Contents lists available at ScienceDirect

Digital Investigation

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

Forensic analysis of video file formats

Thomas Gloe [a,*], André Fischer [a], Matthias Kirchner [b]

[a] *dence GmbH, Dresden, Germany*
[b] *University of Münster, Department of Information Systems, Münster, Germany*

Keywords:
Digital forensics
File format
Metadata
Video
Authentication

ABSTRACT

Video file format standards define only a limited number of mandatory features and leave room for interpretation. Design decisions of device manufacturers and software vendors are thus a fruitful resource for forensic video authentication. This paper explores AVI and MP4-like video streams of mobile phones and digital cameras in detail. We use customized parsers to extract all file format structures of videos from overall 19 digital camera models, 14 mobile phone models, and 6 video editing toolboxes. We report considerable differences in the choice of container formats, audio and video compression algorithms, acquisition parameters, and internal file structure. In combination, such characteristics can help to authenticate digital video files in forensic settings by distinguishing between original and post-processed videos, verifying the purported source of a file, or identifying the true acquisition device model or the processing software used for video processing.

Introduction

Methods to verify the authenticity of media data are of growing relevance in our digital world. While most consumer devices lack practical authentication support at all, attacks against professional camera authentication systems have demonstrated weaknesses of existing in-device solutions.[1] Forensic techniques to infer the provenance and the processing history of media files *ex post* have thus gained more and more interest among researchers and practitioners. Forensic image analysis has been the main driver of the field (Sencar and Memon, 2013), but also video files have recently been brought to the forefront (Milani et al., 2012b). Resembling the evolution of digital image forensics, an already ample body of literature approaches the problem of video forensics through the analysis of inherent device characteristics or processing artifacts in the video data (Chen et al., 2007; Wang and Farid, 2007; Hsu et al., 2008; Conotter et al., 2011; Stamm et al., 2012; Vázquez-Padín et al., 2012, amongst others).

File format information and *metadata* are another source of forensic evidence, but have generally received less attention. Existing works mainly focus on digital images. Here, basic JPEG (ISO/IEC 10918-1, ITU-T Recommendation T.81, 1992) and EXIF (Japan Electronics and Information Technology Industries Association, 2002) metadata properties have gained major interest. Digital cameras and image processing software (or groups thereof) use customized quantization tables. Differences therein can narrow down the source device of a questioned image (Farid, 2008; Kornblum, 2008). Characteristics of thumbnail images (often saved as JPEG images themselves) have been reported to be another pool of forensically relevant features (Murdoch and Dornseif, 2004; Kee and Farid, 2010). In one of the most elaborate approaches, Kee et al. (2011) combine image and thumbnail compression parameters, image and thumbnail dimensions and the number of EXIF entries into signatures of camera model or processing software configurations. By testing against a reference database, images of unknown or uncertain provenance can be attributed to a

* Corresponding author.
E-mail addresses: thomas.gloe@dence.de (T. Gloe), matthias.kirchner@uni-muenster.de (M. Kirchner).

[1] http://www.elcomsoft.de/nikon.html http://www.elcomsoft.de/canon.html.

http://dx.doi.org/10.1016/j.diin.2014.03.009

class of source configurations. Images are flagged as suspicious if no match is found. In a different approach, Fan et al. (2013) exploit noise characteristics to determine whether image content and EXIF data are consistent. However, as tampering with compression parameters or EXIF entries is only a question of using proper software tools (which are often publicly available), concerns have frequently been expressed that high-level file format information and metadata are easily replaceable and/or forgeable (Sencar and Memon, 2008). On the contrary, existing image processing software and metadata editors do not allow users to access or to modify *core file structures*. Along these lines, Gloe (2012) reports that peculiarities in the specific internal order of JPEG and EXIF structures are particularly valuable and distinctive information for digital image authentication. Such low-level characteristics thus offer a much increased reliability and relax—to some degree—the common assumption that file format and metadata information shall not be trusted *per se*.

