

INTERNATIONAL HDTV CONTENT EXCHANGE

Mike Knee, Snell & Wilcox (UK)



ABSTRACT

The key technical process in international exchange of broadcast content is standards conversion. This continues to be an important process as the world moves rapidly into HDTV broadcasting. This paper reviews the general subject of standards conversion. It then discusses the particular requirements of HDTV standards conversion, arising from new production techniques, workflows and display devices. Particular emphasis is given to the problem of high quality upconversion and de-interlacing of standard definition sources. Techniques of motion estimation and picture building are then presented, with special regard to the ways in which Phase Correlation is applied to HDTV sources. The performance of HDTV standards conversion within a complete broadcast chain is discussed, taking into account downstream compression performance, for which some new results are presented. Finally, the application of the techniques to other motion compensated image processing tasks is briefly discussed.

INTRODUCTION

Standards conversion has been around almost as long as television standards themselves. This paper looks at what is special about conversion between HDTV standards. We start with a review of standards conversion in general, before looking at the particular requirements of HDTV. We then introduce techniques for HDTV standards conversion and discuss their performance. Finally, we take a wider look at where the best HDTV standards conversion technology might lead us.

REVIEW OF STANDARDS CONVERSION

The Fundamental Problem

The main problem of standards conversion arises because we need to change the temporal sampling rate of the video, for example from 50 Hz to 59.94 Hz, by creating information at time instants not represented in the input signal. The need to change the number of lines per field is a secondary problem, though not trivial, especially when interlace is involved.

Why is temporal sampling rate conversion more difficult than spatial sample rate conversion? The main reason is to do with the sampling theorem, which states that, for alias-free reconstruction, the sampling rate of a signal must be at least the Nyquist limit, which is twice the signal's bandwidth. Spatial sampling comes close to meeting this limit because the horizontal and vertical bandwidth of television signals is constrained by the camera optics. Temporal sampling, however, especially for shuttered cameras in fast-moving areas, is far below the Nyquist limit. The eye and brain manage to overcome this problem by tracking the motion and filling in missing information between fields according to where it expects objects to be. If a standards converter, which also has to create missing

information between fields, fails to meet the brain's expectations, the result will be a degraded picture.

Simple Conversion

Figure 1 shows a simplified example of field rate upconversion by a factor of 3:2. The first row of pictures shows three successive input fields of a sequence where a figure moves in front of a house. The second row shows what the brain would expect to perceive at the output frame rate. The third row illustrates the simplest possible field rate conversion, where each output field is copied from the input field nearest in time. Although each output field seems to be free of processing artefacts, the repetition of every second input field would be perceived as motion judder. The final row illustrates the effect of linear temporal interpolation, a method which would work reasonably well if the temporal sampling rate were above the Nyquist limit. Here, however, it leads to the phenomenon of a double image. More sophisticated, multi-tap interpolation filters would not help here – they would simply lead to more multiple images.

