SWGDE Best Practice for Frame Timing Analysis of H.264 Video Stored in ISO Base Media File Formats

The version of this document is in draft form and is being provided for comment by all interested parties for a minimum period of 60 days.

SWGDE encourages stakeholder participation in the preparation of documents. Suggestions for modifications are welcome and must be forwarded to the Secretary in writing at secretary@swgde.org. The following information is required as a part of the response:

a) Submitter's name
b) Affiliation (agency/organization)
c) Address
d) Telephone number and email address
e) Document title and version number
f) Change from (note document section number)
g) Change to (provide suggested text where appropriate; comments not including suggested text will not be considered)
h) Basis for change

Intellectual Property:
Unauthorized use of the SWGDE logo or documents without written permission from SWGDE is a violation of our intellectual property rights.

Individuals may not misstate and/or over represent duties and responsibilities of SWGDE work. This includes claiming oneself as a contributing member without actively participating in SWGDE meetings; claiming oneself as an officer of SWGDE without serving as such; claiming sole authorship of a document; use the SWGDE logo on any material and/or curriculum vitae.

Any mention of specific products within SWGDE documents is for informational purposes only; it does not imply a recommendation or endorsement by SWGDE.

SWGDE Best Practice for Frame Timing Analysis of H.264 Video Stored in ISO Base Media File Formats

Table of Contents

1. Purpose

The purpose of this document is to provide forensic examiners recommendations for determining frame rate and frame interval timing as a part of forensic analysis of digital video.

2. Scope

This document addresses file formats encoded to the *ISO/IEC 14496-12 Information technology — Coding of audiovisual objects — Part 12: ISO base media file format* [1]. Additionally, this document specifically addresses video encoded according to the H.264 specification [2]. The intended audience is forensic examiners with an advanced understanding of digital video file formats and encoding.

This document is not intended to be used as a step-by-step guide for conducting a forensic examination or reaching a conclusion.

3. Limitations

Due to the wide variety of proprietary digital video recording devices and file formats, a singular approach to frame timing analysis cannot be applied to all files. Other multimedia file formats (e.g., .AVI, .MKV) and video coding standards (e.g., MPEG-1 Part 2, MPEG-2 Part 2) may require different approaches than those covered in this document.

Proprietary video files may store metadata with the video and audio data streams in a proprietary container and may not adhere to an encoding standard. Due to the unique nature of their file structure, proprietary video files may result in inaccurate frame rate reporting in many video playback/processing software programs.

Note: It is recommended that examiners acquire both proprietary and open file formats from the source, if available. This allows for both validation of a determined frame time and an additional resource to analyze should the proprietary format not properly decode. When multiple file formats are available, the examiner should exercise caution in identifying the file format with the intended frame rate timing. It is also important to know the source of the video. Regardless of the acquired file format(s), frame timing analysis should not be conducted on transcoded or screen captured video files.

The concepts in this document may be used as part of investigations into determining object speed in recorded video. However, this document does not address the forensic use of photogrammetry, which is an integral part of any speed calculation. See *SWGDE Best Practices for the Forensic Use of Photogrammetry* for more information [3].

4. Introduction

The *ISO/IEC 14496-12 Information technology — Coding of audiovisual objects — Part 12*: *ISO base media file format* was originally derived from the Quicktime format specification and was standardized by ISO and is the foundation for MP4, 3GP, 3G2, and M4V multimedia file formats. Included in the standard is specific encoding language pertaining to the timing and presentation of video frames. Using this information, the specific time intervals between frames can be calculated.

Determining the frame timing within a video file has several applications and may be particularly helpful in determining the accuracy of an unknown variable during an event of interest. The importance of understanding frame timing has been shown in circumstances including calculating vehicle speed and evaluating the use of force where a misunderstanding of timing resulted in incorrect opinions from video files[1,2].

