iTunes Video Best Practices 4.0

2010-02-02

Apple Inc.
1 Infinite Loop
Cupertino, CA 95014
408-996-1010

Simultaneously published in the United States and Canada.

Contents

Video Best Practices

Overview

This document reviews and discusses different types of digital video and provides information to assist providers in creating a quality product.

Revision History

Date/Version	Changes Made
February 3, 2010 - Version 4.0	Removed “Encoding and Delivery” and “Music Video Cover Art” chapters.
June 8, 2007 - Version 3.0	Includes discussion of high definition video. Added recommendations for creation of music video cover art.
April 1, 2007 - Version 1.0	Initial document version.

A Simplified Background On Video

Standards

There are essentially three types of video, two are “Standard Definition” (SDTV) varieties and the rest are, by definition, anything larger than SDTV, or “High Definition” (HDTV).

The two varieties of SDTV are divided into the U.S. NTSC specification and the mainly European PAL specification. NTSC and PAL both occupy a 4:3 display ratio, while HDTV generally occupies a 16:9 widescreen ratio.

NTSC and PAL were defined early in the last century and have very specific guidelines governing how a picture can be transmitted and displayed with standardized analog, tube technology. HDTV was developed for digital distribution on a variety of different display devices. All raw SDTV frames have black “inactive” picture areas around the video frame that is a vestige of the SDTV standard. HDTV is more likely to be clean from edge to edge, with no inactive black areas.

The spacing of SDTV resolution is not designed for “square pixel” display devices like HDTV, so SDTV pixels have to be converted to square to allow the aspect ratios of an SDTV image to display correctly (more below).

Displays

There are two types of video displays in the marketplace: *interlaced* and *progressive*. Generally, anything shot and edited for television up until a few years ago was interlaced to work with tube display technology. Any flat- screen plasma, LCD or computer monitor is a progressive display. Everything iTunes delivers to its customers is geared for computers, iPods, and modern television displays. Interlaced television transmission is scheduled to cease in 2009.

Interlaced vs. progressive

If handled properly, progressive source creates the best video on both progressive and interlaced displays. Interlaced files never look good on progressive displays.

Interlaced video *frames* are created from two *fields*, each field containing half the *frame* information. These fields are never supposed to be displayed simultaneously. "De-interlacing" is a way to create a progressive frame out of a pair of interlaced fields by choosing a dominant field and supplementing it with pieces of its counterpart.

The "supplementing" part of the job is called *interpolation*. In most cases, de-interlaced video **appears** to have the resolution of progressive video except on diagonal lines and circles, where there will be noticeable stair steps resulting from poor interpolation (termed "aliasing").

Understanding Color

iTunes content conforms to the VGA standard of eight bit color expressed as 24 bits per pixel, 256 colors per channel or about 16 million possible colors per pixel. The "color space" used for delivery is "YUV," defined as one channel of luminance (Y) and two color components (U and V). Each channel, or "component," is assigned a set number of bits expressed as "4:2:0" for four bits of luminance, two bits for the channel representing red and no bits at all representing blue. Blue is created from differences between red and luminance channels be- cause human vision doesn't see blue very well.

NTSC and PAL use the YUV model.

Any source intended to be compressed and delivered by iTunes in 4:2:0 color must equal or exceed 4:2:0 color. DV25 (MiniDV and DVCAM) for NTSC is 4:1:1 and is unacceptable because, when converted to 4:2:0, the effective amount of color information will be 4:1:0. DV25 for PAL is 4:2:0, so the color depth is acceptable.

DVCPRO25 in either NTSC or PAL is 4:1:1 and is not acceptable. DVCPRO50 is 4:2:2 and exceeds our requirement.

HDCAM is 3:1:1 but, because the actual recorded resolution is less than the display resolution of 1920x1080 (whereas standard definition is recorded at 100% of its display dimension), the color sample rate isn't directly associated with our 4:2:0 needs. When the HDCAM is compressed from its full resolution, the 3:1:1 is accept- able. HDCAM SR is 4:2:2 or 4:4:4 and is superior.

The general rule is that source content to be delivered to iTunes should be 4:2:0 or better at full resolution.

