

**ISO/IEC JTC 1/SC 29/AG 3**  
**MPEG Liaison and Communication**  
**Convenorship: KATS (Korea, Republic of)**

<b>Document type:</b>	Output Document
<b>Title:</b>	<b>White paper on Low Complexity Enhancement Video Coding (LCEVC)</b>
<b>Status:</b>	Approved
<b>Date of document:</b>	2022-01-23
<b>Source:</b>	ISO/IEC JTC 1/SC 29/AG 3
<b>Expected action:</b>	None
<b>No. of pages:</b>	11 (with cover page)
<b>Email of Convenor:</b>	kyuheonkim@khu.ac.kr
<b>Committee URL:</b>	<a href="https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag3">https://isotc.iso.org/livelink/livelink/open/jtc1sc29ag3</a>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION  
ORGANISATION INTERNATIONALE DE NORMALISATION  
ISO/IEC JTC 1/SC 29/AG 3  
CODING OF MOVING PICTURES AND AUDIO**

**ISO/IEC JTC 1/SC 29/AG 3 N58**  
**Online – January 2022**

<b>Title</b>	<b>White Paper on Low Complexity Enhancement Video Coding (LCEVC)</b>
<b>Source</b>	<b>AG 03, MPEG Liaison and communication</b>
<b>Status</b>	<b>Approved</b>
<b>Serial Number</b>	<b>21273</b>

### **Abstract**

This White Paper provides an overview of MPEG-5 Part 2 Low Complexity Enhancement Video Coding (LCEVC), a novel video coding standard from MPEG ISO WG 04. The codec is designed for use in conjunction with existing video codecs, leveraging on encoder-driven upsampling and specific tools for encoding “residuals”, i.e. the difference between the original video and a predicted rendition. LCEVC can improve the compression efficiency and reduce the overall computational complexity for coding a given resolution and bit depth by using a small number of specialized enhancement tools. This paper provides an outline of the architecture, coding tools, and an overview of the performance of the compression efficiency of LCEVC.

### **1 Introduction**

This White Paper provides an overview of MPEG-5 Part 2 Low Complexity Enhancement Video Coding (LCEVC), a new video coding standard developed by MPEG, which is published as ISO/IEC 23094-2 [1]. LCEVC works by encoding a lower resolution (and potentially also lower bit depth) version of a source video using any existing codec (the “base codec”) and then coding the differences between the lower resolution video and the full-resolution source, up to mathematically lossless coding if needed, using a different compression method (the “enhancement”).

Rather than being a replacement for existing video coding schemes, LCEVC is designed to leverage on existing and future codecs, enhancing their performances whilst reducing their computational complexity. It is not meant to be an alternative to other codecs, but rather a useful complement to any codec. As an enhancement of other codecs, LCEVC enables to find the optimal trade-off between compression and complexity for a given codec and/or application, with the potential in multiple use cases of reducing either bandwidth requirements for a given visual quality or coding energy requirements, or both.

This enhancement is achieved by a combination of processing an input video at a lower resolution with an existing single-layer codec and using a simple and small set of highly specialized tools to correct impairments, upscale and add details to the processed video. LCEVC’s low-complexity requirement meant that the tool definition process accounted for the availability of hardware acceleration for graphics processing available in existing chipsets and it’s especially amenable to optimised software implementations with low power consumption

(e.g., using SIMD CPU, GPU and heterogeneous parallel processing), as well as small-footprint silicon implementations.

The LCEVC specification also enables a stream to signal adaptive dithering post-processing (which reduces banding and aliasing impairments) as well as providing a platform for the inclusion of user data within the bitstream at transform block level. This allows LCEVC to support the adoption of evolving image manipulation techniques within the standard, while still offering an efficient method of encoding residual data and providing up to mathematically lossless picture reconstruction. In addition, tiling options for the residual sublayers enable parallel execution of entropy coding, striped ultra-low-latency use cases and region of interest decoding of the enhancement.

In terms of overall rate control accuracy, LCEVC data is encoded after the base layer, which allows it to compensate at least in part for accuracy issues in the coded size of the base layer (or base layer stripe), with consequent benefits in real time low-latency use cases.