5. Frame Timing in ISO Base Media File Formats

Timing information for digital video in ISO Base Media files is stored in multiple locations throughout the file. Core concepts that are central for understanding the time elements discussed below can be found in *SWGDE Core Technical Concepts for Time-Based Analysis of Digital Video Files* [4]. The following is a hierarchical example of the common locations (i.e., structures) within an ISO Base Media file where timing information is stored[3]:

```
-mvhd
-- timescale
-- duration
-- rate
--tkhd
     -- duration
---mdhd
     -- timescale
     -- duration
----stbl
-----stts (time to sample)
-----ctts (composition time to sample)
```

[1] State of New Hampshire v. Wiley, November 25, 2015, New Hampshire Superior Court, Southern District, Docket No. 226-2014 CR-00568

[2] Leyritz v. State, 93 So. 3d 1156, 2012 Fla. App. LEXIS 12526, 37 Fla. L. Weekly D 1835, 2012 WL 3101493

[3] See *ISO/IEC 14496-12 Information technology — Coding of audiovisual objects — Part 12: ISO base media file format* for more detailed information on the structural components of these file formats.

- The Movie Header Box (mvhd), defines overall information which is media-independent, and relevant to the entire multimedia presentation. Within this box, there are three time-related data elements.
    - **Timescale** specifies the timescale for the entire presentation; this is the number of time units that pass in one second. For example, a time coordinate system that measures time in sixtieths (1/60) of a second has a time scale of 60.
    - **Duration** declares the length of the presentation in the timescale of this mvhd box. This property is derived from the presentation's tracks; the value of this field corresponds to the duration of the longest track in the presentation.
    - **Rate** indicates the preferred rate to play the presentation (e.g., 1 is the equivalent of normal forward playback).
- The Track Header Box (tkhd) specifies the characteristics of a single track. The value of "duration" in this box indicates the duration of this track (in the timescale indicated in the mvhd). In general, the duration is the sum of the sample durations, converted into the timescale in the mvhd box.
- The Media Header Box (mdhd) declares overall information that is media-independent, and relevant to the characteristics of the media in a track. This is media specific, which may differ from the timescale of the overall multimedia file. The value "duration" declares the duration of this media (in the scale of the timescale of this mdhd box).
- The Sample Table Box (stbl) contains all the time and data indexing of the media samples (e.g., frames, in the case of video) in a track. Time to sample boxes contain the composition times (CT) and decoding times (DT) of samples, of which there are two types.
- The Time To Sample (stts) box gives durations for all samples, expressed in the timescale of the mdhd box.
- The Composition Time to Sample table (ctts) provides the offset between decoding time and composition time, in the case that they differ.

6. Frame timing in H.264 video codings

When encapsulated within ISO Base Media Files, H.264 video is packaged according to a specification commonly known as "AVCC," which is the name of the box within the ISO Base Media File that stores supplemental metadata required to re-present H.264 coded video samples. This box stores sequence parameter set (SPS) and picture parameter set (PPS) data that H.264 coding uses to establish the characteristics of the coded video samples. Within the SPS data, timing information can be present, but it is optional. When it is present, it can define the timescale of the media samples and whether or not there is a constant frame rate.[4]

[4] See International Telecommunication Union (ITU) Telecommunication Standardization Sector (ITU-T) Recommendation H.264, *Infrastructure of audiovisual services, Annex E,* for further information about this Video Usability Information (VUI).

When H.264 is found within an ISO Base Media file, the combination of video sample timing information and multimedia file timing information must be evaluated carefully to calculate the duration and timecode for each video frame. For example, MP4 files store frame-by-frame timing data in sample tables, and use header boxes to store overall frame rate and time bases. These are all used to generate the calculations needed for playback devices to represent the frames of the embedded video samples as they were intended to be. The ability to determine frame timing is dependent on the ability to decode the container information. While many software programs and tools allow for the decoding and playback of video codings, they may not fully decode multimedia file metadata in the same manner or in a transparent way for the examiner to see.

For additional information on file formats and applications used for frame timing see *SWGDE Technical Overview of Digital Video Files* and *SWGDE Technical Notes on FFmpeg* [5,6].

7. Timing Determination Using Multimedia File Metadata

Utilizing the open source application *FFmpeg* or other software tools, metadata from ISO Base Media File formats may be decoded to identify specific times at which individual frames are to be displayed. This timing information may account for skipped or dropped frames, variable frame rates, and potential encoding errors. The specific timecode information for each frame can be determined using a *ffprobe* Frame Information Report.

The Frame Information Report is generated within ffprobe utilizing the following command:

```
ffprobe -show_frames -print_format xml input.mp4 > output.xml
```

The report returns time and display information from the video file as shown in Appendix A.

