Film Sequence Detection and Removal
In
DTV Format and Standards Conversion

by
Scott Ackerman
Product Manager
Teranex, Inc.
http://www.teranex.com

Abstract

One of the more challenging tasks in DTV format and standards conversion is the handling of material that was originated on 24-frame film and then converted to a 30-frame video format. This process is known as 3:2 Pull-down, or more appropriately as 2:3 Pull-down.

This type of material presents many interesting challenges to a format or standards conversion process. Format and standards converters create new output information through a process of interpolation. As material with inconsistent and/or mixed film frame pairing sequences is presented to the interpolation process, the problems are compounded, thereby increasing the number of artifacts present in the output.

This paper will discuss the challenges faced in dealing with this type of material. It will discuss some of the methods used to detect and correct these problems so that they do not become "enhanced" in a conversion process. It will also look at the removal of these sequences as an alternative to a temporal rate conversion, when converting to a 24-frame video format such as 1080p24.

Introduction

A great deal of the programming is originated on film. This material ranges from full-length feature films to prime-time dramas and comedies to commercial spots. While some of this programming maybe captured using 30-frame film, or indeed other frame rates, most is captured on 24-frame film for economic reasons.

In the early days of film production, cameras were hand cranked. Even though many different frame rates were used, 18 frames per second (Fps) was a speed that could be maintained by a person comfortably. When the human arm was replaced by the electric motor, it also ran the camera at 18 Fps. This changed when sound was introduced. To achieve reasonable sound quality, the magnetic audio recorder had to move faster than 18 Fps. A rate of 24 Fps was adopted for the audio recorder and similarly for the film camera and has been in use ever since.

The playback of video based material in the United States was originally derived from the frequency of AC power. Since AC power runs at 60 Hz, video ran at 60 fields per second (fps). This allowed every television set to derive its playback rate from AC power, an easily available reference. When color was introduced, the color information was superimposed over the B/W signal to ensure that the television broadcast could still be viewed on the old B/W sets. To make this work, the vertical frequency had to be shifted from 60 fps to 59.94 fps.

The 2:3 Pulldown Process

In order to understand how 24 Fps film is transferred to 59.94 fps video, lets first look at a simpler task: transferring 24 Fps film to 60 fps video.

When 24 Fps film is transferred to video, redundant fields need to be inserted to pad to 30 Fps (60 fps). In this process the first film frame is used to make field 1 and field 2 of video frame 1. The second film frame is used to make field 1 & 2 of video frame 2 and field 1 of video frame 3. The third film frame then makes field 2 of video frame 3 and field 1 of video frame 4. Finally film frame 4 makes field 1 of video frame 4 and fields 1 & 2 of video frame 5. This whole process then repeats itself throughout the film transfer. The four film frames of this sequence are called A, B, C and D. The corresponding video fields are called A1, A2, B1, B2, B3, C1, C2, D1, D2 and D3. An Example of the 2:3 sequence is shown below in Figure 1.1.

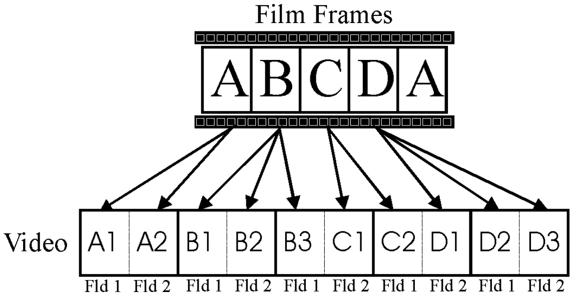

Figure 1.1 – 2:3 Sequence

To transfer film to the actual video rate of 29.97 Fps (59.94 fields per second), the film playback in the telecine is slowed down by 0.1% to 23.976 Fps. All US telecines play 24 Fps film back at 23.976 Fps. 29.97 Fps is 0.1% slower than 30 Fps, and 23.976 Fps is 0.1% slower than 24 Fps. 23.976 Fps will now fit nicely into 29.97 Fps if transferred with the 2:3 pulldown sequence.

It should be noted that this process of 2:3 Pulldown is only necessary for video formats based on 60 Hz. In 50 Hz systems the film is simply transferred from 24 Fps to 25 Fps (50 fields) by running the telecine 4% faster than normal.

Overview of Format and Standards Conversion

Moving images exist in three dimensions. In the horizontal direction they are made up of individual pixels. In the vertical direction they are made up of the lines contained in the field or frame. This is referred to as the spatial domain. Finally there is the number of fields or frames per second, which is referred to as the temporal domain.

The process of format or standards conversion is a form of sample rate conversion in two or three of the above dimensions. It consists of expressing moving images sampled on one three-dimensional sampling lattice to a different lattice.

