10 bit high quality MPEG-4 AVC video compression

Matthew Compton
TANDBERG Television - Part of the Ericsson Group.
Southampton, UK

Abstract

HD MPEG-4 AVC is a standard that is well established in the broadcast industry for over 3 years. It offers premium compression performance at significantly lower bitrates then MPEG-2. However, there are still aspects of the H264 AVC toolset which are not widely used, in particular High 10 Profile (Hi10P) - which supports full resolution 10 bit video encoding. From contribution and distribution (C&D) to satellite direct-to-home there are applications for 10 bit HD operation that can enhance performance.

Direct-to-home broadcasters are competing with the high definition DVD market and therefore require optimum compression performance. In particular, areas for improvement are plain backgrounds, which can suffer from colour contouring etc. This is also becoming increasing relevant as consumer displays migrate to 10 bit technologies, therefore there will be a demand for ultimate dynamic range. Likewise, C&D markets desire 10 bit HD video through the entire broadcast production chain, this can be achieved in addition to 4:2:2 colour processing with the H264 AVC High 4:2:2 Profile (Hi422P).

This paper explores the advantages and disadvantages of HD MPEG-4 10 bit encoding at a variety of bitrates on different types of content. Comparisons are made at the different operating points, demonstrating where gains maybe achieved.

Introduction

The high compression efficiency of MPEG-4 AVC makes it an attractive solution compared to MPEG-2. Broadcasters have the choice of increasing the channel density through a given transmission medium (compared to MPEG-2) or increasing video channel quality. In addition, as technology advances and MPEG-4 AVC compression is better understood, incremental quality improvements are achieved over time.

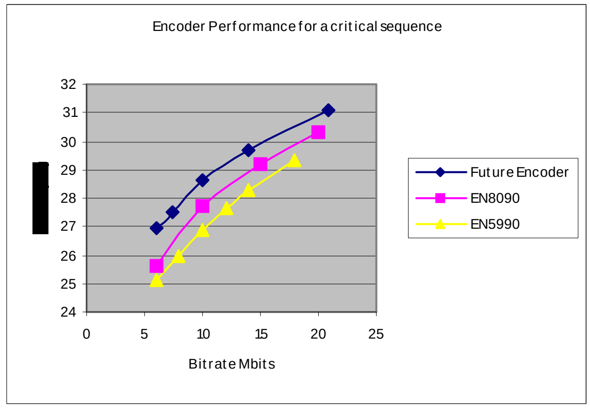

Fig. 1 showing PSNR comparison of MPEG-4 AVC equipment.

Fig. 1 shows the compression performance improvements of existing equipment over the last three years. An indication is also given to what might be expected in the future.

The coding gain of MPEG-4 AVC can be attributed to a number of new coding tools which are not available in MPEG-2. Detailed descriptions of the new coding tools have been published [1, 2]. A list of the most important coding tools is shown below:

- 9 4x4 intra prediction modes
- 4 16x16 intra prediction modes
- 7 inter prediction modes from 16x16 down to 4x4 block sizes
- 1/4 pel motion compensation
- advanced bi-directional (B) pictures
- motion compensation from outside of the picture
- multiple reference pictures
- integer transformation
- in-loop de-blocking filter
- Context-Adaptive Binary Arithmetic Coding (CABAC)

Measuring Image Quality

Perceived video quality can be by its very nature subjective, especially if there is any degradation from the source due to compression or video processing.

In order to make comparisons in this paper, a mechanism needs to be selected to compare sequences compressed at different bitrates and/or with different coding tools.

Common mechanisms are:

- **PSNR** (Peak Signal to Noise Ratio) is most commonly used, and is built into the JVT's reference software (JM 12). In addition it is widely used through the broadcasting industry. PSNR effectively measures pixel by pixel differences that are not necessarily visible to the human eye. This is considered to be an objective measurement

- **JND** (Just Noticeable Difference) is a general psychophysical measurement unit defined as: 'The difference between two stimuli that (under properly controlled experimental conditions) is detected as often as it is undetected' [3] to model human perceived video quality. This is considered to be a subjective measurement.