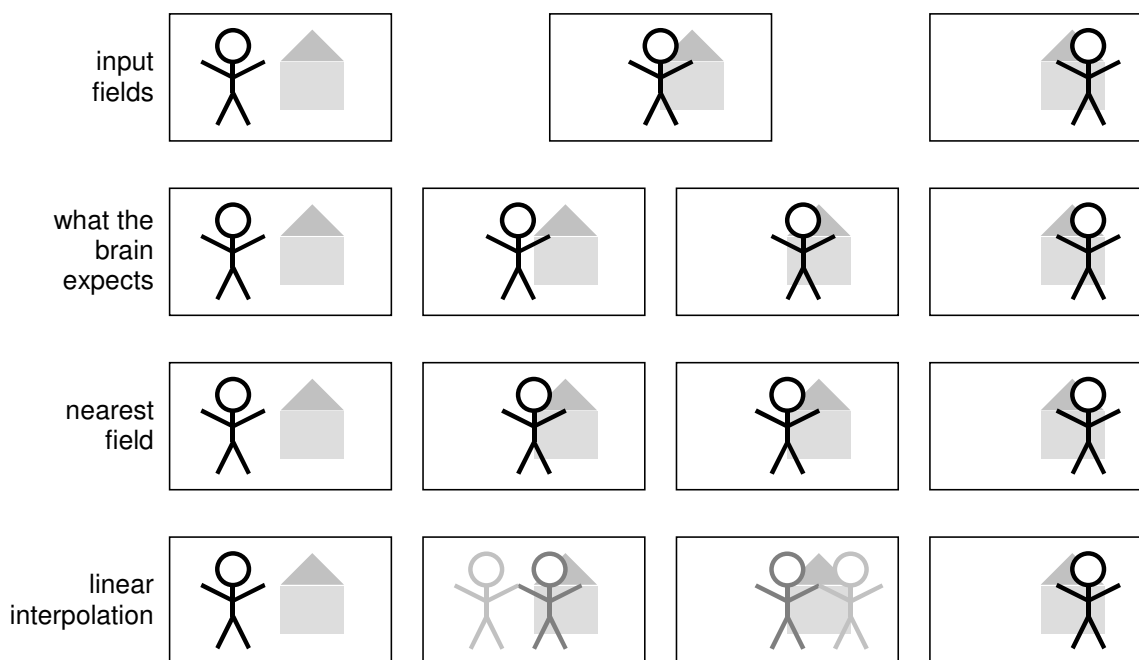


Figure 1 – Standards conversion examples

The simple approaches shown in Figure 1 are unacceptable but they do have two key advantages, which we would like to preserve when developing a better algorithm. They are safe, in the sense that the extent of picture degradation is predictable depending on the input material. And they are free from additional artefacts resulting from attempts to switch between interpolation modes within a picture or sequence.

Motion Compensated Conversion

We have seen how standards conversion fails to provide good perceived picture quality unless the interpolation algorithm is able to track the motion of objects in the scene. Motion compensated standards conversion aims to track this motion so that the interpolated fields can be filled in as the brain would expect. A typical motion compensated converter has two blocks as shown in Figure 2.

The picture builder calculates the value of each pixel in the output picture as a function of the input picture and of the motion vectors. If the motion vectors correctly describe the motion and the picture builder makes correct use of them, then the output sequence in our example above should look like the second row of Figure 1. However, if the motion vectors are incorrect, for example part of the figure is given the static motion belong to the background, then undesirable artefacts can occur which can sometimes look even worse than the results of linear interpolation. Figure 3 gives an example of what might happen, though the exact behaviour depends on how the picture builder makes use of the information before and after the interpolated frame.

