InSync White Paper : Achieving optimal conversions in UHDTV workflows
April 2015

Abstract - *UHDTV 120Hz workflows require careful management of content at existing formats and frame rates, into and out of the workflow. Standards converters are needed to accommodate the huge range of current and future production and distribution standards. UHDTV poses two significant new issues. Presentation of content on large, high resolution displays, requires absolutely perfect picture quality. The standards converter must not introduce visible artifacts, even for multiple conversions which may arise in the workflow. Secondly, there are severe technical challenges with low frame rate input, such as 1080 23.98p, commonly used for film production. Not only is there a 4x conversion in the spatial dimension, there is also a 5x increase in frame rate, and content owners need to preserve specific temporal effects when converting film rate material to 120Hz. In this paper, we present some typical UHDTV workflow scenarios, describe some of the issues in conversion, and present solutions which significantly improve the state of the art in frame rate conversion between HD and UHDTV material.*

Introduction

High frame rates, such as 120Hz, in UHDTV systems enable enhanced motion rendition, which is critical to picture quality at higher resolutions. Since there is a wealth of existing material created at current production standards e.g. HD 50Hz/59.94Hz and UHDTV 23.98Hz, future UHDTV 120Hz workflows will require careful management of content into and out of the workflow. Standards converters will continue to be needed to accommodate the huge range of current and future production and distribution standards.

UHDTV poses two significant new issues. Firstly, as consumers will be watching programs on large, expensive, high resolution displays, picture quality must be absolutely perfect. Sensitivity to picture degradations will be high, so the standards converter must not introduce visible artifacts, and this must hold true even for multiple conversions which may arise in the workflow.

Secondly, frame rate upconversion to UHDTV at 120Hz poses a severe technical challenge to conventional converters when dealing with low frame rate input, such as 1080 23.98p material, commonly used for film production. Not only is there a 4x conversion in the spatial dimension, there is also a 5x increase in frame rate. Furthermore, content owners may wish to preserve specific temporal effects when converting film rate material to 120Hz.

In this paper, we present some typical future UHDTV workflow scenarios, describe some of the issues in conversion into/out of the workflow, and present some solutions which significantly improve the state of the art in frame rate conversion between HD and UHDTV material.

UHDTV workflows

Future UHDTV 120Hz workflows will need to incorporate multiple spatial and temporal conversion processes to ensure that all available content can be correctly handled. An example future workflow is shown in Figure 1, where different conversion processes are shown for each type of source content. We will consider each source separately for the purposes of the following discussion, although in practice, a good quality standards converter would accommodate each source type automatically.
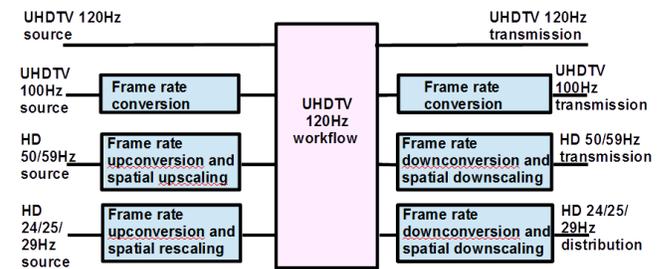
Figure 1 : Future UHDTV 120Hz workflow

In the workflow illustrated in Figure 1, content from existing sources, such as broadcast assets, scanned movies etc will be frame rate and spatially format converted before being passed to the main editing and effects processes. Note that we have omitted SD from this diagram. This is because it is unlikely that SD material upconverted to UHDTV would be commercially acceptable to viewers, but there is no technical reason why an SD source should not be included.

