Why does cadence matter?

Introduction

Video cadence remains one of the most confusing terms within multimedia systems. Whereas certain terms such as 2:3 pull-down are well understood by people familiar with film production, it's easy to be tripped up by unexpected events such as receiving content with inverted field order or discovering wrongly paired interlaced fields due to an error with upstream editing.

In Part 1 of this InSync Technology White Paper, we will establish the terminology associated with video cadence and we'll point out some of the areas in which special care is needed in video processing.

In Part 2 of the paper, we'll go on to look at what happens in practice and how your devices (mobile screens, TV displays and set-top-boxes) handle cadence.

What is cadence?

Film and TV video systems create the illusion of motion by presenting the viewer with a sequence of images, each captured at regular intervals in time (referred to as the frame rate). Cadence is often used to describe a situation where content with a lower frame rate is packed into a video stream transport with a higher frame rate. It's also used to describe the number of repetitions of a field or frame in a sequence.

Frame repetition has been used since the earliest days of cinema where repeated opening and closing of the shutter was used to reduce flickering. When movies were first transferred for TV viewing in the US, the need to change from 24 fps to 59.94 fps was solved using a scanning pattern of 2 fields then 3 fields, which is termed “2:3 cadence”.

Most of the time we don't realise that there is a cadence associated with the video we're watching. For example, US TV audiences have become used to watching 2:3 material, whilst European audiences are frequently offered 2:2 content, in this case a 24 frames per second film is broadcast as interlaced video at 50 fields per second. The solution in this case is to run the movie 4% faster than shown in theatres, mapping each film frame to 2 video fields.

What cadences are used?

1:1

Interlaced video is a sequence of "odd" then "even" fields, where two fields comprise a video frame. For example in 625 50i systems:

- odd lines of the frame are scanned in field 1 at a time T

- even lines of the frame are scanned in field 2 at a time T + field interval

Remaining common in broadcast systems, the cadence of interlaced video is referred to as 1:1. This denotes that the image is refreshed every field i.e. that there are no repeats.

2:3

As explained previously, in 59.94Hz systems where material is transferred from 24Hz film, a cadence pattern of 2:3 is used to avoid the need for interpolation when showing films on domestic TV sets. Figure 1 shows the scanning pattern when a 24Hz film is converted for 59.94Hz viewing.

For each film frame (either scanned or from telecine), Si, a series of intermediate interlaced fields are created, fij, where 2 or 3 interlaced fields are generated from each source frame Si in a consistent, alternating 2:3 pattern. For example, source frame 'S_0' in Figure 1 is used to generate two intermediate interlaced fields ('f_{01}' and 'f_{02}'), whereby source frame 'S_0' uses a field 1 sample grid to generate the intermediate field 'f_{01}', and the same source frame 'S_0' is then sampled again using a field 2 sample grid to generate the intermediate field 'f_{02}'.

The second source frame 'S_1' is used to generate three intermediate interlaced fields ('f_{11}', 'f_{12}' and 'f_{11}'), whereby image content 'S_1' is sampled using the field 1 sample grid, generating intermediate field 'f_{11}', then 'S_1' is sampled again using the field 2 sample grid, to create 'f_{12}', and 'S_1' is again sampled using the field 1 sample grid, generating intermediate field 'f_{11}'.

The output frame sequence is then created using the intermediate fields. Irrespective of how many fields are generated from a given source frame Si, the sampled intermediate fields fi must always alternate between a field 1 and a field 2 as per normal interlaced video. Interlaced video is recorded and broadcast as a whole number of interleaved frames comprising a field 1 and a subsequent field 2. For example, as illustrated in Figure 1, output fields 'R_{11}' and 'R_{12}' comprise one 30Hz output frame.

