SWGDE Technical Overview of Digital Video Files

Disclaimer:
As a condition to the use of this document and the information contained therein, the SWGDE requests notification by e-mail before or contemporaneous to the introduction of this document, or any portion thereof, as a marked exhibit offered for or moved into evidence in any judicial, administrative, legislative or adjudicatory hearing or other proceeding (including discovery proceedings) in the United States or any Foreign country. Such notification shall include: 1) The formal name of the proceeding, including docket number or similar identifier; 2) the name and location of the body conducting the hearing or proceeding; 3) subsequent to the use of this document in a formal proceeding please notify SWGDE as to its use and outcome; 4) the name, mailing address (if available) and contact information of the party offering or moving the document into evidence. Notifications should be sent to secretary@swgde.org.

It is the reader's responsibility to ensure they have the most current version of this document. It is recommended that previous versions be archived.

Redistribution Policy:
SWGDE grants permission for redistribution and use of all publicly posted documents created by SWGDE, provided that the following conditions are met:

1. Redistribution of documents or parts of documents must retain the SWGDE cover page containing the disclaimer.
2. Neither the name of SWGDE nor the names of contributors may be used to endorse or promote products derived from its documents.
3. Any reference or quote from a SWGDE document must include the version number (or create date) of the document and mention if the document is in a draft status.

Requests for Modification:
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 Technical Overview of Digital Video Files

Table of Contents

Table of Figures

1. Introduction

This document provides a foundation of knowledge of file formats, encoding standards, and compression algorithms used in digital video. It does not cover still image compression algorithms or file formats[1]. Understanding these elements, including the advantages and disadvantages of the options within each element, allows organizations to make informed decisions about the handling of digital video evidence. For a comprehensive understanding, the reader is encouraged to seek out additional sources[2].

1.1 General Concepts

There are several important terms and concepts related to digital video that may seem difficult to understand. As an example, practitioners often conflate video file formats with video encoding, when, in practice, these terms describe unique elements of digital video. The terms ***file format***[3], ***wrapper***, and ***container*** are used interchangeably and represent the same concept, which is a standardized structural method to store the variety of elements necessary to represent video and audio information. An ***encoding format*** is an algorithm applied to video data to reduce its size for lower impact on storage and/or bandwidth requirements, while retaining as much image fidelity as possible.

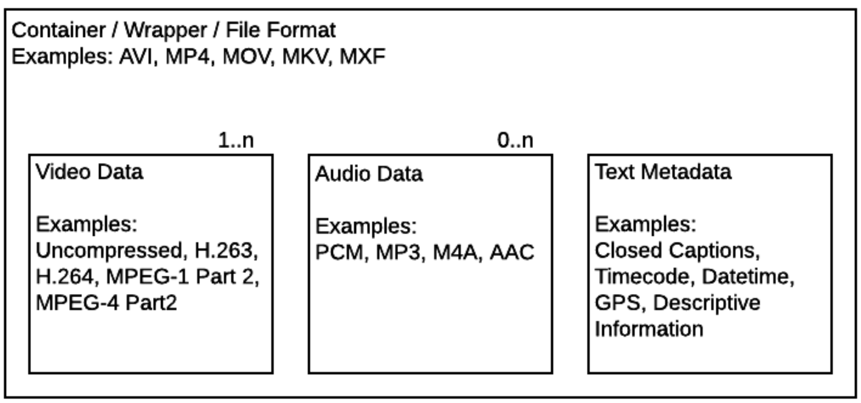

Figure 1. Simplified illustration of the relationship between digital file containers and video, audio, and metadata information stored within

The illustration above simplifies the anatomy of digital video, demonstrating that the container is the holder of at least one encoded video stream, zero or more audio streams, and any additional textual information that is included in the file.

[1] See *SWGDE Digital Image Compression and File Formats Guidelines* for detailed information on digital still images. [1]

[2] See ***Section 8. Additional Resources*** at the end of this document for suggested sources.

