Using AVC/H.264 and H.265 expertise to boost MPEG-2 efficiency and make the 6-in-6 concept a reality

A Technical Paper prepared for the Society of Cable Telecommunications Engineers
By

Anais Painchault
Field Application Engineer
ATEME
3400 Coral Way, suite 303, Miami FL 33129
305 433 4244
a.painchault@ateme.com

Overview

After its standardization in the mid 90s, the MPEG-2/H.262 video compression format rapidly achieved great market success, enabling the digitalization of TV Services around the world, including satellite, cable, IPTV and terrestrial. It became widely deployed and video service providers achieved a huge installed base of subscribers using MPEG-2 decoders.

However, video compression continued to evolve over the years and when MPEG-4 AVC/H.264 was introduced a few years later it gradually stole MPEG-2's thunder, allowing video service providers to enhance the distribution of content.

For the past decade, the industry players have been challenged to improve the efficiency and use of this new codec to address the new video distribution landscape and gain market shares: squeezing more channels per MHz; migrating the services from SD to HD with either the same bandwidth or a better video quality; increasing the user experience for the existing services.

In order to stay competitive, new approaches and optimizations have emerged and the different codec vendors have been forced to redesign their implementations several times. Today, their encoding expertise and creativity is once again challenged with the introduction of HEVC/H. as this involves new research tools and approaches.

However because of the millions of Set top boxes and TVs deployed in the field, no migration from MPEG-2 to any other technology can be considered in the short-term. At the same time, service providers still deeply need to improve the compression efficiency to address the growth of services and the higher video resolutions, or to free up some QAMs to continue their migration to more recent compression technologies.

The goal of this presentation is to show how some of the expertise gained from AVC/H.264 and H.265 can be re-injected into MPEG-2 without impacting the legacy receivers; and how these new developments can both improve the video quality, expand the number of video channels and extend the useful life of older MPEG-2 networks, most notably by the ability to boost the number of video channels for today's general standard of two/three/ HD channels per QAM all the way to five and potentially six HD channels per QAM.

Contents

Video codecs overview and history

MPEG-2/H.262 video standard was published by the ISO/IEC and the ITU in 1994 just before the apparition of the first digital television standards, with for example the introduction of DVB-C in 1994 and ATSC in 1998. As a consequence, MPEG-2 video codec was the state-of-the-art when the service providers moved to digital and became the *de facto* video codec for digital television world-wide. As a result, service providers deployed and maintained tens of millions of digital set-top boxes (STBs) and/or integrated digital televisions (DTVs) to allow their subscribers to receive and decode television broadcasts in the field

MPEG-4 AVC/H.264 followed a few years later, with the first version of the standard jointly developed by the ISO/IEC and the ITU in 2003. MPEG-4 AVC was designed to provide high efficiency compression. Even if this new codec adds a more comprehensive and complex tool kit for compression as shown in Table 1, the overall concept remains the same and it absorbs most of the features and approaches present in MPEG-2.

	MPEG-2	H.264/AVC	HEVC
Partition size	Macroblock 16x16	Macroblock 16x16	Coding Unit 8x8 to 64x64
Partitioning	Inter 16x8, Intra 8x8	Sub-block down to 4x4	Prediction Unit Quadtree down to 4x4 Square, symmetric and asymmetric (only square for intra)
Transform	Floating DCT	Integer DCT 8x8, 4x4	Transform Unit square IDCT from 32x32 to 4x4 + DST Luma Intra 4x4
Intra prediction	DC predictor	Up to 9 predictors	35 predictors
Motion prediction	Vector from one Neighbor	Spatial Median (3 blocks)	Advanced Motion Vector Prediction AMVP (spatial + temporal)
« Motion-copy »	/	Direct mode	Merge mode
Motion precision	½ Pixel bilinear	½ Pixel 6-tap, ¼ Pixel bilinear	¼ Pixel 7or 8-tap, 1/8 Pixel 4-tap chroma
Entropy coding	VLC	CABAC, CAVLC	CABAC
Filters	/	Deblocking Filter	Deblocking Filter Sample Adaptive Offset

Table 1: MPEG-2/MPEG-4/HEVC toolsets differences

Compared to MPEG-2 side-by-side, MPEG-4 shows dramatic video quality gain and the use of the codec allows important bandwidth saving – between 30% and 60%, depending on the content and the resolution.

