https://doi.org/10.2352/EI.2023.35.10.HVEI-260

User Perception for Dynamic Video Resolution Change Using VVC

Sachin Deshpande, Philip Cowan; Sharp Corporation; WA, USA

Abstract

We performed experiments that measured user perception when video resolution changes dynamically. Versatile Video Coding (VVC) standard was recently finalized and it includes a reference picture resampling (RPR) tool. VVC RPR supports changing spatial resolution in a coded video sequence on a per picture basis. VVC RPR defines the downsampling and upsampling filters to be used when changing resolution. This paper provides results from subjective evaluation when VVC RPR is used for part of the video sequence to dynamically change resolution. The experiments use different QP values (or bitrates), different RPR scale factors and different highest original spatial resolutions. The results compare how users perceive video coded using VVC RPR for some pictures compared to an anchor which does not use RPR. In addition to the subjective results, we also describe performance of various metrics including PSNR, VMAF and MS-SSIM. Our results can help choose the highest RPR scale factor that can be used to achieve/ maintain certain perceived quality when using RPR (for example for bitrate reduction). The study also confirms that MS-SSIM and VMAF match subjective test results more closely compared to PSNR.

Introduction

Versatile Video Coding (VVC) [1], also known as ITU-T H.266 | ISO/IEC 23090-3, is the next generation video coding standard that has been recently finalized by the Joint Video Experts Team (JVET) of ITU-T VCEG and ISO/IEC MPEG. VVC includes a reference picture resampling (RPR) tool [2]. VVC RPR enables picture resolution change within a coded video sequence at any picture without encoding an Intra Random Access Picture (IRAP) picture which is completely intra coded. Thus, VVC RPR supports changing spatial resolution of a coded picture in a coded video sequence even on a per picture basis. This is different than previous standards (AVC, HEVC) where the spatial resolution of pictures can only change by starting a new sequence with an IRAP picture and a new sequence parameter set (SPS). VVC RPR defines the downsampling and upsampling filters to be used when changing resolution.

We define experiments to measure user perception of reference picture resampling (RPR) of VVC. Our results can help choose the highest RPR downsample scale factor that can be used to achieve/ maintain certain perceived quality when using RPR. The study also investigates the question of which metrics (MS-SSIM, VMAF, PSNR) match RPR subjective test results more closely.

RPR tool is introduced in VVC mainly to handle network bandwidth changes. For example, when network bandwidth drops, the video resolution could be lowered compared to the original video resolution. The video resolution could be reverted back to the original resolution when network bandwidth returns to nominal level. Recent studies in JVET have found that VVC RPR tool can also be used as a compression efficiency tool. When used as a compression tool, JVET standardization documents mainly only report PSNR results. However, the change in spatial video resolution may be perceived by the user differently, which may not be captured by the PSNR metric. As a result, we design subjective evaluation method to assess the user perception when video resolution dynamically changes in the sequence. Also, we evaluate VMAF and MS-SSIM metrics which are asserted to provide better understanding of user perception for video with dynamic resolution change. We perform evaluation at several different RPR scale factors and different quantization parameter (QP) values.

Subjective Test Method

We use video sequences which include different types of motion and different original spatial resolutions. The sequences used include both Joint Video Experts Group (JVET) test sequences (Racehorses) and common sequences used for DASH streaming (Big Buck Bunny – at two different original resolutions 1080P, 720P). The VVC RPR encoding with downsample scale factors ranging from 1.1 to 1.9 in steps of 0.2, and 2.0 are used. Each sequence is 10 seconds long. When using VVC RPR, initial video pictures (first 5 seconds of video data) are downsampled by the specific RPR scale factor followed by the video switching to full original resolution.

