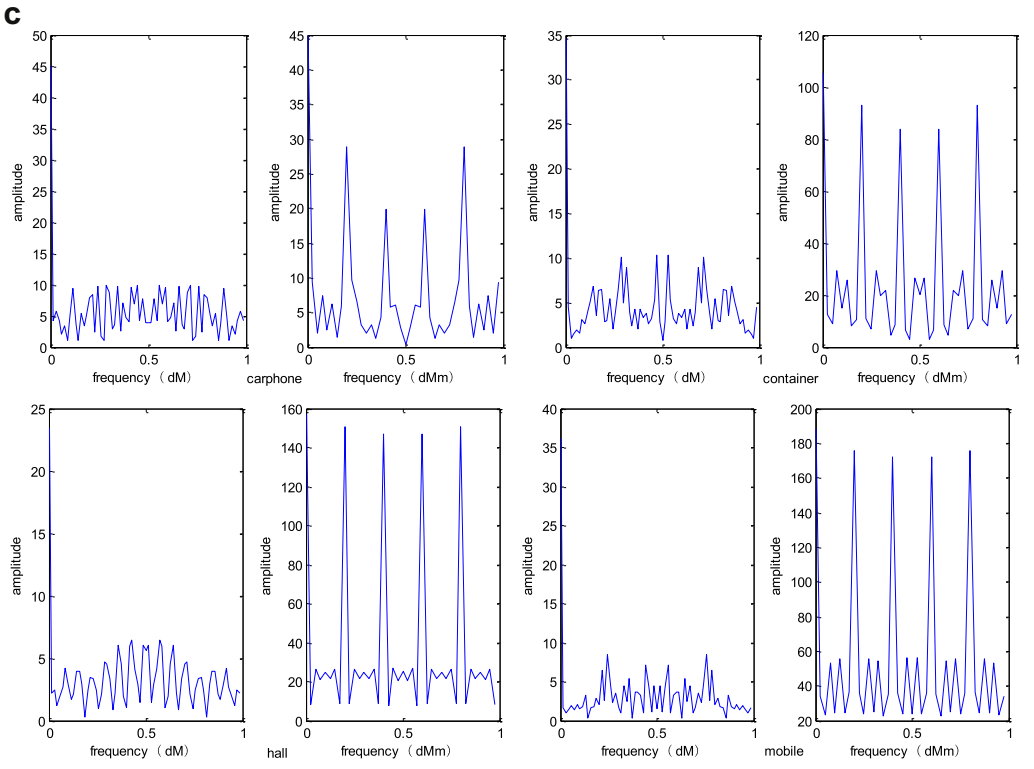
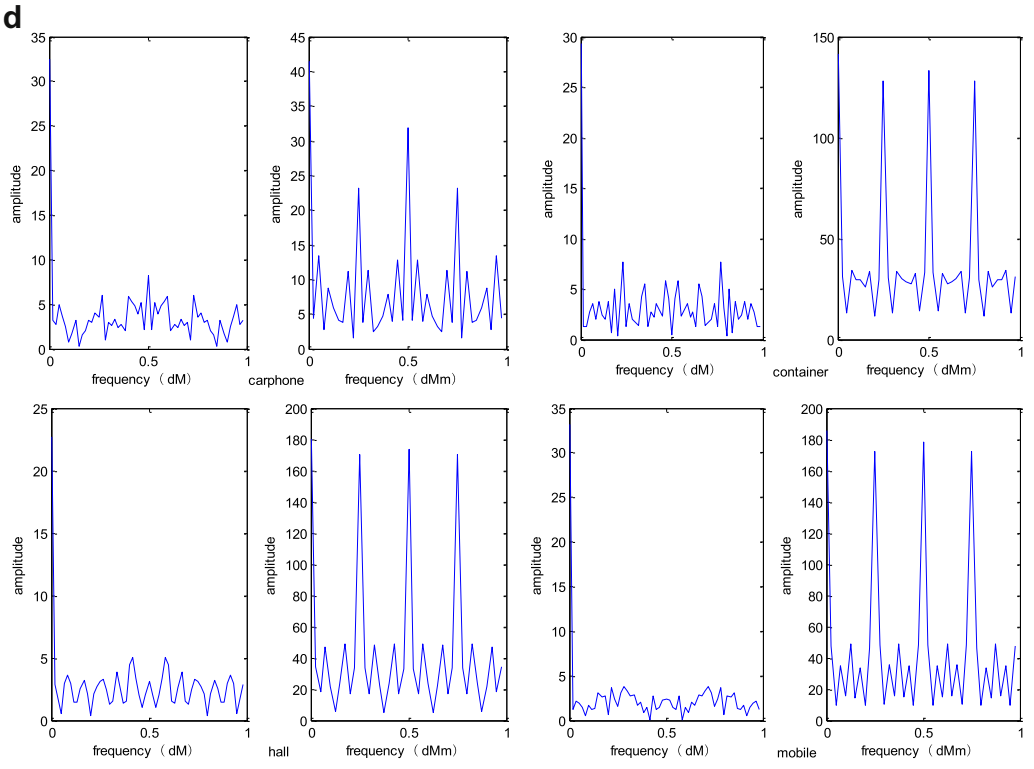


Fig. 10. The magnitude of MCEA different value of FFT of changing GOP structure re-compression.