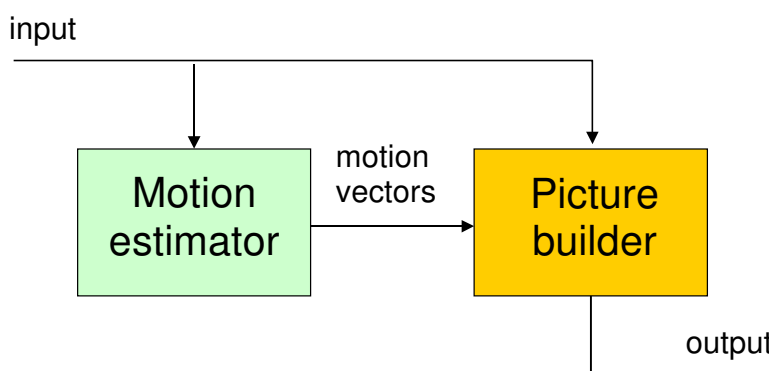


Figure 2 – Motion compensated standards converter

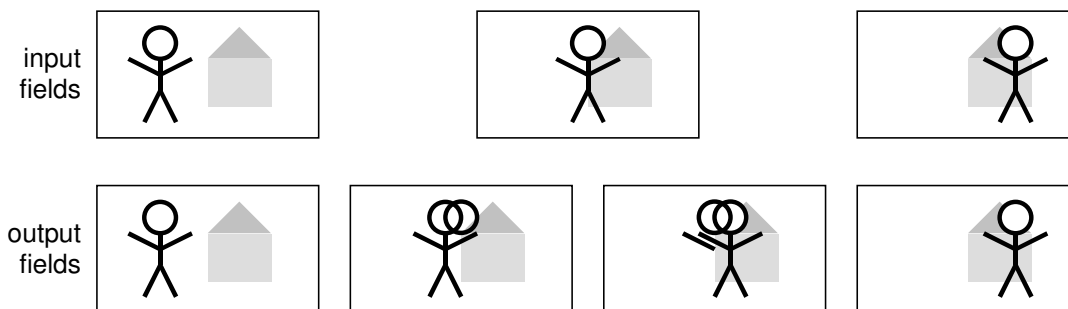


Figure 3 – Example of bad motion-compensated conversion

Phase Correlation

Most of the problems of motion estimation encountered in standards conversion can be overcome by using a technique that aims to measure “true motion” in the scene rather than simply obtaining a good match. One such technique is Phase Correlation (PhC), shown in Figure 4.

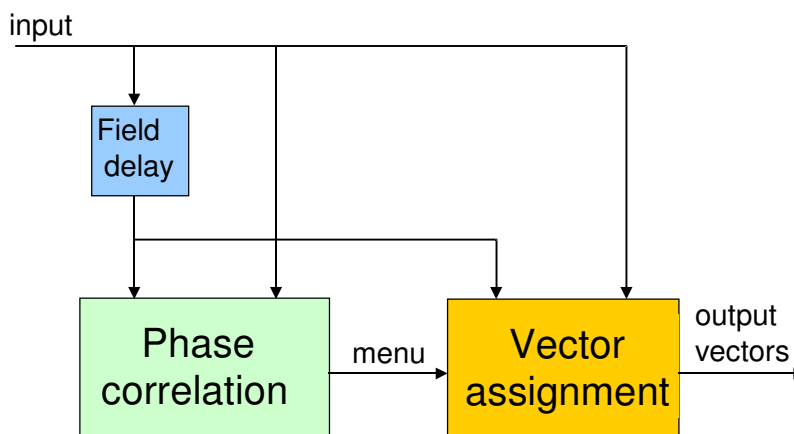


Figure 4 – Motion estimation by PhC

PhC works by breaking the picture up into large blocks and using Fast Fourier Transforms (FFTs) to calculate the phase difference between corresponding blocks in consecutive fields. The inverse FFT of the phase difference is a **correlation surface**, which has peaks corresponding to the prevalent displacements between the two fields, leading to a **menu** of candidate vectors. Each pixel is then **assigned** one of the menu vectors by a matching technique. The action of limiting vector choices to menu vectors that have resulted from a larger-scale analysis of the picture helps to bring about a better approximation to true motion than most other techniques.

The **range** of motion vectors that can be estimated using phase correlation is typically about $\frac{3}{4}$ of the size of the FFT blocks. As blocks of between 64x64 and 256x256 are perfectly feasible in hardware, the motion vector range is often many times greater than might be achieved using other techniques.

The **accuracy** of motion vectors depends on how the peaks in the correlation surface are detected but is typically a small fraction of a pixel, again in contrast to other methods which are fundamentally limited to integer accuracy unless explicit interpolation is used on the input picture to obtain, say, half or quarter-pixel accuracy.

A fundamental choice in implementing phase correlation is the sample density of the pictures to be used. Reducing the sampling density brings about a trade-off between actual motion vector range for a given block size and the output sample density of the motion vectors. Other factors come into play here, for example the consistency and noise immunity of the output motion vectors so that, even if the FFT block size is not an issue in terms of hardware complexity, a degree of downconversion may be desirable.

REQUIREMENTS OF HDTV STANDARDS CONVERSION

The techniques described above were developed for standard definition (SD) signals. We now look at the particular requirements of HDTV.

Why HDTV is Different

In many applications, the only important point to note when comparing HDTV to SDTV is the

INTERNATIONAL HDTV CONTENT EXCHANGE

Mike Knee, Snell & Wilcox (UK)



increase, typically by a factor of 5, in the sampling rate. For uncompressed links and storage, that is all that matters.

But HDTV does not just mean more pixels! The next few paragraphs describe more substantial differences between HDTV and SDTV.

Upconversion

For many years to come, SD sources will form an important part of the input to HDTV broadcast networks. These sources need to be **upconverted** to HDTV. Ironically, as the proportion of true HD sources increases, the need for good quality upconversion of remaining SD sources will also increase because viewers will get used to HDTV quality and will become less tolerant of inferior SD pictures.