With the introduction of LCEVC, it is possible to efficiently extend the features of a video service also when a portion of the target devices does not support LCEVC decoding. In fact, as soon as LCEVC is supported by a portion of target devices, a video service can already introduce new features (such as UHD, HDR, etc.) without disruption for devices that only support the base codec and without requiring to duplicate video workflows to separately serve old and new devices. Further, when possible to retrofit LCEVC via software onto a portion of existing target devices, the video service would benefit even more from the possibility to upgrade service quality without having to wait for a full replacement of the installed base of target devices.

## 2 Background and Requirements

In October 2018, 28 industry signatories outlined market needs and requirements for a software-based capability extension on top of existing and future video codecs. They claimed that “despite surging demand for video, it is often difficult – or prohibitively costly – to deliver the high video quality that most end users expect”.

They added that the combination of legacy devices and long replacement cycles “makes it difficult to upgrade video services to higher resolutions, bit-depths and frame rates without either ignoring customers with legacy video devices or creating duplicate services for new devices. The consequent low availability of higher resolution services reduces the demand for newer decoder devices, in a vicious cycle.”

Signatories were industry leaders active in live and VoD streaming, videoconferencing, VR, broadcast video and real-time video feeds for industrial applications. In response to that request, LCEVC performance requirements were detailed in w18098 [2]:

*“When enhancing an  $n$ -th generation MPEG codec (e.g. AVC), compression efficiency for the aggregate stream is appreciably higher than that of the  $n$ -th generation MPEG codec used at full resolution and as close as possible to that of the  $(n+1)$ -th generation MPEG codec (e.g. HEVC) used at full resolution, at bandwidths and operating conditions relevant to mass market distribution;*

and

*Encoding and decoding complexity for the aggregate full resolution video (i.e. base plus*

enhancement) shall be comparable with that of the base encoder or decoder, respectively, when used alone at full resolution.”

### **3 Coding structure**

#### **3.1 Encoder**

The encoding process to create an LCEVC conformant bitstream is shown in Figure 1 and can be depicted in three major steps.

##### **3.1.1 Base codec**

Firstly, the input sequence is fed into two consecutive non-normative downscalers and is processed according to the chosen scaling modes. Any combination of the three available options (2-dimensional scaling, 1-dimensional scaling in the horizontal direction only or no scaling) can be used. The output then invokes the base codec which produces a base bitstream according to its own specification. This encoded base can be included as part of the LCEVC bitstream.

##### **3.1.2 Enhancement sub-layer 1**

The reconstructed base picture may be upsampled to undo the downscaling process and is then subtracted from the first-order downsampled input sequence in order to generate the sub-layer 1 (L-1) residuals. These residuals form the starting point of the encoding process of the first enhancement sub-layer. A number of coding tools, which will be described further in the following section, process the input and generate entropy encoded quantized transform coefficients.

##### **3.1.3 Enhancement sub-layer 2**

As a last step of the encoding process, the enhancement data for sub-layer 2 (L-2) needs to be generated. In order to create the residuals, the coefficients from sub-layer 1 are processed by an in-loop LCEVC decoder to achieve the corresponding reconstructed picture. Depending on the chosen scaling mode, the reconstructed picture is processed by an upscaler. Finally, the residuals are calculated by a subtraction of the input sequence and the upsampled reconstruction.

Similar to sub-layer 1, the samples are processed by a few coding tools. In addition, a temporal prediction can be applied on the transform coefficients in order to achieve a better removal of redundant information. The entropy encoded quantized transform coefficients of sub-layer 2, as well as a temporal layer specifying the use of the temporal prediction on a block basis, are included in the LCEVC bitstream.

#### **3.2 Decoder**

For the creation of the output sequence, the decoder analyses the LCEVC conformant bitstream. As can be seen in Figure 2, the process can again be divided into three parts.

##### **3.2.1 Base codec**

In order to generate the *Decoded Base Picture (Layer 0)* the base decoder is fed with the extracted base bitstream. According to the chosen scaling mode in the configuration, this reconstructed picture might be upsampled and is afterwards called *Preliminary Intermediate Picture*.

##### **3.2.2 Enhancement sub-layer 1**

Following the base layer, the enhancement part needs to be decoded. Firstly, the coefficients belonging to enhancement sub-layer 1 are decoded using the inverse tools of the encoding



## 4 Bitstream structure

The LCEVC bitstream contains a base layer, which may be at a lower resolution, and an enhancement layer consisting of up to two sub-layers. This subsection briefly explains the structure of this bitstream and how the information can be extracted.