The first hybrid STBs decoding both MPEG-2 and MPEG-4 started to be manufactured in the years 2005/2006 and gradually deployed the following years. However, even today millions of MPEG-2 STBs are still present on the field, making it cost prohibitive for service providers to stop their MPEG-2 distribution and completely migrate to a newer technology. As shown in Fig.1 the majority of channels are still distributed using MPEG-2 compression today.

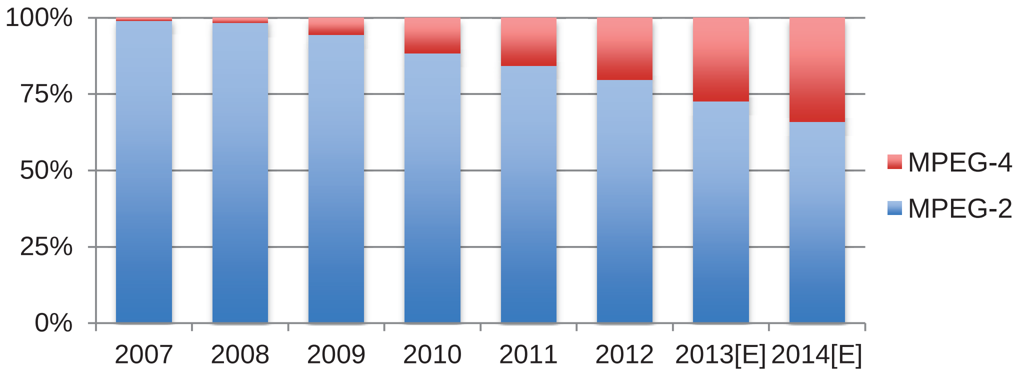


Figure 1: Video Encoders for Cable TV Head-ends in Americas

H.265, also known as HEVC (High Efficiency Video Coding) is the newly born addition to the list of video codecs and promises a compression efficiency gain of around 50% compared to its predecessor MPEG-4. Once again this considerable improvement is achieved by introducing more complex tools in the coding scheme as shown in Table 1, but the codec itself is absorbing most of MPEG-4 and MPEG-2 features.

Digital cable overview – the 6-in-6 concept

Today in the United States, digital cable networks as operated by the cable Multi System Operators (MSOs) are based on quadrature amplitude modulation (QAM) systems comprised of 6MHz channels. HDTV-capable cable set-top boxes operate with 256QAM modulation which allows approximately 38.8Mb/s of bandwidth in a 6MHz channel. Traditionally, using MPEG-2 compression the operators are able to fit 2 to 3 HD channels per 6MHz QAM, allowing a bandwidth between 12 to 18 Mb/s per video.

However, the pressure from the market to make better use of their finite bandwidth for more TV and more broadband services is forcing the service providers to free up additional bandwidth to address this changing distribution landscape: they need to adapt to the always increasing number of HD channels, to address more and more VOD/OTT services and to have the ability to free up some QAM infrastructure to continue their migration to newest technologies (MPEG-4 or even HEVC in the future).

The 6-in-6 (5-in-6) notion refers to the concept of being able to put six (five) 1080i HD MPEG-2 channels in a 6MHz QAM. This concept is extremely attractive in terms of bandwidth savings, however it is very challenging in terms of video compression. As shown on Fig.2, when taking into consideration audio and data tracks, it leaves only 5.7 Mb/s of bandwidth per video track, which means a very high compression ratio when compared to the bitrates used today in the cable industry.