- **DMOS** (Differential Mean Opinion Score) likewise this is a measurement based on the human vision system and is considered to be subjective [3].

Primarily, this paper is considering the best quality and dynamic range that can be achieved during the compression process. It is not expected that there will be any obviously compromising artefacts. Therefore, for the moment PSNR and visual inspection are the sole measurements.

In addition, the PSNR metric may consider chroma if calculated using the following weightings:

$$PSNR = (0.8xY_{PSNR}) + (0.1xU_{PSNR}) + (0.1xV_{PSNR})$$

This is used for all curve comparisons in this paper.

Levels, Profiles and Operating Points

ISO/IEC 14496-10 otherwise known as MPEG-4 AVC defines a number of profiles applicable to video conferencing, broadcast and streaming applications:

- The **Baseline Profile** is mainly intended for video conferencing and streaming to mobile devices. Its simplistic applications mean that it does not support bi-directional frames, interlace or CABAC entropy encoding.
- The **Main Profile** allows bi-directionally predicted (B) frames with two direct modes: spatial and temporal and weighted predictions. Furthermore, it supports all interlace coding tools including Picture Adaptive Field/Frame coding (PAFF) and Macro-Block Adaptive Field/Frame coding (MBAFF) as well as CABAC. Main profile is most widely used for HD and SD broadcasting applications.
- The coding tools of MPEG-4 profiles which go beyond Main Profile are summarised as Fidelity Range Extensions [4]. In particular the **High Profile** allows adaptive 8x8 integer transforms, intra 8x8 predictions modes and scaling lists.
- The **High 10 Profile** allows coding of 4:2:0 video signals with 10 bit accuracy and
- the **High 4:2:2 Profile** allows coding of 4:2:2 video signals with 10 or 8 bit accuracy.

This paper considers two applications for **HDTV** MPEG-4 AVC. Here is a description of how they generally operate today:

Direct-To-Home

Typically direct-to-home satellite broadcasters (DTH) use **Main Profile** (i.e. 4:2:0 8bit video) at bitrates ranging from 6-20Mbits. Delay is not a critical issue for these applications, therefore a 15Mbit encoding buffer and 3x B pictures are considered quite acceptable.

Contribution and Distribution

Contribution and Distribution applications are varied and cannot be generalised. However, there is a common characteristic; in that end-to-end delay is critical. Therefore the number of B pictures and encoding buffer size is kept to a minimum. Bitrates can vary in the range of 8-90Mbits depending on the quality point that is required. In addition, **Hi 422 Profile** is used to achieve 4:2:2 8 bit video compression.

In both cases, 10bit processing is explored to demonstrate the gains that maybe achieved.

The importance of 4:2:2 video compression

Traditionally C&D applications have demanded 4:2:2 MPEG-2 video compression. This prevents the need to repeatedly down-sample and up-sample the chroma channels. As mis-matches in spatial positioning of the chroma filters and/or soft filter roll-offs can significantly degrade the chroma quality.

In addition, repeated concatenation of video encoding processes can impede compression performance through the later stages in the chain. Therefore, preserving optimum video quality is desirable from the very first stage.

Furthermore, 10 bit 4:2:2 is the most commonly used studio and production format. Therefore, from a compression efficiency prospective, it is important to know at what bit rate the picture quality of 10 bit 4:2:2 would be noticeably better than 8 bit 4:2:2.

Comparison of 8 and 10 bit encoding for DTH applications

Main Profile is compared with High 10 Profile for a set of sequences. Both use the following common tools:

- CABAC
- Inter prediction modes 16x16, 16x8, 8x16, 8x8
- Intra prediction modes 16x16, 4x4 (8x8 in Hi10P)
- 3x B pictures
- Field coding
- De-blocking filter on
- RDO (Rate Distortion Optimisation)

A software model is used to compress the sequences, which represents a viable realtime encoder.

The sequences selected are reasonably clean from noise and have large plain areas as well as detailed areas. They contain fade-to-black or subtle movement which can cause contouring or posterisation in the plain regions.

