A Quick Guide to Digital Video Files

Digital Preservation Office at the University of Michigan
August 2010

What we think of as a video file is frequently a "wrapper" surrounding two different parts: the audio stream and the video stream. This can add an added degree of complexity to both the creation and the preservation of a video file. It is therefore, important to pay attention to things such as resolution and bit rate when creating a video in order to balance the desired quality with the file size. Paying close attention to the details can help you make videos that are easier for viewers to watch and to preserve in the long term.

Source

An important thing to consider when creating a video file is the source you're using. This can have an effect on the quality of the audio and video of the file. Take the example of converting an older VHS tape into a digital video. Despite choosing a HD (High Definition) resolution, such as 1920x1080, you will not create a video that looks the same as an HD broadcast, simply because the resolution of the original video was much lower. Try to balance the original source quality with the other elements you can control when creating your video file, such as file size, rather than just arbitrarily deciding on a certain level of quality.

Resolution and Aspect Ratio

As stated earlier, it's best to consider the resolution of your source when converting to digital for things such as DVDs, VHS tapes, etc. If you are encoding far above the resolution of your original source, chances are you're just likely to get a very large file with picture quality not much different than the original. When creating a born digital file, it is best to consider the intended file size and the amount of resolution required for the content. Below are some points of reference for common resolutions that are used today.

- 1920x1080 = Also known as 1080p. Used for high definition broadcast streams and also as the resolution for the 'high' profile setting in MPEG-2. To complicate matters, 1080i also exists, which are two 1920x540 fields that are interlaced (combined) to form a 1920x1080 image.
- 1280x720 = Also known as 720p. Used in high definition broadcasts.
- 720x483 = Standard component/composite. Used in the past for broadcast transmissions as well as for VHS tapes, both NTSC (American standard) & PAL (European standard).
- 720x480 = Digital Standard Definition. Commonly known as 480i (when interlaced) or 480p (when progressive). Also used as the resolution for the 'main' profile setting in MPEG-2.
- 352x240 = The 'low' level profile setting resolution used in MPEG-2.

The Aspect ratio of a video is the ratio of the width to the height. There are two common aspect ratios currently in use.

- 4:3 has been the aspect ratio commonly used in the past for broadcast television and VHS tapes.

- 16:9 is what we think of as a 'widescreen' picture. All standard High Definition video formats are this aspect ratio.

Bit rate

Bit rate (also known as data rate) is the maximum amount of information that can be processed/unit of time. A higher bit rate means more information is being captured in each frame of the video. They also increase the quality of the video as well as the size of the file.

Format[i]

Unfortunately, a single format does not exist that will meet every need related to digital video. When deciding on a format to use, there are a few questions that can be used as a guidelines. A few of the common formats are listed below with a short overview of each one.

- Is this format currently in common use and well supported by playback software?
    - It is always best not to use an obscure or seldom used format when creating a video file. The more widely supported a format is now, the more interest there is in maintaining it over time, and the greater chance there is for the video to be usable in the future.
- Is this format a proprietary format? Does it have open documentation?
    - Proprietary formats are less likely to have open documentation and source materials available than non-proprietary formats. If, in the future, a video was no longer able to be decoded, it would likely be easier to reconstruct a software program that could decode an open format than a proprietary one, due to available documentation. Open documentation allows for the (re)creation of new software programs that can open and use the files.
- What are the current preservation standards?
    - Use sites such as the Library of Congress's Sustainability of Digital Formats[1] to see information on currently used formats.
    - Research what formats are being used by archival organizations (Library of Congress, National Archives, your University library system, etc.)