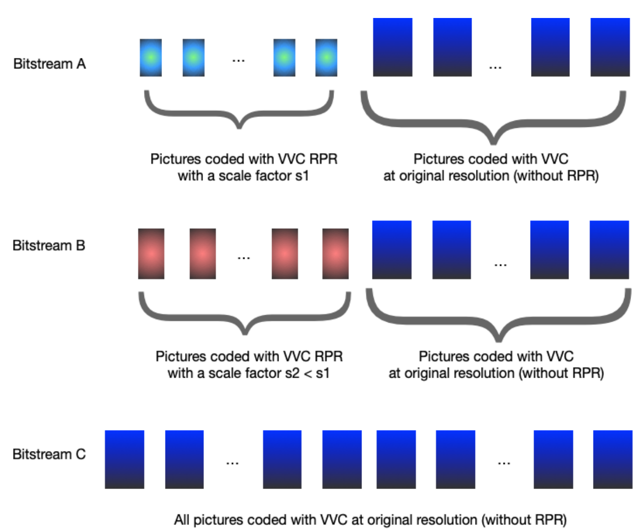

Figure 1: Bitstreams used for subjective tests. Example shows three bitstreams. Bitstream A and B use RPR (with different RPR scale factors S1 and S2 respectively) for initial pictures followed by original resolution pictures. Bitstream C uses all original resolution pictures.

It is asserted that this closely resembles scenario that typically happens when a user starts watching a streaming video. In this case to reduce the initial buffering delay, initial pictures are coded with fewer bits because of use of VVC RPR. This allows the client buffer to be filled quickly to the target buffer occupancy (in seconds) such that the playback can start with less initial delay. After initial duration the video resolution switches back to the original resolution for the video sequence. For the subjective visual test, the RPR downscaled pictures are upscaled to original resolution for display. This is illustrated in Figure 2 below. We used FFMPEG [11] bicubic scaling filter for upsampling. The input data to FFMPEG was yuv420 10bit planar and the output was the same format.

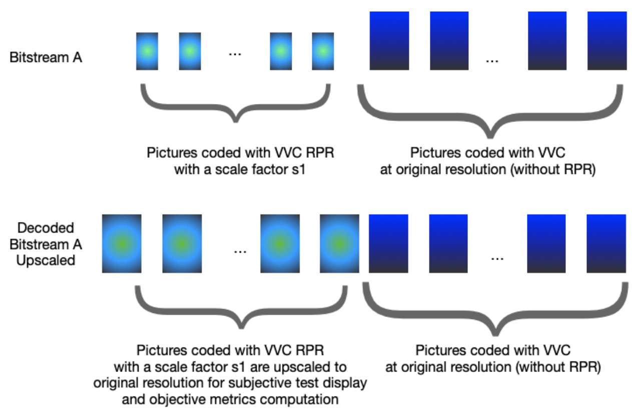

Figure 2: RPR Bitstream and corresponding decoded bitstream upscaled for subjective tests and objective metrics

We conducted subjective tests using ITU-R BT.500 recommendation's [5] double-stimulus impairment scale (DSIS) method to determine subjective quality of RPR VVC bitstreams compared with non-RPR VVC bitstreams. Fifteen subjects conducted the user study. Three video sequences of different resolution and different motion characteristics as described in Table 1 were used for the subjective tests. The video display used had a 1080P – Full HD resolution. The viewing distance for a subject was set to 3 picture heights (standard viewing distance for HD) in each case. For all the subjective quality tests the viewing conditions were set based on Annex 1 of Recommendation ITU-R BT.710-4 [6]

Subjective visual quality evaluation was conducted by playing back original resolution and RPR videos using procedure as follows.