Since users will transition to UHDTV production over a period of time, due to the high capital investment needed for new production and transmission equipment, many organisations will continue to operate with a 59Hz or 50Hz workflow. Integration of new high frame rate UHDTV cameras into today's workflows also requires conversion interfaces, as illustrated in Figure 2. It may be seen that the same types of conversion are required as illustrated in Figure 1, to get content into and out of the workflow.
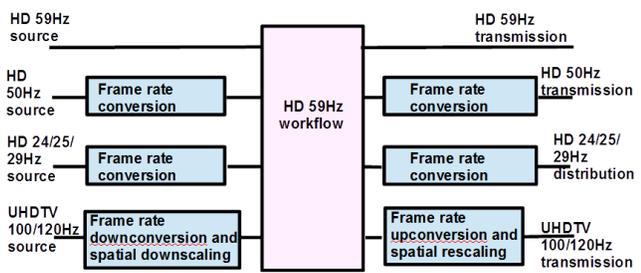
Figure 2 : Interim hybrid HD/UHDTV workflow

Conversion issues : deinterlacing

As explained in [1], a typical HD to UHDTV conversion system, includes steps of

- deinterlacing
- temporal rate conversion
- rescaling

As deinterlacing is the first step in the conversion, it is essential that this is of the highest possible quality. The deinterlacer must preserve the maximum possible picture resolution and introduce zero (or in practice a minimal level of) artifacts. Since the rescaling stage to UHDTV will effectively magnify any tiny picture defects, introduction of undesirable effects at the deinterlacing stage could have a huge impact on visual quality of the UHDTV output.

In recent years, deinterlacing technology has been improved through application of a number of techniques, mainly focusing on non-linear, adaptive methods. More computationally intensive methods such as full motion compensation can be applied to the problem, but it is not always possible to use such methods affordably and robustly.

The problem of deinterlacing is further complicated by the diversity of content that will be encountered, ranging from noisy/grainy, unsteady legacy content, that may exhibit any number of encoding/decoding artifacts, through all manner of film cadences to the sharpest, short shuttered synthetic computer generated imagery.

Typical problems with deinterlacers include over-filtering, leading to a very soft output, ringing, and poor static area/moving area adaption. Figures 3 to 5 show some typical examples of deinterlacing artifacts which include jagged diagonals, loss of picture detail and edge artifacts (ringing). Upconversion of these errors leads to extremely visible picture defects.

Historically, deinterlacers also had problems when processing film-originated content originally sourced at 23.98Hz contained in a 59.94Hz 2:3 cadence sequence. In such content, incorrect cadence detection can actually lead to generation of more picture defects than if the processing ignored the source cadence. Modern deinterlacers are generally cadence-aware so such problems are rare, but they are occasionally encountered in less sophisticated converters which do not have modes to manage mixed cadence material e.g. where 2:3 material is inserted into part of a video sequence, or where video captions and graphics are overlaid on 2:3 content.

Figure 3 : Deinterlacing artifact : jagged diagonals

One issue which is difficult to show in a printed document is poor stationary/moving adaptation. A good deinterlacer will apply different processing in moving and stationary areas in order to obtain the maximum possible output resolution.

Figure 4 : Comparison of two deinterlacers (a) Preserving resolution (left) (b) Loss of picture detail (right)

Figure 5 : Deinterlacing artifacts : ringing visible on edges

Problems can also arise if the deinterlacer fails to identify a picture area correctly e.g. a very slowly moving area is identified as static, or a stationary area is

identified as moving because it is close to an area in motion. Weak analysis can also generate errors in periodic structures (such as posts in a fence or stripes in wallpaper) where a static area is classified as moving.

In general, deinterlacing errors are often not visible in HD material when presented on today's domestic, flat-screen monitors. However, when rescaled to UHDTV resolutions, any small defect becomes immediately more visible. An example is shown in Figure 6 where poor adaptation has led to distortion of the edges of the captions in an HD sequence. When upconverted to UHDTV, the distortions are very visible, as shown in Figure 7.

Figure 6 : Deinterlacing artifact: incorrect moving/static adaptation

Figure 7 : Magnification of Figure 6

Minimising the degree to which any one of these numerous potential artifacts degrade the upconverted UHDTV picture is therefore reliant on the use of the highest possible quality deinterlacing. A deinterlacer meeting this requirement must be robust and repeatable in its identification of still versus moving content such that optimum tuned filters may be used to preserve maximum possible detail.