[3] See *SWGDE Digital and Multimedia Evidence Glossary* for a succinct definition of this concept. [2]

2. Container

A multimedia container (referred to in this document as a container) is a digital file format that is used as a wrapper for data files. The specification for a container—for example, the AVI format specification, the ISO Base Media File format (MP4), or even a proprietary container specification—describes how different elements of data co-exist within the file. Some containers are simple, designed for only a single type of audio or video data. Others are much more advanced and can support a variety of audio and video file types, as well as subtitles, chapter information, metadata, and synchronization details necessary for proper playback. When choosing the most appropriate container to use, it is important to consider its compatibility with commonly available media players, as some files require proprietary applications for playback and/or retrieval of all available metadata.

2.1 Examples of Common Multimedia Containers

Name	Description
3GP	Used mostly for mobile phone recordings
ASF	(Advanced Systems Format) Originally used for Microsoft .WMA and .WMV files
AVI	Part of the RIFF family of formats, this is a standard Microsoft Windows container
FLV/F4V	(Flash Video) Developed by Adobe Systems for flash video, also includes SWF extensions
MKV	(Matroska) An open standard container that can hold almost any file format
MOV/ QuickTime	Standard video container from Apple, Inc.
MPEG	Standard container for MPEG-1 and MPEG-2 streams used on DVD-Video discs and some others
MPEG-2/ MPEG-TS	Used for digital broadcasting, and on Blu-ray Disc
MPEG-4/MP4	A standard audio and video container for the MPEG-4 and H.264

Table 1. Selected examples of common video file format[4]

There are many other multimedia container formats, and this list should not be considered exhaustive.

[4] This table was derived from a list originally located at the Library of Congress. [3]

2.2 Notes on Atypical Data Streams and Containers

In some applications (e.g., DVRs, proprietary playback systems), the video samples, audio samples, and/or text metadata are stored separately on the file system and is compiled into a sequence. In these cases, the elements are not typically stored in a container format. The proprietary application has internal code that compiles the elements for playback within the system. In order to extract the video data from these applications, either the system must support an export function that will organize the video samples, audio samples, and/or text metadata into a standard container; the system will export the video samples, audio samples, and/or text metadata into a proprietary container (still requiring the proprietary player to render the video object); or an examiner will need to obtain direct access to the storage media to retrieve a bit-for-bit copy of the contents in order to locate, carve, and recompile the video samples, audio samples, and/or text metadata manually[5].

3. Codec[6]

A codec is an algorithm used to encode or decode a stream of digital data or signals according to a specific encoding format. Video codecs use encoding formats to compress data for more efficient transmission or storage of recordings. Decoding extracts digital video data from a previously encoded file, converting it into a displayable, decompressed form for playback or editing.

It is important to be aware of the following considerations concerning video codecs:

- There are many different encoding formats, and the amount of compression achieved can vary dramatically between the various encoding formats and even between different versions or implementations of the same codec.
- In general, higher compression can be achieved at the expense of reducing the quality of the decoded video.
- The original and decoded video may or may not be identical. If the output of the decoder is identical to the original video, the compression process is lossless. If the two videos are not identical, the compression process is lossy. It is not possible to recover the original data from lossy compressed video [5]. (See ***Section 6*** below for more information on lossless and lossy compression.)

[5] See *SWGDE Proposed Techniques for Advanced Data Recovery from Security DVRs Containing H.264 Data* for more in-depth discussion of this advanced process. [4]

[6] See [2], "Codec," for succinct definition of this concept.

The following table lists a sampling of common video encoding formats.

Standard	ISO/IEC	Common Uses
H.261	23002-1	Video-conferencing
MPEG-1 Part 2	11172	Video-CD
H.262/MPEG-2 Part 2	13818-2	DVD-Video, Blu-Ray, Digital Video Broadcasting, SVCD
H.263	14496(old)	Videoconferencing, Video Telephony, Video on Mobile Phones (3GP)
MPEG-4 Part 2	14496-2	Video on Internet (DivX, Xvid ...)
H.264/MPEG-4 AVC	14496-10	Blu-ray, Streaming Video, Digital Video Broadcasting, iPod Video, Apple TV
VC-2 (Dirac)	*SMPTE Std	Video on Internet, HDTV broadcast, UHDTV
H.265 (HEVC)	23008-2	4k streaming services and UHD Video on Demand