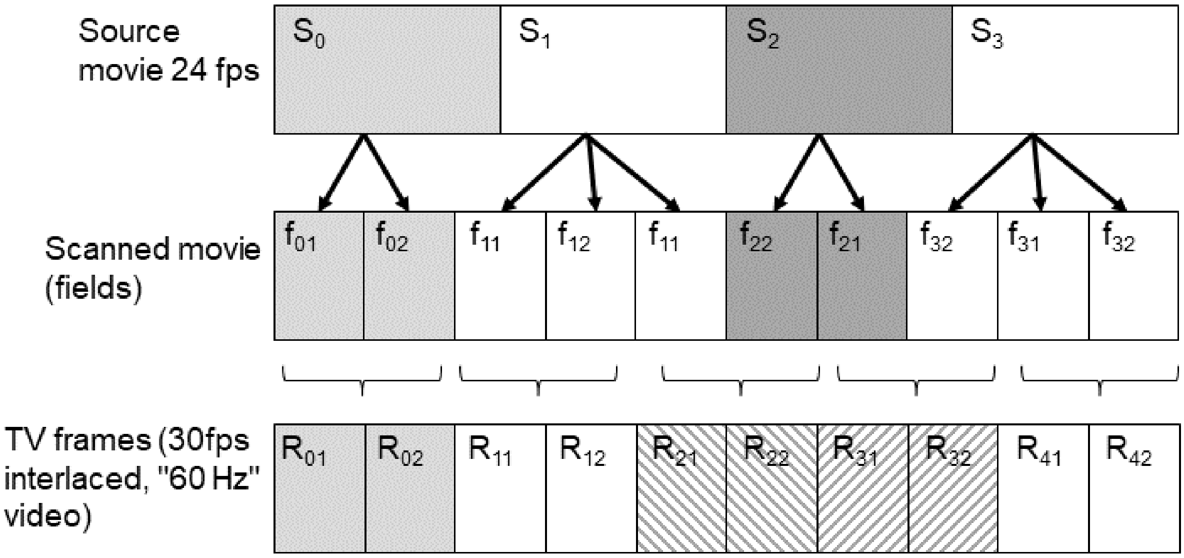

Figure 1: illustration of 2:3 cadence

2:2

A cadence pattern commonly used in TV productions is the progressive segmented frame (psf) format, which is often referred to as 2:2. Unlike true progressive format video, where each video frame represents the entire image at the time interval, a psf frame comprises two fields ("segments"), each of which contains half the lines in the frame, scanned alternate odd and even lines. The two fields are transmitted separated in time by half a frame interval, so that they can be carried within a transport designed for interlaced video. The receiving device must be told (e.g. via metadata) that the content is psf so that the fields can be stored and re-assembled correctly. Since there is no motion between the fields in a psf frame, misinterpretation of a psf source as interlaced fields or as true progressive video would result in severe motion artifacts. Live sources can use SMPTE ST352 payload identification, and video files can use i/p/psf metadata, to enable downstream devices and displays to correctly identify psf material.

2:3:3:2 and 2:2:2:4

Examining Figure 1 above, it is clear that certain output frames contain two fields which originate from different source frames, for example the frame containing 'R_{21}' and 'R_{22}'. These are sometimes referred to as mixed frames.

If an edit system or compression encoder recognises that the video is interlaced 2:3, then this presents no problem. Issues arise in situations where an edit is made on a frame boundary between mixed frames, for example, between the frame comprising 'R_{21}/R_{22}' and 'R_{31}/R_{32}'. The resulting sequence now has a discontinuous cadence 2:2:1:3:2:3, where the "1" field is referred to as an "orphan" field.

If this sequence is subsequently frame rate converted to 50Hz, or encoded via a compression encoder which pairs adjacent fields, there is a danger that incorrect pairing occurs, which will lead to visible degradation e.g. unpleasant content blending. Therefore downstream processing such as frame rate conversion (standards conversion) must include sophisticated cadence detection tools to ensure that such anomalies are identified, and the correct processing is applied.

If we think about the reasoning behind 2:3 cadence, we can see quite quickly that other cadence patterns could achieve the same objective. Essentially, we are trying to present 24Hz content at 59.94 (or 60) Hz, so for each input frame, we need "2.5" output frames. Presenting each input frame twice gives us 48Hz, and three times would be 72Hz, so the 2:3 pattern provides the required 60Hz frame rate (which can then be slowed down to 59.94Hz for TV transmission without any visible degradation).