- The original resolution video encoded with VVC, with no RPR coding was used as unimpaired reference. Videos encoded with RPR for initial pictures in the video and followed by non-RPR for remainder of the video was used as impaired videos.
- Different RPR scale factors: 1.1, 1.3, 1.5, 1.7, 1.9 and 2.0 were used to generate various impaired videos.
- The subjects are shown the reference video and one of the impaired videos. The subject then uses a 5-point impairment scale (shown below) for assessment of impaired video with respect to the reference video.
- The above step is repeated for each of the RPR videos with different scale factors as an impairment video in random order.
- The following 5-point impairment scale was used by the subjects for ranking:
    - 5= imperceptible
    - 4= perceptible, but not annoying
    - 3= slightly annoying
    - 2= annoying
    - 1= very annoying

Table 1: Video test sequences used for the subjective tests

Sequence Name	Resolution	Duration (seconds) & frame rate (Hz)	Source
Big Buck Bunny 1080P	1920x1080	10 sec @ 24 Hz	https://peach.blender.org [1]
Big Buck Bunny 720P (Referred to as Small Buck Bunny)	1280x720	10 sec @ 24 Hz	https://peach.blender.org [1]
Racehorses	832x 480	10 sec @ 30 Hz	JVET [4]

Subjective Test Results

Our subjective test results show user perception (in a 5-grade scale) for video which uses VVC RPR (for some of the pictures) compared to an anchor video which is at original resolution, as function of RPR scale factor. Our subjective results show that the user perception for RPR depends upon the quantization parameter (QP), i.e. the overall bitrate and quality of the video. In the mean, for QP values of 22 and 27, subjects rate RPR video to be no worse than score of 4 (i.e. perceptible, but not annoying) for all RPR scale factors (1.1-2.0). In the mean, for QP 32, for scale factors up to 1.7, subjects rate RPR video to be no worse than score of 4 (i.e. perceptible, but not annoying). In the mean at QP 37, RPR performs significantly worse than the anchor for scale factor of 1.7 and above, where users on the average rate it lower than a score of 4. Also, as expected the subjective scores drop as RPR scale factor increases.

Figures 3 (a)-(c) show mean subjective quality test results respectively for Big Buck Bunny (1080P), Small Buck Bunny (720P) and Racehorses sequences. Figures show mean scores with confidence intervals on Y axis and the RPR scale factor on X axis.

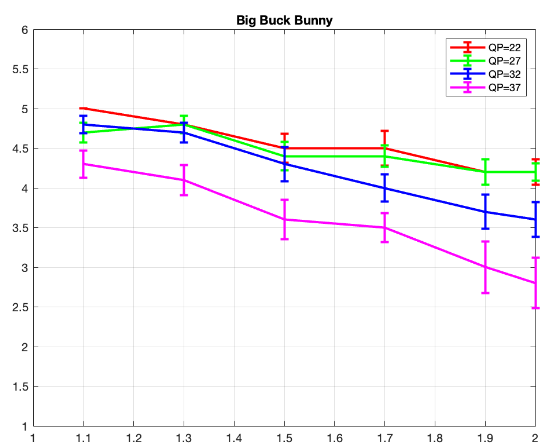

Figure 3 (a)

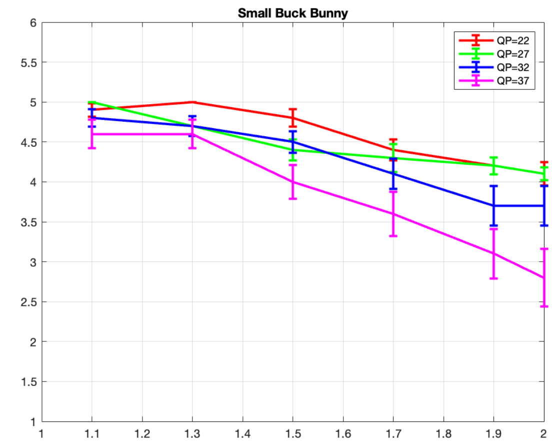

Figure 3 (b)

