AN OVERVIEW ON VIDEO FORENSICS

P. Bestagini♯, M. Fontani♭,▽, S. Milani♯, M. Barni♭,▽, A. Piva◇,▽, M. Tagliasacchi♯, S. Tubaro♯

♯Politecnico di Milano, Dipartimento di Elettronica e Informazione, Milano, Italy
♭ University of Siena, Dept. of Information Engineering, Siena, Italy
◇ University of Florence, Dept. of Electronics and Telecommunications, Florence, Italy
▽ National Inter-University Consortium for Telecommunications (CNIT), Florence, Italy

ABSTRACT

Validating a given multimedia content is nowadays quite a hard task because of the huge amount of possible alterations that could have been operated on it. In order to face this problem, image and video experts have proposed a wide set of solutions to reconstruct the processing history of a given multimedia signal. These strategies rely on the fact that non-reversible operations applied to a signal leave some traces ("footprints") that can be identified and classified in order to reconstruct the possible alterations that have been operated on the original source. These solutions permit also to identify which source generated a specific image or video content given some device-related peculiarities.

The paper aims at providing an overview of the existing video processing techniques, considering all the possible alterations that can be operated on a single signal and also the possibility of identifying the traces that could reveal important information about its origin and use.

Index Terms— video forensics, forgery detection, double compression, processing history estimation

1. INTRODUCTION

Nowadays authenticating a given multimedia content has become more and more difficult because of the possible diverse origins and the potential alterations that could have been operated on it. This is due to the availability of inexpensive and easily-operable digital multimedia devices (such as cameras, mobile-phones, digital recorders, etc.), together with the flourishing of high-quality data processing tools and algorithms, has made signal acquisition and processing accessible to a wide range of users. As a consequence, a single image or video could have been processed and altered many times by different users. This fact has severe implications when the digital content is used to support legal evidences since its origin and integrity cannot be assured [1]. Important details could be maliciously hidden or erased from the recorded scene, and the true original source of the multimedia material can be concealed. Moreover, the detection of copyright infringements and the validation of the legal property of multimedia data may be difficult since there is no way to identify the original owner. This fact can be exploited to redistribute the original signal without the owner's permission or to pretend on its characteristics (e.g., low-quality contents re-encoded at high quality) [2, 3].

From these premises, a significant research effort has been recently devoted to the forensic analysis of multimedia data. These solutions rely on the consideration that many processing steps are not reversible and leave some traces in the resulting signal (hereby called "footprints"). Detecting and analyzing these footprints allows the reconstruction of the chain of processing steps. In other words, the detection of these footprints allows a sort of reverse engineering of digital content, in order to identify the type and order of the processing steps that a digital content has undergone, from its first generation to its actual form.

A large part of the research activities in this field are devoted to the analysis of still images. However, scientific research has been recently focusing on the forensics issues related to video signals because of their peculiarities and the wide range of possible alterations that can be applied to them. All the possible operations that can be applied to images can be operated on the different frames of a video sequence as well. Moreover, these modifications can be replicated similarly in the temporal dimension increasing the number of degrees of freedom in the alterations of the signal. As a result, video forensics proves to be extremely harder than its counterpart on still images since the recovering of the signal processing history could be much more complex. In addition to this difficulty, video data is practically always available in compressed formats and strong compression ratios may cancel or fatally compromise the existing footprints so that the processing history is, entirely or in part, no longer recoverable.

The original contribution of this paper relies in providing an overview of the main techniques that have been designed to

The project REWIND acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number:268478.

© EURASIP, 2012 - ISSN 2076-1465

recover the processing history of a given video content. Section 2 deals with the identification of the device that captured a given video content. Section 3 considers the traces left by video coding. Video doctoring is addressed in Section 4. Finally Section 5 concludes the survey, indicating open issues in the field of video forensics.

2. FORENSIC TOOLS FOR VIDEO ACQUISITION ANALYSIS

The analysis of image acquisition is one of the earliest problems that emerged in multimedia forensics, being very similar to the "classical" forensic technique of ballistic fingerprinting. Its basic goal is to understand the very first steps of the history of content, namely identifying the originating device.