The 2:3:3:2 pattern provides the same frame ratio as 2:3 but using the 2:3:3:2 pattern avoids one of the mixed frames, as illustrated in Figure 2.

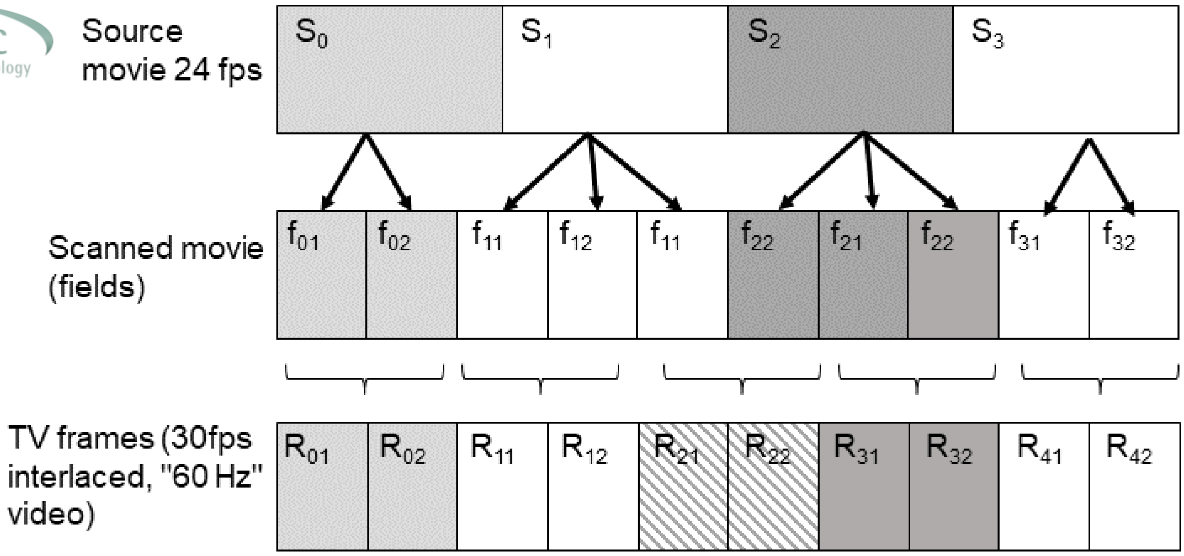

Figure 2: example of 2:3:3:2 cadence

An equally timed result could be achieved with a 2:2:2:4 pattern, which avoids all mixed frames, as shown in Figure 3.