Typical methods for detecting stationary areas or image similarity use field to field (or frame to frame) absolute differences, with simple noise coring. This method has a number of weaknesses which include errors in areas of high vertical frequency due to aliasing, over enhanced sources causing resolution variation which is falsely interpreted as motion or where there are illumination changes between subsequent images. See, for example [2].

More complex methods exist, for example [3] in which dynamic models of the scene are built up over a large number of frames. However, these types of method have the significant drawback of requiring long processing times which are not applicable to live video applications, and generally have an assumption of a static camera, which is not the case in typical video sequences.

Ideally, the static area detection method should be efficient, robust to noise and aliasing, and independent of lighting and other typical scene effects. A high quality solution will contain specific algorithm refinements to directly mitigate away the substantial false negatives that would otherwise be obtained when attempting to identify still areas in the presence of interlace alias. A robust solution should be equally capable of rejecting the frequent false positives obtained when attempting to identify still areas in the presence of spatially repetitive moving image elements.

Other filtering and interpolation artifacts such as ringing may be avoided, at no loss of resolution and often a higher *perceived* resolution, through the use of non-linear analysis and filtering techniques leading to significant advances over linear filtering methods employed previously or in lower quality solutions.
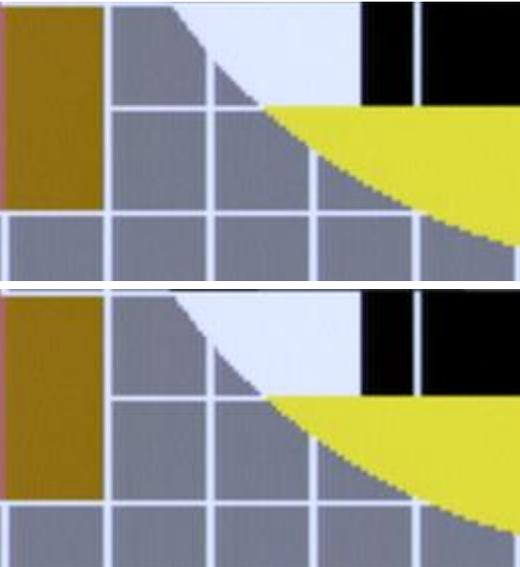
Figure 8 : Two frames of a test card using a simple deinterlacer

3
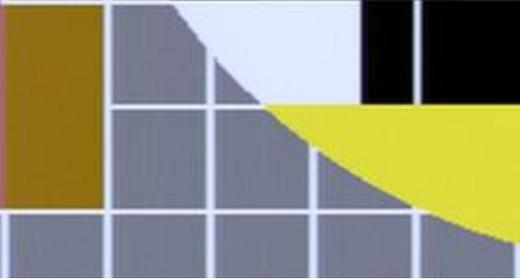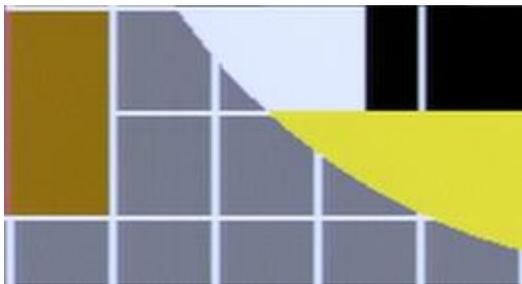
Figure 9 : Two frames of a test card using an improved deinterlacer

In extremis, the interlace sampling format has complex and subtle implications. Techniques described thus far aim to recover and exploit the maximum possible information content from the interlaced sampled signal. Unfortunately, due to the fundamental nature of the sampling method, there are occasions where aliases in the source cannot be recovered. This occurs where image content moves vertically through the sampling structure such that some image detail persistently evades sampling.