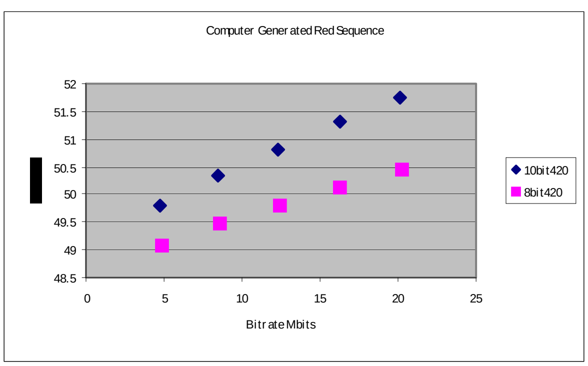

Fig. 2 a computer generated red landscape that subtly changes.

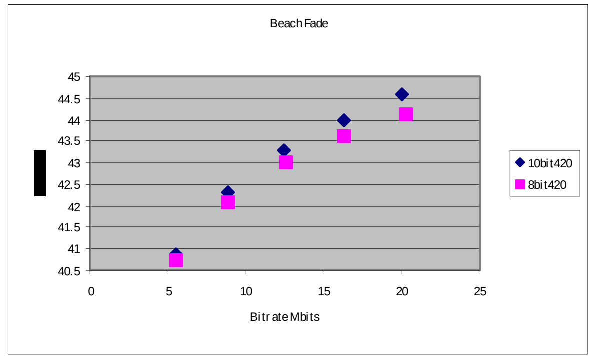

Fig. 3 a beach scene with blue sky, running horses and fade to black.

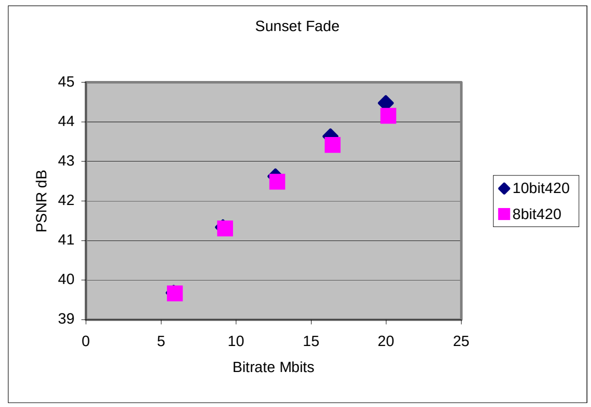

Fig. 4 a sunset scene with palm trees and fade to black.

Clearly all three fig. 2,3,4 show an improvement in terms of PSNR. But also there are subjective improvements which are visible on LCD and CRT screens. The differences are within the plain areas and contain less blocking or stripe artefacts in the 10bit version compared to the 8bit.

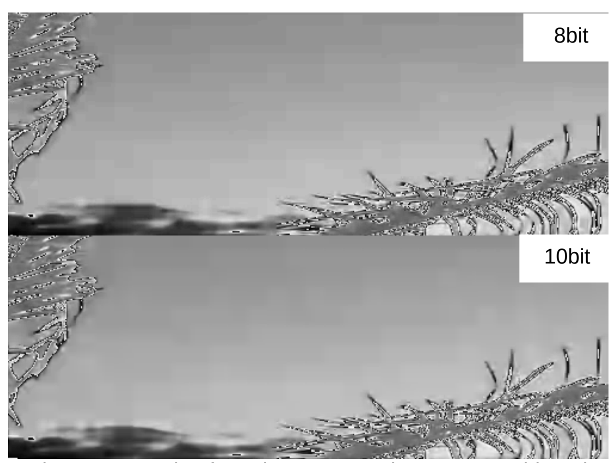

Fig. 5 a x4 section from the SunsetFade sequence 8bit and 10bit.

In order to appreciate the visual differences, fig. 5 shows an amplified (x4) section of the sky from the 9Mbit point of the SunsetFade sequence. The macroblock boundaries are much more noticeable in the 8bit version.

Further inspection reveals that there is only a luma level difference of 1 or 2 (in 8 bit terms) between neighbouring blocks, but because the entire macroblock is so plain the eye is very sensitive to this change. When this data is presented to the in-loop de-blocking filter, there is nothing the filter can do to hide this boundary. However, in the 10 bit case, the in loop de-blocking filter has two extra bits of precision, therefore intermediate levels can be created and hence soften the macroblock boundary. This can be verified by switching the de-blocking filter off in the 10 bit experiment.