A key component of a good SD to HD upconverter is high quality **de-interlacing**. Artefacts and shortcomings of poor de-interlacing, such as loss of resolution, line twitter, “barber’s poles” on diagonals, and “mouse teeth” and double images on moving edges, would all be magnified in an HD display.

Other important elements in upconversion are interpolation filters that are sharp and free of ringing, and noise reduction.

HDTV production techniques

Turning to true HD sources, it is important to recognize that an HD production will not just produce the same pictures as SD but with more resolution. An SD camera following live sport will often zoom in quite tightly to the action. An HD camera, providing a wider “window on the world”, will be more likely to remain zoomed out, allowing the action to move across the screen, as illustrated in Figure 5.

The wider field of view of the camera means that the maximum speed in picture widths per second of HD shots might not be as high as in SD. However, this is offset by the greater resolution of the HD raster. And fast panning is still a possibility, so in reality we must assume that the maximum speed in HDTV pixels per second is significantly higher than for SDTV.

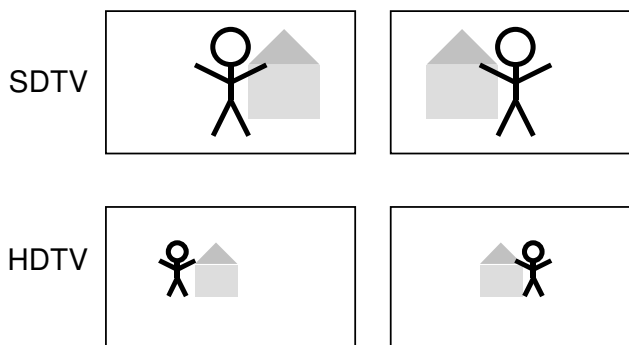


Figure 5 – SDTV and HDTV camera techniques

A more important consequence of HDTV camera techniques is that artefacts due to bad motion compensation are more likely to show up in the foreground. In SDTV, most of the motion, and thus any motion-related artefacts, will be in the background as the camera tracks the action. In HDTV, the foreground is more likely to be moving and will therefore be more susceptible to degradation.

INTERNATIONAL HDTV CONTENT EXCHANGE

Mike Knee, Snell & Wilcox (UK)



Upstream compression

Compression has long been a feature of broadcast television networks throughout the chain, not just in the final transmission link. The higher raw data rate of HDTV means that compression is a more attractive option and will be used more frequently. For example, digital video tape recorders for SDTV may use compressed or uncompressed formats, but for HDTV they will almost certainly be compressed. Care must be taken in the design of motion estimators for HDTV image processing that they are not fooled by compression artefacts, including those that may not be visually annoying.

Downstream compression

Any compression coder that is downstream of a standards converter may have difficulties when it encounters standards conversion artefacts. When assessing the effect of different standards converters on downstream compression quality, it is important to consider the complete chain, rather than the performance of the compression process alone. For example, a poor quality standards converter that worked by field or frame repetition would produce a result that appears quite “friendly” to a compression coder. A converter that produced very soft pictures in order to mask motion artefacts would possibly seem even easier to compress. But in both cases the overall performance of the chain would be poor.

Later in this paper we will show that it is possible to develop standards conversion algorithms that perform well in themselves while also being compression-friendly.

HDTV displays

Many flat-screen consumer HDTV displays have poor motion performance, partly because of LCD switching speeds and partly because of the sample-and-hold operation of the display. Such displays suffer from motion blur, which can be quite effective in masking motion artefacts in standards conversion. However, this phenomenon could lead to complacency about conversion quality. The motion performance of HDTV displays will undoubtedly improve, for example through the use of pulsed backlight to reduce the effective lag of the display, or through intelligent display processing designed to compensate for motion blur. LCD switching times are also set to decrease, from the typical values of 12 to 15 ms today to a projected value of 4ms in 2008/9. Such improvements may well have the effect of “unmasking” motion-related artefacts in upstream standards conversion.

So it is not enough for converted pictures to look acceptable on the current generation of consumer HDTV displays; they must also look good on large CRT displays, which remain an excellent reference for viewing the effects of moving-image processing algorithms.