Under these circumstances one might imagine that no good solution is possible. In earlier generations of technology with limited resources and the option of only linear filtering, this was true. Non-linear techniques, previously considered too complex to be commercially viable, may now be used to obtain often perceptually flawless results in regions where image edges and textures can be interpolated with the anticipated qualities of directionality and continuity.

Figure 8 shows the output of a simple deinterlacing process on two consecutive frames of a standard test card. Note the jagged diagonals and reduced resolution of the output. Also by comparing the two frames, it can be seen that the tops of the horizontal lines are different. On a video monitor, this would be visible as line flicker or line twitter, and is quite visually distracting for viewers.

Figure 9 shows the same two frames as Figure 8 but with the techniques discussed above implemented. Notice that the horizontal lines are consistent, the diagonals are smooth, and the overall image is much sharper.

Conversion issues : temporal rate conversion

Frame rate conversion requires creation of new picture content in temporal locations which do not necessarily exist in the source. Conversion of material from a lower frame rate such as 23.98Hz or 50Hz to a higher frame rate such as 100Hz or 120Hz requires interpolation of new frames based on the available source frames. Clearly the lower the source frame rate, the larger the nominal amount of motion between frames, and the harder it becomes to accurately predict the content in intermediate frames.

Conversion of low frame rate material is a problem already known to broadcasters and content owners, who have to convert movie content to TV frame rates for national and international distribution. However, the problem is compounded when frame rate upconverting to 100Hz and 120Hz as further intermediate frames need to be made, and therefore visibility of any conversion anomalies will be greater.

In addition, material sourced at UHDTV will tend to have a greater number of objects and much more complex motion than HD, simply due to the increased production freedom associated with UHDTV. Conversion of such content therefore requires management of many more object models, and the probability of occlusions and revealed areas is significantly higher than when converting HD material. Conventional standards converters limited to an HD perspective will have difficulties when attempting to scale to UHDTV. In particular, precision of the conversion at any level of detail would need to be higher for the quality expectations for a UHDTV source.

In Figure 10 two consecutive frames from a conversion from 48Hz to 60Hz are shown. In Figure 11, the same source has been converted to 120Hz, so additional interpolated frames are required. Since the source contains moderately moving objects in good definition, interpolation of the additional frames poses only a small degree of difficulty for the standards converter.

Figure 10 : Detail from two consecutive frames of conversion 48Hz to 60Hz
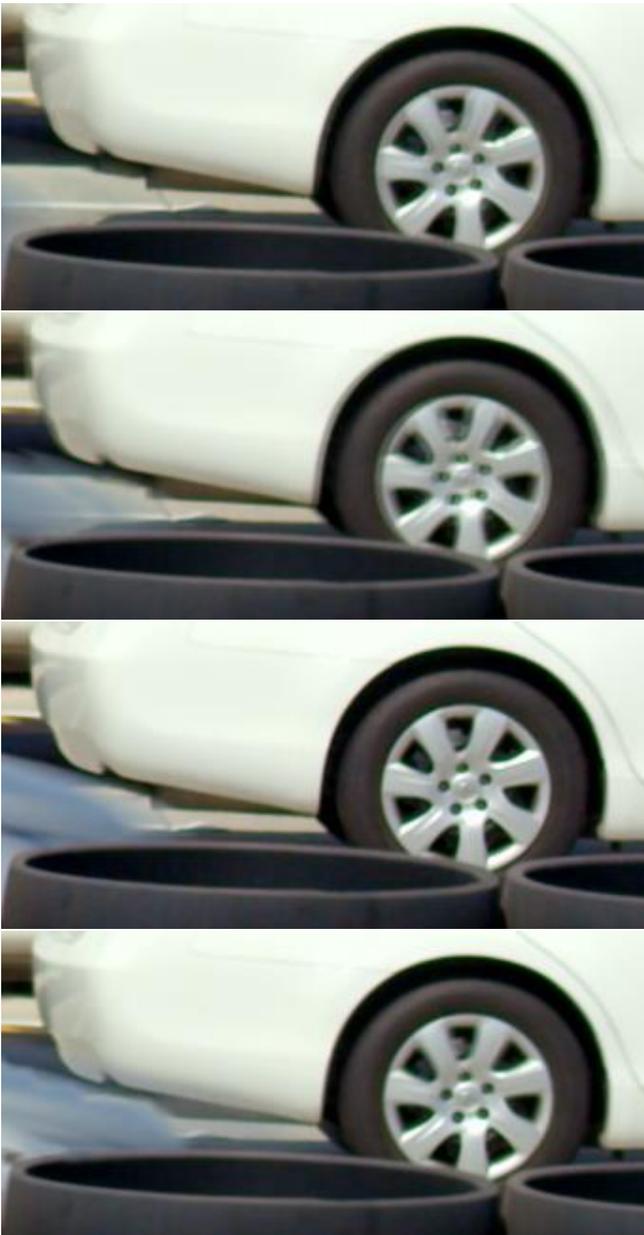
**Figure 11 : Detail from four consecutive frames
of conversion 48Hz to 120Hz**