7.1 Presentation Frame Timing

When determining frame timing, timecode for each frame is reported in the packet presentation time (pkt_pts_time) column, decoded as a calculation of pkt_pts / the timescale of the media. The packet presentation time, displayed in seconds, is the exact time that a particular frame is to be displayed. It should be noted that the initial start point of timecode within a video stream may not start at 0 seconds, rather it may be an arbitrary number. By looking at the difference between pkt_pts_time values for sequential frames, examiners can determine the time that has elapsed between those frames. This is different from frame rate as it identifies the time between each frame rather than the amount of frames displayed in one second.[5]

[5] Also reported in the Frame Information Report is packet duration time (pkt_duration_time). Packet duration time displays the total time that an individual frame is to be displayed (the value is expressed in the timescale of the media). Examiners should validate this time against packet presentation time before use in an examination. Due to the nature of how certain video files are encoded, differences between packet presentation times of sequential frames and packet duration time may occur.

The distinction between elapsed time between frames and frame rate is an important one. The delta of packet presentation times identifies the elapsed time between two specific frames, whereas frame rate identifies a total number of frames displayed in a second. Frame rate does not address the specific time between frames within that second.

7.2 Decode Frame Timing

Packet decode time (packet_decode_time) is also reported in the Frame Information Report. This is the specific time at which a frame is to be decoded by the playback software. It is important to note that the aforementioned *ffprobe* command is deriving information from the video file that has not been decoded. For that reason, the packet decode timestamp (pkt_dts_timestamp) and the packet presentation timestamp (pkt_pts_timestamp) will most likely be the same value.

Understanding packet presentation and packet decode times is an important distinction in video files that contain bi-directional frames (b-frames). However, the display order is not determined until after it is decoded (see *SWGDE Technical Overview of Digital Video Files*, Section 6.4.2*)* [5]. When b-frames are noted in the video file, an *ffmpeg* command to decode the file and determine the appropriate display order and timing of the b-frames first is more appropriate.

Packet decode time and packet presentation time can be generated after decoding a file utilizing the following command, which will generate a log file in the same directory as the input file and will include timing information:

```
ffmpeg -i input.dvr -dump -map 0:v -f null - -report -loglevel quiet
```

Note that in the example below the packet decode time may be prior to the presentation time due to the b-frames' reliance on frames presented before and after them. Also note that the values for duration, pts, and dts are expressed in seconds.

```
stream #0:
  keyframe=0
  duration=0.033
  dts=0.167  pts=0.167
  size=17424
stream #0:
  keyframe=0
  duration=0.033
  dts=0.200  pts=0.300
  size=36907
stream #0:
  keyframe=0
  duration=0.033
  dts=0.234  pts=0.234
  size=11742
```

A combination of the ffprobe and *ffmpeg* analysis of individual frame information can be used to verify frame timing outputs. If pts and dts values are found to be the same for each frame, the ffprobe report (described earlier in this section) may be preferred as it is easier to view the information in spreadsheet form as opposed to the created text log file.

8. Frame Timing Determination Using an External Timing Source

When ISO Base Media File metadata analysis cannot be conducted, an external timing source can be used to evaluate frame timing. Even when there is metadata available, one should still validate findings against a known time source. While deriving frame time from file metadata may have more precision than other methods, there are occurrences when that is not an option (e.g., inability to decode proprietary container information).