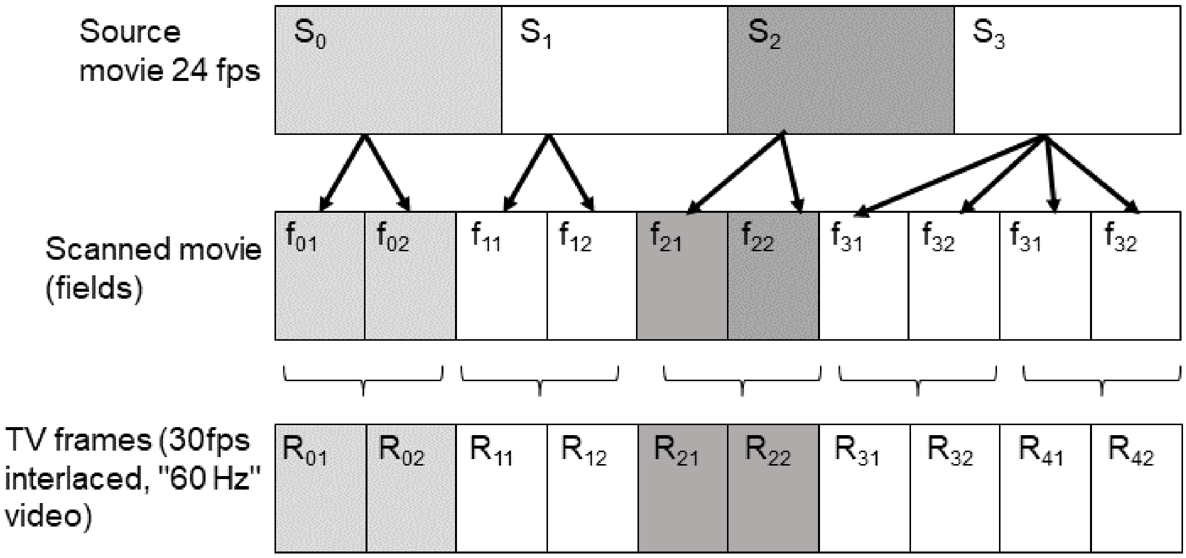

Figure 3: example of 2:2:2:4 cadence

2:3:3:2 and 2:2:2:4 cadence patterns reduce or avoid blending content from different temporal intervals, thereby making life easier for video editors, and have the advantage that if a user watching a streamed sequence pauses the video, they will see a clean frame. However, any non-unform field repetition will produce visible motion discontinuities, sometimes referred to as "cadence judder", and both 2:3:3:2 and 2:2:2:4 patterns will exhibit this problem.

5:5, 6:4 and 8:7 - "animation cadences"

Animations are created from artists' drawings and computer graphics which are then scanned and composed into the movie or programme. Because of the time required for such a process, images are often repeated in longer patterns in a 24Hz or 59.94Hz sequence. Therefore cadences of 5:5, 6:4 and 8:7 are commonly found.

The unusual motion effects created by such cadences are part of the artistic intent when animated sequences are produced. Storytelling using animated characters is a specialist craft which deliberately uses non-natural effects to enhance the viewer's immersion in the story.

Mixed cadence content

Not all content uses a consistent cadence. Where a programme is composed from multiple sources, different cadences may be encountered over the length of the sequence. For example, a studio segment shot at 59.94Hz (interlaced, i.e. 1:1) may be followed by an clip transferred from a movie, which may have had 2:3 insertion (see above).

It's also possible to find mixed cadence within a video frame. For example, a programme may be mastered at 25p (for example movie content at 24Hz slightly sped up to 25Hz), then graphics from a video graphics generator at 50Hz interlaced may then be overlaid.

It is therefore essential for any downstream processing to monitor the incoming video and identify the incoming cadence. In the case of mixed cadence within a frame, the user must decide which cadence should be considered dominant when processing such as frame rate conversion is applied.

What else do we need to consider?

Cadence phase

In the case of 2:3 or 2:3:3:2 cadence patterns, an edit system must be able to identify where mixed frames are located to avoid the creation of ‘orphan fields’, as mentioned previously. This knowledge is also important in frame rate converters where fields from different temporal positions will be used to construct new frames in the output sequence.

To understand cadence phase, we need to revise Figure 1, which is actually an over-simplified illustration. In Figure 4, we see a situation where the 2:3 pull-down starts on the first frame of the source sequence. We would refer to the scanned field sequence as