HDTV STANDARDS CONVERSION TECHNIQUES

Upconversion

We shall concentrate on de-interlacing, the most demanding aspect of upconversion.

Broadly, de-interlacing algorithms fall into three categories: (a) spatial, (b) motion adaptive and (c) motion compensated. In addition, spatial de-interlacing may include elements of adaptation to diagonal content.

In order to obtain the best possible de-interlaced picture quality, motion compensated

processing is essential. This is because, although there is always some information present in the input signal at each output time instant, the full resolution can only be obtained by access to other fields. If there is any motion present in the scene, such access must be along the line of motion.

Motion compensated processing may be necessary, but it may also be more “dangerous” than simple spatial processing because the artefacts resulting from bad motion vectors may be more objectionable than commonly accepted features of simple spatial interpolation.

Multi-dimensional adaptive de-interlacing

The algorithm we have developed overcomes these problems by adapting smoothly between multiple de-interlacing modes that cover all the categories listed above. Figure 6 shows a generic block diagram of such a multi-dimensional adaptive de-interlacer.

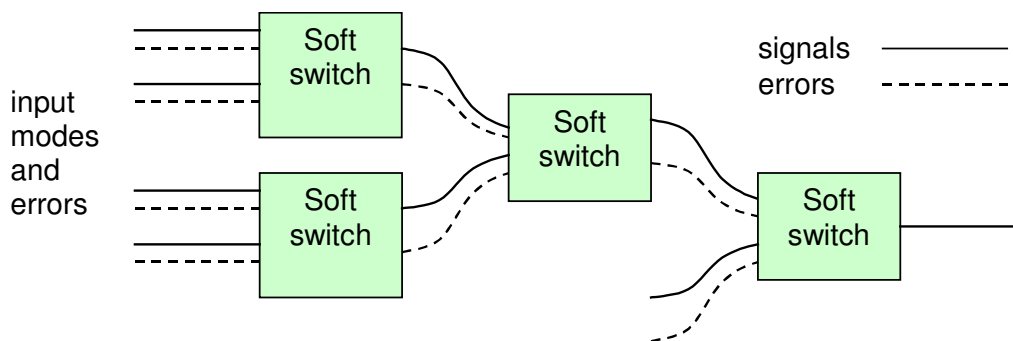


Figure 6 – Multi-dimensional adaptive de-interlacing

Each soft switch adapts between two de-interlacing modes according to an associated error signal. Its outputs are the switched main signal output and a new representative error signal to pass on to the next stage. The parameters of all the switches have been optimized over a very wide range of critical source material. The use of soft switching avoids switching artefacts which can otherwise look more objectionable than the original interlace artefacts.

Motion Estimation

We now turn from upconversion to the main HDTV signal path. From now on, we are dealing with an HDTV signal, whether upconverted or HDTV-originated.

Picture conditioning

Even though we are beginning with an HDTV signal, we still have the problem that the input may be interlaced. We prefer to work with progressive signals, so there is a separate de-interlacer at the HDTV input, split into two stages. First, spatial de-interlacing is applied to obtain an input to the main motion estimator, and then motion compensated de-interlacing similar to that described above is applied to obtain the input to the picture builder.

Phase correlation

The motion estimator for HDTV standards conversion continues to be based around phase correlation. We have already seen that there are tradeoffs involving block size, sample density and the range, accuracy, localization and noise immunity of motion vectors. The final choice of block size and sample density exploits these tradeoffs, taking into account the specific requirements and properties of HDTV discussed above.

Motion vector post-processing

Following phase correlation, post-processing is carried out on the motion vectors, modifying them to reflect a measure of confidence obtained during the phase correlation process. There is also a process in which the spatial resolution of the motion vector field is increased so that the edges of objects are followed as accurately as possible, while avoiding the noise that would be associated with other pixel-based motion estimation schemes.

Picture Building

We recall that, in general in standards conversion, we wish to generate a picture at a point in time somewhere between two input pictures. The main principle of the picture builder is to apply forward motion compensation to the previous picture and backward motion compensation to the next picture, and then to combine the two motion compensated pictures to produce the final output picture.

Read-side or write-side?