Table 2. Selected examples of common video encodings [7]

4. Frame Rate

Frame rate is a measure of the video display rate in frames per second (FPS). The higher the FPS, the smoother the motion appears. In general, the minimum FPS needed to provide an appearance of continuous/smooth motion is approximately 24 FPS. Below are the two most common television broadcast FPS standards:

- **NTSC** (National Television Standards Committee) – The NTSC is responsible for setting television and video standards in the United States. The NTSC standard for television defines a video framerate of 29.97 FPS. The NTSC standard also requires that these frames be interlaced.
- **PAL** (Phase Alternating Line) – The dominant television standard in Europe. The PAL standard delivers 25 FPS.

[7] This list of common encoding formats was adapted from a list originally located at the Library of Congress. [3]

5. Resolution

Resolution is the pixel dimensions of an image or video[8]. It is typically expressed as the number of pixels captured horizontally and vertically. Common video resolutions are 320x240 (some CCTV), 640x480 (Standard Definition Video), 1280x720 (720 High Definition [HD] Video), 1920x1080 (1080 HD Video), 1920x1200 (Computer Monitor) and 4096x2196 (4K or UHD Video). These are seen in the image below for a reference.

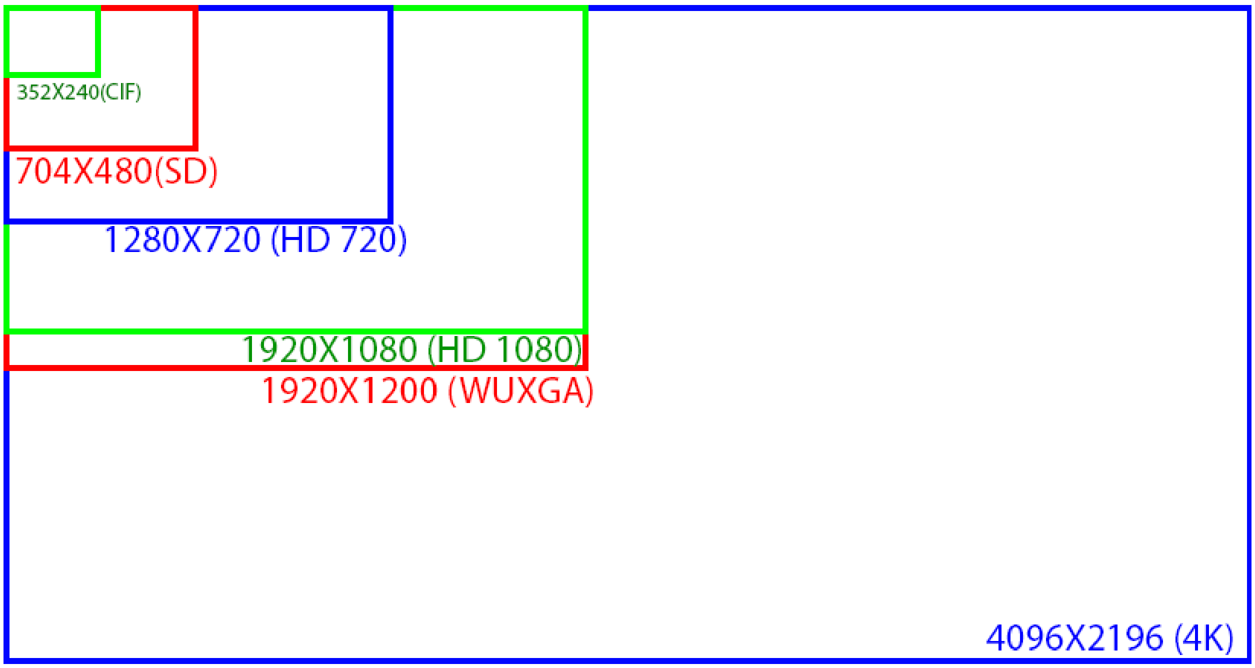

Figure 2. Comparison of common image dimensions used in video data

[8] See [2], "Resolution," for further detail on the facets of resolution as an image-related concept.