While the base layer can be created using any video encoder and is not specified further in the LCEVC specification, the enhancement layer must follow the structure as specified. Similar to other MPEG codecs [3][4], the syntax elements are encapsulated in network abstraction layer (NAL) units which also help synchronize the enhancement layer information with the base layer decoded information. Depending on the position of the frame within a group of pictures (GOP), additional data specifying the global configuration and controlling the decoder may be present.

The data of one enhancement picture is encoded into several chunks. These data chunks are hierarchically organized as shown in Figure 3. For each processed plane (nPlanes), up to two enhancement sub-layers (nLevels) are extracted. Each of them again unfolds into numerous coefficient groups of entropy encoded transform coefficients. The amount depends on the chosen type of transform (nLayers). Additionally, if the temporal prediction is used, for each processed plane an additional chunk with temporal data for Enhancement sub-layer 2 is present.

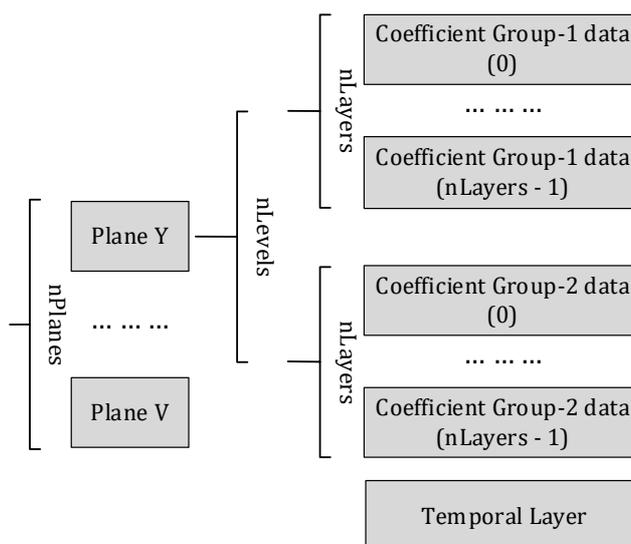


Figure 3 – Encoded enhancement picture data chunk structure

## 5 Coding tools

This section provides a brief overview of the coding tools which are available in accordance with the LCEVC specification. For simplicity, the description is mostly focussed on an encoder.

### 5.1 Down- and upscaler

Two non-normative downscalers can be used to downscale the input sequence to a lower resolution. The downscaling can be done either in both vertical and horizontal directions, only in the horizontal direction or alternatively cannot be applied. Two upscalers are available reconstructing the sequence at a higher resolution. One of four specified four-tap upscaling kernels can be used. Additionally, a customizable four-tap kernel is available whose

coefficients can be signalled in the LCEVC bitstream.

After the upscaling process, an additional predicted residuals coding block can be processed. For low bitrates, the tool can improve visual quality by correcting upscaling average errors at the quad level where otherwise the bitrate to signal non-zero average coefficients wouldn't be available. For higher bitrates, it reduces the number of average coefficients needed to be signalled, reducing the overall achieved bitrate.

## **5.2 Transform**

LCEVC allows the usage of two different transforms. Both operate with a small kernel of size  $2 \times 2$  or  $4 \times 4$ . In case the upscaling process is performed in the horizontal direction only, the transform kernels are slightly modified to better reflect the preceding 1-dimensional upscaling.

The transforms are chosen to have a small kernel to optimize the coding of sparse residuals. In particular, sharp edges can efficiently be transformed.

## **5.3 Quantization**

The transform coefficients are quantized using a linear quantizer whose quantization step width is signalled on a picture basis. The linear quantizer may use a dead zone whose size changes relative to the quantization step. The quantization can be configured independently for both enhancement sub-layers to give more flexibility to the encoder. Additionally, the level of quantization can be modified for transform blocks that use temporal prediction compared to intra-coded transform blocks. For chroma planes, the quantization can be scaled separately from the luma plane.

## **5.4 L-1 filter**

The L-1 filter can be applied on the sub-layer 1 residuals if the transform with the larger kernel size ( $4 \times 4$ ) is used. The aim of this filter is to reduce the blocking artefact which the transform applied at a lower resolution can create. The residuals on the outer boundary of the transform block are multiplied with a coefficient between 0 and 1. The value of these coefficients can be signalled independently for edges and corners.

