THE EMERGING MPEG-5 EVC STANDARD - APPLICATIONS, TECHNOLOGY AND RESULTS

Jonatan Samuelsson[1], Kiho Choi[2], Jianle Chen[3], Dmytro Rusanovskyy[4]

[1]Divideon, [2]Samsung, [3]Futurewei, [4]Qualcomm

ABSTRACT

For more than 30 years, digitalization in the media industry has enabled new services, increased reach of conventional services and brought a continuously improving TV experience to consumers all over the world. One of the most central technology components in digital media distribution and consumption is *compression* and in particular *video compression*, since video data represents by far the largest amount of data in most media services. Over the years, multiple different video coding standards and proprietary codecs have emerged and the compression performance of new generations of codecs is constantly improving. However, the compression efficiency is not the only factor that determines how well a codec is suited for usage in different application areas – and eventually – how widely deployed the codec will become. The MPEG-5 Essential Video Coding (EVC) standard, presented in this paper, is being developed by MPEG using a novel process targeted at addressing business requirements, including licensing, as well as technical requirements, so as to facilitate rapid deployment and widespread use of the standard throughout the media industry.

INTRODUCTION

Development of a standard typically starts with an analysis of the requirements: What are the problems that will be addressed by this standard? What are the desirable technical properties of this standard? How will this standard interact with other components of the eco-system? What are the target applications and use cases?

For video coding standards, the core problem has remained the same over the years: reduce the size of transmitted video data as much as possible while keeping the visual quality as close as possible to the original video. The convention when it comes to video coding standards has been to only define the bitstream format and the decoder (not the encoder). This allows for cross-industry compatibility of the most critical component (decoder) and at the same time it allows for flexibility in the design of the encoding process, for example to meet requirements of latency and availability of computational resources.

Even though the input format to an encoder and the output format from a decoder (uncompressed samples of video data) has remained the same over the years, the applications of different generations of video coding standards has been quite different. MPEG-1 Video [1], completed in 1993, saw great deployment with Video CD. MPEG-2 Video [2], completed in 1995, became immensely popular with the dawn of DVD and digital

TV broadcasting. For MPEG-4 Advanced Video Coding (AVC/H.264) [3], completed in 2003, the primary applications were HD TV and IP based video services such as Over The Top (OTT) Video On Demand (VOD). MPEG-H High Efficiency Video Coding (HEVC/H.265) [4], completed in 2013, has mainly seen deployment for 4K and HDR/WCG video applications.

Since April 2018, MPEG (ISO/IEC JTC 1/SC 29/WG 11) has together with VCEG (ITU-T SG 16/Q 6), worked on a new project; Versatile Video Coding (VVC) which is expected to see deployment in applications providing an immersive video experience such as Virtual Reality streaming, 8K broadcasting and High Dynamic Range (HDR) video.

It should be noted that from a technical perspective, and from a standards point of view, there is nothing preventing old standards being used for new applications. Even though there might be no decoders to support it and there is no level explicitly defined for it, you could very well apply the compression methods from MPEG-1 to 4K video. However, the bandwidth requirements for such a system would go through the roof and it would not be *cost efficient* to run such a service. By using the latest and most advanced video coding standard it is possible to minimize the resources needed for distributing encoded videos and provide as high quality as possible at all different operating points. The net result is a win-win situation for both service providers and consumers – provided that the cost for deploying and licensing the standard itself is reasonable.

The HEVC standard

The HEVC video coding standard was developed in a joint standardization project of MPEG and VCEG with participation of organizations from all across the industry. The standardization phase for the first version of HEVC went on between 2010 and 2013 with 4 physical meetings per year, close to 500 participants at the meetings and with up to (and once even exceeding) 1000 input contributions per meeting. It is difficult to say exactly how many organizations were involved in one way or another, but it should be safe to say that more than 50 companies and research institutes contributed to the development of the HEVC standard. A broad participation has multiple advantages: there is a lot of innovative technology brought forward for consideration, there is a lot of resources to review all the different parts of the standard and there is a lot of organizations backing the standard in the industry and supporting its deployment in different applications. However, one potential downside to such broad participation with many contributors is that the final standard is understood to contain patented technology from a large number of patent holders. Anyone that is looking to implement or use such a standard must ensure that they have the right license to do so. For earlier generations of MPEG video coding standards, a single licensing instance (patent pool) has been available, offering a single license, covering a vast majority of the technology in the standard, with licensing terms publicly available. The situation became different for HEVC, where there are three different patent pools and many patent holders that do not offer their patents through any of the patent pools. Figure 1 shows an illustration of some of the organizations that have declared to hold patents which would be required to be used to implement HEVC.