The more pixels there are in a given image, the higher the resolution, and the more detail that can potentially be captured and examined. In the example below, the same license plate is captured at varying resolutions. Note that at the lower resolutions, the information on the license plate is unreadable.

Figure 3. Illustration of the information loss as pixel resolution decreases

6. Compression

Compression is the process of reducing the size of a data file, utilizing algorithms to minimize redundancy and rearrange the way information is organized within the file.[9] Compression is often used to facilitate the storage, transfer, and/or streaming of large digital video files. Compression algorithms that retain all the original information are referred to as *lossless*, while those resulting in a loss of data are *lossy*.

SWGDE Digital Image Compression and File Formats Guidelines notes three methods used in standard image compression: run-length encoding (lossless), lexicographic encoding (lossless), and quantization encoding (lossy).[10] Additionally, moving image compression algorithms employ methods such as Group of Pictures (GOP) or chroma subsampling; both are lossy forms of compression. Other factors, such as bit rate settings, can be used in conjunction with encoding formats to limit quality or file size.

When deciding upon digital video encodings that use compression methods, decision makers should carefully consider cost, workflow, time, storage demands, available bandwidth, network infrastructure, and video quality. Lossy compression can make forensic analysis more difficult, even when, during motion playback, a recording seems to have properly captured events as they occurred. Also, when received files have been compressed, care should be taken not to compress them any further. If additional processing is required, it is preferable to save a copy of the file in an uncompressed format. Work can then continue as needed and can be saved with no compression or using a lossless encoding.

6.1 Lossless Compression

When using lossless compression, no information is lost, but the compressed file uses fewer bits to represent the information. When the file is decompressed, the original pre-compressed data is reconstructed completely. Generally, lossless compression can achieve compression at a ratio of about 2:1 (thus reducing the file size of the original file by half). Selection of a lossless compression option will result in the preservation of all data; however, the resulting file size may be significantly higher than what can be achieved using lossy compression methods. Lempel-Ziv-Welch (LZW) algorithm is an example of lossless compression.[11]

6.2 Lossy Compression

To achieve higher compression ratios, lossy compression algorithms reorganize or remove sections in video files that contain redundant or irrelevant data. This compression reduces the number of bytes needed to represent the original image. Some video data will be lost during this process, and the original pre-compressed video data will not be able to be reconstructed in its original form.

The average user of commercially available software will have limited control on how lossy compression algorithms are deployed. This lack of control must be taken into consideration as a

[9] See [1], Section 2.7.

[10] See [1], Section 2.

[11] This section borrows language from [1].

compression method is being selected. Additionally, lossy compression algorithms will continue to compress the video data in each subsequent re-encoding of a video file, resulting in additional irretrievable loss of video data.

Lossy compression may produce compression artifacts, which are noticeable distortions of images, audio, and video.[12] These include, but are not limited to, blocking or pixelation, clipping of high or low audio frequencies, or jerky motion.

Figure 4. Example of the visual distortion that can be the result of high degrees of lossy compression

6.2.1 Spatial and Temporal Compression

Because video data consists of a sequence of still images reproduced in real time, compression can take place within a single image frame of video as well as across a set of related image frames. Spatial compression (intraframe compression) refers to compression methods that reduce the data contained within a single video frame by eliminating redundancy within areas of similar color. This compression decreases the file size for each frame of spatially compressed video. The adverse effects of this are that in grouping similar pixel values, high-frequency information or details can be lost, such as in Abraham Lincoln's beard in ***Figure 4***. This effect is further discussed in the section ***6.2.3 Macroblocks***.

Temporal compression (interframe compression) reduces the data contained within a single video frame by eliminating redundancy between similar areas in adjacent frames. The adverse effects of this are similar areas can be copied or moved from previous frames, making it possible to miss small changes within an individual frame of video. This topic is further discussed in the section ***6.2.4 GOP (Group of Pictures) Structure***.

6.2.2 Bit Rate

The bit rate of a video file is the size of the data stream when the video is being rendered in real time, often expressed in kilobits (Kbps) or megabits per second (Mbps). It represents the quality at which a video is transmitted. Bit rate specifies the minimum capabilities needed to play a video without interruption. With higher bit rates, a particular encoding format can support a larger frame size, higher frame rate, less compression per frame, or some combination of each. With a lower bit rate, one or more of these video signal characteristics will be reduced.