Before deepening the discussion, we introduce in Figure 1 a simplified model of the acquisition chain, when a standard camcorder is adopted. First, the sensed scene is distorted by optical lenses and then mosaiced by an RGB Color Filter Array (CFA). Pixel values are stored on the internal CCD/CMOS array, and then further processed by the in-camera software. The last step usually consists in lossy encoding the resulting frames, typically using MPEG-x or H.26x codecs for cameras and 3GP codecs for mobile phones. The captured images are then either displayed/projected on screen or printed, and can be potentially recaptured with another camera.

Each of these elements leaves some footprints in the resulting signal that can be used to detect the originating device or verify if the analyzed content has been reacquired by different devices. In the following, these two possibilities are explored.

2.1. Identification of acquisition device

In the field of image forensics, many approaches have been developed to identify the specific device that originated a given content.

Kurosawa et al. [4] were the first to introduce the problem of camcorder fingerprinting. After this initial pioneering work, research in image forensics demonstrated that Photo Response Non Uniformity (PRNU) noise could provide a much more strong and reliable fingerprint of a CCD array.

Given a noise free image I_0, the image I acquired by the sensor is modeled as:

$$I = I_0 + \gamma I_0 K + N, \tag{1}$$

where γ is a multiplicative factor, K is the PRNU noise and N models all the other additive noise sources. Note that all operations are intended element-wise.

Once K is obtained for a device, checking if a query image S has been generated from that device reduces to evaluating the correlation between the noise component of the query image and the reference noise of the device. The first work about camcorder identification was proposed by Chen et al. [5]. However, coding bit rate has a significant impact on the efficiency of the algorithm since the lower the quality of the video the longer the sequence to be analyzed has to be.

The challenging problem of video source identification from low quality videos has been deeply explored by van Houten et al. in several works (see [6] as one of the most recent ones). In these cases, however, the identification of the acquisition device could also be based on the identification the codec, leveraging the techniques described in Section 3

2.2. Detection of (illegal) reproduction of videos

An important problem in copyright protection is the proliferation of bootleg videos. A great deal of these fake copies is produced by recording films with camcorders in cinemas. Video forensics contributes to facing these problems by: i) detecting re-projected videos; ii) providing video retrieval techniques based on device fingerprinting.

To deal with the re-acquisition problem, in the literature, some approaches were proposed based on active watermarking [7]. Recently, blind techniques are also emerging. Wang et al. [2] developed the most significant work in this field, exploiting the principles of multiple view geometry. Lee et al. [8] addressed the problem of detecting if an image might be a screenshot re-captured from an interlaced video.

As for video copy detection, the most common approach is to extract salient features from visual content that do not depend on the device used to capture the video. However, in [9], Bayram et al. pointed out that robust content-based signatures may hinder the capability of distinguishing between videos which are similar, although they are not copies of each other. This issue might arise, e.g., in the case of recordings of the same scene taken by different users. For this reason, they proposed to use source device characteristics extracted from videos (e.g., the PRNU fingerprints of camcorders involved in the generation of the video).

3. FORENSIC TOOLS FOR VIDEO COMPRESSION

Video content is typically available in a lossy compression format due to the large bit rate that is necessary to represent uncompressed motion pictures. Lossy compression leaves characteristic footprints, which might be detected by the forensic analyst. At the same time, coding operations have the potential effect of erasing the footprints left by previous manipulations. In this way, the processing history cannot be recovered anymore. Moreover, the wide set of video coding architectures that have been standardized during the last two decades introduces several degrees of freedom. As such, the codec adopted to compress a video sequence represents a distinctive connotative element. Therefore, if detected, it can be useful for the identification of the acquisition device, as well as for revealing possible manipulations.