There may be different opinions regarding to what extent the HEVC licensing situation has affected the deployment of the standard, but it is generally understood that earlier licensing announcements, more consolidated licenses and more transparency related to licensing would have provided HEVC better opportunities for widespread market adoption. The convenor of MPEG described the situation as a “crisis” in one of his blog posts [11], and it should be clear that there are application areas, such as internet video streaming, where adoption of HEVC has been quite modest. For example, the site Encoding.com reports in their Global Media Formats Report for 2019 [12] (six years after the HEVC standard was completed) that only 12% of their encodes use HEVC, while 82% still use the AVC standard from 2003.

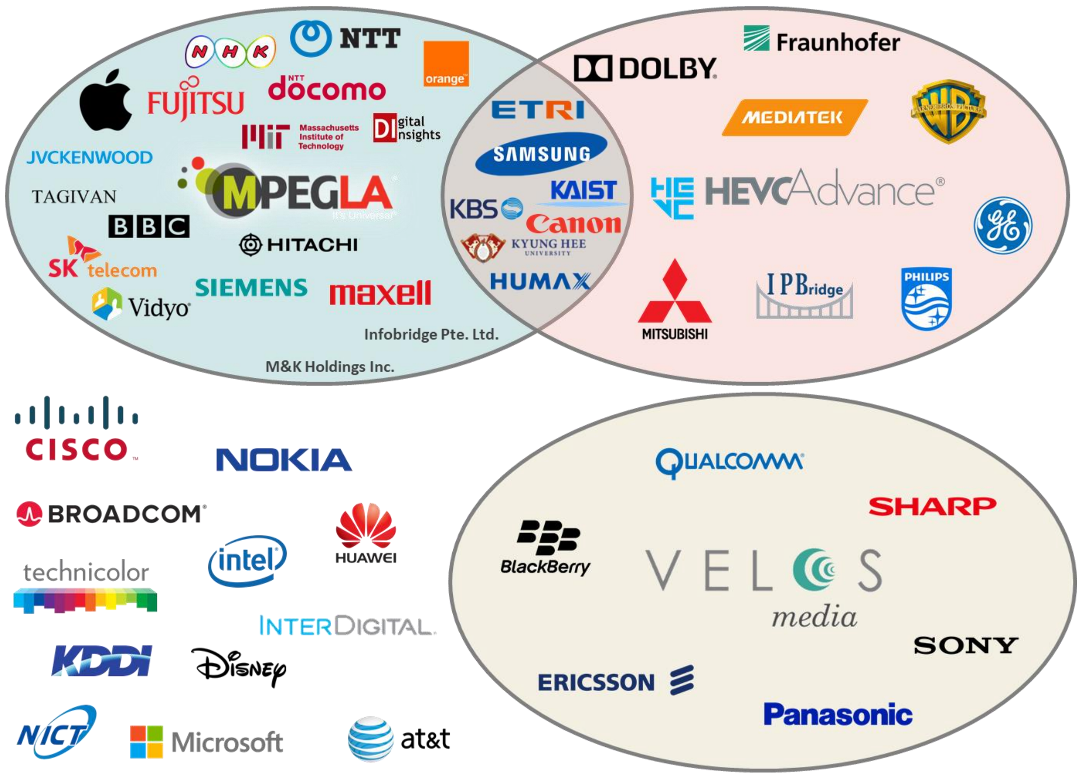

Figure 1 – A subset of the organizations that have declared to hold HEVC essential patents. The figure is based on public information available from the patent databases of ISO [5], IEC [6], and ITU-T [7] and the patent pools of MPEG LA [8], HEVC Advance [9], and Velos Media.

