White Paper : Achieving synthetic slow-motion in UHDTV

InSync Technology Ltd, UK

ABSTRACT

High speed cameras used for slow motion playback are ubiquitous in sports productions, but their high cost, and the cost and complexity of the associated equipment for storage and management of the content limits their use. Replicating typical current slow-motion system designs in UHDTV, with parallel deployment of high-speed and conventional cameras, becomes prohibitively expensive for many broadcasters, especially considering the 8x higher data bandwidth of UHDTV compared to interlaced HD.

Current synthetic slow motion systems may not be acceptable to consumers when deployed in UHDTV programming, as even the very best of today's motion compensated systems cannot create the flawless interpolated frames needed for perfect motion rendition. UHDTV sourced at 100Hz or 120Hz opens up the opportunity to create synthetic slow motion effects which can be applied in post production, and therefore can apply to content captured with any of the main production cameras. Such processing does not need special on-site equipment, and if sophisticated motion processing techniques are used, creation of a sequence at one third or one quarter actual speed can be implemented very efficiently, thereby meeting the requirement for live productions to have slow motion clips available straight after the action has occurred.

INTRODUCTION

Synthetic slow motion systems are an attractive proposition for broadcasters wishing to offer slow motion effects in a variety of programmes without the expense of dedicated high-speed cameras and specialised slow motion equipment. The quality of slow motion effects shown on TV and in cinemas is now so good that consumers would not expect anything less from a synthetic slow motion system. However, current synthetic slow motion systems are limited, as even the very best of today's motion compensated systems cannot create the flawless interpolated frames needed for perfect motion rendition. Playback of a 50Hz or 59.94Hz clip at one third or one quarter speed requires creation of 3 or 4 output frames for every input frame. Applying such interpolation to very fast moving content, as is required for sports slow motion playback, inevitably leads to visible picture defects.

Broadcasters are already experimenting with UHDTV sourced at 100Hz or 120Hz, especially for sports productions. In such deployments, we can start to imagine new opportunities to create synthetic slow motion effects which can be applied in post production. Synthetic slow motion effects can be applied to content captured with any of the main production cameras, which reduces the need for special on-site equipment. In this paper, we will show that, by application of state of the art motion processing techniques, it is possible to implement a very efficient system which offers good quality synthetic slow motion playback of sports action at one third or one quarter actual speed. Not only does such a system help broadcasters reduce production costs, the efficient implementation leads to shorter processing times, thereby also meeting the requirement for live productions to have slow motion clips available straight after the action has occurred.

Starting with a higher temporal sampling rate is clearly advantageous, but we still need to advance the state of the art in motion compensated processing in order to create additional interpolated frames which will meet the requirements for high quality video on domestic UHDTV sets. In this paper, we will illustrate the problems of creating synthetic slow motion from 50Hz and 59Hz content. We will go on to describe some advanced motion analysis methods which enable improved motion compensated interpolation of content, which, when applied to creation of production-quality synthetic slow motion from 120Hz UHDTV material, offer broadcasters the opportunity to have low cost slow-motion effects for any content.

SLOW MOTION SYSTEMS

Conventional slow motion systems

Slow motion systems in current popular use are based around high speed cameras that enable very high frame rate capture of events. The frames are stored to disk, then can be played back at a range of slower speeds, as desired by the programme director. The advantage of such systems is that picture quality will be excellent, since high resolution, fast shuttered video cameras are used for acquisition.

A complete slow motion system, used for e.g. a sporting event or film special effects, will include multiple high speed cameras, large, specialised storage devices, and dedicated user interaction equipment to enable slow motion replay of the chosen action within a few seconds of the action occurring.

The disadvantage of such systems is that they tend to be quite expensive, since specialised equipment is needed. Therefore they are usually found only in high-end broadcast workflows. Costs will continue to be an issue when production moves over to full UHDTV, since both the cameras and the storage will be more expensive.

A further disadvantage is that only the action captured with the high speed cameras can be shown in slow motion. In a typical sports event, there will be multiple cameras around the stadium, but only those with the very high speed capability can be selected for slow motion playback.