Common File Formats

Format	Overview
AVI[ii]	File extension is .avi. Proprietary format developed by Microsoft and IBM for Windows 3.1. Has been used for video production and in the past was used for projects such as the Library of Congress' American Memory Project.
MPEG-2[iii]	File extension is usually seen as .mpeg or mpg. Open standard format developed by ISO. Required by ATSC for digital non-satellite broadcasting and also used by organizations such as the Library of Congress.
MPEG-4[ivv]	File extension is .mp4. Open standard format developed by ISO which is based

[1] Library of Congress. (2007, May 21). *Sustainability of Digital Formats: Planning for the Library of Congress Collections*. Retrieved August 2, 2010, from http://www.digitalpreservation.gov/formats/

	on the Quicktime file format. Used in DVD systems and for closed system broadcasting such as satellite television and mobile phone broadcasting.
Quicktime[vi]	File extension is .mov. Proprietary format developed by Apple. Used with Apple products such as the iPod and also is the basis for MPEG-4.
WMV[vii]	File extension is .wmv. Proprietary format developed by Microsoft. Software to play and make WMVs is included with the Windows operating system.

At this time, MPEG-2 and MPEG-4 stand out as the best formats for long term preservation. They are open formats, and are used both for commercial broadcast and by government entities such as the Library of Congress. This helps to ensure that these formats will likely have ways to preserve them in the future as opposed to proprietary formats that are not used as widely.

File Size

File size is heavily dependent on the previous elements discussed in this document. The size of the video file is determined by the video codec (**co**mpressor/**dec**ompressor) selected, the file format, the amount of compression used, the resolution, and the bit rate. A high quality video – using a high resolution, low compression rate, and high bit rate, for example – will have a large file size. A large file might not matter for only a single video, but it can quickly add up given the many files that can be in a collection. It is therefore best to balance the desired video quality with the desired file size. Think of what you are trying to convey to your potential viewers. Is it important that the video be able to quickly stream over the Internet, or do your viewers need to see minute details?

Metadata

Metadata is, essentially, 'data about data.' It is the information stored to describe aspects of a file, and the contents of the file. This can include everything from technical specifications, such as resolution and bit rate, to bibliographic information such as title and author. It is important to include as much information about the file you are creating as you can. This information will give future users or archivists a context for your file as well as the ability to help preserve it. Metadata for the objects can be embedded within the file in the header, as it is with AVI[viii] or QuickTime[ix] files or stored externally as it is with MPEG (MPEG bibliographic metadata can be stored in MPEG-7 files)[x]. Metadata can also stored in separate databases or spreadsheets that the video creator or archivist creates and maintains.

Keep a backup copy!

One last thing that should be done is the creation of a backup copy (or several copies). Put these backup copies on external media such as DVDs or a thumb drive in case the original is lost.

---

[i] Library of Congress. (2007, March 7). *Sustainability Factors*. Retrieved July 27, 2010, from http://www.digitalpreservation.gov/formats/sustain/sustain.shtml

[ii] Library of Congress. (2010, May 13). *AVI*. Retrieved July 28, 2010, from http://www.digitalpreservation.gov/formats/fdd/fdd000059.shtml

[iii] Library of Congress. (2007, March 7). *MPEG-2 Video Encoding (H.262)*. Retrieved July 28, 2010, from http://www.digitalpreservation.gov/formats/fdd/fdd000028.shtml

[iv] Library of Congress. (2007, March 7). *MPEG-4, Advanced Video Coding (Part 10) (H.264)*. Retrieved July 28, 2010, from http://www.digitalpreservation.gov/formats/fdd/fdd000081.shtml

v Library of Congress. (2007, March 7). *MPEG-4, Visual Coding (Part 2) (H.263)*. Retrieved July 28, 2010, from http://www.digitalpreservation.gov/formats/fdd/fdd000080.shtml

[vi] Library of Congress. (2007, August 21). *Quicktime File Format*. Retrieved July 28, 2010, from http://www.digitalpreservation.gov/formats/fdd/fdd000052.shtml

[vii] Library of Congress. (2007, August 20). *WMA (Windows Media Video) File Format*. Retrieved July 28, 2010, from http://www.digitalpreservation.gov/formats/fdd/fdd000091.shtml

[viii] Library of Congress. (2010, May 13). *AVI*.

[ix] Library of Congress. (2007, August 21). *Quicktime File Format*.

[x] Library of Congress. (2007, March 7). *MPEG-4, Advanced Video Coding (Part 10) (H.264)*.