The process of interpolation is used to convert between these various spaces. Interpolation is defined as computing the value of a sample or samples, which lie off the sampling matrix of the source signal. In other words it is the process of computing the values of output samples that lie between the input samples.

The term format conversion is generally associated with changing the number of pixels and lines in a format. Most format conversions do not involve temporal rate changes. Examples of format conversion include:

480i59.94 to 720p59.94
480i59.94 to 1080i59.94
576i50 to 1080i50

If we look at a format conversion from 480i59.94 to 1080i59.94 we need to change two of the three parameters of the signal. The first is the number of pixels in each. A 480i signal has 720 pixels per line while the 1080i signal has 1920 pixels.

The second parameter that needs to be changed is the number of lines in each field (or frame). The 480i signal has 240 active video lines per field while the 1080i signal has 540 lines, as shown in Figure 2.1.

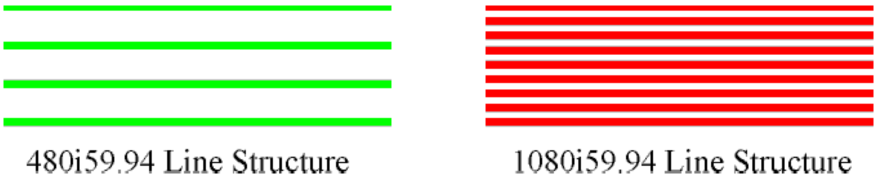

Figure 2.1 – 480i vs. 1080i Line Structure

The third parameter, the number of fields (or frames) per second does not change between the 480i signal and the 1080i signal.

Standards conversions are generally concerned with changing the number of lines and fields (or frames) per seconds in an image. Examples of standards conversions include:

480i59.94 to 576i50
720p59.94 to 1080i50
1080i59.94 to 1080i50

If we look at a standards conversion from 480i59.94 to 576i50 we need to change two of the three parameters of the signal. The first is the number of lines in each field. A 480i signal has 240 lines per field while the 576i signal has 288 lines, as shown in Figure 2.2.

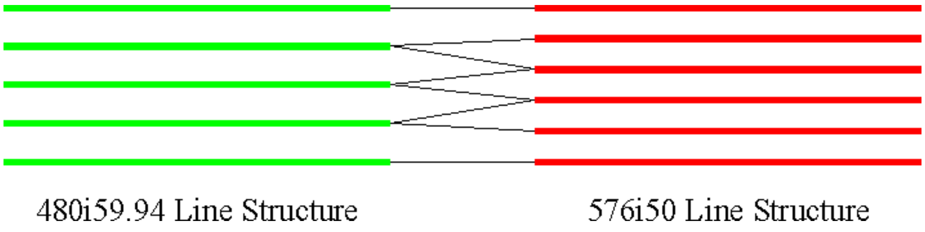

Figure 2.2 – 480i vs. 576i Line Structure

The second parameter, which needs to be changed, is the number of fields per second. The 480i signal has 59.94 fields per second while the 576i signal has 50, as shown in Figure 2.3.

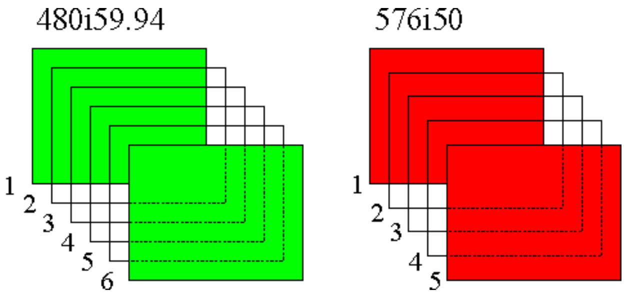

Figure 2.3 – 480i59.94 vs. 576i50 Field Structure

The third parameter, the number of pixels on each active picture line does not change significantly between 480i and 576i.

There are also cases where all three domains must be changed. Examples include:

480i59.94 to 1080i50
1080i59.94 to 576i50
480i59.94 to 1080p23.98-sf

Issues Faced in the Conversion of 2:3 Material

There are two main issues faced in the conversion of film-originated material. The first is motion artifacts due to the low temporal rate of the original film material and the second is the possibility of mixed frames caused by a video frame being made up of two completely different fields.

The first problem of motion artifacts, due to the low temporal rate of the original film material, can be explained in terms of the sampling theorem. If an object is in motion it will be in a different place in each successive field. Interpolating between four fields gives four images of the object in the output field. The position of the dominant image will not move smoothly, so it will be seen to judder. If the field rate is 60Hz as in normal video based material, then by the sampling theorem, the maximum movement frequency allowable in the signal being sampled is 30Hz. We are, however, dealing with film-originated material which means the original sampling rate of the signal is only 24 Hz. Unfortunately, objects move a lot faster than this, so temporal aliasing almost always occurs. When this temporal aliasing is presented to a converter, the converter cannot tell the difference between the original object and the ones created by the aliasing, so it resamples both. These multiple ‘alias’ images are the cause of the perceived judder.