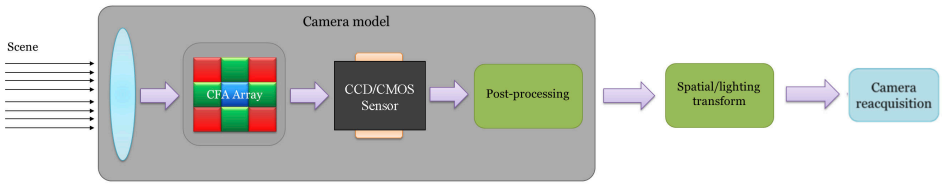

Fig. 1: Typical acquisition pipeline.

The most relevant footprints left by compression operations are those related to quantization. Many image and video coding strategies adopt a transform-based coding scheme, where the input image is divided into blocks, transformed, and the values of the resulting coefficients are quantized. This operation is not invertible and is the main source for information loss. Other important footprints are the choice of block sizes and motion vectors that are used to characterize the movement in the scene.

The analysis of coding-based footprints might be leveraged to: i) infer details about the encoder (e.g. coding standard, coding parameters, non-normative tools); ii) reveal the number of compressions that have been applied; or iii) assess the quality of a sequence in a no-reference framework to study the characteristics of the channel used to transmit the sequence.

In the following, these three aspects are discussed.

3.1. Video coding parameter identification

In image and video coding architectures, the choice of the coding parameters is driven by non-normative tools. In the case of video compression, the number of coding parameters that can be adjusted is significantly wider than in image compression. As a consequence, the forensic analyst needs to take into account a larger number of degrees of freedom.

In the literature, the methods aiming at estimating different coding parameters and syntax elements characterizing the adopted codec can be grouped into three main categories. A first class consists of those approaches detecting block boundaries [10]. A second class consider the possibility of estimating the quantization parameters [11] (mainly exploiting the typical comb-like distribution of transform coefficients). A third class relies on the identification of motion vectors for temporal prediction. In [12], it is shown how to estimate, at the decoder, the motion vectors originally adopted by the encoder, also when the bitstream is missing.

3.2. Video re-encoding

Every time a video sequence that has already been compressed is edited (e.g., scaling, local manipulation, etc.), it has to be re-compressed. Studying processing chains consisting of multiple compression steps is useful, e.g., for tampering detection or to identify the original encoder being used.

In [13], Lukáš and Fridrich show how double compression introduces characteristic peaks in the histogram, which alter the original statistics. More precisely, the authors highlight how peaks can be more or less evident depending on the relationship between the two step sizes, and propose a strategy to identify double compression. Another widely-adopted strategy for the detection of double compression relies on the so-called Benford's law or first digit law [14]. In a nutshell, it relies on the analysis of the distribution of the most significant decimal digit of the absolute value of quantized transformed coefficients. By analying the probability mass function of the first digit value, it is possible to detect the number of coding stages whenever multiple codings can be applied on a image [15]. In [3], the authors address the problem of estimating the traces of double compression of an MPEG coded video.

3.3. Network footprints identification

Video transmission over a noisy channel leaves characteristic footprints in the reconstructed video content. Indeed, packet losses and errors might affect the received bitstream. As a consequence, some of the coded data will be missing or corrupted. Error concealment is designed to take care of this, trying to recover the correct information and mitigate the channel-induced distortion. However, this operation introduces some artifacts in the reconstructed video, which can be detected to infer the underlying loss (or error) pattern. The specific loss pattern permits the identification of the characteristics of the channel that was employed during the transmission of the coded video. In [16], the authors present an algorithm based on several quality assessment metrics to estimate the packet loss impairment in the reconstructed video.

4. FORENSIC TOOLS FOR VIDEO DOCTORING DETECTION

Although being more complicated than for images, creating a forged video is now easier than before, due to the availability of video editing suites. At the same time, videos are

extensively used for surveillance, and they are usually considered a much stronger proof than a single shot. There are many different ways of tampering with a video: one may be interested in replacing or removing some frames (e.g., from a video-surveillance recording), replicating a set of frames, introducing, duplicating or removing some objects from the scene.

It is possible to classify both video forgery and video forensic techniques as intra-frame (attack/analysis is performed frame-wise, considering one frame at a time), or inter-frame (relationships between adjacent frames are considered).