Following this trail, this paper extends the idea of file format forensics to popular digital video data container formats, for which—to the best of our knowledge—no systematic exploration of file format and/or metadata specifics has been reported in the forensic literature so far. Similar to differences in the JPEG file structure, we identify manufacturer- and model-specific video file format characteristics and point to traces left by processing software. Such traces can be used to authenticate digital video streams and to attribute recordings of unknown or questionable provenance to (groups of) video camera models. Note that this differs from video file carving (Pal and Memon, 2009; Lewis, 2012), where it is usually sufficient to find *valid* video streams in (fragmented) mass storage dumps. Based on a description of our (Test setup) and (General observations) on video file format forensics, the following sections demonstrate that peculiarities of the (AVI Container format), of (Quicktime and related container formats (MP4, 3GP)), and of (MJPEG Compression parameters) can yield important insights about provenance and processing history of digital videos. The paper closes with a (Summary and concluding remarks).

Test setup

We report findings from the examination of each one device of overall 19 digital camera models and 14 mobile phone models, all of them equipped with video capturing functionality. We acquired 3 to 14 videos per device by iterating over all available video quality settings (e.g., frame size and frame rate). Mobile phones were also switched between regular and MMS (Multimedia Message Service) mode where available. All devices were subject to slight motion during the video capturing process. Table 1 summarizes our test setup.

For a selected number of camera models, video editing software was used to cut short sequences (length: 10 s) from the recorded video streams. All software in our tests supports non-intrusive ('lossless') video editing, i.e., we saved files without re-compressing the original stream. Hence, the edited videos are presumably not detectable by means of double compression artifacts (Wang and Farid, 2006; Milani et al., 2012a). The 'Adobe Premiere' toolbox was a sole exception in our test set in this regard. The commercial software is one of the major professional video editing tools, but does not support lossless processing.

We have written our own customized file parser(s) to read and extract *all* available file format information and metadata from the videos in our database.[2] As it is impossible to detail all model- or vendor-specific singularities within the scope of this paper, the following sections focus on selected results and observations that we believe are particularly relevant for practical forensic analyses of common video container formats.

General observations

The majority of *digital cameras* in our database stores videos in the AVI container format. Only a few of the test devices use Apple Quicktime MOV containers. We found that most digital cameras compress video data using Motion JPEG (MJPEG), where every video frame is handled as independently JPEG-compressed image. Only three camera models in our sample use more sophisticated and efficient compression algorithms (DivX, Xvid or H.264). Before compression, frames are generally converted to the YUV color space. We encountered 4:2:2 and 4:2:0 subsampling to reduce the resolution of chroma channels. MJPEG compressed video streams utilize the full intensity range of 256 intensity levels for 8-bit encoded frames (yuvj422p or yuvj420p), whereas camera models with DivX or Xvid support (yuv420p) use only a reduced number of 220/225 intensity levels for the luminance/chrominance channel(s) (ITU-R Recommendation BT.601-7, 2011). Audio data in the video container is usually stored as raw data (PCM), using linear (8-/16-bit) or logarithmic (μ-law) quantization.

All *mobile phones* in our database store video data in MOV-based container formats (MOV, 3GP, MP4). The LG KU990 camera phone is an exception and also supports AVI containers. Interestingly, none of the mobile phones uses MJPEG compression. Instead, more sophisticated compression algorithms find application, e.g., H.263, H.264, MPEG-4 video (simple profile) or DivX. Subsampling always follows a 4:2:0 scheme. In contrast to videos from digital cameras, the audio track of mobile phone videos is typically also subject to lossy compression. We found MPEG-based audio compression or the Adaptive Multi-Rate audio codec (AMR-NB) to be most common. The latter is a standard optimized for speech coding (3rd Generation Partnership Project, 1999), which is very common in mobile phones designed for GSM- and UMTS-networks.

AVI Container format

Microsoft introduced AVI (Audio Video Interleave) in 1992 as a multimedia container format, which can contain

[2] Also the free software `exiftool` (available at: http://www.sno.phy.queensu.ca/~phil/exiftool) can be used to extract metadata and high-level file format information, but it does not provide access to all information that is of interest here.