It is important to note that the difference between 8 bit and 10 bit pixel precision is not noticeable when comparing the source. The justification for this is that 8 bit source images contain enough noise to hide any contouring which might otherwise be visible in such areas (i.e. a dithering effect). However, the inherent noise reducing properties of video compression can expose the luma and colour graduations and lead to a posterisation effect, even at relatively high bit rates.

Interestingly, there are two mechanisms in MPEG-2 which unintentionally help to hide such artefacts: DCT inaccuracy and mismatch control. Mismatch control is required as a result of DCT inaccuracy. The combined effect of these two mechanisms is to inject a small amount of noise at the output of MPEG-2 decoders. This noise acts like a dither signal. Therefore, although MPEG-2, like most compression algorithms, eliminates some of the original source noise due to the low-pass characteristics of its weighting matrices, it reintroduces enough DCT noise at the decoder output to reduce posterisation.

The disadvantage of MPEG-2 DCT inaccuracies is, of course, decoder drift. This limits the number of predictions that can be made from previous predictions before a refresh is required, i.e. it limits how many P frames can be coded between I frames. MPEG-4 AVC does not suffer from DCT inaccuracies because its integer transformations have no rounding differences between encoder and decoder. Therefore, it has no need for mismatch control and there is no limit to the number of predictions that can be made from previous predictions.

Finally, common to all three sequences, the PSNR difference (8bit vs. 10bit) approaches >0.25 dB when the average encoding QP for the sequence is in the order of <20.

Comparison of 8 and 10 bit encoding for C&D applications

High 4:2:2 profile 8 bit is compared with High 4:2:2 profile 10bit for a set of sequences. Both use the following common tools:

- CABAC
- Inter prediction modes 16x16, 16x8, 8x16, 8x8
- Intra prediction modes 16x16, 8x8, 4x4
- 1x B pictures
- Field coding
- De-blocking filter on
- RDO (Rate Distortion Optimisation)

The first sequence is a 10bit CrowdRun scene provided by SVT [5]. It contains an area of high motion, and a plain blue sky. In addition, the sequence is relatively noisy. The second sequence is 10bit cricket scene with grass and a plain sky, this is relatively free from noise.

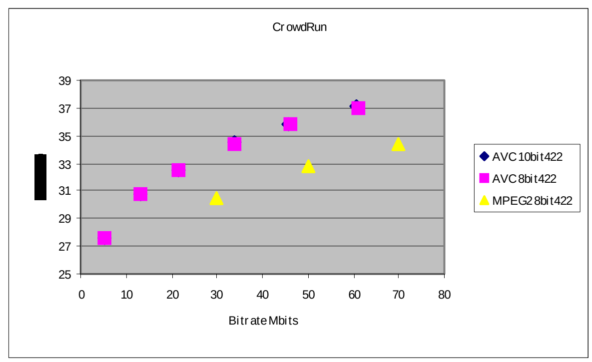

Fig. 6 a crowd of people running in a marathon, with a blue sky.

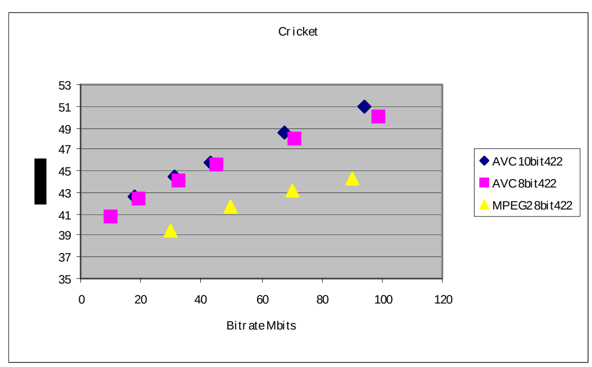

Fig. 7 a cricket scene with a plain sky.

The CrowdRun sequence in Fig. 6 shows little gain using 10bit precision. Of course, the PSNR measurement used is including the likeness of the noise between source and compressed sequences.