In the following subsections we survey existing techniques for video doctoring detection clustered as: i) camera-based techniques; ii) coding-based techniques iii) geometrical/physical inconsistencies exploiting techniques. Finally, we analyze the problem of identifying copy-move forgeries involving frames or parts of them.

4.1. Camera based editing detection

As discussed in Section 2, camcorders usually leave a characteristic fingerprint in recorded videos. Although these kinds of artifacts are usually exploited just for device identification, some works leverage on them also for tampering detection.

Mondaini et al. [17] proposed a direct application of the PRNU fingerprinting technique (see Section 2.1) to video sequences.

Hsu et al. [18] adopted a technique based on temporal correlation of noise residues, where the "noise residue" of a frame is defined as what remains after subtracting from the frame its denoised version.

Another camera-based approach is the one from Kobayashi et al. [19]: they proposed to detect suspicious regions in video recorded from a *static scene* by using noise characteristics of the acquisition device.

4.2. Detection based on coding artifacts

From what emerged in the previous Section, video encoding strongly hinders the performances of camera based detection techniques. On the other hand, however, coding itself introduces artifacts that can be leveraged to investigate the integrity of the content. In the last years, some forensic researchers investigated the presence or the inconsistencies of these artifacts to assess the integrity of a video, and to localize which regions are not original.

The first approach in this direction was from Wang and Farid [3], focusing on MPEG compressed videos, where two phenomena are explored, one static (intra-frame) and one temporal (inter-frame).

Quantization artifacts are not the only effect that have been exploited for video doctoring detection: Wang and Farid proposed another approach [20] for detecting tampering in interlaced and de-interlaced video.

4.3. Detection based on inconsistencies in content

It is very difficult to understand whether the geometry or the physical/lighting properties of a scene are consistent. In particular, it is very hard to do so unless some assistance from the analyst is provided, whose effort would be prohibitive to check geometric consistencies in video on a frame-by-frame basis. Existing works usually exploit phenomena connected to motion in order to detect editing. So far, two approaches have been proposed: i) the one in [21], based on artifacts introduced by video inpainting, ii) the one in [22], that reveal inconsistencies in the motion of objects in free-flight.

4.4. Copy-move detection in videos

Copy-move attacks are defined for video both as intra- and inter- frame techniques. An intra-frame copy-move attack is conceptually identical to the one for still images, and consists in replicating a portion of the frame in the frame itself (the goal is usually to hide or replicate some object). An inter-frame copy-move, instead, consists in replacing some frames with a copy of previous ones, usually to hide something that entered the scene in the original video. To the best of our knowledge, there is only one work authored by Wang and Farid [23] that targets copy-move detection directly in video. The method uses a kind of divide-and-conquer approach: the whole video is split in subparts, and different kinds of correlation coefficients are computed in order to highlight similarities between different parts of the sequence.

5. CONCLUSIONS AND FUTURE WORKS

As it has been shown in the previous sections, video forensics is nowadays a hot research issue in the signal processing world, opening new problems and investigation threads.

Despite several techniques have been borrowed from image forensics, video signals pose new challenges in the forensic application world because of the amount and the complexity of data to be processed and the wide employment of compression techniques, which may alter or erase footprints left by previous signal modifications.

Current results show that it is possible to reconstruct simple processing chains (i.e., acquisition followed by compression, double compression, etc.) under the assumption that each processing step does not introduce an excessive amount of distortion on the signal. This proves to be reasonable since a severe deterioration of the quality of the signal would make it useless.

Future research has still to investigate more complex processing chains, where each operation on the signal may be iterated multiple times. In order to tackle with this problem, also the robustness of the analysis in presence of multiple aggressive encodings should be enhanced.

6. REFERENCES

[1] Hany Farid, "Exposing digital forgeries in scientific images," in *Proceedings of the 8th workshop on Multimedia and security (MM&Sec 2006)*, 2006.

[2] Weihong Wang and Hany Farid, "Detecting re-projected video," in *Information Hiding*, 2008.