upconverting to 120Hz, more frames need to be interpolated, so there are more chances of possible visible artifacts and interpolation errors. Any tiny defect in the standards converter will also be apparent in the additional frames at the higher frame rate as Figure 14 shows.

Figure 12 : Source frame at 24Hz

**Figure 13 : Detail from two consecutive frames
of conversion 24Hz to 60Hz**

However, Figures 13 and 14, showing consecutive frames from a 24Hz to 120Hz conversion, illustrate the additional difficulty of frame rate upconverting to a higher frame rate where the source (Figure 12) contains fast movement. There are two important considerations when converting film rate content with fast movement. Firstly, longer shutter duration results in a high degree of "film blur" along the path of motion which increases the difficulty of deriving accurate motion vectors, and secondly, objects will have relatively more spatial displacement from frame to frame requiring much larger search areas for valid vectors. When frame rate

may be encountered when using UHDTV 120Hz material in today's workflows, or when managing future UHDTV 120Hz workflows, sophisticated motion analysis is needed.

Various motion analysis techniques, e.g. [4], have been used successfully in standards conversion for many years, and have been successively refined by various manufacturers. However, certain types of content tend to present challenges to commonly used motion analysis methods, leading to inaccurate and sometimes erroneous motion vectors, which then cause visible picture artifacts.

Even very sophisticated motion analysis methods can fail in the presence of particularly difficult material. This can include objects in very fast motion (e.g. fast moving cars), fast out-of-focus pans (often used in film productions to give a sense of speed), rotating objects (very common in dance sequences), and scenes with large differences in perspective (often used to create dramatic effect in movies).

Transparency can be particularly challenging e.g. where there are objects with one direction of motion overlaid with semi-transparent foreground or background objects which have a different motion (e.g. a person running behind a fountain). Figure 15 illustrates this type of error encountered with a typical motion compensated converter when faced with such material. In Figure 15, a video sequence sourced at UHDTV 48Hz has been frame rate upconverted to UHDTV at 120Hz. It can be seen that the motion analysis has been unable to differentiate areas which appear very similar e.g. the highlights on the child's head are difficult to distinguish from the flowing water. This has led to incorrect reconstruction of the head, as well as defects in the flowing water, in the output frame.

Figure 15 : Detail from conversion of 48Hz source to 120Hz with error seen on the second converted frame

Figure 14 : Detail from four consecutive frames of conversion 24Hz to 120Hz

In order to more correctly manage the large temporal frame rate upconversions and downconversions which

There is no shortage of proposals for advanced motion analysis techniques such as [5] and [6]. However, such methods significantly add to computational complexity, and the real challenge for standards conversion is to be able to implement more sophisticated motion models

efficiently and robustly such that they can be applied to real-time conversion of non-ideal UHDTV content. "Non ideal" means real world content which may have various characteristics including noise, excessive camera integration, mixed natural and synthetic elements, or defects introduced as a result of previous conversions etc. The real-time consideration should not be underestimated, as each UHDTV frame at 120Hz requires processing of four times the amount of data as HD in around 8 msec.