Proprietary video codecs

The primary focus of this paper is on video coding standards, which means specifications for compressed bitstreams and decoding processes, that are defined by international Standards Developing Organizations (SDOs), such as MPEG (part of ISO/IEC). Codecs that have been defined by SDOs typically come with certain properties for example related to FRAND (Fair, Reasonable and Non-Discriminatory) commitments, availability of specification text and reference software, and maintenance.

However, there are other video codecs from individual companies and consortia of companies such as the xvc codec by Divideon [13], RealMediaHD by RealNetworks [14], Perseus by V-Nova [15], and AV1 by Alliance for Open Media (AOM) [16] available on the market. Out of these, the AV1 codec is currently attracting a lot of interest, in part because of its influential backers (AOM founders include Amazon, Cisco, Google, Intel, Microsoft, Mozilla and Netflix) and in part because of its royalty-free claims. The jury is still out on this latter part and it can be mentioned that there have been statements around third-party patents in AV1 [17] and there has even been a licensing program launched for AV1. In general, it remains to be seen to what extent proprietary codecs will be able to fulfil requirements set by different industry groups and gain traction for applications that have traditionally relied on standardized codecs.

BACKGROUND

Conventionally, the process for developing video coding standards in MPEG has been quite straightforward. The standard has been constructed with the technology that brings the best performance in terms of compression efficiency, as long as it is possible to efficiently implement and run in both software and hardware implementations. No considerations have been taken on licensing aspects of the technology other than a requirement for FRAND commitment by the contributors. Commercial aspects, and in particular, licensing aspects have been handled externally and independently of MPEG.

Discussions around a different process to address the concerns related to unclear and complicated licensing of MPEG standards started during 2018 and lead to an attempt to employ a modified standardization process [19] that can be summarized as follows:

1. A standard is defined with a base set of tools that only includes technologies that are more than 20 years old or submitted with a royalty-free declaration.
2. Additional tools that each provide a significant improvement in terms of compression performance are defined on top of the base toolset.
3. Each additional tool is isolated so that it can be switched off independently of other tools.
4. Contributors are encouraged to submit voluntary declarations related to the licensing terms and their publication.
5. A profiling mechanism is defined so as to allow for different tools to be included in different profiles e.g. depending on the licensing declarations.

The defined process is expected to provide much better insight into the ownership of technology, offer a safe set of fallback tools and still provide good compression performance thanks to the additional tools on top. It should be noted that with this process, it is expected that there is no need to explicitly define any particular licensing terms or licensing model and there is no discrimination of technology based on non-technical factors. The profiling mechanisms may be utilized within MPEG or externally by organizations using the standard.

Call for Proposals

The MPEG-5 Essential Video Coding (EVC) project started in January 2019 after having evaluated the responses to the Call for Proposals (CfP) and a first Working Draft and Test Model were defined as a result of this evaluation. Divideon, Samsung, Huawei and Qualcomm all responded to the CfP and are actively contributing to the standardization work, with the four authors of this paper being the MPEG-5 EVC Working Draft editors.

Applications

Based on the process defined above and with a combination of appealing properties – both technical and commercial – it is expected that MPEG-5 EVC can become ubiquitously deployed in wide variety of applications. The requirements document for the project [19] specifically emphasizes the importance of real time encoding for live OTT streaming and offline encoding for streaming VOD but other applications such as video conferencing and traditional broadcasting are also expected to be efficiently supported by MPEG-5 EVC. The standard will support video resolutions of (at least) up to 8K and High Frame Rate (HFR) of at least 120 frames per second. In order to provide highest possible quality on today's and future displays the standard will support High Dynamic Range (HDR) and Wide Colour Gamut (WCG) with 10 bits precision.