[3] Weihong Wang and Hany Farid, "Exposing digital forgeries in video by detecting double MPEG compression," in *MM&Sec*, 2006.

[4] K. Kurosawa, K. Kuroki, and N. Saitoh, "CCD fingerprint method-identification of a video camera from videotaped images," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, 1999.

[5] M. Chen, J. Fridrich, M. Goljan, and J. Lukás, "Source digital camcorder identification using sensor photo response non-uniformity," in *Proceedings of SPIE*, 2007.

[6] Wiger van Houten, Zeno J. M. H. Geradts, Katrin Franke, and Cor J. Veenman, "Verification of video source camera competition (camcom 2010)," in *ICPR Contests*, 2010.

[7] Min-Jeong Lee, Kyung-Su Kim, and Heung-Kyu Lee, "Digital cinema watermarking for estimating the position of the pirate," *IEEE Transactions on Multimedia*, vol. 12, pp. 605–621, 2010.

[8] Ji-Won Lee, Min-Jeong Lee, Tae-Woo Oh, Seung-Jin Ryu, and Heung-Kyu Lee, "Screenshot identification using combing artifact from interlaced video," in *Proceedings of the 12th ACM workshop on Multimedia and security*, 2010.

[9] Sevinc Bayram, Husrev T. Sencar, and Nasir D. Memon, "Video copy detection based on source device characteristics: a complementary approach to content-based methods," in *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, 2008.

[10] Zhigang Fan and Ricardo L. de Queiroz, "Identification of bitmap compression history: JPEG detection and quantizer estimation," *IEEE Transactions on Image Processing*, vol. 12, pp. 230–235, 2003.

[11] M. Tagliasacchi and S. Tubaro, "Blind estimation of the QP parameter in H.264/AVC decoded video," in *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*, 2010.

[12] Giuseppe Valenzise, Marco Tagliasacchi, and Stefano Tubaro, "Estimating QP and motion vectors in H.264/AVC video from decoded pixels," in *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*, 2010.

[13] Jan Lukás and Jessica Fridrich, "Estimation of primary quantization matrix in double compressed jpeg images," in *Proc. of DFRWS*, 2003.

[14] Dongdong Fu, Yun Q Shi, and Wei Su, "A generalized Benford's law for JPEG coefficients and its applications in image forensics," *Proceedings of SPIE*, vol. 6505, pp. 65051L–65051L–11, 2007.

[15] S. Milani, M. Tagliasacchi, and M. Tubaro, "Discriminating multiple jpeg compression using first digit features," in *to appear on Proc. of the* 37^{th} *International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012)*, 2012.

[16] Amy R. Reibman and David Poole, "Characterizing packet-loss impairments in compressed video," in *ICIP (5)*, 2007.

[17] N. Mondaini, R. Caldelli, A. Piva, M. Barni, and V. Cappellini, "Detection of malevolent changes in digital video for forensic applications," in *Proceedings of SPIE, Security, Steganography, and Watermarking of Multimedia Contents IX*, 2007.

[18] Chih-Chung Hsu, Tzu-Yi Hung, Chia-Wen Lin, and Chiou-Ting Hsu, "Video forgery detection using correlation of noise residue," in *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, 2008.

[19] Michihiro Kobayashi, Takahiro Okabe, and Yoichi Sato, "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.

[20] Weihong Wang and Hany Farid, "Exposing digital forgeries in interlaced and deinterlaced video," *IEEE Transactions on Information Forensics and Security*, vol. 2, pp. 438–449, 2007.

[21] Jing Zhang, Yuting Su, and Mingyu Zhang, "Exposing digital video forgery by ghost shadow artifact," in *Proceedings of the First ACM workshop on Multimedia in forensics*, 2009.

[22] V. Conotter, J. O'Brien, and H. Farid, "Exposing digital forgeries in ballistic motion," *Information Forensics and Security, IEEE Transactions on*, vol. PP, no. 99, pp. 1, 2011.

[23] Weihong Wang and Hany Farid, "Exposing digital forgeries in video by detecting duplication," in *MM&Sec*, 2007.