There are two basic approaches to motion compensated picture building, which we refer to as **read-side** and **write-side**; see Figure 7. In read-side picture building, each output pixel value is calculated by applying a motion vector to its position and **reading** the value of the pixel pointed to in the input picture. In write-side picture building, each input pixel value is projected according to its motion vector and **written** to the location pointed to in the output picture. Each method has advantages and disadvantages. Read-side picture building guarantees that there will be a value for each output pixel, but relies on the availability of a motion vector for each output pixel, whereas the motion vectors have been calculated on input pictures. Conversely, write-side picture building uses the vectors that have been calculated on the input picture grid, but may leave gaps (“holes”) or conflicts (“multiple hits”) in the output picture.

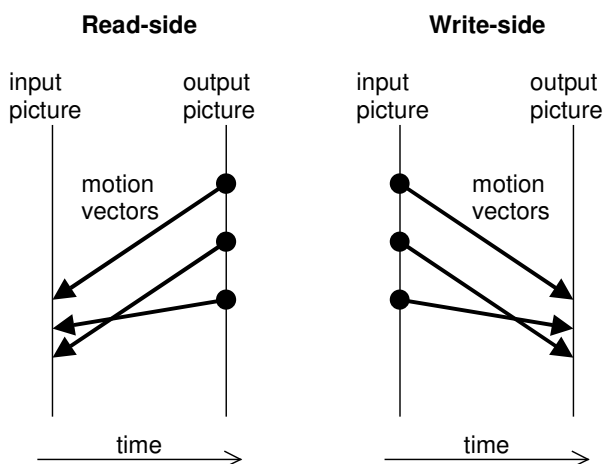


Figure 7 – Read-side and write-side picture building

Occlusions

We have chosen to use write-side picture building, so we have to solve the problem of holes and multiple hits. If the motion vector field is well-behaved, then these phenomena will be due to **occlusions** – areas of background that are being revealed or obscured by foreground. For example, the moving ellipse in Figure 8 will produce holes in its trailing edge and double hits in its leading edge in the forward built picture, and *vice versa* for the backward built picture.

By combining the forward and backward built pictures taking into account the “hit count” for each pixel, it is possible to ensure that occluded areas are filled in cleanly from information in the opposite temporal direction, producing the result shown at the bottom of the figure.

The HDTV picture builder includes provision for sub-pixel interpolation accuracy and for partial occlusions.

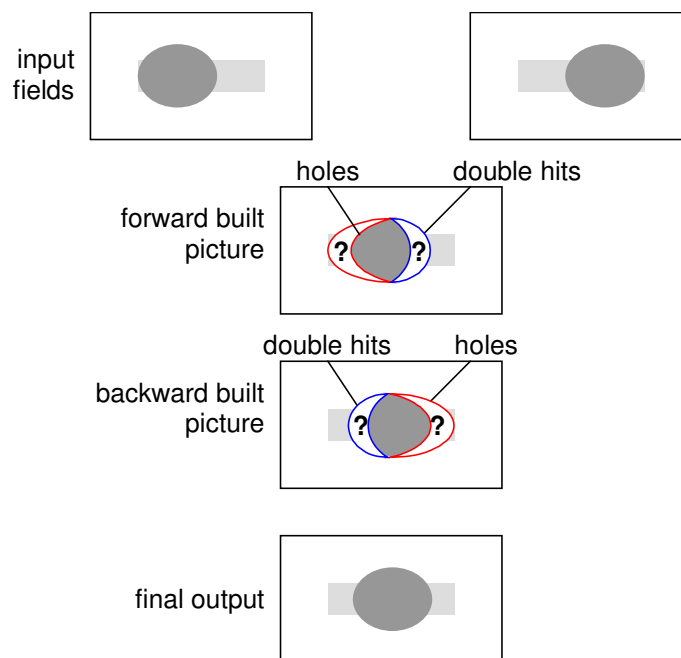


Figure 8 – Occlusions

Picture Quality Assessment

Two kinds of picture quality assessment have been used in the development of the algorithm and the fine tuning of its parameters. For certain test sequences, it is possible to generate a “perfect” or “ground truth” output with which to compare the actual output, optimizing numerical parameters to minimize an objective measure such as mean-squared error. It is important to note that objective measurement systems, including JND and PQA, cannot be used with ordinary picture material to test standards conversion because no reference signal is available. The most important method of picture quality assessment therefore continues to be subjective evaluation with critical test material on a variety of display devices.