TECHNOLOGY

The MPEG-5 EVC draft standard is based on the CfP response by Samsung, Huawei and Qualcomm, with reference picture management and other high-level syntax aspects brought in from the CfP response by Divideon. At the core of MPEG-5 EVC is the Baseline profile that only includes technology that is assessed to be older than 20 years or that has been submitted with a royalty-free license declaration. The additional tools are included in a different profile called Main profile. Figure 2 shows a simplified overview of the MPEG-5 EVC standard with the 20 additional coding or HLS features (compared to the Baseline profile) that each can be individually disabled. The figure and the description in this section is based on the current state of the MPEG-5 EVC project as of July 2019. It should be noted that tools may be added, removed or modified before the final standard is submitted for publication.

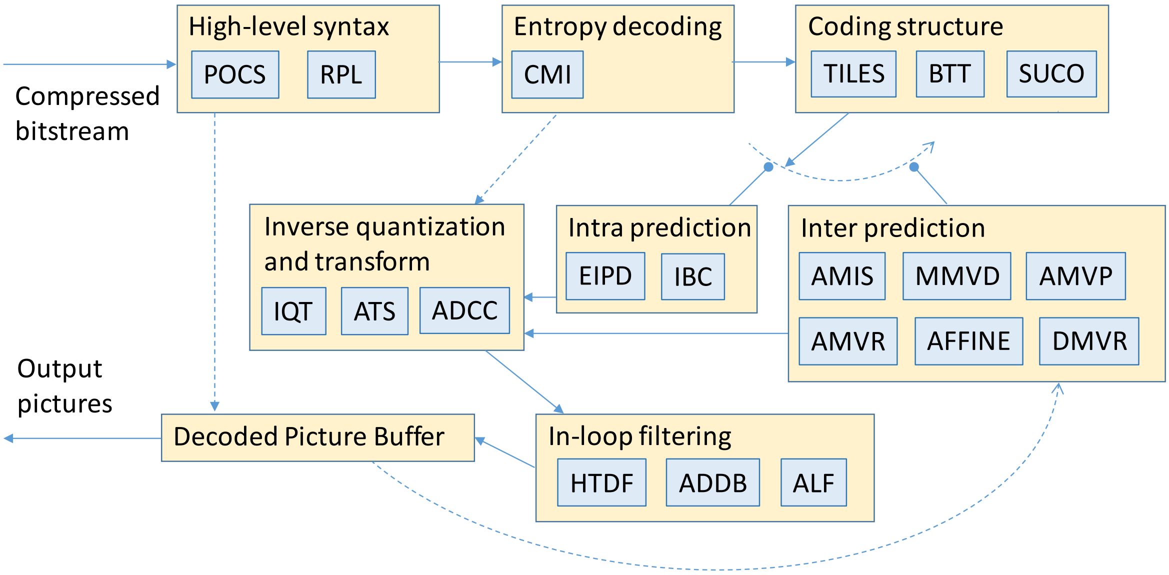

Figure 2 – Overview of the MPEG-5 EVC standard.

High-level syntax

An MPEG-5 EVC bitstream consists of a number of Network Abstraction Layer (NAL) units with a small NAL unit header that indicates some properties of the NAL unit, such as the type of data and the Temporal ID. A Sequence Parameter Set (SPS) contains parameters that apply to an entire Coded Video Sequence (CVS), a Picture Parameter Set (PPS) contains data that applies to one or more pictures of a CVS, and an Adaptation Parameter Set (APS) contains data that applies to one or more parts of one or more pictures of a CVS. In the Baseline profile, picture identifiers and reference picture structures are derived directly from information in the SPS and the NAL unit headers, while the Main profile allows for flexibility in picture identifiers, Picture Order Count Signalling (POCS), and Reference Picture Lists (RPL), signalled at the picture level.

Entropy decoding

MPEG-5 EVC uses the same binary arithmetic coding scheme as JPEG, Annex D [20]. The scheme includes a binarization step and probability update defined by a look-up table. In the Main profile, a Context Modelling and Initialization (CMI) process allows for more efficient modelling of probabilities, through derivation processes based on syntax elements from neighbouring blocks.

Coding structure

At the core of the block structure in MPEG-5 EVC is a conventional quad-tree coding structure, partitioning the coded picture into blocks of up to 128x128 luma samples that can be recursively split into smaller and smaller square-shaped blocks. The Main profile adds two advanced coding structure tools: The Binary Ternary Tree (BTT) which allows for non-square coding units and the Split Unit Coding Order (SUCO) which allows the processing order of split units to change from the conventional left-to-right scanning order processing, to instead use right-to-left scanning order processing in cases where that provides a benefit. In the Main profile, pictures can be split into rectangular TILES that can be encoded independently in parallel.