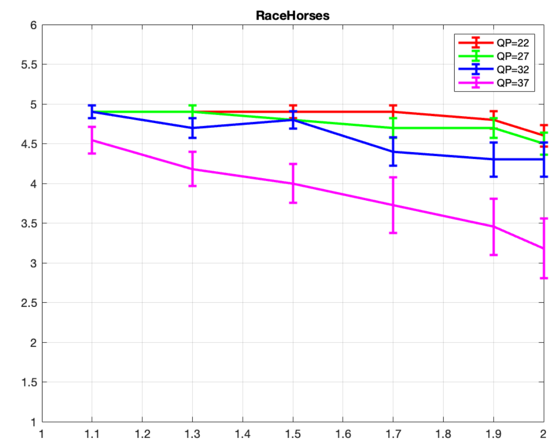

Figure 3 (c)

Figure 3 (a)-(c) : Mean subjective quality test results

Objective Metrics

In addition to the subjective tests, we also calculated PSNR, MS-SSIM [7], VMAF [8] objective metrics. We used the code from Github repository [9] for computation of the metrics psnr_y, psnr_cb, psncr_cr, float_ms_ssim, and vmaf. The objective metrics are calculated using the original resolution video as a reference (or anchor) video. For the video under test, each picture which is using VVC RPR is first upscaled to the original video resolution before calculating the objective metric. It should be noted that we used JVET VTM software [10] for encoding VVC video. For non RPR bitrate for Big Buck Bunny ranges from 255 Kbps to 3.39 Mbps, for small buck Bunny from 126 Kbps to 1.45 Mbps and for RaceHorses from 332 Kbps to 4 Mbps. For RPR the bitrate for Big Buck Bunny ranges from 205 Kbps to 2.9 Mbps, for small buck Bunny from 109 Kbps to 1.3 Mbps and for RaceHorses from 236 Kbps to 3.12 Mbps.

Figure 4 (a)-(c) respectively show PSNR (for luma) results for Big Buck Bunny (1080P), Small Buck Bunny (720P) and Racehorses sequences.

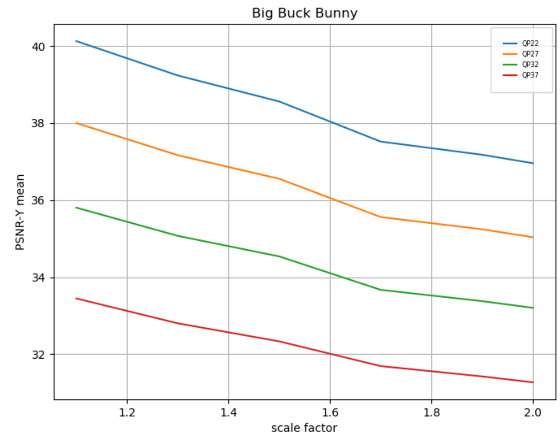

Figure 4 (a)

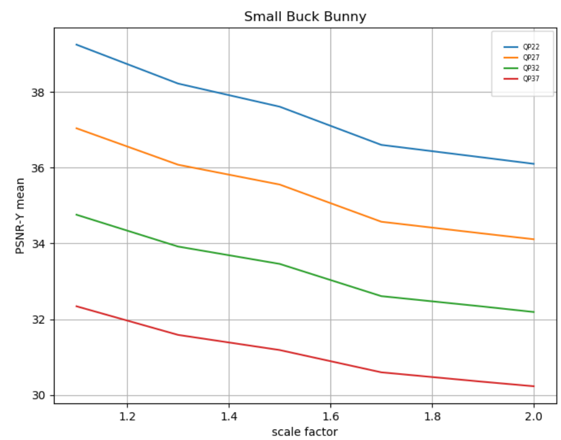

Figure 4 (b)

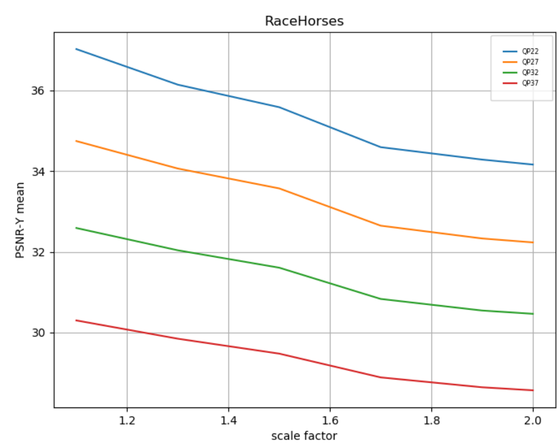

Figure 4 (c)