---

[12] See [1], Section 2.5, for more detailed coverage of this topic.

Most tools that can edit or transcode video can also offer the user the ability to set the bit rate for any resultant file. Options include constant bit rate or variable bit rate, which offers the software the flexibility to add more compression based on changes in the image data.

6.2.3 Macroblocks

Each codec has a set way of handling the encoding and decoding of pixel information. Many codecs utilize macroblocking, which takes the information in a defined region and examines each pixel in relation to the other pixels in that region. Depending on the codec, this can be done in 4x4, 8x8, or 16x16 pixel blocks. More modern codecs, such as H.264 and H.265, have the ability to utilize dynamic macroblocks (sometimes called Coding Tree Units). This allows blocks to change between 4x4 and 64x64 pixel regions to preserve details.

6.2.3.1 Chroma Subsampling. Within a macroblock, each pixel has a specific brightness (called *luminance*, usually signaled as *Y)* and color information (called *chrominance*, usually signaled as *Cb* and *Cr)*. If the luminance information is isolated without the chroma information, the resulting image will show each pixel in a shade of gray. Chroma subsampling involves the reduction of the color information for neighboring sets of pixels to reduce the overall storage size and bit rate of a video file. Chroma subsampling is used in multiple image compression codecs. How much compression happens is often expressed as a set of numbers that define a pixel set, such as 4:4:4, 4:4:2, or 4:2:0, where the first value describes the ratio of *luminance (Y)* values in a sample, and the second and third values describe the ratio of *chrominance* (*Cb* and *Cr*) values that will be used to describe the neighboring pixels in the set.