Synthetic slow motion systems

A synthetic slow motion system is one in which material shot at standard broadcast frame rates, e.g. 50Hz or 59.94Hz, is used to create slow motion effects by interpolation of frames representing the slower (playback) rate. For example, if the director wishes to play back at one third speed some action which has been captured at 50Hz, the synthetic slow motion system will frame rate upconvert to a 150Hz frame rate, then the file system will play back the sequence at 50Hz. This is illustrated in Figure 1, where source frames are captured at 50Hz (S_i) then are frame rate upconverted to three times the rate (U_i) i.e. 150Hz. If the frame rate upconverted sequence is now played out at 50Hz, the effect will be motion at one third speed.

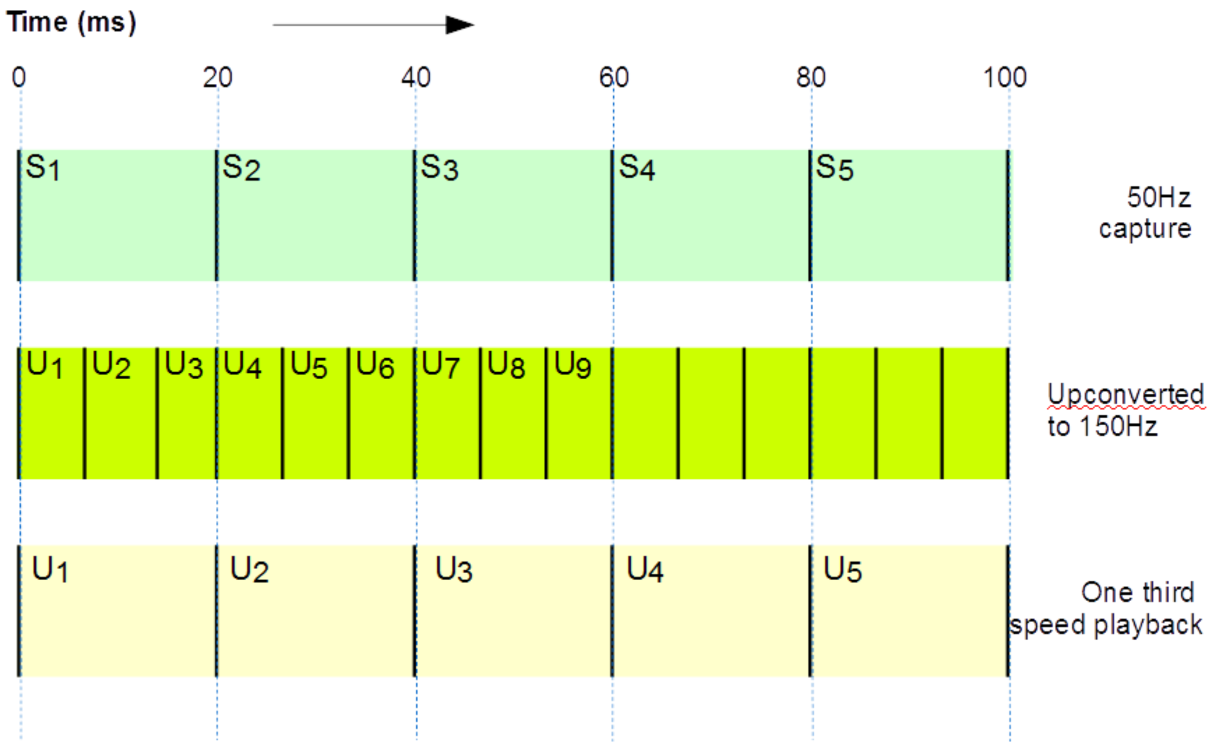

Figure 1 - Illustration of synthetic slow motion using 50Hz source

As Figure 1 illustrates, generation of the synthetic slow motion sequence requires creation of two new frames per input frame for one-third speed, and this would be three new frames per input frame for one-quarter speed. Given the high level of sophistication of current motion compensated standards converters, it is tempting to assume that one may simply use motion compensation to interpolate the intermediate frames, and achieve the desired slow motion effect.

However, content for which slow motion replay is required typically contains fast moving action, i.e. exactly that which poses problems for motion estimation algorithms. Small objects with fast, complex and possibly contrary motion, are difficult to track accurately, and the results of motion compensated processing are not always high enough quality to meet the needs of the home viewing audience.

Figure 2 shows an example of using a motion compensated converter to interpolate 60Hz content to 240Hz, for playback at quarter speed. It can be seen that even though much of the picture is of good quality, there are places where the motion analysis and frame interpolation has not been accurate enough to obtain a good rendition of the person's arm in rotary motion.