Figure 4 (a)-(c) : Luma PSNR metric results

Figure 5 (a)-(c) respectively show PSNR (for Chroma – Cb and Cr) results for Big Buck Bunny (1080P), Small Buck Bunny (720P) and Racehorses sequences.

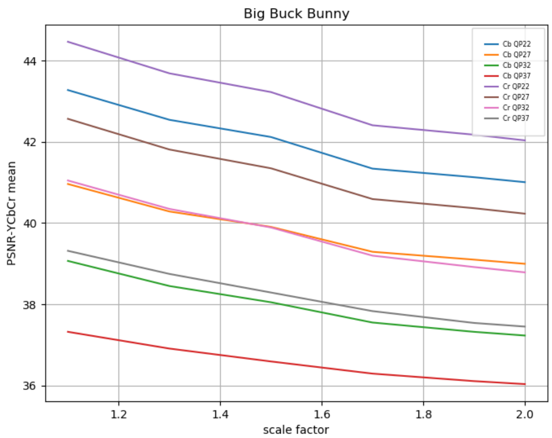

Figure 5 (a)

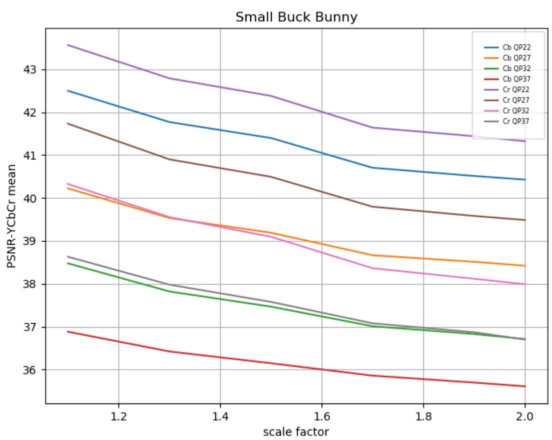

Figure 5 (b)

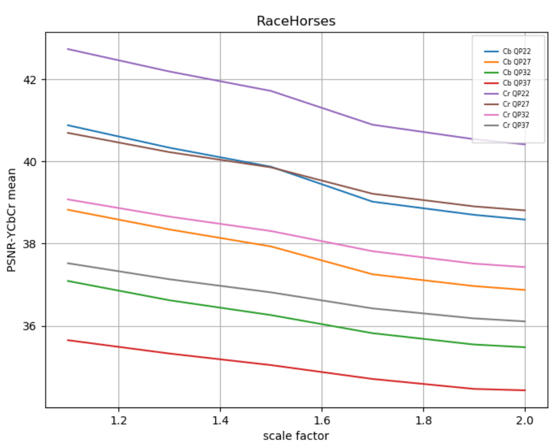

Figure 5 (c)

Figure 5 (a)-(c) : Chroma PSNR metric results

Figure 6 (a)-(c) respectively show MS-SSIM [7] results for Big Buck Bunny (1080P), Small Buck Bunny (720P) and Racehorses sequences.