The second issue of mixed frames is caused when a scene change has occurred between two film frames. As mentioned earlier, most standards conversions use a four-field sampling aperture to create the new output information. If a scene change occurs between a B and C frame, as shown in Figure 3.1, or between a C and D frame then the potential exists for a mixed frame to occur in video frames 3 and 4 of the 5-frame sequence.

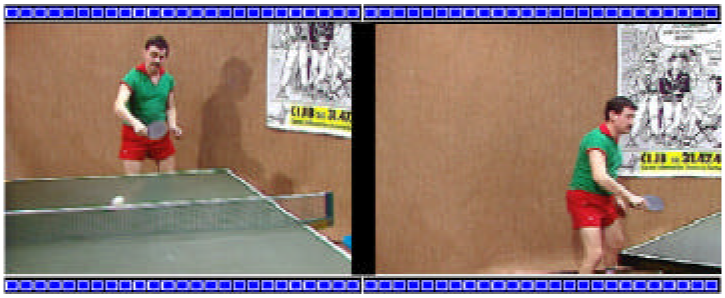

Figure 3.1 – Shows a scene change between film frame B and C.

The result of this mixed frame, shown in Figure 3.2, is a video frame made up of two different film frames as each field would have come from a different film frame, one on each side of the scene change.

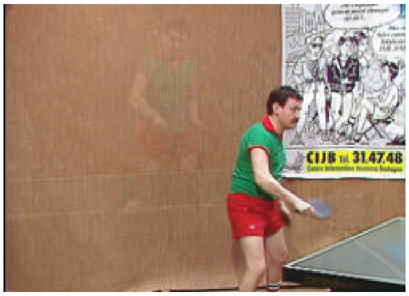

Figure 3.2 - The two fields that make up this video frame each came from a different film frame. Because a scene change occurred between the two film frames a mixed frame was created.

The process of electronic editing can further aggravate these problems. It is quite common, in electronic editing, for little or no attention to be paid to the underlying 2:3 sequence when material is being edited. This creates discontinuities in the 2:3 sequences that compound the problems already created in the film transfer.

2:3 Sequence Removal vs. Temporal rate conversion

One method which can be used to overcome the artifacts created in the conversion of 2:3 pulldown material to 50 Hz is the removal of the original 2:3 sequence. If the sequence could be detected and removed from the film originated material it would then be possible to recover the original 24 film frames. This 24 Fps video would then be spatially converted from 480-lines to 576-lines and then recorded at 24 Fps on specially modified "Slow PAL" VTRs. The 24-frame, 576-line video would then be played back at 25 Fps in a standard PAL VTR. This is similar to the process mentioned earlier where a 24-frame film is converted to 50 Hz video by running the telecine is run 4% fast to arrive at 25 Fps (50 fields). This process eliminates the need for a temporal rate conversion and thus eliminates the motion artifacts and mixed frames that can occur when converting this type of material.

This same approach can also be used when converting film-originated material (with 2:3 pulldown) to 1080p24. Once the 2:3 sequence has been removed from the 480i60 film original, the resultant 480sf24 signal is then spatially converted in the horizontal and vertical domain to 1080p.

Film Sequence Detection

The ability to reliably detect film vs. video based material has a great number of advantages in certain format and standards conversion processes. As discussed previously it can be used to eliminate temporal artifacts in a conversion of film-based material from 59.94Hz to 50Hz. It can also be used in conversions of film-based material to 1080p24. In an up-conversion the ability to detect film based material can help in the de-interlacing process because film originated material came from a progressive format and does not require de-interlacing.

There is no standard method used for 2:3 sequence detection. Each manufacturer must develop their own unique method for differentiating between film and video based material. The foundation for almost all detection schemes is field comparison. When a 2:3 sequence is present there should be 2 identical fields, followed by 3 identical fields, followed by 2 identical fields, etc. Because this detection must occur in real-time, along with the processing of the material, it is difficult to react to changes in the sequence causes by discontinuities. If 2:3 sequences where continuous and uninterrupted the detection process would be a simple one.

There are many methods used in modern post production that can corrupt or change the 2:3 sequence. These processes include electronic editing, insertion of video effects and electronic titles, video fades, film compositing, and animation sequences. Each of these processes, and their effect on the sequence, is outlined below.

Electronic Editing

Electronic editing is the most common cause of discontinuities in the 2:3 sequence. If all edits occurred on an ‘A’ frame there would be no problems. It is not, however uncommon for edits to be done without regard to the underlying 2:3 sequence. Whenever an edit is done on something other than an ‘A’ frame there will be a discontinuity in the 2:3 sequence. These discontinuities make it difficult to detect film sequences.