AABBBCCDDD

and thus the 5 output frames derived from the 4 source frames would be denoted

AA

BB

BC

CD

DD

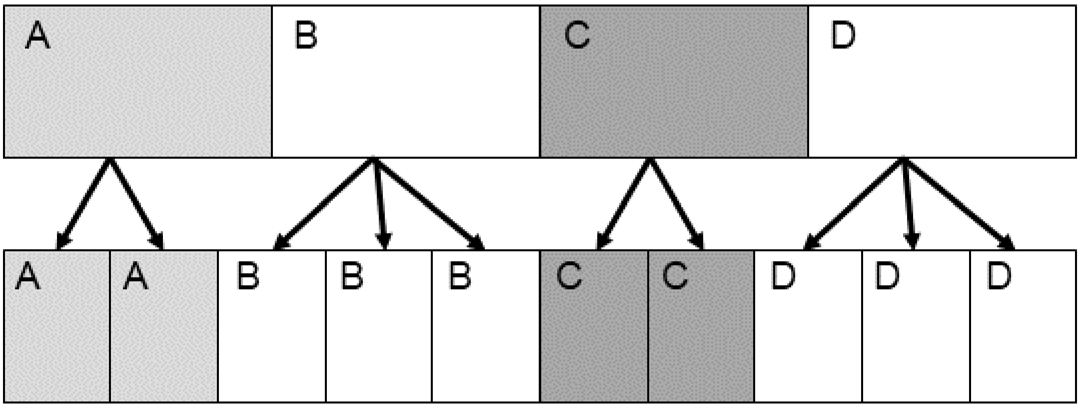

Figure 4: illustration of 2:3 phase

If the sampling had started on the 3 field scan, the sequence would have been
BBBCCDDDAA
and so the 5 output frames derived from the 4 source frames would be denoted
BB
BC
CD
DD
AA
Figure 5 shows the situation for a 2:3:3:2 cadence pattern, where the output frames would be denoted
AA
BB
BC
CC
DD

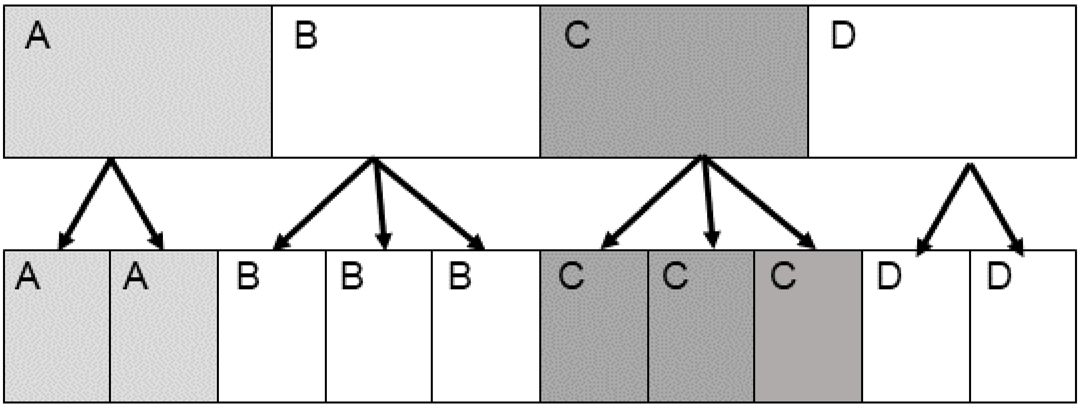

Figure 5: illustration of 2:3:3:2 phase

Given the different locations of the mixed frame in the above illustrations, it is clearly important for downstream processing to correctly understand the starting phase of the sequence.

Field order

Phase (or field order) is similarly important in psf (progressive segmented frame) video sequences. As explained previously, a psf frame comprises two fields ("segments"), each of which contains half the lines in the frame, scanned alternately odd and even lines. Scanning the odd lines creates a field denoted ‘F1’, and scanning the even lines creates a field denoted ‘F2’.

‘F1’ and ‘F2’ are transmitted separated in time by half a frame interval, so that they can be carried within a transport designed for interlaced video. If ‘F1’ is transmitted in the first half time interval, and ‘F2’ follows in the second half of the frame interval, we refer to ‘F1’ as the dominant field, and denote this as ‘F1/F2’ field order. If the ‘F2’ field is transmitted first, we would say the sequence had an ‘F2/F1’ field order.

Downstream processing must be informed of the field order e.g. via file metadata, as reconstructing a frame correctly depends entirely on knowing which field was derived from which source lines.

Field order is also important in PAL and NTSC standard definition stored assets. Interlaced PAL sequences are usually transmitted "top field first". If we denote the frame interval as ‘T’, a field representing the scanned odd numbered lines would be transmitted starting at time 0, and a field representing the even numbered lines scanned at time T/2 would be transmitted starting at time T/2. It's important to remember that interlaced video sends two fields per frame, where the fields are derived at different time intervals, spaced by T/2.