Intra prediction

Intra prediction in MPEG-5 EVC is performed from neighbouring samples for any coding unit in the split structure. For the Baseline profile, all coding units are square-shaped and five different prediction modes exists; DC (average value of the neighbours), horizontal, vertical, and two different diagonal directions. In the Main profile, prediction can be applied to any rectangular coding unit and 28 additional directional modes are available in what is called Enhanced Intra Prediction Directions (EIPD). In the Main profile it is also possible to use Intra Block Copy (IBC) to reference a block of previously coded samples in the same picture.

Inter prediction

The basis for the inter prediction in MPEG-5 EVC is motion compensation with interpolation filters of quarter sample resolution. In the Baseline profile, motion vectors are signalled using one out of three spatially neighbouring motion vectors and one temporally collocated motion vector as predictor. A motion vector difference may be signalled relative to the selected predictor but a specific mode, called skip mode, exists for the case where no motion vector difference is signalled and there is no residual data in the block. The Main profile includes six additional tools for providing improved inter prediction. With Advanced Motion Interpolation and Signalling (AMIS) it is possible to conceptually merge neighbouring blocks to indicate that they use the same motion, but also to use a more advanced scheme for creating a list of candidate predictors compared to the predictors in the Baseline profile. The Merge with Motion Vector Difference (MMVD) tool uses a process similar to the conceptual merging of neighbouring blocks but additionally allows signalling of a motion vector using an expression that includes a starting point, a motion magnitude, and a motion direction.

With Advanced Motion Vector Prediction (ADMVP), a larger number of candidate motion vector predictors for a block is derived from neighbouring blocks in the same picture as well as from the collocated block in a reference picture. The Adaptive Motion Vector Resolution (AMVR) tool provides a means to reduce the precision of motion vectors from quarter sample, to half sample, full sample, double sample or quadruple sample, which can provide an efficiency benefit for example when signalling large motion vector differences. With the AFFINE prediction mode, it is possible to represent motions that are not purely translational. The AFFINE prediction mode can either use a merge process or be signalled with 3 motion vector differences (for a 6-parameter model) or with 2 motion vector differences (for a 4-parameter model). The Main profile also includes Decoder-side Motion Vector Refinement (DMVR) which uses a bilateral template matching process for refining motion vectors in bi-prediction mode.

Inverse quantization and transform

When a block of data has been predicted using either intra prediction or inter prediction, residual data is generally added to the predicted block (unless there are no coded coefficients). The residual data is acquired by applying an inverse quantization process followed by an inverse transformation. The convention is to call these processes “inverse” in the decoder since they are the inverse of the transformation and quantization processes that are generally applied in the encoder. MPEG-5 EVC includes an integer Discrete Cosine Transform (DCT2) together with a scalar quantization (with a scaling factor acquired from a look-up table). For the Main Profile, the Improved Quantization and Transform (IQT) uses different mapping and clipping functions for quantization to provide better performance. The residual data for a block is scanned in an inverse zig-zag scan order starting from bottom right corner, since the transformed coefficient generally have higher absolute values in the top left corner. The Advanced Coefficient Coding (ADCC) in Main profile enables more efficient signalling of coefficient values for example through a last non-zero coefficient indication. In Main profile it is also possible to use Adaptive Transform Selection (ATS) to apply integer versions of DST7 or DCT8 instead of just DCT2.

In-loop filters

The Baseline profile of MPEG-5 EVC uses a deblocking filter that is defined in H.263 Annex J [21]. In the Main profile it is possible to instead use an Advanced Deblocking Filter (ADDB), which can further reduce blocking artefacts, compare to the deblocking filter in the Baseline profile. The Main profile also defines two additional in-loop filters that can be used to improve the quality of decoded pictures before outputting them and/or using them for inter prediction. The Hadamard Transform Domain Filter (HTDF) is applied to luma samples before deblocking and uses a scanning process to determine 4 neighbouring samples to be used for filtering. The Adaptive Loop Filter (ALF) allows signalling of up to 25 different filters for the luma component and optimal filters can be selected through a classification process for each 4x4 block. Filter parameters for the ALF filter are signalled in the APS data structure.