Vari-speed

Vari-speed is the process of increasing or decreasing the speed at which a film is played back on the telecine. This process is typically used to time compress or expand a film segment to fit in to a particular time slot. For example, a 15 second clip, shot at 24 Fps that needs to be placed in a time slot which is only 12 seconds in duration. In order to achieve this compression the telecine is set to play the clip back at a rate faster than the normal 24 Fps rate. This has the effect of compressing the material to the desired length but also changes the 2:3 sequence to a variable sequence.

Video Effects & Titles and Video Fades

Most video effects, video generated titles, and video fades occur at a rate of 30 Fps (59.94 fps). When these elements are laid over a film background it is done at a field rate as is appropriate for an interlaced video format. As seen previously when film is transferred to video it is done through a repetition of film frames. When the video elements, which change on a field basis are laid over the film originated material they cause a difference to occur in each field. Even though the underlying structure is repeated film frames the fact of having an interlace structure on top makes each field different thus making it difficult to determine the film sequence.

Film Compositing

Film compositing is a method of building complex shots by combining multiple film elements together. This compositing is typically done in the electronic domain. Ideally each of the elements would start on an ‘A’ frame, which is the first film frame in the 2:3 sequence. If, however one element starts on an ‘A’ frame and the other starts on a ‘B’ frame then the resultant composite would contain two different sequences, thus making automatic detection difficult.

Animation

In animated elements, because of the time and cost of rendering high quality images, it is not uncommon for the animation to be done at 12 Fps. To get this 12-frame format to the 30-frame video domain each frame is first doubled to create a 24-frame sequence. Then a 2:3 sequence is inserted to create arrive at 30 Fps. This has the effect of creating a 5:5 sequence as seen in Figure 4.1

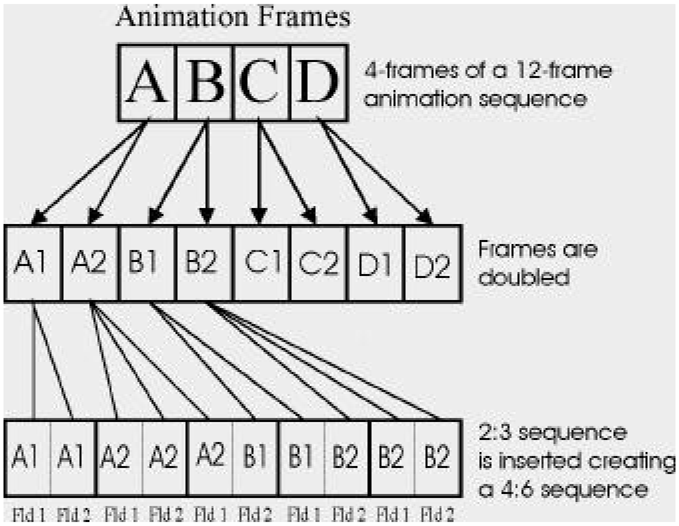

Figure 4.1 – 5:5 Animation Sequence

Slow or Non-Moving Film

While the presence of slow or still film images is not an effect of the post production process it still has an impact on the detection process. If the original film sequence portrays a slow or non-moving scene it become difficult to determine the difference between one film frame and the next. The film will still have a 2:3 sequence, but as nothing is changing from one film frame to the next it becomes difficult to detect the transition from a 2 to a 3 and so on.

Teranex Solution

Teranex has developed an advanced 2:3 sequence detection system. The system looks at an individual field and makes a comparison between it and the previous field. If a match cannot be made the field in question is then compared with the next field to see if a progressive match can be made. Because the system is only concerned with the comparison of two fields at a time as opposed to looking for a complete 2-field, 3-field sequence, it is not susceptible to discontinuities in a longer sequence. The matching of the fields is achieved through the use of proprietary algorithms and a unique parallel processing environment which makes these comparison very effective.

The parallel processing environment, which assigns a processor to each pixel in the image, enables whole fields to be compared in an extremely short period of time. This then allows for the detection of film vs. video-based material and the application of the appropriate filter to occur in real-time.

Conclusion

There are a number of processes used in modern post production environments that can cause discontinuities in the 2:3 sequence. Each of these discontinuities cause the detection system to have to readjust to the new sequence. If the interruption is brief, a field or two in duration, the problem can often be overcome with minimal effort. When the interruptions are longer in duration the problem of recovery and reestablishment of the sequence becomes more difficult.

The benefits associated with the correct detection of film vs. video based material are great. Fortunately the Teranex parallel processing platform with advanced 2:3 sequence detection and handling can be used to detect these sequences more reliably, in real-time, than any other system; thus maximizing picture quality and allowing users to take advantage of the benefits discussed.