In NTSC, the transmission order is usually "bottom field first" i.e. the field containing the even numbered lines would be sent starting at time 0, with the field containing the odd numbered lines transmitted starting at time T/2.

As in the psf case, downstream processing must know the field order to be able to correctly present the frame on a display or carry out field-based editing. Given the time interval between fields in interlaced content, presentation of video with incorrect field order results in highly distorted pictures where moving objects alternately move forwards and backwards.

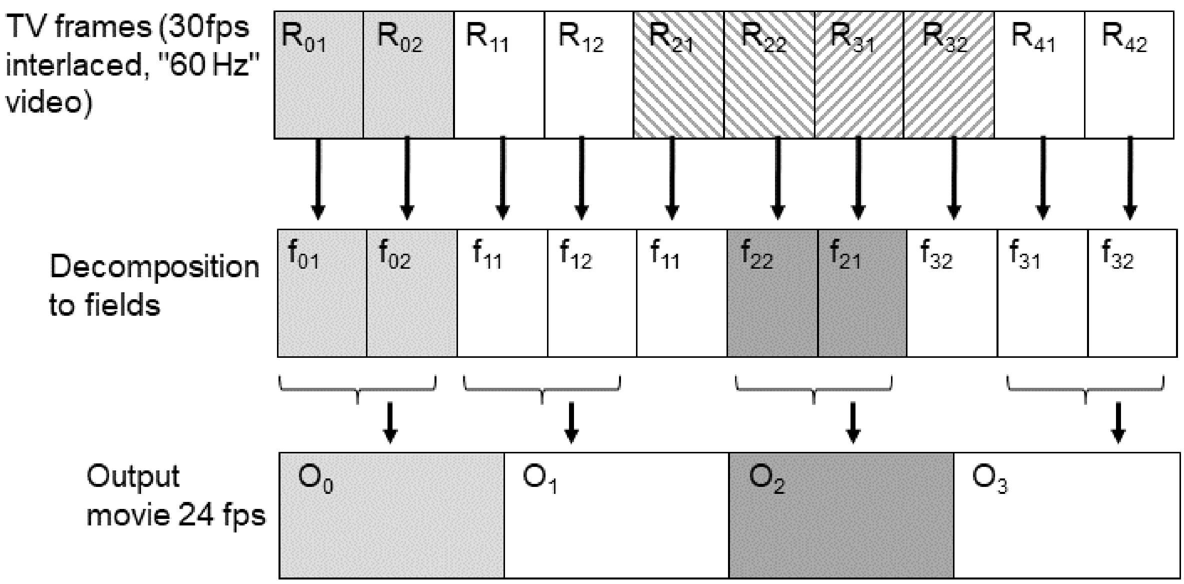

Figure 6: Illustration of inverse telecine

Reverse telecine

The phrase "reverse telecine" or "inverse telecine" is used to describe the process of removal of the third frame in 2:3 video, to revert a movie converted to 60Hz back to 24Hz. This is illustrated in Figure 6.

As Figure 6 shows, the additional third frame is removed and the remaining interlaced fields are paired to recreate the 24Hz sequence. Referring back to the description of phase above, it becomes clear that information about the cadence phase is essential for correct field pairing.

Figure 7 illustrates the case of correct and incorrect inverse telecine. In the top part of Figure 7, we see the correct removal of the third field combined with correct field pairing, thus restoring the original movie frames.