Film vs. video

Even though progressive video is the wave of the future, it also has roots in the past. Film remains the best form of progressive source. The good news is that the great majority of TV series produced prior to 1980 were shot and edited in film, and transferred in one pass to videotape using a process called a 3:2 pulldown. In this process, 24 frames per second (fps) film is converted to ~30 fps video by sharing fields with other frames. In- stead of two fields per frame, some have three fields. The result is acceptable on an interlaced display, but blurry and difficult to compress in progressive environments. The fact that an entire program was transferred from film to tape in one pass (called a "cadence") means that reversing the transfer back to 24 fps is relatively easy with the right tools. Anything that was shot at 24 fps should be delivered to the progressive-screen viewer at 24 fps (or 23.98 fps if the video was previously recorded on tape).

A vast majority of music videos are shot on film. Unfortunately, the common approach after the film is processed is to transfer it all to video in one cadence and edit the videotape. Video editing totally ignores the film cadence and mixes the odd and even telecined (interlaced) fields from scene to scene. It's much more difficult to return this kind of video back to ~24 fps, and simply de-interlacing makes this type of video unsharp and difficult to compress.

Aspect Ratio

When SDTV video is digitized, the file is displayed in an aspect ratio that appears too wide (720 rectangular pixels horizontal) when viewed on a progressive screen. Circles are squashed and people look fatter and shorter than they should be. To return the stretched look to normal, the 720 rectangular pixels have to be squeezed down to 640 square pixels horizontal. The vertical size is unaffected by this change, so squashed circles be- come round again.

Aspect ratio is even more complicated by the use of 16:9 inside the SDTV standard of 4:3. This has been an artistic method of providing a cinematic experience while often reducing the actual active picture area by a significant amount. There are two ways that the SDTV standard can offer a 16:9 image and they each have to be treated differently when creating a file for iTunes.

Letterbox

The easy way is to edit a video within a boundary that may be constrained by an edit system timeline. Every 4:3 scene is moved so that its critical action fits inside the boundary, or mask. This is easy because no special cam- era equipment is required to create the cinematic look. The unfortunate result is that a 640x480 image is now downsized to 640x360, 640x320, even 640x240, leaving less than half the original available resolution. The SDTV standard doesn't change, so all of the black pixels in the masked area (the difference in pixels between the vertical dimension of, say, 360 and 480) have to be eliminated. For example, 120 pixels have to be "cropped" out, leaving a new file with a true, active picture area of 640x360.

Video files in iTunes benefit from not carrying all of those black pixels as excess baggage. All black pixels—top and bottom, left and right—have to be cropped out and the resulting file has to end up 640 wide by whatever vertical dimension best returns the file to its correct aspect ratio. MPEG is based on 16 pixel macro-blocks so files should be divisible by 16 in both dimensions. 360 is a common vertical dimension for 16:9 content and MPEG represents this size without any problems. We don't recommend odd dimensions. At the minimum, we recommend divisibility by eight.

Anamorphic

The second way to create 16:9 inside of the SDTV standard starts with the camera or film transfer. Knowing that all of those black pixels are wasted in the letterbox method, it was determined that a wider chip would allow more image to be gathered horizontally while leaving the required vertical dimension in compliance with the SDTV standard. The wider chip gathers more information, but when displayed at 640x480, everybody looks really skinny. This type of content is called anamorphic, and it is absolutely the best SDTV source for a progressive screen.

When preparing an anamorphic file for iTunes, crop as you would any other 4:3 file and treat it for frame rate as suggested in this document, with one difference: when you compress the file, assign 360 as the vertical dimension. This will squeeze the image back into the correct 640x360 ratio while preserving information normally lost in the de-interlace process (i.e. no stair step aliasing on diagonal lines).

Mixed aspect ratios

Graphics must be the same size as the program content unless there is a clear artistic reason for mixing aspect ratios. 4:3 televisions are becoming vestiges of a bygone era, so creating 4:3 bumpers and opening graphics will force the entire delivery file to be 4:3. The larger axis (vertical or horizontal) will take precedence on the display, resulting in big graphics and a smaller presentation of the body of the show, with black bars on all four sides. On a small display it's an especially disappointing result.

Difficult Files

Mixed frame rate

Some content originating on video is inter-cut or composited with content originating on film. Masters exhibiting this issue make it impossible to apply an inverse-telecine process to remove the 3:2 pull-down and return the video to its native frame rate for digital playback. Though the file will never be flawless, the only solution is to simply crop and de-interlace the file. It is important we inform video creators that film should be shot at 30 fps to accommodate the video portions, or video should be edited with film in a 24 fps timeline.