60 Hz source frames

240Hz output frames

Figure 2 - Illustration of interpolation from 60Hz to 240Hz

IMPROVING SYNTHETIC SLOW MOTION

A slow motion system based on 120Hz acquisition

Higher frame rates such as 100Hz and 120Hz, have been proposed for UHDTV production and delivery systems in order to create a step change in viewing experience. Faster acquisition is highly beneficial in improving viewer enjoyment of sports events, and Figure 3 shows how such a system could be used to create synthetic slow motion for both the 120Hz UHDTV viewer and those watching via traditional HD services.

In Figure 3, cameras at the sports event acquire frames at 120Hz. The content is stored ready for motion compensated frame rate upconversion when the director requires it. The slow motion content is then made available to the vision mixer for integration into the programme. Support for HD 60Hz services is provided via spatial and frame rate downconversion which includes synthetic aperture processing to enable the 60Hz viewer to obtain a smooth motion effect.

The system of Figure 3 can help overcome the problems seen in Figure 2, as using high frame rate capture at the initial stage provides more input frames from which to estimate the motion in the scene. Using higher frame rate source material has the obvious benefit of smaller increments of motion between each frame. This enables a motion compensating interpolator to calculate a more accurate motion vector, which then leads to a higher likelihood of being able to generate high quality intermediate frames. In addition, if sophisticated motion models are employed, the best possible quality output can be obtained from the input content.

In the example of detail from a horse race, Figure 4, a 120Hz source is converted to 400Hz using motion compensated interpolation. If played back at 100Hz in a possible future European UHDTV 100Hz service, such processing would allow the 400Hz result to be played back at quarter speed, with excellent rendition of the horses' movement in slow motion. If additional synthetic aperture is applied and the content is spatially and frame rate downconverted, it could also be delivered via a conventional HD 50Hz service.