## **5.5 Temporal prediction**

The temporal prediction uses a zero-motion vector prediction with a temporal buffer which stores residuals from the previous frame only. The decision, where to use temporal prediction, is done on a transform block basis. Additionally, an entire tile of  $32 \times 32$  residuals can be signalled to be used without temporal prediction, reducing the signalling overhead for, e.g., a fast-moving part of the sequence.

## **5.6 Entropy encoding**

The two coefficient layers and the temporal layer are processed independently by an entropy encoder before the encapsulation into the bitstream. The entropy coding process consists of two components: a Run Length Encoder (RLE) and a Prefix Coding encoder. Additionally, it is possible to only use the RLE for the entire data within a coefficient group.

Tiling options enable parallel execution of entropy coding, including the possibility to define custom tile sizes for residual sublayer data. It is also possible to embed user data into some of the coefficients at transform block level.

## 6 Performance results

MPEG carried out a formal subjective verification test of the LCEVC Main profile for standard dynamic range (SDR) content testing AVC, HEVC, EVC, and VVC as base and anchor codecs [5]. The purpose of the verification test was to confirm that the coding efficiency objective for the LCEVC standard has been met achieving a bit-rate reduction at a similar level of subjective visual quality relative to a single-layer video codec.

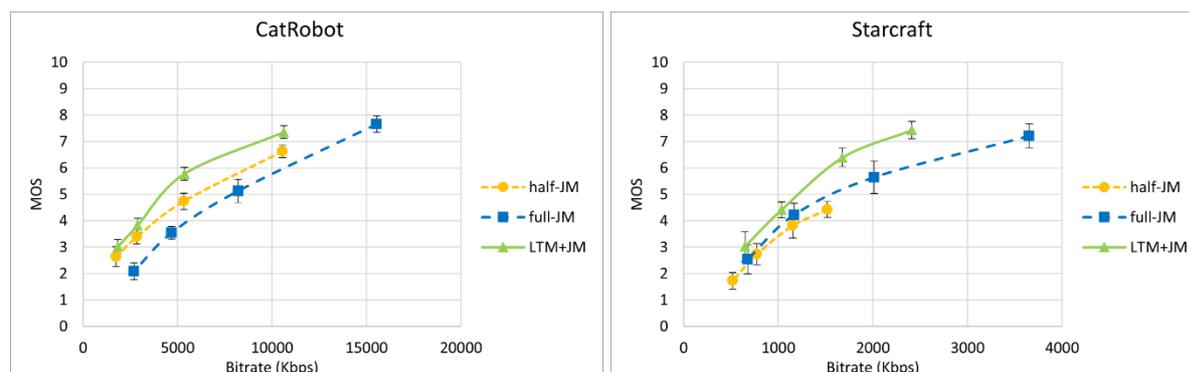
Verification tests compared full-resolution LCEVC-enhanced encoded sequences both with full-resolution single layer anchors and with half-resolution anchors upsampled to full-resolution with Lanczos upsampling. The comparison with half-resolution anchors was used to validate that enhancing a base codec with LCEVC is preferable both with respect to encoding with the base codec alone at full resolution and with respect to encoding with the base codec alone at a lower resolution and then relying on unguided upsampling at the decoder device.

### 6.1 Rate-distortion results

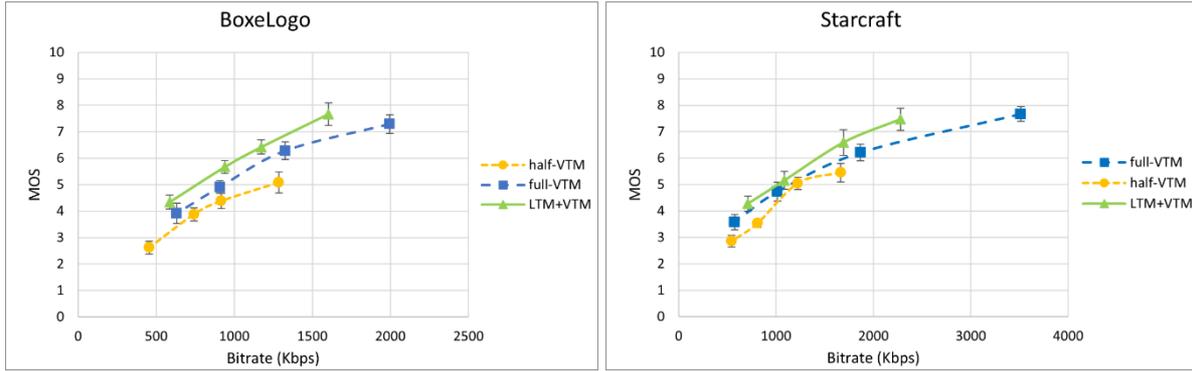
The anchors have been generated using AVC (JM 19.0), HEVC (HM 16.20), EVC (ETM 6 rc 1), and VVC (VTM 11). LCEVC encoded sequences were generated using the software model LTM 5.1.