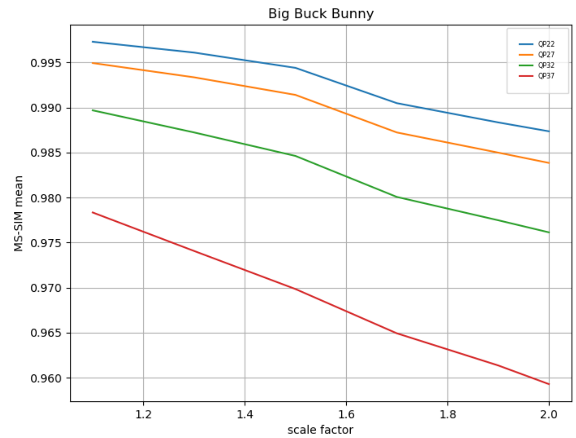

Figure 6 (a)

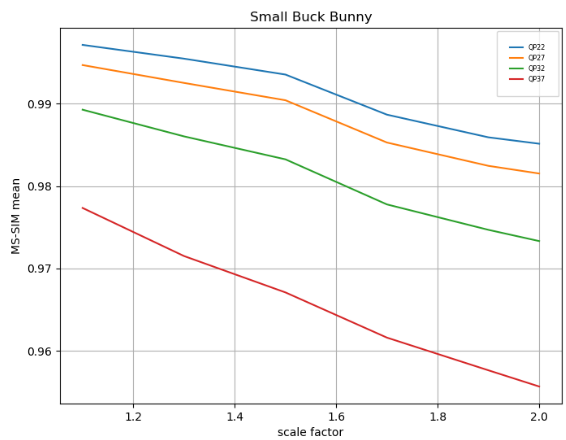

Figure 6 (b)

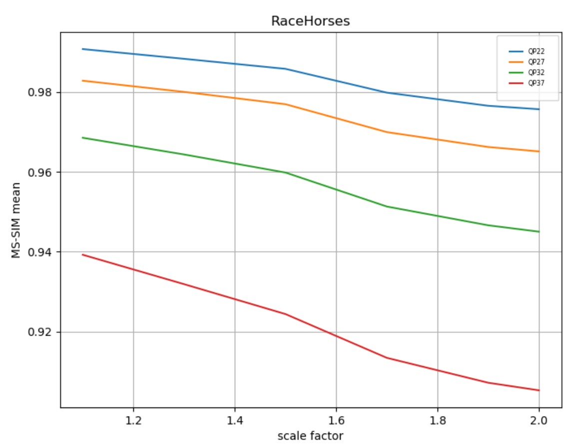

Figure 6 (c)

Figure 6 (a)-(c) : MS-SSIM metric results

Figure 7 (a)-(c) respectively show VMAF [8] results for Big Buck Bunny (1080P), Small Buck Bunny (720P) and Racehorses sequences.

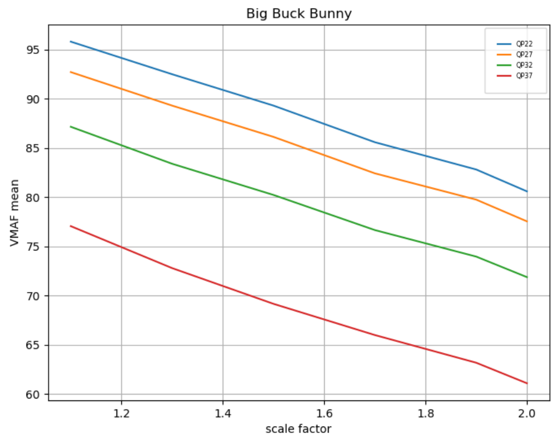

Figure 7 (a)

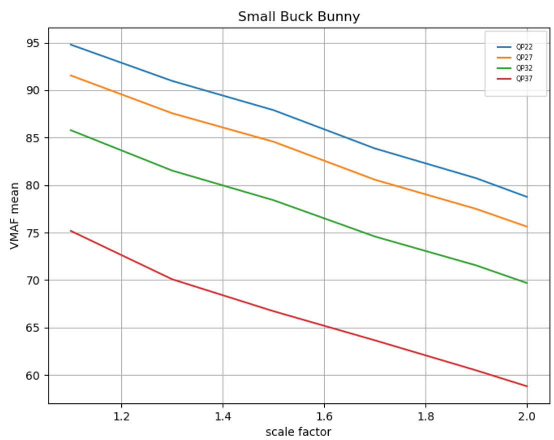

Figure 7 (b)