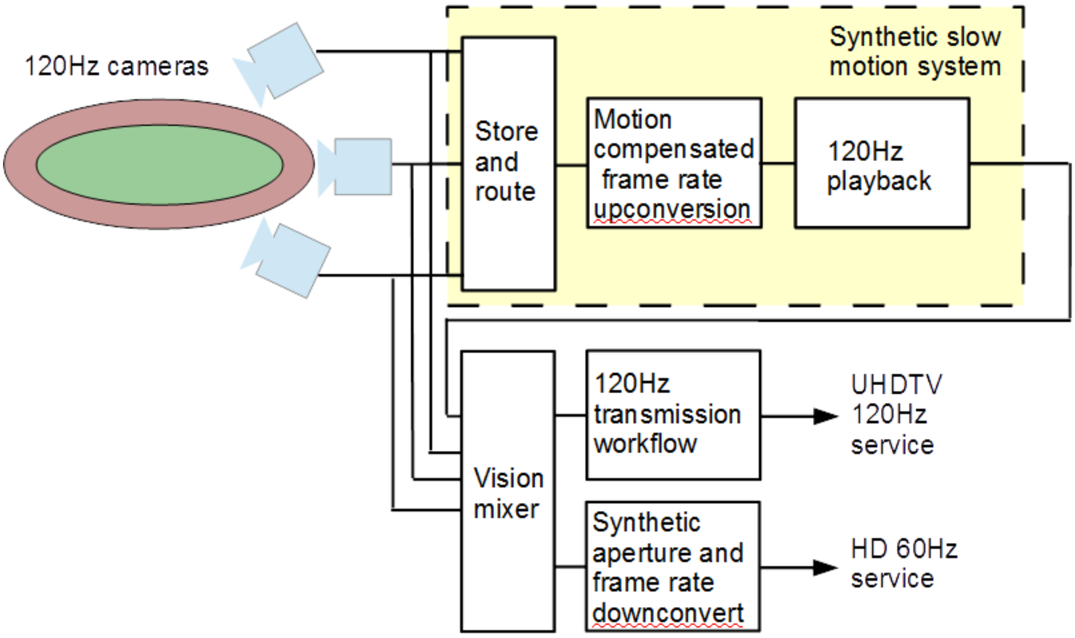

Figure 3 : Synthetic slow motion system based on 120Hz UHDTV cameras

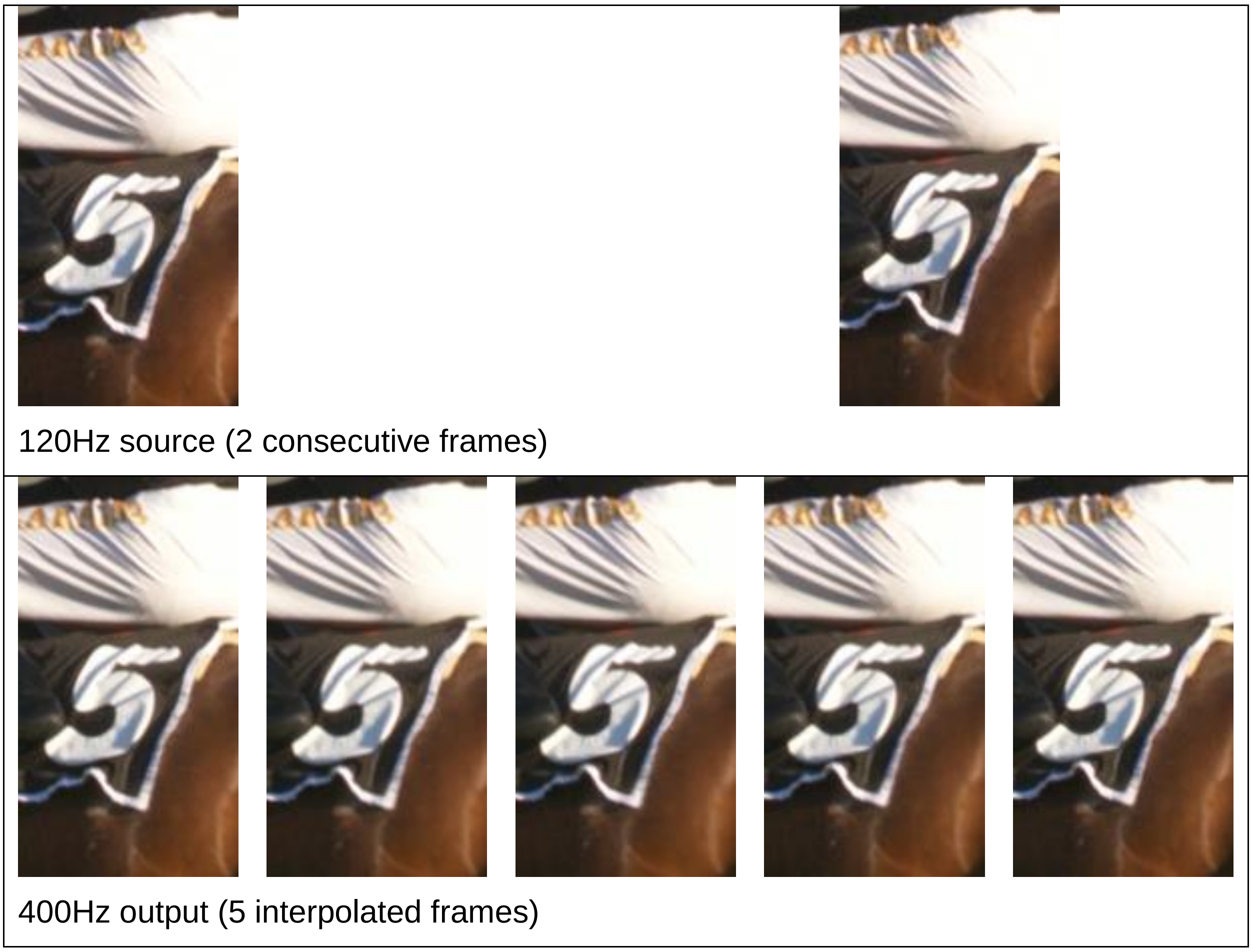

Figure 4 - Illustration of motion compensated interpolation from 120Hz to 400Hz

Improved motion processing

Achieving really excellent synthetic slow motion reproduction requires improved motion compensated processing in addition to higher frame rate acquisition at 100 or 120Hz, which on its own does not create the superb quality required. When watching a slow motion replay in a UHDTV viewing environment, the consumer has an expectation of high quality and a desire to watch the action very carefully to see all the detail of the situation. Any artifacts introduced during the motion compensated interpolation process will be highly visible. These problems are compounded by the fact that slow motion effects are required for fast moving content (such as sports) which are well known for being the most difficult types of content for motion estimation.