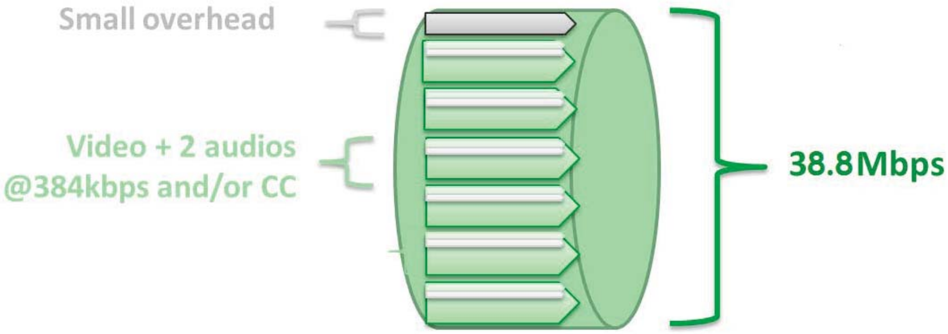


Figure 2: bandwidth requirements for 6-in-6

When considering the average number of bytes available to transmit in each macroblock (as shown in Formula 1) we can see that each macroblock needs to be coded in less than 3 bytes. Given that one macroblock in 4:2:0 is 384 bytes big, that means we need to apply a colossal compression ratio of 130:1. We're in a scenario where every bit counts; and no matter how the performance of the encoding is, the pixel-by-pixel comparison between the original block and the encoded one will always hugely differ. In other words, we are in a very unusual scenario for video compression where the Peak signal-to-noise ratio (PSNR) values will always be quite low and objective measurements quality tools will not be very relevant. In this approach, what needs to be improved is the overall visual image coherence.

$$\frac{\overset{\text{Number of macroblocks}}{\boxed{\frac{1920\times 1080}{256}}}\times\overset{\text{Per second}}{\boxed{29.97\,fps}}}{5.7\,Mbps}\approx 23\ bits\ per\ macroblock$$

Formula 1: average bandwidth available per macroblock in 6-in-6 scenario

Preliminary considerations to maximize bandwidth savings

Given this very challenging scenario where every bit counts it’s important to keep in mind that to achieve such a big compression ratio all the available tools need to be used. This is also true for the MPEG-2 encoding tools such as inverse telecine, PAFF, adaptive GOP structure aligned on scene cut, variable number of B frames to adapt to the content and long GOPs. While the use of some of these tools is restricted by CableLabs specifications, the operators need to carefully investigate what they can and cannot use within the extended MPEG-2 toolset to determine whether 6-in-6 is a viable objective within their design and quality constraints.

Without the complete toolset, maybe only a certain type of less challenging content (like movies) might be eligible for a 6-in-6 model, but more difficult content like live sports and high motion dancing/singing shows might be too demanding. Going for 5-in-6 could be the solution in this case.

The quality of the source channels will also have a direct impact on the level of quality that can be achieved after the compression stage. The sources need be delivered with pristine video quality otherwise the artifacts they carry will be added to the ones created by the new compression stage. Considering the use of better compression technologies such as 4:2:2 10 bits at a primary distribution stage could help increasing the overall quality of the channels.

Another system tool that can bring significant help is the use of statistical multiplexing. Statistical multiplexing allows taking into account the complex variations between successive picture frames to optimize quality within a given bandwidth. For example; a fast moving touchdown during a sporting event will certainly be more complex to encode than a static talking head reporter, and require more bandwidth to maintain a given picture quality. That means that the bandwidth required by each channel to maintain a given level of quality is constantly varying (Fig. 3). This creates an opportunity to provide bandwidth optimization across the multiple video channels by re-allocating bandwidth not required by one channel at a given time to another one which at that instantaneous moment is displaying more complex pictures with bigger differences between successive frames and simply requires more of the bandwidth as illustrated in figure 4 [1].