Throughout this section, the following terminology is used to refer to different types of encoded sequences (xM indicates a generic test model):

- full-xM: anchor encoded at full resolution using xM as test model
- half-xM: anchor encoded at half resolution in both horizontal and vertical direction, using xM as test model, and then upsampled with Lanczos filter [6] to full resolution using FFmpeg 4.3.1
- LTM+xM: LCEVC encoded at full resolution using LTM and enhancing xM used for encoding the base layer



**Figure 4** – Results comparing JM vs. LTM+JM for UHD and HD sequences



**Figure 5** – Results comparing VTM vs. LTM+VTM for UHD and HD sequences

## 6.2 BD-Rate comparisons

In this section, the average bit rate savings of LCEVC profiles compared to references for each sequence were computed from the MOS vs. bit rate data in the same manner that was done in [7] to further quantify the bit rate savings achieved. The bit rate savings are averaged over the whole range where the same MOS scores for LCEVC, AVC (Full Resolution), AVC (Half Resolution), HEVC (Full Resolution), HEVC (Half Resolution), EVC (Half Resolution), and VVC (Half Resolution) can be interpolated from subjective test results.

Numerical analysis of the average benefit of LCEVC and its statistical significance compared to the corresponding full resolution EVC or VVC codec was more difficult to interpret, due to several test points having overlapping confidence intervals, as such the MOS BD-rate savings are not reported. However, the test results tend to indicate an overall benefit when using LCEVC with these two codecs.

For the tests indicated as Half Resolution, the half-resolution anchors were used as the base layer for LCEVC and hence not all curves may overlap as much as would be ideal when calculating a BD-rate.

Tables 1 and 2 show the MOS BD-rate for the sequences in this test. The BD-rate measure described in [8] is used with MOS scores taking the place of PSNR. A piece wise cubic interpolation used in the LCEVC common conditions spreadsheet is used.

**Table 1** – MOS BD-rate – LTM-enhanced JM and HM vs JM and HM anchors at full resolution

LTM 5.1 vs JM 19.0		BD-rate
UHD	LupoPuppet	-53.98%
	CatRobot	-43.83%
	DrivingPOVLogo	-30.19%
	BoxeLogo	-55.61%
<b>Average</b>		<b>-45.90%</b>

LTM 5.1 vs HM 16.20		BD-rate
UHD	LupoPuppet	-31.14%
	CatRobot	-41.88%
	DrivingPOVLogo	-26.01%
	BoxeLogo	-24.44%
<b>Average</b>		<b>-30.87%</b>

LTM 5.1 vs JM 19.0		BD-rate
HD	TrafficLogo	-30.18%
	Starcraft	-26.75%
<b>Average</b>		<b>-28.47%</b>

LTM 5.1 vs HM 16.20		BD-rate
HD	TrafficLogo	-18.52%
	Starcraft	-29.76%
<b>Average</b>		<b>-24.14%</b>

**Table 2** – MOS BD-rate – LTM-enhanced JM, HM, ETM, VTM vs JM, HM, ETM, VTM upsampled half-resolution anchors

LTM 5.1 vs JM 19.0		BD-rate	LTM 5.1 vs HM 16.20		BD-rate
UHD	LupoPuppet	-22.66%	UHD	LupoPuppet	-20.64%
	CatRobot	-26.25%		CatRobot	-50.77%
	DrivingPOVLogo	-26.95%		DrivingPOVLogo	-33.96%
	BoxeLogo	-34.20%		BoxeLogo	-29.23%
<b>Average</b>		<b>-27.52%</b>	<b>Average</b>		<b>-33.65%</b>