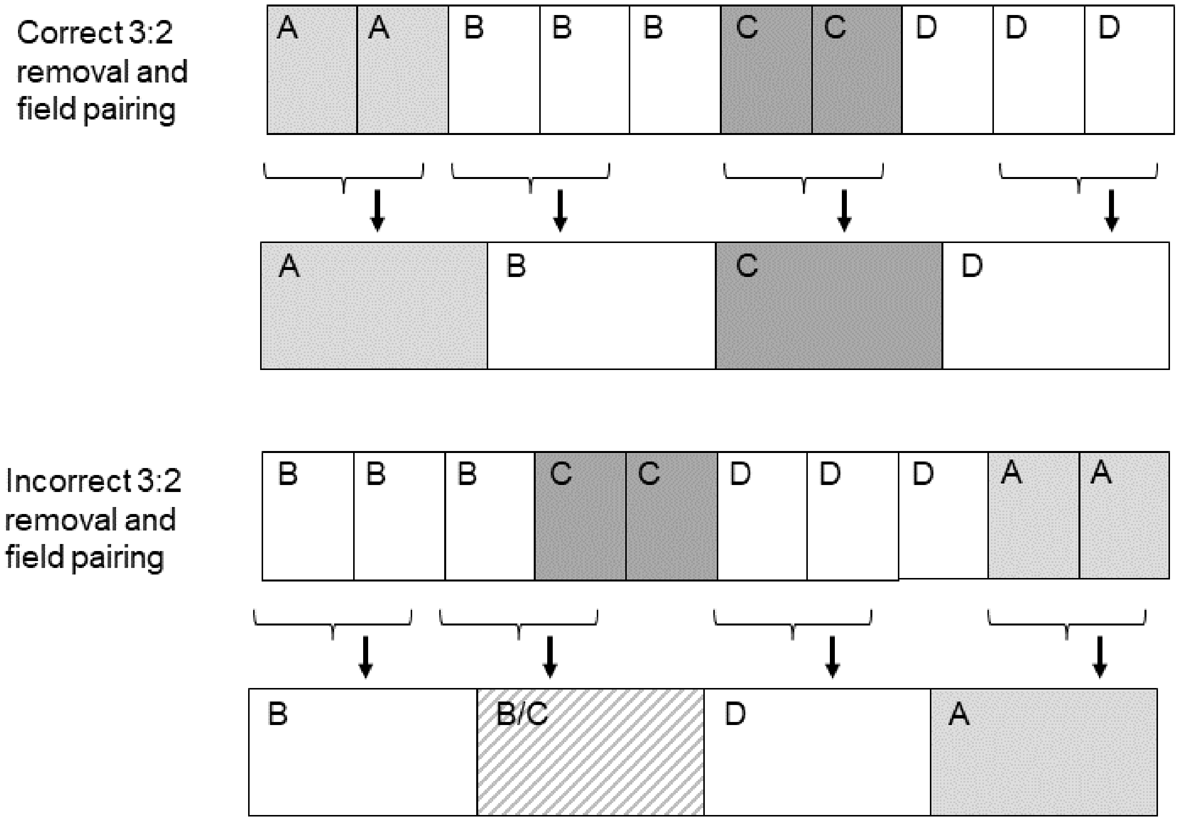

Figure 7: Correct and incorrect reverse telecine

In the lower half of Figure 7, we imagine a situation where incorrect metadata were available, thus leading to mislabelling of the cadence phase. The result is a blended frame where information from two different points in time are combined. This will become obvious on viewing as a visible picture distortion.

Note that the problem illustrated in the lower part of Figure 7 can also arise from incorrect editing. In the event of a movie being converted to 60Hz with 2:3 pull-down, a frame or series of frames being removed during editing and the metadata not updated; then the sequence could start BBB instead of AA, with the result as shown in Figure 7.

Frame rate conversion

When content is processed for international distribution, frame rate conversion is necessary. For example, a US comedy show watched at home by European audiences would need conversion from 59.94Hz to 50Hz. If the US comedy show had been produced at 23.98Hz, then typically 2:3 insertion would be needed to create the 59.94hz version shown to domestic audiences.

Ideally, the 23.98Hz master would be made available to the European network wishing to air the show, and a 50Hz version could be made via a slight speed-up to 25Hz, with a 25psf version being shown on European 50Hz interlaced networks.