With application of more sophisticated motion analysis algorithms, taking into account rotations, transparency and complex motion models involving multiple motions, we have been able to overcome some of the problems associated with UHDTV frame rate upconversion at high frame rates, with only a moderate increase in system complexity. As shown in Figure 16, application of our more advanced conversion method leads to correct reconstruction of the child's head when converting from 48Hz to 120Hz.

Figure 16 : Detail from conversion of 48Hz source to 120Hz using InSync method

Conversion issues : film rate material

Content originally produced for movie theatre presentation generally follows production conventions somewhat different to material shot for domestic TV broadcast. Motion may be deliberately jerky, or objects in motion may be blurred. In any one production, the director may use a range of shutter angles from 90° to 270° in different scenes to obtain the desired artistic effect.

Frame rate upconversion to 120Hz could give this material an entirely "video" look, for example where all scenes have smooth, consistent motion, and motion blur has been reduced in post-production (not currently feasible on a regular basis, but techniques used to deblur long-exposure still images, e.g. [7] or [8] among others, could be modified to be applicable to movie material in future systems with access to very large

computational resources). While home movie audiences might become used to this type of presentation if they watch a lot of content sourced at 120Hz in the future, this modifies the content away from the director's original intent.

Similarly, where material is shot at 120Hz with a short shutter, but is intended to be converted to lower frame rates to be integrated into a film production, extreme care is needed in processing to ensure that the desired effect of film integration is created. This can include addition of motion blur, proportional to, and along the direction of, objects in motion, see [9] for an example of one typical method.

However, even if the standards converter has been able to create the required motion profile when integrating 120Hz content into a 24Hz production, unless the consumer has a mode available in their domestic TV set to disable the inbuilt upconversion, all careful post-production processing will be ineffective in preserving the "film look".

Conclusions

In this paper, we have discussed the special issues relating to UHDTV conversion which arise from the increased spatial resolution and expected higher frame rates. In particular, we have explained that conventional conversion technology applicable to HD material may not be suitable for the full range of UHDTV material which will be created, and we have demonstrated that specific techniques are needed to preserve the intrinsic value of the UHDTV content when carrying out frame rate and spatial conversions.

Acknowledgements

References

[1] Hobson P, "High frame rate video conversion", *SMPTE 2014 Annual Technical Conference & Exhibition,* October 2014.

[2] Huang Q, et al, "Adaptive Deinterlacing for Real-Time Applications" in *Advances in Multimedia Processing, PCM 2005*, Part II, LNCS 3768, 2005, pp 550 - 560.

[3] Smitha H and Palanisamy V, "Detection of Stationary Foreground Objects in Region of Interest from Traffic Video Sequences", *IJCSI*

International Journal of Computer Science Issues, Vol. 9, Issue 2, No 2, March 2012.

[4] Thomas G, "TV Picture Motion Vector Measurement by Correlation of Pictures", *US patent 4890160,* 1989.

[5] Lu Q, et al, "Frame Rate Upconversion for Depth-Based 3D Video", *2012 IEEE Conference on Multimedia and Expo,* 2012, pp 598 - 603.

[6] Jacobson N, et al, "Motion Vector refinement for FRUC Using Saliency and Segmentation", *2010 IEEE Conference on Multimedia and Expo,* 2010, pp 778 - 783.

[7] Joshi N, et al, "PSF Estimation Using Sharp Edge Prediction", *IEEE Conference on Computer Vision and Pattern Recognition,* 2008*.*

[8] Bunyak Y, et al, "Blind PSF estimation and methods of deconvolution optimization", *Computer Vision and Pattern Recognition*, 2012.

[9] Zheng Y, et al, "Enhanced Motion Blur Calculation with Optical Flow", *Vision, Modeling, and Visualization 2006: Proceedings, November 22-24,* 2006, pp 253 - 258.