Effects and graphics at a different frame rate

Often a show is shot and edited at ~24 fps and the overlaying titles or effects are added at ~30 fps. While the show is edited in cadence and a very good inverse-telecine can be accomplished, the graphics or the effects suffer from missing frames. This is often a problem with animation created at 12 or 24 fps with 29.97 pans, zooms, or tilts. Usually shows contaminated with mixed frame rates like this have to be delivered at 29.97 which reduces sharpness and creates compression artifacts.

Videos that appear to be ripped from a DVD

These files may be the result of a video master being misplaced or lost, but video content that is ripped from a DVD and then transcoded usually exhibits unacceptable compression artifacts, unrecoverable frame rate issues, and generally substandard video quality. Sometimes the DVD authoring master file is offered, but these often have the same frame rate issues as well.

Upconverted data rates and dimensions

Source content that is smaller in size than the iTunes specification should not be increased in size ("scaled up") to meet the specification. There is no benefit to increasing display size from standard definition to high definition just as there is no benefit to recompressing a low data rate file to a 15 megabits per second (Mbps) MPEG- 2 file. Every transition a file endures will have some negative effects and these will become apparent when an iTunes delivery file is created from this kind of content.

Video that has been through a PAL to NTSC conversion process

Files that have undergone these conversions will result in unrecoverable frame rate issues that produce motion artifacts in the subsequent video file. It is necessary to seek source from the original edit timeline.

Legacy video originating on obsolete master

Sometimes the only available master is a legacy format that does not meet today's quality expectations. In particular, any video format that stores video in composite form (luminance and chroma combined) can be a problem. This includes masters on D2, 3/4 inch U-matic, VHS, and 1-inch (Type C) masters. The composite artifacts exhibit themselves as "vibrating color" around edges and details that may become blocky when compressed.

Video content with unequal aspect ratios

Some videos combine 4:3 with other dimensions; others neglect to align frames sizes consistently on the timeline. This results in the inactive pixel area moving into the active picture area unpredictably. Such mixed aspect ratios are difficult to crop, resulting in an unusual dimension (640x464 for instance); or worse yet, a file that requires we scale up because it has been cropped so much. The most extreme example of this problem is described in the paragraph above titled "Mixed aspect ratios."

Special treatments in editing

Post-production processes such as FilmLook® and similar treatments attempt to make video look as though it was shot on film when viewed on tube technology. These processes should be avoided when developing source for digital master files. If such a process is to be used for broadcast, be sure to retain a *preprocessed* video master for encoding purposes.

Your Archive

Every provider has a different method of archiving video assets depending on the video's age, size, and the immediacy with which the asset generally needs to be accessed. Some archives consist of Dub Master tapes created by the production company responsible for the project. In this case, the tape must be carefully captured and compressed to preserve all of the quality possible within the guidelines of the iTunes Video Delivery Specification. Some providers send their Dub Masters to an encoding house which delivers the completed files to iTunes.

As NTSC and PAL SDTV television heads off into the sunset, it is important that a greater emphasis be placed on the preservation of video assets. Up until a decade ago, it wasn't easy to store full-resolution video without a high degree of compression. Tape was truly the cheapest and easiest way to archive all of the hard work involved in a project. Eight Mbps and 15 Mbps MPEG-2 files were industry standards that don't hold up well to captures from the best tape formats of the day, and in many cases those tapes can't be found. We need to encourage video creators to provide contracting companies with a hard drive of the final edit file—in addition to the requisite broadcast format—so that decisions required to fulfill present needs don't restrict future uses. Storage is cheap, and archives should be designed now to preserve assets that can be repurposed in ways we've never considered.

New Content in High Definition

High definition source offers new challenges and opportunities for delivery to iTunes. There are no legacy tape formats nor are there analog back-compatibility burdens, but there are many sub-standards and methods of creating large-dimension video that aren't as easy to categorize as NTSC and PAL have been for the last half a century.

The number-one requirement for acceptance to iTunes is frame independence. When an encoding facility de- livers a file to iTunes, no frame is allowed share fields from neighboring frames (often called "ghosting"). If a performance is shot at a given frame rate, all cameras have to be shooting at that frame rate, the edit has to take place at that frame rate, and the resulting file you send to us should be at that frame rate. Shooting at ~24 fps, editing at ~30, and then exporting a file from the timeline with ~24 fps as the frame rate will create a mess of overlaid fields that is irreversibly doomed to merged fields and unpleasant motion. iTunes can accept interlaced or progressive HDTV source as long as a simple de-interlace process will result in frame independence.