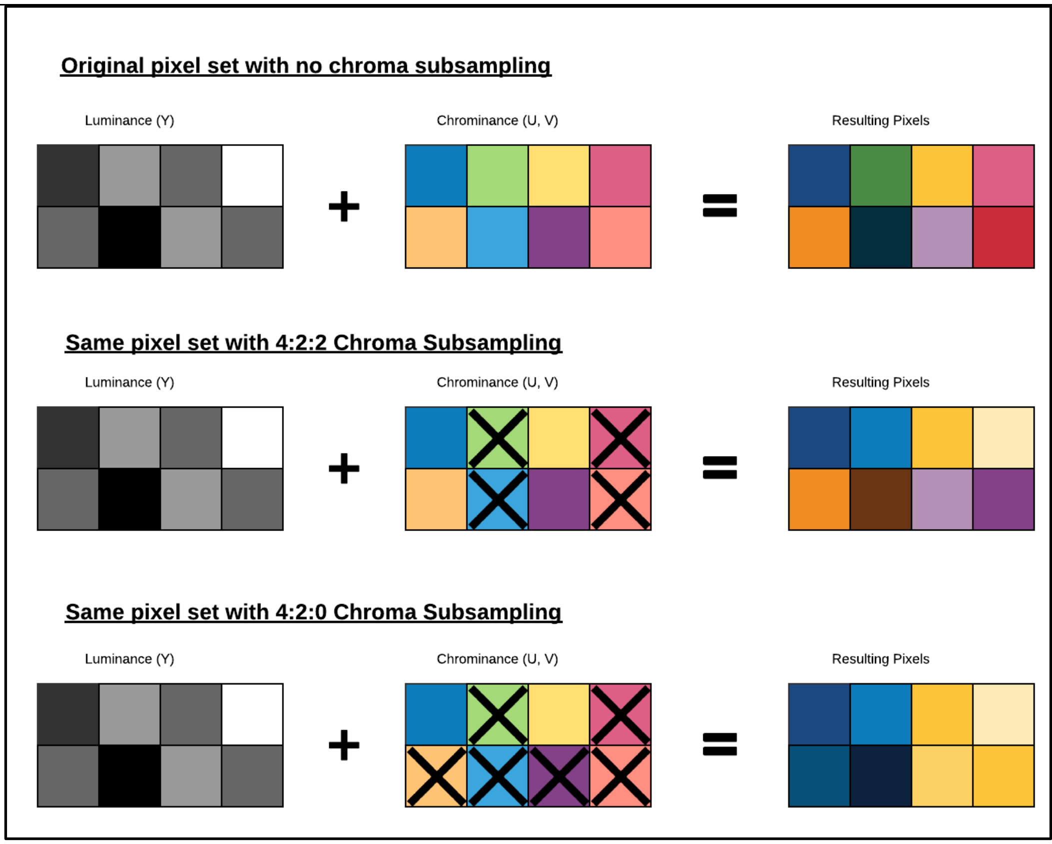

Figure 5. Theoretical illustration of the color information loss that is the result of chroma subsampling

With each level of chroma subsampling, further color information is irretrievably lost from the video data. This is one factor that affects how accurately color may be captured or represented in the video.

Common chroma subsampling ratios include the following:

- **4:4:4** is the highest quality; it effectively has no subsampling, because each pixel is represented and retains its luminance and chroma values.
- **4:2:2** samples two pixels from both the top and bottom rows of a pixel set, reducing the chroma information to 50 percent of the uncompressed source chroma.
- **4:2:0** takes two chroma samples from the top row of a pixel set and none from the bottom row, reducing the overall chroma information to approximately 25 percent of the uncompressed chroma.
- **4:1:1** (not shown above) takes one sample from the top row of a pixel set and one sample from the bottom row, reducing the overall chroma information by 25 percent of the uncompressed chroma.

6.2.4 GOP (Group of Pictures) Structure

A GOP structure is a term to describe both the order and number of a group of frames, each made up of an Intra Frame (I-frame), plus Predicted Frames (P-frames), and Bi-Directional Frames (B-frames), before the next I-frame. Visible data is generated by adding and moving data from the surrounding frames leading up to the next I-frame. When the next I-frame is displayed, the decoder is then instructed to start a new GOP, and previous data no longer needs to be retained. This process allows a file to save space by reducing the amount of redundant information.

6.2.4.1 **I-frames** (also known as a *reference* or *key* frames) are frames that contain all newly encoded information. These frames generally have the most information in the GOP. These frames also contain the most accurate information, as no section will come from other frames. Every GOP contains one I-frame.

6.2.4.2 **P-frames** are predictive frames. They contain information relating to changes from the previous I-frame, and newly encoded information only when the amount of change exceeds a threshold set by the encoder. As such, information similar to that in the I-frame may be ignored or moved to accommodate changes.

6.2.4.3 **B-frames** are encoded based on interpolation from the nearest I- and/or P-frames. Again, these frames may contain newly encoded information. Typically, though, this is less new data than I-frames or P-frames, because both an I-frame and/or P-frame can be used to adjust information before and/or after B-frames.

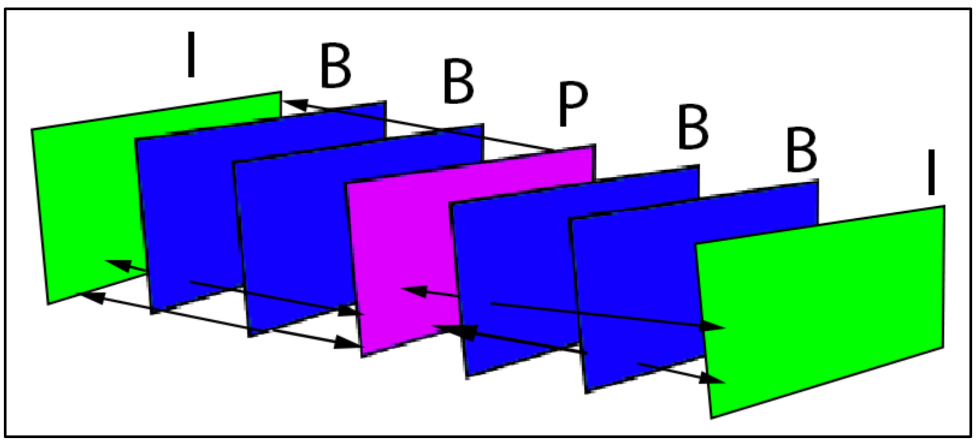

Figure 6. Illustration of the relationship between I-frames, P-frames, and B-frames

6.2.4.4 **GOP Length** is a measure of the number of predicted frames (P- or B-) that exist in a video stream between I-frames. Longer GOP lengths result in more efficient video encoding but do not capture quick transitions as effectively. Maximum GOP length is dependent upon playback specifications.