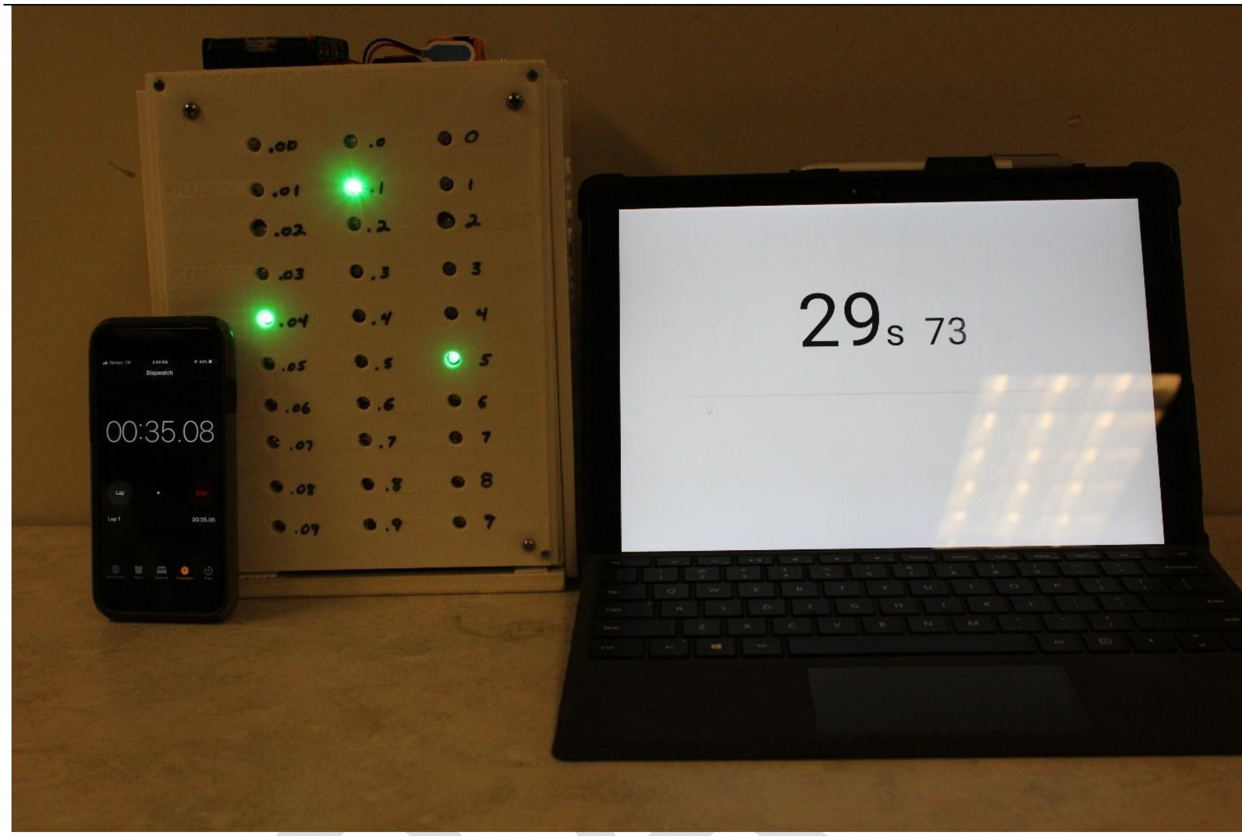
Determining frame timing information for a multimedia file using an external timing source can be accomplished through recording a test video of a timing device and using the newly generated video frames to determine frame timing. When recording a test video, conditions of the recording device's stressors (e.g., amount of motion, number of cameras connected) for that test video should be similar, or worse, than the evidence video. Those stressors may affect the ability of the recording device to process and encode data accurately and should be evaluated in worst case conditions. For example, if the evidence video has 16 cameras recording a store with shoppers walking around, the test video should have all 16 cameras recording with equal or greater motion in each camera.

There are a number of methods that can be used to display time for a test recording. It is important to use a timer where the examiner can accurately discern the correct time with adequate precision, minimally displaying .01 of a second. Considerations should be made to properly resolve any displayed timer with minimal motion blur. Techniques to accomplish this can include placing the timer as close as possible to the camera and using a larger sized display (e.g., tablet or computer displayed timer).

Using the known timing information generated by an external timing device, a margin of error can be calculated for frame timing. By using the external time display, the delta between subsequent frames can be calculated. Those intervals between frames can also be plotted and a minimum and maximum deviation calculated. This information can then be used to determine minimum and maximum timing between frames. It is important that the test video be of sufficient length and that calculations be made at various points within the file to ensure accuracy.

An LED light timing device may assist in determining time in individual frames. The use of an LED timer allows for precise determination of smaller increments of time through the use of an individual LED opposed to a digital display. If an LED box is employed to determine time, the box should be oriented in both a vertical and horizontal fashion to account for any variance in cameras using a rolling shutter.

Prior to use, any timer should be calibrated according to methods recommended by NIST [7].