However, if only the 59.94Hz 2:3 version is available to the European broadcaster, then care is needed in frame rate conversion. Processing must first apply inverse telecine to revert back to the original fields before frame rate conversion is applied. If this is carried out with an incorrect phase, as shown in the lower half of Figure 7, then blended frames will result, leading to a very poor quality output.

Alternatively, incorrect definition of the 59.94Hz content as interlaced video (i.e. 1:1) would lead to an extremely distorted result, as illustrated in Figure 8. As can be seen in Figure 8, content from frames 1 and 2 (separated in time by the frame interval) become combined in an erroneous frame. When this is eventually frame rate converted to 50Hz, the distortion is propagated and it is not possible to remove it.

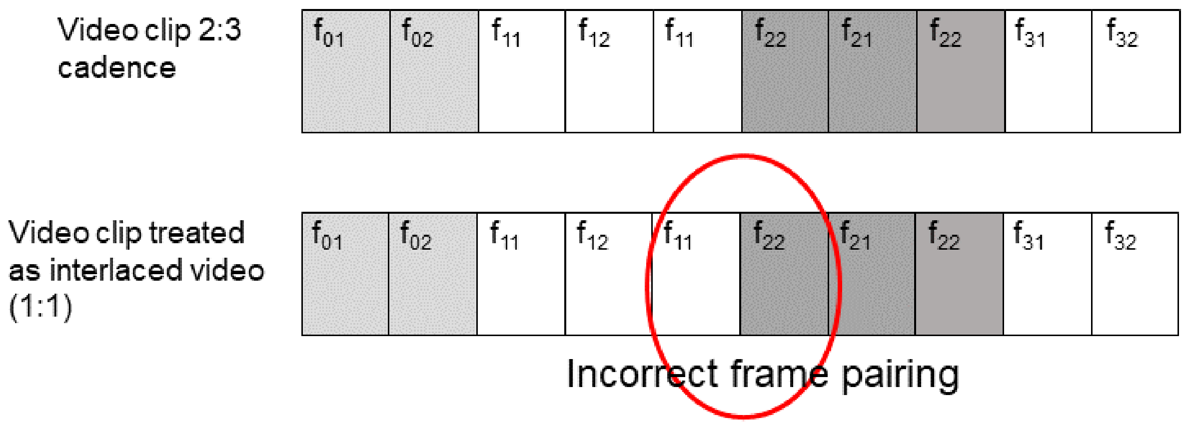

Figure 8: Treating 2:3 material as 1:1 interlaced video

Furthermore, we must remember that a 60Hz 2:3 movie has been derived from a 24Hz source, so that fields ‘f_{01}’ and ‘f_{02}’ actually relate to the same temporal position. If the content is treated as 59.94Hz 1:1 interlaced video, the converter would assume that ‘f_{01}’ and ‘f_{02}’ contain data which are spaced in time by one field interval.

If a video frame rate conversion process has been carried out without an understanding of the correct cadence and its phase, it will not be possible to recover the original content, and therefore quality is irrecoverably lost. Thus, it is essential to choose a frame rate converter which is sensitive to cadence and/or which allows the user to select incoming and outgoing cadence patterns.

High quality frame rate converters, such as FrameFormer by InSync Technology, will monitor incoming content to make an assessment of the cadence, and apply the appropriate processing to avoid the problems outlined above. Tools are also provided for the user to apply in the case of erroneous or missing metadata, or for situations of mixed cadence where the user wishes to achieve a specific result.

Setting parameters for outgoing cadence are particularly important where there is a specific downstream process which the output of the frame rate converter will feed.

Conclusions

Video content available to mobile and home viewers may have been produced with one of many cadence patterns. Any processing within the distribution chain must recognise and correctly deal with the cadence.

In Part 2 of this white paper, we will examine typical distribution systems and will explain what happens when cadence isn't properly considered.

About us

FIND OUT MORE

InSync Technology is a successful and growing employee-owned business. Since our inception in 2003, InSync has specialised in development of highly efficient signal processing and standards conversion products, with a focus on motion compensated frame rate conversion

Find us on:

www.insync.tv