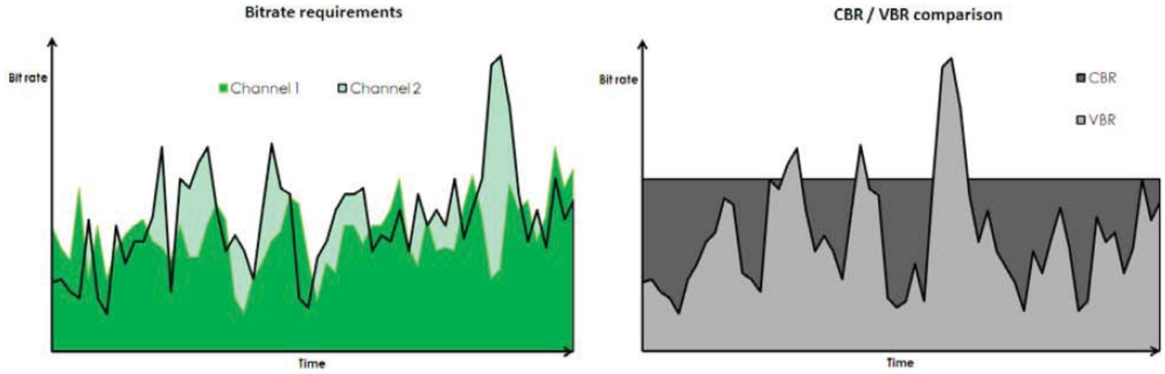


Figure 3: CBR vs. VBR bandwidth requirements

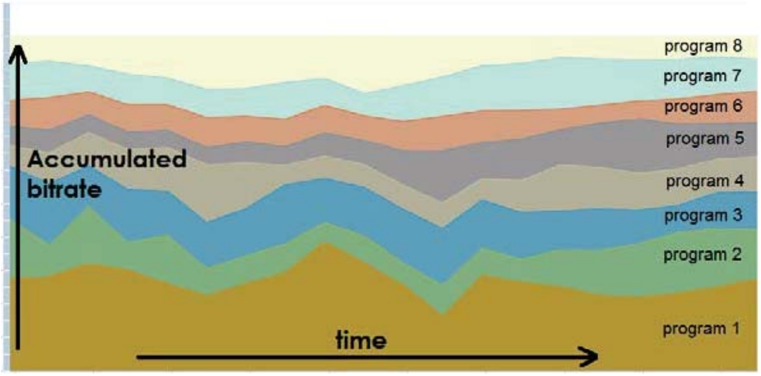


Figure 4: Example of 8 channels statistically multiplexed into a CBR pipe

Today in 2-in-6 / 3-in-6 architecture the channels are usually encoded using Constant Bit Rate (CBR) rate control because statistical multiplexing will not provide much of an improvement on a small number of channels. However, when considering multiplexing 6 channels together at the low bitrates required by 6-in-6, the bandwidth savings can be quite significant: the tests we have run in the ATEME lab have shown a 30% quality improvement when using statistical multiplexing compared to CBR, which relieves some of the pressure on the codec efficiency optimization.

An additional interesting aspect that should be taken into consideration is the impact of resolution on bandwidth. In the broadcasting world some content can be quite challenging with no temporal stability. For example the use of flashes, light effects or water splashing make the number of coefficients and vectors to transmit extremely important. In this situation 23 bits per macroblock might not be sufficient to ensure a satisfying visual quality level.

An interesting alternative would be to use 1440x1080 resolution instead of the usual 1920x1080 resolution for HD. The visual difference for the viewer is imperceptible, and sources from HDCAM and even some live events already come in at 1440x1080 already. By reducing the resolution in this fashion, the number of macroblocks to transmit drops by 30%, which allows significant additional bandwidth for the video codec to work with. The scope of this paper is not to tell what should or should not be done by the operators because they might be confronted to legal restrictions, but this is one choice that could be made to big major improvements on the technical and visual domains.