The image above shows an LED light timer as compared to a cell phone and an internet-based stopwatch. Multiple still images are captured and the time differences for each device in those images is calculated. The time differences between each frame should be consistent, as shown below:

Image	LED Timer	LED Difference	Cell Phone	Cell Difference	Internet	Internet Difference
1	3.24		33.18		27.83	
2	5.14	1.9	35.08	1.9	29.73	1.9
3	7.76	2.62	37.7	2.62	32.35	2.62
4	9.77	2.01	39.71	2.01	34.36	2.01
5	12.82	3.05	42.76	3.05	37.41	3.05

9. Verification

Any findings should be verified as there are numerous variables involving recording manufacturers, codecs, and containers. This can be accomplished by verifying frame timing in the ffprobe frame information report by manual subtraction of packet presentation times, comparing file metadata with a timing source, or comparing the overall frame rate with specific frame timings to look at any potential variances. When comparing metadata to a visual timing source, multiple samples from a wide cross section of the video file with a sufficient number of frames should be analyzed. This should help ensure an accurate analysis of LED displayed time and metadata time.

10. Example

An examiner is provided with an MP4 video file that is native to the DVR it was acquired from. The examiner is asked to determine the elapsed time between a vehicle's brake lights activating and its impact with another vehicle. The examiner generates sequential still images from the file and identifies the frame where the brake lights are activated as well as the frame where the vehicles impact. A determination is then made that no b-frames are present in the video file. The examiner then generates a frame information report with ffprobe to determine the packet presentation time of those frames. Using the packet presentation times, frame timing was calculated for the examination. Prior to reporting the findings, the examiner returns to the scene with a LED light timer and records an exemplar video. A frame information report is then generated of the exemplar video file and the frame timing derived from the metadata is confirmed with the visual display of the light box.

11. References

[1] *Information technology — Coding of audiovisual objects — Part 12: ISO base media file format*, ISO/IEC 14496:2012.

[2] *Infrastructure of audiovisual services – Coding of moving video*, ITU-T Recommendation H.264.

[3] Scientific Working Group on Digital Evidence, "SWGDE Best Practices for the Forensic Use of Photogrammetry," 2015. [Online]. https://www.swgde.org/documents

[4] Scientific Working Group on Digital Evidence, "SWGDE Core Technical Concepts for Time-Based Analysis of Digital Video Files," *Public Draft*, 2019. [Online]. https://www.swgde.org/documents/draftsForPublicComment

[5] Scientific Working Group on Digital Evidence, "SWGDE Technical Overview of Digital Video Files," 2017. [Online]. https://www.swgde.org/documents

[6] Scientific Working Group on Digital Evidence, "SWGDE Technical Notes on FFmpeg," 2018. [Online]. https://www.swgde.org/documents

[7] J. Gust, R. Graham, and M. Lombardi, "Stopwatch and Timer Calibrations," *NIST Special Publication 960-12*, 2009. [Online]. https://ws680.nist.gov/publication/get_pdf.cfm?pub_id=50659