Decoded Picture Buffer

Previously decoded pictures can be stored in a Decoded Picture Buffer (DPB) to be used for prediction for pictures following them in decoding order. In the Baseline profile, the management of the DPB (i.e. addition and removal of reference pictures) is controlled by information in the SPS. For the Main Profile, the DPB management may be controlled by information signalled at the picture level, if the RPL scheme is used.

RESULTS

The MPEG-5 EVC is being developed with a specific set of test condition and test sequences used in the development process. These conditions are called the Common Test Conditions (CTC) and they provide a means for monitoring the progress of the project and comparing different technical proposals with each other. The reference implementation of HEVC (HM) and AVC (JM) are used as anchors in the CTC to provide an indication for how the MPEG-5 EVC test model compares to existing codecs. Other video coding standards and proprietary codecs can also be compared using the same test conditions.

Table 1 shows comparisons of the MPEG-5 EVC test model in the Main profile configuration against the HEVC anchor (HM16.16). Table 2 shows MPEG-5 EVC in the Baseline profile configuration compared against the AVC anchor. Comparative results in Table 1 and Table 2 are shown for Random Access configurations and for UHD sequences (4K) and HD test sequences (1080p) comprising the EVC CTC. The reported numbers are the Bjontegaard-Delta Rate (BD-Rate) savings (bitrate reduction) for luma (Y) and chroma (U and V) measured using 4 rate points and PSNR as a metric. In short it can be said that BD-Rate gives an indication of how much the bitrate can be reduced while keeping the same objective

	Y	U	V	EncT	DecT
Tango2	-26,62%	-19,60%	-23,47%	394%	249%
FoodMarket4	-26,67%	-17,52%	-18,06%	316%	242%
CatRobot1	-33,40%	-30,32%	-26,46%	461%	238%
DaylightRoad2	-33,40%	-22,56%	-24,55%	485%	227%
ParkRunning3	-26,08%	-16,28%	-15,19%	685%	222%
MarketPlace	-22,18%	-10,19%	-11,66%	509%	208%
RitualDance	-21,30%	-17,33%	-18,78%	481%	198%
Cactus	-24,77%	-20,12%	-16,78%	599%	203%
BasketballDrive	-24,48%	-27,01%	-28,24%	575%	219%
BQTerrace	-20,67%	-16,97%	2,41%	570%	200%
Overall	-25,96%	-19,79%	-18,08%	497%	220%

Table 1 – Results for MPEG-5 EVC Main (ETM) compared to HEVC (HM16.16).

visual quality. The EncT and DecT columns represent relative encoding time and decoding time, respectively.

	Y	U	V	EncT	DecT
Tango2	-48,51%	-31,03%	-37,31%	42%	102%
FoodMarket4	-52,00%	-38,61%	-41,42%	47%	101%
CatRobot1	-31,95%	-34,52%	-41,74%	50%	107%
DaylightRoad2	-38,74%	-31,08%	-43,00%	41%	98%
ParkRunning3	-17,77%	-32,10%	-26,92%	40%	101%
MarketPlace	-30,99%	-25,57%	-27,87%	34%	97%
RitualDance	-23,89%	-20,75%	-20,66%	37%	105%
Cactus	-22,70%	-24,91%	-29,72%	40%	114%
BasketballDrive	-20,90%	-26,44%	-31,36%	37%	103%
BQTerrace	-25,35%	-42,53%	-26,33%	39%	97%
Overall	-31,28%	-30,75%	-32,63%	40%	102%

Table 2 – Results for MPEG-5 EVC Baseline (ETM) compared to AVC (JM19).