Enhanced Motion Estimation

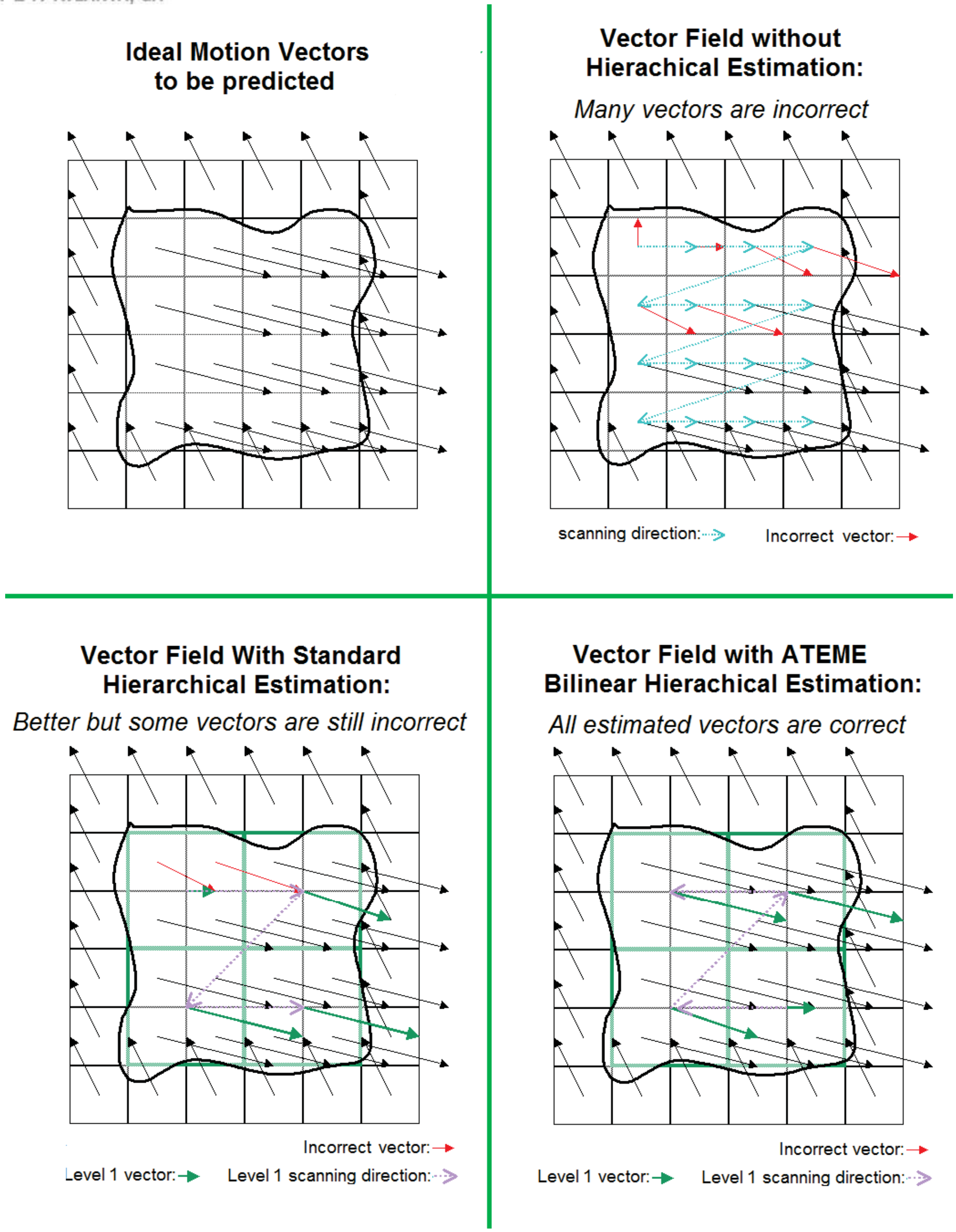
Motion estimation is not a new topic as it appeared on the scene with MPEG-2 almost two decades ago. However, thanks to the introduction of new codecs such as MPEG-4 and HEVC, there is still a lot that can be done and the optimization process is not over yet. Because of the heavy partitioning introduced in MPEG-4, a hierarchical bilinear approach had been developed by ATEME to optimize the motion estimation process.

Motion estimation consists in trying to predict the movement of a macroblock in the form of a vector by looking at the similarities in adjacent images. A very common technique is to reduce the search window by focusing on the most probable motion vectors. These vectors are usually determined by looking at the neighboring blocks that already have an estimate as well as the vectors that the same block had in the previous images.

Hierarchical motion estimation allows adding another very good candidate and reducing even more the search window at low computing cost. The idea is the same as the previous but applied to sub-partitions. The smallest sub-partition level is estimated first, and the vectors found for this sub-partition can then be re-used as candidates for the lower level. This can be applied recursively until we reach the non-partitioned image.

One of the limitations of this technique is that the raster scanning of the blocks induces some asymmetry in the motion estimation. ATEME has patented an approach that proposes bilinear hierarchical motion estimation, where the direction of the scanning for the motion estimation changes for each hierarchical level. As shown in figure 5, this technique shows a better accuracy for motion vector and improves the video quality significantly.

Re-injecting this approach into MPEG-2 also produces visible quality improvements as some degraded sequences with inaccurate motion vectors show a much better visual coherence because of improved homogeneity. Such a technique will also be re-used for HEVC with the hierarchical Quad-tree approach and might even be optimized again over the coming years to get re-injected in MPEG-2 for even more quality improvements.


Ideal Motion Vectors to be predicted

Vector Field without Hierachical Estimation:

Many vectors are incorrect

Vector Field With Standard Hierarchical Estimation:

Better but some vectors are still incorrect

Vector Field with ATEME Bilinear Hierachical Estimation:

All estimated vectors are correct

Figure 5: Better Vector Prediction with Bilinear Hierarchical Motion Estimation

Rate-Distortion Optimization

The Rate-Distortion Optimization method is not new. It was introduced by G. Sullivan [2] back in 1998 and is used by the AVC/H.264 reference encoder. For this reason, this approach is widely used in MPEG-4 implementations but wasn't applied to MPEG-2 until recently [3] [4].