LTM 5.1 vs JM 19.0		BD-rate	LTM 5.1 vs HM 16.20		BD-rate
HD	TrafficLogo	-27.80%	HD	TrafficLogo	-22.51%
	Starcraft	-26.28%		Starcraft	-29.10%
<b>Average</b>		<b>-27.04%</b>	<b>Average</b>		<b>-25.80%</b>

LTM 5.1 vs ETM 6 rc1		BD-rate	LTM 5.1 vs VTM 11		BD-rate
UHD	BarScene	-40.54%	UHD	Marathon	-30.44%
	CatRobot	-39.64%		MountainBay2	-22.52%
	DrivingPOVLogo	-46.67%		DrivingPOVLogo	-43.42%
	BoxeLogo	-24.58%		BoxeLogo	-37.07%
<b>Average</b>		<b>-37.86%</b>	<b>Average</b>		<b>-33.36%</b>

LTM 5.1 vs ETM 6 rc1		BD-rate	LTM 5.1 vs VTM 11		BD-rate
HD	TrafficLogo	-30.91%	HD	TrafficLogo	-21.11%
	Starcraft	-11.28%		Starcraft	-19.96%
<b>Average</b>		<b>-21.09%</b>	<b>Average</b>		<b>-20.53%</b>

### 6.3 Conclusion

The first set of tests compared full-resolution LCEVC-enhanced encoded sequences with full-resolution single-layer anchors. The average bit rate savings for LCEVC when enhancing AVC were determined to be approximately 46% for UHD and 28% for HD. The average bit rate savings for LCEVC when enhancing HEVC were determined to be approximately 31% for UHD and 24% for HD. The test results tend to indicate an overall benefit also when using LCEVC with EVC and VVC.

The second set of tests aimed to confirm that LCEVC provided a more efficient means of resolution enhancement of half resolution anchors than unguided up-sampling. For these tests, the test sequences were coded using AVC, HEVC, EVC, or VVC at half resolution in both horizontal and vertical direction. For anchor generation, the half resolution encoded sequences were upsampled with Lanczos filters to full resolution for visual assessment. The same half resolution encoded sequences were also used as base layers for LCEVC and hence not all curves may overlap as much as would be ideal when calculating a BD-rate. Comparing LCEVC full-resolution encoded sequences with the up-sampled half-resolution anchors, the average bit-rate savings when using LCEVC with AVC, HEVC, EVC, and VVC were calculated to be approximately 28%, 34%, 38%, and 33% respectively for UHD, and 27%, 26%, 21%, and 21% respectively for HD.

Importantly, the above performance results come in combination with significant reduction of overall processing complexity (base layer plus LCEVC enhancement vs. base codec alone used at full resolution), the possibility for LCEVC-unaware decoder devices to still decode the base

layer as well as the possibility to process LCEVC also by means of hardware components that are typically available in general purpose video-capable devices, such as GPUs, scalars, DSPs, etc.

## References

- [1] “Information technology — General video coding — Part 2: Low complexity enhancement video coding”, ISO/IEC 23094-2
- [2] “Requirements for Low Complexity Video Coding Enhancements”, ISO/IEC JTC 1/SC 29/WG 11 w18098, October 2018, Available: <https://mpeg.chiariglione.org/standards/exploration/low-complexity-video-coding-enhancements/requirements-low-complexity-video>
- [3] “Information technology — High efficiency coding and media delivery in heterogeneous environments — Part 2: High efficiency video coding”, ISO/IEC 23008-2, ITU-T H.265
- [4] “Information technology — Coded representation of immersive media — Part 3: Versatile video coding”, ISO/IEC 23090-3, ITU-T H.266
- [5] “Verification Test Report on the Compression Performance of Low Complexity Enhancement Video Coding”, ISO/IEC JTC 1/SC 29/WG 04 N0076, May 2021, Available: <https://www.mpegstandards.org/standards/MPEG-5/2/>
- [6] “FFmpeg Filters Documentation”, The FFmpeg developers, Available: <https://ffmpeg.org/ffmpeg-filters.html#scale> (Retrieved: 2021/05/10)
- [7] “Report on Essential Video Coding Compression Performance Verification Testing for SDR Content”, ISO/IEC JTC 1/SC 29/WG 04 N0047 (w20000), January 2021
- [8] “Information technology — MPEG video technologies — Part 8: Working practices using objective metrics for evaluation of video coding efficiency experiments”, ISO/IEC TR 23002-8, ITU-T HSTP-VID-WPOM