As discussed previously, the significance of noise in the source material is important. A) it provides a dither in the plain areas that may hide step changes in luma. B) the transform and quantisation process used in the MPEG-4 AVC algorithm removes noise at higher QPs.

Based on these points, there is no need to have 10bit precision in the de-blocking filter, and there is enough dither in the picture. Secondly, the QP would have to go very low (i.e. high bitrate) in order to preserve the exact noise pattern in the source and hence not showing the benefit of 10 precision in terms of PSNR (for the given bitrate operating range). It is quite noticeable visually that across the bitrate range, the MPEG-4 AVC sequences have less noise than the source (at 60Mbits the average QP is 23).

The MPEG-2 curve is added for comparison and is significantly under the MPEG-4 AVC curves. Visually, there are some MPEG-2 compression artefacts in the high motion area. Fig. 8 shows an example of processing noise at 50Mbits.

Fig. 8 artefact comparison between MPEG-2 and MPEG-4 AVC at 50Mbits.

Conversely the cricket sequence in Fig. 7 is similar to the sequences used for the DTH experiments. The difference in performance can be measured at bitrates within the operating range. In PSNR terms, MPEG-4 AVC 10bit yields better performance than MPEG-2 and MPEG-4 AVC 8bit.

In summary, the advantage of using 10bit compression in C&D applications, is that: when noise is minimal, there is opportunity for 10bit precision to be preserved during the compression process. This will be useful for downstream processing (e.g. gamma or colour correction) where luma/chroma levels are changed in a non-linear fashion Hence, the extra precision could potentially mask any perceivable artefacts.

LCD Screen technology

Currently, most LCD screens available are 8bit resolution. However, several leading consumer electronics manufactures are producing 10bit capable technology.

The DTH sequences discussed in this paper were also viewed on a Sony 8bit LCD screen, and there is still a perceived difference between 8bit and 10bit MPEG-4 AVC compression. It can only be assumed that whist the final display is 8bits, intermediate processing (i.e. non linear gamma and colour correction) are on the full 10bit resolution. Hence, avoiding large step changes in 8bit terms.

In addition, 10bit LCD screen owners may further benefit from 10bit broadcast Television transmission.

Conclusions

In this paper, advantages of 10bit MPEG-4 AVC for C&D and DTH applications have been illustrated.

In the case of C&D, it has been shown that the High 4:2:2 Profile of MPEG-4 AVC can deliver a level of picture quality (within a practical bitrate range) that could not have been achieved in MPEG-2 (since MPEG-2 does not support 10 bit coding) and MPEG-4 AVC 8bit. Also, at the points where 10bit coding does not offer any advantage, there is not a degradation in compression performance. This makes High 4:2:2 10bit profile an attractive and safe solution.

In the case of DTH, it is unlikely that the bitrate ranges used will show a PSNR difference between 8bit and 10bit precision. However, the extra precision that 10bit offers gives more scope for the de-blocking filter to hide macroblock and transform boundary processing.

Further work maybe to investigate if it is possible to manipulate the MPEG-4 AVC encoding process to choose the appropriate coding mode, QP and/or transform - that maximises the amount of noise (i.e. dither) that gets through the compression process for a given macroblock.

References

[1] T. Wiegand, et al, 'Overview of the H.264 AVC Video Coding Standard', IEEE Transactions on Circuits and Systems for Video Technology, July 2003.

[2] M. Goldman, A.M. Bock, 'Advanced Compression Technologies for High Definition', IEE Conference 'IT to HD', London, Nov. 2005.

[3] K. Ferguson, 'An Adaptable Human Vision Model for Subjective Video Quality Rating Prediction Among CIF, SD, HD and E-Cinema', International Broadcasting Convention, Amsterdam, Sept. 2007.

[4] D. Marpe, et al, 'H.264/MPEG-4 AVC Fidelity Range Extensions: Tools, Profiles, Performance, and Application Areas', Proc. ICIP, Genova, Sept. 2005.

[5] L. Haglund. (2006, The SVT high definition multi format test set). Swedish Television, Stockholm.

Acknowledgements

Many thanks to Alois Bock and Greg Baggott for assisting with the preparation and analysis of data for this paper.