The goal of rate distortion optimization is to test as many encoding modes as possible and choose the one that gives the better rate-distortion compromise as shown in figure 6. By reusing the very powerful machine that has been developed and optimized over the years for MPEG-4 today and applying it to the much simpler codec that is MPEG-2, and by taking full advantage of the available processing power that can be found today, it is now possible to achieve massive RDO tests where almost 90 modes are tested for each macroblock. Combined with the better motion estimation approach, that means that all the possible modes (INTRA/INTER/SKIP, partitioning, motion vectors…) can be tested before choosing the most efficient one (see Fig.6).

This massive RDO allows achieving quite significant quality gains on MPEG-2 at low bitrates as shown by the results obtained on whale show sequence on Fig.7.

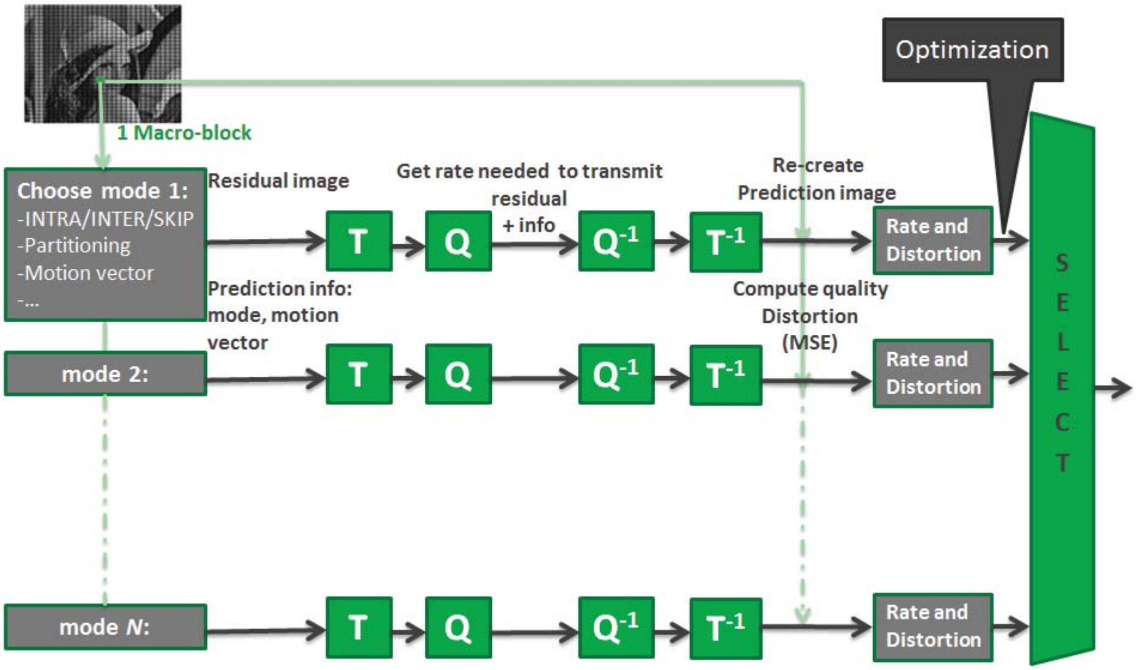


Figure 6: Rate Distortion Optimization steps

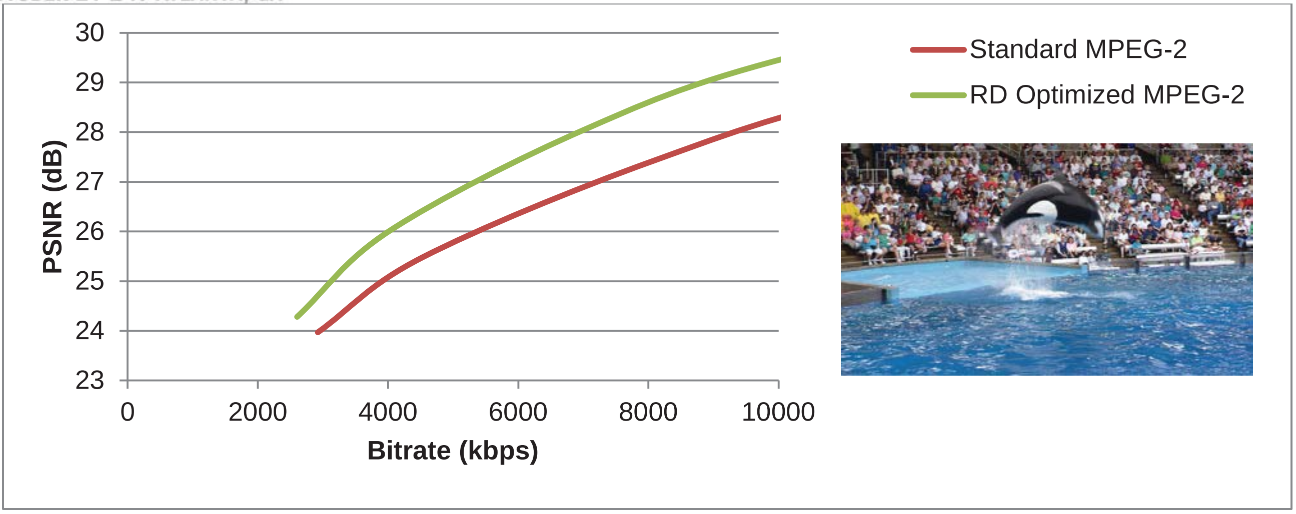


Figure 7: RD-Optimized MPEG-2 quality improvements on Whale Show sequence

Smart use of SKIP mode

In the very challenging context of 6-in-6 where every bit counts, the only solution is to try to use as much as possible modes with low data rate. A SKIP block is very economical in terms of bandwidth utilization because neither motion vector nor residual is sent to the decoder. The only information sent is the mode itself telling the decoder to reuse the block from the previous image.

It's quite easy to understand why such a mode would perform extremely well in a RDO algorithm and be elected quite often. However an over-use of SKIP blocks can create a very bad visual rendering not taken into account during RDO. As a matter of fact, when SKIP mode is too frequently used some blocks don't refresh often enough and seem to be anchored in the image. This creates an artifact that can be detected very easily by the viewer. That means that on top of RDO, SKIP mode needs special attention to be selected carefully.

This concern also affects MPEG-4 and over the years new techniques were developed to validate the use of SKIP modes on other criteria than RDO. The main idea is that the motion vector needs to be coherent with the neighboring blocks vectors. When applying these new techniques to MPEG-4 they show an important visual quality gain.