Appendix A – Frame Information Report

media_type	stream_index	coded_picture_number	key_frame	pict_type	repeat_pict	sample_aspect_ratio	height	width	pix_fmt	pkt_dts	pkt_dts_time	pkt_pts	pkt_pts_time	best_effort_timestamp	best_effort_timestamp_time	pkt_duration	pkt_duration_time	pkt_pos	pkt_size	channel_layout	channels	nb_samples	sample_fmt
video	0	0	1	I	0	1:1	1080	1920	yuv420p	0	0	0	0	0	0	1001	0.016683	58502	149863				
audio	1		1							0	0	0	0	0	0	1024	0.021333	273623	9	stereo	2	1024	fltp
video	0	2	0	B	0	1:1	1080	1920	yuv420p	1001	0.016683	1001	0.016683	1001	0.016683	1001	0.016683	263747	9876				
audio	1		1							1024	0.021333	1024	0.021333	1024	0.021333	1024	0.021333	329444	9	stereo	2	1024	fltp
video	0	1	0	P	0	1:1	1080	1920	yuv420p	2002	0.033367	2002	0.033367	2002	0.033367	1001	0.016683	208391	55356				
audio	1		1							2048	0.042667	2048	0.042667	2048	0.042667	1024	0.021333	341478	308	stereo	2	1024	fltp
video	0	4	0	B	0	1:1	1080	1920	yuv420p	3003	0.05005	3003	0.05005	3003	0.05005	1001	0.016683	329453	12025				
audio	1		1							3072	0.064	3072	0.064	3072	0.064	1024	0.021333	400513	405	stereo	2	1024	fltp
video	0	3	0	P	0	1:1	1080	1920	yuv420p	4004	0.066733	4004	0.066733	4004	0.066733	1001	0.016683	273632	55812				
video	0	7	0	B	0	1:1	1080	1920	yuv420p	5005	0.083417	5005	0.083417	5005	0.083417	1001	0.016683	414029	12258				
audio	1		1							4096	0.085333	4096	0.085333	4096	0.085333	1024	0.021333	426287	377	stereo	2	1024	fltp
video	0	6	0	B	0	1:1	1080	1920	yuv420p	6006	0.1001	6006	0.1001	6006	0.1001	1001	0.016683	400918	13111				
audio	1		1							5120	0.106667	5120	0.106667	5120	0.106667	1024	0.021333	452473	376	stereo	2	1024	fltp
video	0	5	0	P	0	1:1	1080	1920	yuv420p	7007	0.116783	7007	0.116783	7007	0.116783	1001	0.016683	341786	58727				
audio	1		1							6144	0.128	6144	0.128	6144	0.128	1024	0.021333	460643	367	stereo	2	1024	fltp
video	0	10	0	B	0	1:1	1080	1920	yuv420p	8008	0.133467	8008	0.133467	8008	0.133467	1001	0.016683	461010	8672				
audio	1		1							7168	0.149333	7168	0.149333	7168	0.149333	1024	0.021333	469682	422	stereo	2	1024	fltp
video	0	9	0	B	0	1:1	1080	1920	yuv420p	9009	0.15015	9009	0.15015	9009	0.15015	1001	0.016683	452849	7794				
video	0	8	0	P	0	1:1	1080	1920	yuv420p	10010	0.166833	10010	0.166833	10010	0.166833	1001	0.016683	426664	25809				
audio	1		1							8192	0.170667	8192	0.170667	8192	0.170667	1024	0.021333	488349	534	stereo	2	1024	fltp
video	0	13	0	B	0	1:1	1080	1920	yuv420p	11011	0.183517	11011	0.183517	11011	0.183517	1001	0.016683	488883	5294				
audio	1		1							9216	0.192	9216	0.192	9216	0.192	1024	0.021333	494177	489	stereo	2	1024	fltp
video	0	12	0	B	0	1:1	1080	1920	yuv420p	12012	0.2002	12012	0.2002	12012	0.2002	1001	0.016683	484102	4247				
audio	1		1							10240	0.213333	10240	0.213333	10240	0.213333	1024	0.021333	515780	468	stereo	2	1024	fltp
video	0	11	0	P	0	1:1	1080	1920	yuv420p	13013	0.216883	13013	0.216883	13013	0.216883	1001	0.016683	470104	13998				
video	0	16	0	B	0	1:1	1080	1920	yuv420p	14014	0.233567	14014	0.233567	14014	0.233567	1001	0.016683	519817	3372				
audio	1		1							11264	0.234667	11264	0.234667	11264	0.234667	1024	0.021333	523189	445	stereo	2	1024	fltp
video	0	15	0	B	0	1:1	1080	1920	yuv420p	15015	0.25025	15015	0.25025	15015	0.25025	1001	0.016683	516248	3569				
audio	1		1							12288	0.256	12288	0.256	12288	0.256	1024	0.021333	548217	432	stereo	2	1024	fltp
video	0	14	0	P	0	1:1	1080	1920	yuv420p	16016	0.266933	16016	0.266933	16016	0.266933	1001	0.016683	494666	21114				
audio	1		1							13312	0.277333	13312	0.277333	13312	0.277333	1024	0.021333	552638	427	stereo	2	1024	fltp

SWGDE Best Practice for Frame Timing Analysis of H.264 Video Stored in ISO Base Media File Formats

History

Revision	Issue Date	Section	History
1.0 DRAFT	2019-06-06	All	Initial draft created and voted by SWGDE for release as a Draft for Public Comment.
1.0 DRAFT	2019-07-16	All	Formatting and technical edit performed for release as a Draft for Public Comment.