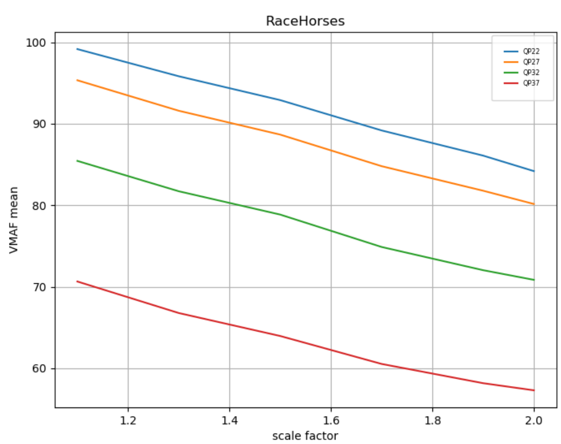

Figure 7 (c)

Figure 7 (a)-(c) : VMAF metric results

From the computed objective metrics, we observe the following:

- All three metrics decrease monotonically as RPR scale factor increases.
- It is observed than PSNR drops more uniformly as QP changes from 22 to 37. Thus, PSNR curves are mostly equally spaced.
- In comparison for both MS-SSIM and VMAF, the metric value drops more significantly for QP 37 compared to other QP values. This is consistent with the trend seen in the subjective quality test results for QP 37 compared to other QP values.

This leads us to conclude that MS-SSIM and VMAF maybe considered better metrics and match subjective test results better compared to PSNR. Furthermore, on an absolute scale basis VMAF appears to be able to capture the notably lower subjective quality at QP 37 somewhat better compared to MS-SSIM.

Conclusion

To our knowledge VVC RPR performance has not been subjectively evaluated where RPR is used for part of the video followed by full resolution video. Also, typical JVET studies use PSNR as the metric. Compared to this we also compute MS-SSIM and VMAF metric. Our results can help choose the highest RPR downsample scale factor that can be used to achieve/ maintain certain perceived quality when using RPR.

References

[1] Rec. ITU-T H.266 | ISO/IEC 23090-3, “Versatile Video Coding”, 2020.

[2] Y.-K. Wang, R. Skupin, M. Hannuksela, S. Deshpande, Hendry, V. Drugeon, R. Sjoberg, B.Choi, V. Seregin, Y. Sanchez, J. Boyce, W. Wan, G. Sullivan , “The High-Level Syntax of the Versatile Video Coding (VVC) Standard,” IEEE Transactions on Circuits and Systems for Video Technology. 2021, 10.1109/TCSVT.2021.3070860.

[3] Big Buck Bunny Test Sequence, “https://peach.blender.org”.

[4] J. Boyce, K. Suehring, X. Li, Xiang, V. Seregin, “JVET-J1010: JVET common test conditions and software reference configurations,” 2018

[5] ITU-R Recommendation BT.500-14, “Methodology for the subjective assessment of the quality of television images,” 2019

[6] ITU-R Recommendation BT.710-4, “subjective Assessment Methods for Image Quality in High-Definition Television,” 1998.

[7] Z. Wang, E. Simoncelli, A. Bovik, “Multiscale structural similarity for image quality assessment,” The Asilomar Conference on Signals, Systems and Computers, 2003.

[8] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, M. Manohara,, Toward A Practical Perceptual Video Quality Metric,” 2016

[9] VMAF software, https://github.com/Netflix/vmaf

[10] VVC VTM Reference Software, “https://vcgit.hhi.fraunhofer.de/jvet/VVCSoftware_VTM”

[11] FFMPEG Software, https://ffmpeg.org

Acknowledgement
These research results were obtained from the commissioned research (No.04201) by National Institute of Information and Communications Technology (NICT), Japan.