Therefore, in addition to shooting at 100 or 120Hz, we need to adopt advanced, yet efficient, motion processing methods. Efficiency (meaning low computational time) is needed as the synthetic slow motion effect must be applied as quickly as possible, since viewers want to see the replay of the action as soon as possible after the event has occurred.

Motion compensated processing has been well covered in the literature over many years, so we will not repeat here descriptions of the basic techniques used for motion vector generation and picture interpolation. However, in order to meet the higher quality demands of synthetic slow motion, the process must exceed the best existing interpolation as performed by today's high quality standards converters. An optimised method focuses on interpolation precision more than other performance metrics such as the range of motion speeds accommodated. The shorter sampling intervals found in 100Hz and 120Hz standards effectively reduces the magnitude of differences but not their complexity.

To obtain the required quality of conversion to meet viewer expectations when creating synthetic slow motion effects in UHDTV content, we need to create more sophisticated motion models allowing for multiple different motions, object rotations, and changes in field of view, as well as the usual "hazards" of motion analysis e.g. dealing with periodic structures and semi-transparent objects.

There are two key techniques which can bring a step change in conversion quality whilst gaining in computational efficiency. The first is highly accurate moving/static area delineation. Pre-processing to identify static areas which do not need motion vectors to be calculated not only provides efficiency gains, but also enables the highest resolution to be maintained in those stationary areas.

Typical methods for detecting stationary areas or image similarity use field to field (or frame to frame) absolute differences, with simple noise coring, which tend to fail in areas of high frequency due to aliasing, or where there are illumination changes between subsequent images. See, for example Huang (1). In general, methods relying on the mean of absolute difference between frames has the disadvantage of being susceptible to signal noise and issues arising from aliasing and periodic structures in the video sequence, leading to errors of false positives and false negatives. More complex methods exist e.g. Smitha (2), in which dynamic models of the scene are built up over a large number of frames. These latter methods have the significant drawback of requiring long processing times and have an assumption of a static camera, which is not the case in typical video sequences.

120Hz source (2 consecutive frames)

400Hz output (5 interpolated frames)

Figure 5 - Illustration of motion compensated interpolation from 120Hz to 400Hz

We have found that more successful and reliable static area detection can be achieved using methods which apply analysis in multiple orientations. As a pre-processing stage the image is convolved with an aperture designed to attenuate high frequency texture content while preserving mid and low frequency texture content. The purpose of this filtering operation is to attenuate the highest frequency components in the signal which are most likely to exhibit a poor signal-to-noise ratio and well as unhelpful signal properties due to previous processing such as enhancement.

Following the pre-processing, analysis in the chosen multiple orientations leads to a set of results which are then post-processed to obtain a consistent decision (static vs moving). Normalisation of the results reduces the impact of outliers (e.g. due to noise) and using a combination of multiple analyses avoids issues of potential confusion between static periodic structures, and movement of edges in one orientation.

Addressing complexity is not trivial. Irrespective of the innumerable approaches to motion compensated interpolation, a solution interpolating with the required precision will need to support higher order motion models capable of measuring and describing the complexity of motion vectors fields that will maintain key image metrics such as resolution, texture, boundary, shape and scale as perceived by the end viewer. Fortunately, motion model complexity does not have to mean higher processing cost. On the contrary, a more complex model may permit extremely elegant, efficient and precise internal datasets affording orders of magnitude less bandwidth to describe when compare to the coarse approach employed by some converters.

Figure 5 shows further results of the frame interpolation from our synthetic slow motion system, applied to a 120Hz source.

CONCLUSIONS

In this paper, we have illustrated a method for creation of synthetic slow motion effects from higher frame rate (100 or 120Hz) material, enabling broadcasters to save on costs of conventional slow motion systems, which will be very high for UHDTV production systems.

REFERENCES

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

2. 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.

ACKNOWLEDGEMENTS

The author would like to thank James Shepherd and Barry Flannaghan of InSync Technology for their contributions to this paper.

We also thank Digiturk for the 120Hz test sequences used in our experiments, which we obtained from http://ultravideo.cs.tut.fi/#testsequences under Creative Commons Attribution-NonCommercial 3.0 Unported http://creativecommons.org/licenses/by-nc/3.0/deed.en_US.