As seen from Table 1 and Table 2, MPEG-5 EVC Main profile outperforms the HEVC anchor by around 26% of bitrate reduction, whereas MPEG-5 EVC Baseline profile shows around 31% of bitrate reduction vs. AVC. For the complexity estimates, MPEG-5 EVC Test Model in the Main profile shows around 5x and 2x run time relative to the HEVC anchor for encoder and decoder, respectively. The MPEG-5 EVC Baseline profile shows 60% encoding time reduction with approximately the same decoding time as the AVC anchor. It should be noted, however, that the EVC test model software is currently under development and that further improvements in encoding speed and decoding speed are expected.

ROADMAP FOR MPEG-5 EVC

The standards developed by MPEG are produced and published according to the processes defined for ISO (International Organization for Standardization) and includes several stages. Moving from one stage to the next involves approval by a certain number of National Bodies (member countries) and the National Bodies may provide comments and send feedback back to the group developing the standard. For the MPEG-5 EVC standard, the plan is to finalize the technical elements in the standard to be sent out for approval in October 2019 (DIS stage) and that the Final Draft International Standard (FDIS) will be completed in April 2020 and thereafter published.

CONCLUSION

This paper has presented the new video coding standard MPEG-5 Essential Video Coding (EVC), that is currently under development in MPEG. The MPEG-5 EVC standard is developed using a novel process that has been designed to provide increased insight into the ownership of the technology included in the standard and provide a better foundation for licensing of the codec. The draft standard is built on technology submitted by Samsung, Huawei, Qualcomm and Divideon, and includes a Baseline profile only consisting of technology available under royalty free terms. All the technology components included in the Main profile can be individually turned off, thereby providing users of the codec with the ability to avoid using technology that they are unable to license. The results presented in this paper shows that the MPEG-5 EVC standard can reach the same video quality as HEVC at 26% lower bitrate on average, over a set of UHD and HD sequences. Combined with the licensing aspects, these results put MPEG-5 EVC in a good position for widespread adoption, in particular for applications and services that are still based on the AVC standard. The MPEG-5 EVC standard is expected to be finalized and published in 2020.

REFERENCES

[1] ISO/IEC 11172-2: Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s: Video

[2] ISO/IEC 13818-2 / ITU-T H.262: Generic coding of moving pictures and associated audio information: Video

[3] ISO/IEC 14496-10 / ITU-T H.264: Advanced Video Coding for generic audiovisual services

[4] ISO/IEC 23008-2 / ITU-T H.265: High Efficiency Video Coding

[5] ISO Standards and Patents site https://www.iso.org/iso-standards-and-patents.html

[6] IEC patents database http://patents.iec.ch/

[7] IPR in ITU Recommendations site http://www.itu.int/net4/ipr/search.aspx

[8] MPEG LA HEVC http://www.mpegla.com/main/programs/HEVC/Pages/Intro.aspx

[9] HEVC Advance https://www.hevcadvance.com/

[10] Velos Media http://velosmedia.com/

[11] Chiariglione, L. “A crisis, the causes and a solution” http://blog.chiariglione.org/a-crisis-the-causes-and-a-solution/

[12] Encoding.com, Global media formats report https://1yy04i3k9fyt3vqjsf2mv610yvm-wpengine.netdna-ssl.com/files/2019-Global-Media-Formats-Report.pdf

[13] The xvc video codec by Divideon https://xvc.io/

[14] RealMedia HD by RealNetworks https://www.realnetworks.com/realmediaHD

[15] Perseus by V-Nova https://www.v-nova.com/perseus-video-compression-technology/

[16] AV1 by Alliance for Open Media https://aomedia.org/

[17] Velos Media Q & A: “Why pay royalties on HEVC when other compression technologies like VP9 and AV1 are royalty-free?” http://velosmedia.com/technology/q-and-a/

[18] Sisvel Video Coding Licensing Platform, http://www.sisvel.com/licensing-programs/audio-and-video-coding-decoding/video-coding-platform/introduction

[19] Requirements for a new Video Coding Standard https://mpeg.chiariglione.org/standards/exploration/future-video-coding/requirements-a-new-video-coding-standard

[20] ISO/IEC 10918-1 / ITU-T T.81: Coding of still pictures, Annex D

[21] ITU-T H.263: Video coding for low bit rate communication, Annex J