6.2.4.5 **GOP Pattern** is the arrangement of P- and B-frames within a GOP. These are typically expressed as IBBP or IP. The pattern after the I-frame describes the separation between P-frames and is not a full descriptor of the GOP. For example, an IP pattern defines a GOP with no B-frames, and IBBP shows that there are two B-frames between each P-frame. Smaller GOP patterns with shorter GOP lengths are more efficient for use with a video that includes rapid transitions, although they do not offer a high compression ratio.

7. Conclusion

With a comprehensive understanding of digital video file components and structure, administrators will be better equipped to make informed procurement decisions, and practitioners can ensure devices and recording equipment are deployed in the most efficient and effective manner. Careful considerations of the container, codec, frame rate, resolution, and compression can assist in defining project requirements and properly configuring systems for their intended use.

It is recommended users of this document test and evaluate video systems for technical configuration and digital video output files to make informed decisions on optimal operational use.

8. Additional Resources

- C. Lacinak, "A Primer on Codecs for Moving Image and Sound Archives & 10 Recommendations for Codec Selection and Management," AudioVisual Preservation Solutions, Inc., New York, April 8, 2010. [Online]. http://www.avpreserve.com/wp-content/uploads/2010/04/AVPS_Codec_Primer.pdf
- Federal Agencies Digitization Guidelines Initiative (FADGI) Audio-Visual Working Group. (2012, December) Creating and Archiving Born Digital Video. [Online]. http://www.digitizationguidelines.gov/guidelines/video_bornDigital.html
- C. Poynton. (2008) Chroma subsampling notation. [Online]. http://www.poynton.com/PDFs/Chroma_subsampling_notation.pdf
- The International Association For Identification and The Law Enforcement/Emergency Services Video Association International, Inc., "Forensic Imaging And Multi-media Glossary Covering Computer Evidence Recovery (CER), Forensic Audio (FA), Forensic Photography (FP), And Forensic Video (FV)," July 15, 2006. [Online]. https://leva.org/pdfs/GlossaryV7.pdf
- Memoriav, Association for the Preservation of the Audiovisual Heritage of Switzerland, "Digital Archiving of Film and Video: Principles and Gudiance," English version 1.0 April 2016. [Online]. http://memoriav.ch/recommendations-digital-archiving-film-video/

9. References

[1] Scientific Working Group on Digital Evidence, "SWGDE Digital Image Compression and File Formats Guidelines,". [Online]. https://www.swgde.org/documents/documents

[2] Scientific Working Group on Digital Evidence, "SWGDE Digital & Multimedia Evidence Glossary". [Online]. https://www.swgde.org/documents

[3] Library of Congress. (2017, June) Sustainability of Digital Formats: Planning for Library of Congress Collections. [Online]. http://www.loc.gov/preservation/digital/formats/fdd/browse_list.shtml

[4] Scientific Working Group on Digital Evidence, "SWGDE Proposed Techniques for Advanced Data Recovery from Security Digital Video Recorders Containing H.264 Data,". [Online]. https://www.swgde.org/documents

[5] I. Richardson, "Video Compression Codecs: A Survival Guide," *International Association of Sound and Audiovisual Archives (IASA) Journal*, no. 47, January 2017.

SWGDE Technical Overview of Digital Video Files

History

Revision	Issue Date	Section	History
1.0 DRAFT	2017-01-12	All	Initial draft created and SWGDE voted to release as a Draft for Public Comment.
1.0 DRAFT	2017-02-21	All	Formatted and technical edit performed for release as a Draft for Public Comment.
1.0 DRAFT	2017-06-22	All	Updated all sections of the document in response to comments received. SWGDE voted to approve as an Approved document.
1.0	2017-07-18	All	Formatted and published as Approved version 1.0.