HDTV STANDARDS CONVERSION PERFORMANCE

Reference Quality

Careful design and optimization of the motion estimator and picture builder have ensured that the algorithm is capable in the best case of delivering pictures with resolution, colour and grey-scale rendition and noise levels visually indistinguishable from a source generated at the output frame rate, subject to the fundamental resolution constraints



Motion Portrayal

Obviously, a motion compensated standards converter will be judged not only by its reference quality but also by how well it portrays complex motion. Test material can always be found which will give numerically poor results in some parts of the picture. However, the subjective testing element of the optimization process has ensured that the visibility of such errors at the output frame rate is minimized.

Downstream compression performance

Figure 9 shows a rate-distortion comparison on a short but typically demanding sports clip between an excellent existing motion compensated standards conversion algorithm and an algorithm developed according to the techniques described in this paper. The encoding used was long-GOP MPEG-2 using a leading high-quality encoder. In order to overcome the problem of comparing different source sequences at the input to the encoder, mild spatial spectrum equalization was applied to the sequences so that they generated equal bit rates around the centre of the graph when I-frame-only encoding was used.

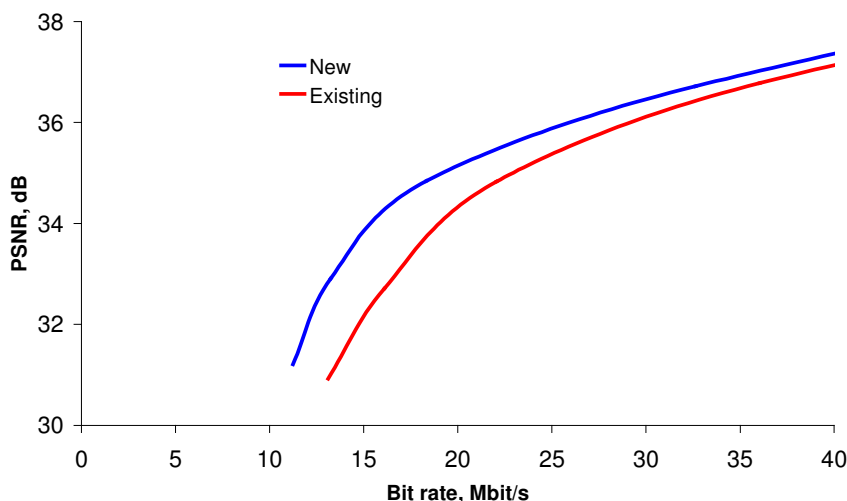


Figure 9 – Downstream compression performance

This result shows that good quality motion estimation and picture building has the added effect of improving downstream compression performance, by up to 1.5 dB or 4 Mbit/s in the example given here.

BEYOND STANDARDS CONVERSION

Film Motion – Changing Attitudes

Until recently, 24Hz film material has been transmitted using patterns of field repetition, 3:2 pulldown in 60 Hz countries and 2:2 (at 25 Hz) in 50 Hz countries. This results in motion judder, but the argument has been that this is no worse than the judder that results from double or triple-flash illumination in cinemas. But as displays become bigger and brighter, there is evidence of increasing intolerance to this effect. There is therefore likely to be a growing market in true standards conversion from 24 Hz to 50 or 59.94 Hz, so that the motion portrayal is smoother than has been accepted in the past. The algorithm presented

INTERNATIONAL HDTV CONTENT EXCHANGE

Mike Knee, Snell & Wilcox (UK)



in this paper is well placed to be adapted to this task.

Rate Changing

There is a great deal of interest in changing the frame rate of material in post-production applications, for example for slow motion or to reproduce the effect of the variable frame rate which has always been a feature of film cameras. Again, the algorithm presented here is suitable for this task, though some parameters may have to be re-optimized to maximize subjective quality when viewed at the display frame rate.

CONCLUSIONS

This paper has explained what is special about HDTV standards conversion and has described techniques for motion compensated HDTV conversion that provide excellent picture quality and performance in a digital television broadcast chain.