Re-applying this approach that improved MPEG-4, we can also enhance the performance of MPEG-2 quite significantly.

Backtracking

Techniques that are designed to optimize the decision process between the available encoding tools like RDO rely on the postulate that all coding decisions are independent. Those methods implicitly consider that a coding decision at one macroblock level will not affect the decisions on surrounding macroblocks. Thus determining which tool to use can be done independently, ignoring any spatial or temporal context. Unfortunately, this is not true and a higher efficiency can be achieved by taking coding dependencies into account.

Treillis quantization is a good illustration of such an algorithm. It allows computing quantized coefficients considering not only their own values, but the context of the whole block. This technique is used in the HEVC reference software, and it shows a bitrate gain of over 10% in some cases. This method can be extrapolated to optimize many of the encoding decisions like picture types (I, P or B), macroblock types (inter, intra or SKIP) or even quantizers.

The idea of Backtracking consists in evaluating the consequences of a coding decision not only on the element itself (coefficient, macroblock, picture), but also on the elements that have not been coded yet. The ultimate goal is to optimize a criterion not on a single element but on the full context. Therefore, the full set of elements has to be evaluated to make a single decision. This is similar to finding an optimal path in a graph, and backtracking is shown to be an efficient algorithm to solve this problem.

While also not new, these techniques are extremely challenging when it comes to putting them into practice in a video encoder because of their very high computing power requirements. For this reason, it's only very recently that simplified versions were implemented in commercial AVC/H.264 software encoders; but they are still out of reach for hardware encoders.

Compared to AVC/H.264 or HEVC, MPEG-2 is a simpler standard with coding tools that require significantly less operations than those used in their newer counterparts. Consequently, leveraging the full benefit of backtracking techniques into MPEG-2 can be envisioned. Figure 8 shows that about 30% bitrate saving could be obtained, on top of an extensive RDO decision algorithm.