GOP(12,3)-GOP(9,3)



GOP(12,3)-GOP(9,3)

Fig. 10. (*Continued*)

5. Experimental results

In order to verify the effectiveness of the proposed approach, four typical test video sequences are selected for experiments (Testing samples). They are *Carphone*, *Container*, *Hall* and *Mobile* (in CIF and QCIF format). Among them, *Container* and *Hall* represent those video sequences with nearly static background or simple motions, whereas *Carphone* and *Mobile* represents those video sequences with acute motion. The MPEG-2 codec by MPEG Software Simulation Group (MSSG) http://www.mpeg.org/MPEG/video is utilized in our experiments to generate MPEG streams.

The original GOP structure is GOP (15, 3). Then, the 3rd, 6th, 9th, 12th and 15th frames are deleted from original video and re-encoded with the same GOP structure. The FFT spectrums of MCEA difference sequence are computed for both true video and tampered video, which is shown in Fig. 9. There is an obvious spike appear at a particular frequency $1/(N_p - 1)$ for the frame deleting at integer multiples of M. It should note that for the artifacts left by deleting N frames, there are no spikes because no P frames is moved from one GOP to another.

In addition, we do some experiments for the changes of GOP structure, to further prove the effectiveness of the proposed approach. There are four groups of experiments: GOP (12, 3) and GOP (9, 3) to GOP (15, 3), GOP (12, 3) and GOP (15, 3) to GOP (9, 3), respectively. The experimental results are shown in Fig. 10. Apparently, all the spikes occur at multiple times of spectrum frequency. The original GOP structure can be deduced by finding the greatest common divisor by Formula (8). The results are summarized in Table 1. Please note that for the prediction of original GOP structure, the M in the GOP structure should be the same.

For all the experiment results in Figs. 9 and 10, it is found that the amplitudes of spikes are usually at least 3 times bigger than the adjacent amplitude in Fourier spectrum. Therefore, we are motivated to use a simple but effective threshold selection method. If there is amplitude 3 times bigger than the average, we will judge it as a spike.

Compare to the Su's work (Su and Zhang, 2009), our approach can overcome the shortcoming of hard threshold selection. Moreover, it can detect the double MPEG compression with different GOP structure.

6. Conclusions

In this paper, a MCEA-based passive forensics scheme is proposed for frame-based video tampering. It exploits the MCEA difference between adjacent P frames, and judges whether there are any spikes in the Fourier transform domain after double MPEG compression. Experimental results on several test sequences show that the proposed scheme is effective for the forensics of deleting integer multiple M in GOP (N, M) and the GOP structure change in double MPEG compression. Furthermore, it can deduce the original GOP structure. For the future research, the influence of B frame to the MCEAs of P frames in double compressed video is not investigated in this paper. It remains an open issue to be further resolved. In addition, the proposed approach can be theoretically extended to other video coding standard such as H.264/AVC, since they all utilize the block-based hybrid coding framework. Yet, special coding tools and features such as multiple inter/intra prediction modes and integer transform are introduced into H.264/AVC. It is worthy of further investigation of the MCEA distribution in the H.264/AVC video stream.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (61072122) and the Special Prophase Project on National Basic Research Program of China (2010CB334706), Key Project of Hunan Provincial Natural Science Foundation (11JJ2053) and the Program for New Century Excellent Talents in University (NCET-11-0134).

References

Chuang Weihong, Su Hui, Wu Min. Exploiting compression effects for improved source camera identification using strongly compressed video. In: Proceedings of IEEE Conference on Image Processing (ICIP), 1953–1956 Brussels, Belgium; 2011.

Leontaris Athanasios, Cosman Pamela C, Reibman Amy R. Quality evaluation of motion-compensated edge artifacts in compressed video. IEEE Transactions on Image Processing 2007;16(4):943–56.

Luo Weiqi, Wu Min, Huang Jiwu. MPEG recompression detection based on block artifacts. In: Proceedings of the SPIE (on security, forensics steganography and watermarking of multimedia contents); 2008. p. 1–12.

Qin Yunlong, Sun Guanglin, Zhang Xinpeng. Blind detection of video sequence montage based on GOP abnormality. Chinese Journal of Electronic 2010;38(7):227–33.

Rocha Anderson, Scheirer Walter, Boult Terrance, Goldenstein Siome. Vision of the unseen: current trends and challenges in digital image and video forensics. ACM Computing Surveys 2011;43(4):26–40.

Su Yuting, Zhang Jing. Exposing digital video forgery by detecting motion-compensated edge artifact. In: International conference on computational intelligence and software engineering; 2009. p. 1–4. Wuhan, China.

Testing samples available at: http://trace.eas.asu.edu/yuv/index.html.

Wang Weihong, Farid H. Exposing digital forgeries in video by detecting double mpeg compression. In: Proceedings of ACM multimedia and security workshop; 2006. p. 34–47. Geneva, Switzerland.

Wang Junwen, Liu Guangjie, Zhang Zhan, et al. Detection of forgery in digital video based on pattern noise [J]. Journal of Southeast University (Natural Science Edition) 2008;38(A2):13–7.

Xiong Xiao, Huang Zheng, Xu Che, Shi Shao-pei, Yang Xu. Digital video forgeries detection based on prediction error [J]. China Information Security 2008;5(12):128–30.