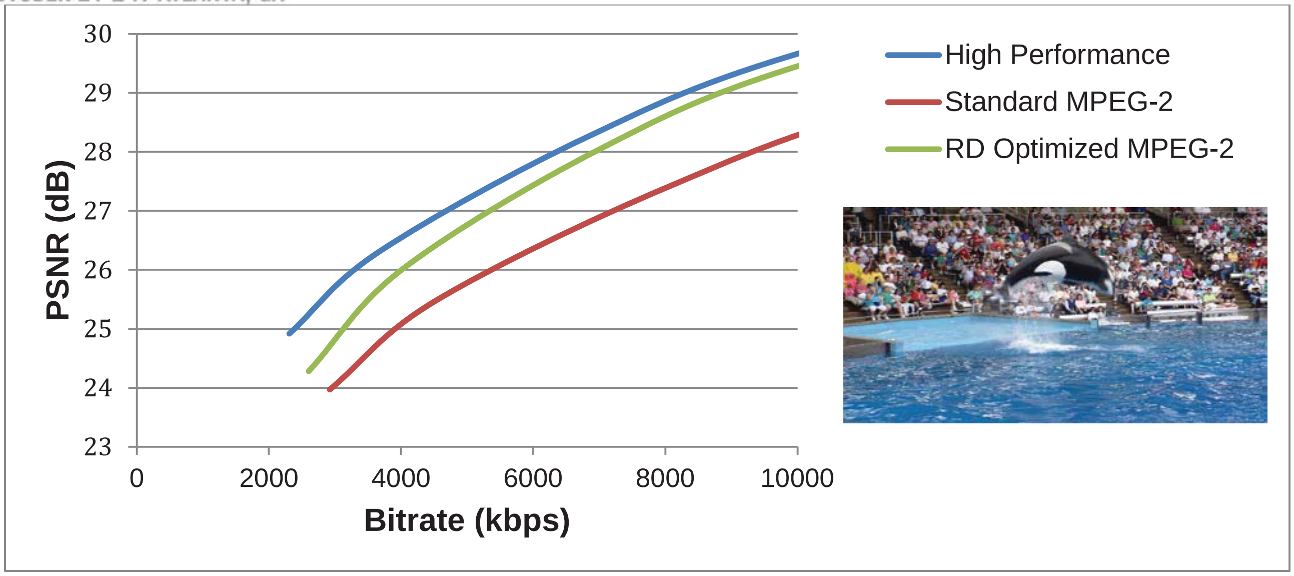


Figure 8: High Performance MPEG-2 including backtracking and all previous approaches quality improvements compared to standard MPEG-2 implementations on Whale Show sequence

This last approach is decisive and is the key to bridging the gap that could allow us to reach the very demanding compression efficiency required for 6-in-6 services.

Conclusion

By re-injecting the expertise and the new approaches introduced by the newer AVC/H.264 and H.265 codecs, it is possible to make substantial improvements in MPEG-2 coding efficiency. Combining all these beneficial techniques together the bandwidth efficiency is very promising, and most especially to achieve the "El Dorado", "Holy Grail" or the "Promised Land" of six HD channel in 6 MHz QAM systems.

Bibliography

[1] “Introducing SmartRate SR1000”, Nicolas Moreau, January 2012
[2] “Rate Distortion Optimization for Video Compression”, G. Sullivan, November 1998
[3] “It’s Not Dead Yet! - MPEG-2 Video Coding Efficiency Improvements”, Matthew S. Goldman, April 2009
[4] “Doubling the Number of ATSC HD Channels” - J.L Diascorn April 2010

Abbreviations and Acronyms

ATSC	Advanced Television Systems Committee standards
AVC	Advanced Video Coding
CBR	Constant Bit Rate
DTV	Digital Television
DVB	Digital Video Broadcasting
DVB-C	Digital Video Broadcasting for Cable
GOP	Group Of Pictures
HEVC	High Efficiency Video Coding
HD	High Definition
IEC	International Electro technical Commission
IPTV	Internet Protocol Television
ISO	International Organization for Standardization
ITU	International Telecommunication Union
kb/s	Kilo bit per second
Mb/s	Mega bit per second
MHz	Mega Hertz
MPEG	Moving Picture Experts Group
OTT	Over The Top
PAFF	Picture-adaptive frame-field coding
PSNR	Peak signal-to-noise ratio
QAM	Quadrature Amplitude Modulation
RDO	Rate Distortion Optimization
SD	Standard Definition
STB	Set-Top Box
TV	Television
VBR	Variable Bit Rate
VOD	Video On Demand