

# Compression Performance of the Versatile Video Coding: HD and UHD Visual Quality Monitoring

Naty Sidaty, Wassim Hamidouche and Olivier Déforges  
Univ Rennes, INSA Rennes, CNRS, IETR - UMR 6164  
Rennes, France  
wassim.hamidouche@insa-rennes.fr

Pierrick Philippe and Jérôme Fournier  
Orange & b<>com  
Cesson-Sévigné, France  
pierrick.philippe@orange.com

**Abstract**—Video compression and content quality have become one of the most research topic in the recent years. Predominantly, trends obviously signpost that the video usage over the Internet is on the upsurge. Simultaneously, users' requirement for enlarged resolution and higher quality is rising. Consequently, a huge effort has been made for video coding technologies and quality monitoring. In this paper, we present a subjective-based comparison as well as an objective measurement between the newest Versatile Video Coding (VVC) and the well-known High Efficiency Video Coding (HEVC) standards. Several videos of various content are selected as tested sequences. Both High Definition (HD) and Ultra High Definition (UHD) resolutions are used in this experiment. An extensive range of bit-rates from low to high bit-rates were selected. These sequences are encoded using both HEVC reference software (HM-16.2) and the latest reference software of VVC (VTM-5.0). Obtained results have shown that VVC outperforms consistently HEVC, for realistic bit rates and quality levels, in the range of 40% on the subjective scale. For the objective measurements, using PSNR, SSIM and VMAF as quality metrics, the quality enhancement of VVC over HEVC is ranging from 31% to 40%, depending on video content and spatial resolution.

**Index Terms**—Quality enhancement, HEVC, VVC, subjective assessment.

## I. INTRODUCTION

During the last decade, multimedia services and video applications have significantly increased due to the huge progress in digital technologies. The emerging video applications and image representation offer an immersive and more natural viewing experience. However, these new services require both higher quality and resolution (4K, 8K) to satisfy the quality of service required by the end users. The latest well-known video coding standard HEVC, for High Efficiency Video Coding standard, has been gradually adopted in many application systems, since it allows up to 50% bit-rate reduction for equal subjective quality compared with its predecessor H.264/AVC video coding [1], [2]. In 2015, a new coding tools have been developed under the Joint Exploration Model (JEM) software with the main goal to provide a high bit rate saving compared to the High Efficiency Video Coding (HEVC) standard and encouraging results were obtained [3], [4]. However, these quality improvements give rise to a considerable complexity [3]. To meet with these challenges, the Joint Video Experts Team (JVET) has released in 2017 a call for proposals in order to propose a new video coding standard. Initially named

Future Video Coding (FVC), it is now known as Versatile Video Coding (VVC). VVC will aid the deployment of higher-quality video services and emerging applications, such as High Dynamic Range (HDR), High Frame Rate (HFR) and 360-degree omnidirectional immersive multimedia. Planned to be finalized by the end of 2020, the primary objective of VVC is to provide a significant improvement in compression performance over the existing HEVC standard. The main objective is to provide 40% or more bitrate savings in terms of PSNR over HEVC with reasonable complexity increase at both encoder and decoder. The new standard is expected to enable the delivery of Ultra High Definition (4K) services at bit rates that actually are used to carry HDTV. In other word, using VVC would enable twice as much multimedia content to be stored on a server or sent through a streaming service.

In this paper, we present a subjective quality assessment study in order to evaluate the performance of this new emerging standard in comparison with HEVC standard. The VVC reference software version 5 (VTM-5.0) is compared with the HEVC reference software (HM-16.2) in Random Access (RA) coding configuration. Several video contents, at different bit-rates, and two spatial resolutions, HD and UHD have been used in this study.

The remainder of this paper is organized as follows. Section II introduces the VVC standard; its architecture and the most important coding tools introduced in the VTM-5.0. In section III, we describe the subjective quality assessment setup and illustrate the test material, environment and used methodologies. Obtained results, and associated analysis are given in Section IV. Finally, this paper ends with some conclusions and ideas about future work.

## II. VERSATILE VIDEO CODING DESCRIPTION

As HEVC, VVC has a block-based hybrid coding architecture, combining Inter and Intra Predictions with Transform Coding. Based on the HEVC standard, VVC refined existing technologies but also added novel coding tools. In this section, we describe briefly the main VVC coding tools present in the VTM-5.0 codec, including frame Partitioning, Multiple Transform Selection (MTS), In-Loop Filtering, Intra & Inter Predictions and Entropy Coding. Most of the gains in VVC are provided by a set of tools, given in “Table I”, with their

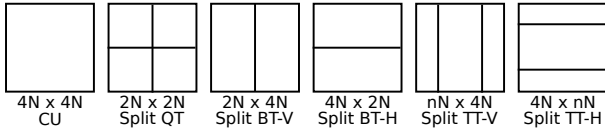


Fig. 1. Available split included in VTM-5.0 of a  $4N \times 4N$  CU.

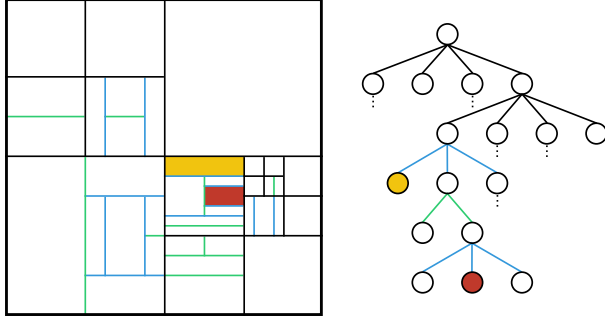


Fig. 2. Example of a CTU partitioning with a part of its corresponding tree representation.

individual loss when these tools are disabled in the VTM-5.0. A detailed description of these coding tools is given in [9].

#### A. Frame Partitioning

In addition to the recursive Quad-Tree (QT) partitioning used in HEVC, VTM-5.0 integrates a nested recursive Multi-Type Tree (MTT) partitioning, i.e. Binary-Tree (BT) and Ternary-Tree (TT) splits. Fig. 1 illustrates all available split in VTM-5.0 for a  $4N \times 4N$  Coding Unit (CU). The BT partitioning consists of symmetric horizontal splitting (BT-H) and symmetric vertical splitting (BT-V) while the TT partitioning allows horizontal triple-tree splitting (TT-H) and vertical triple-tree splitting (TT-V) corresponding to split the CU in three blocks with the middle blocks size equal to the half of the CU. Once BT or TT split is performed on a CU, QT split is not allowed any more on its sub-CUs. Fig. 2 presents on the left an example of a Coding Tree Unit (CTU) split into several CUs after performing the VTM-5.0 CTU partitioning with QT and MTT splits and on the right a part of CTU partitioning corresponding tree representation. Dark, green and blue lines represent QT split, BT split and TT split, respectively. Yellow and red background show two examples of the corresponding CUs on the CTU partitioning and corresponding tree.

#### B. Multiple Transform Selection

The MTS concept in VVC defines three trigonometrical transform types including Discrete Cosine Transform (DCT)-II, VIII and Discrete Sine Transform (DST)-VII. As illustrated in Fig. 3, the MTS concept selects, for Luma blocks of size lower than 64, the set of transforms that minimizes the rate distortion cost among five transform sets and the skip configuration. However, only DCT-II is considered for chroma components and Luma blocks of size 64. The MTS solution brings a significant coding gain of respectively 0.84% and

0.33% in All Intra (AI) and RA coding configurations [6] compared to HEVC.

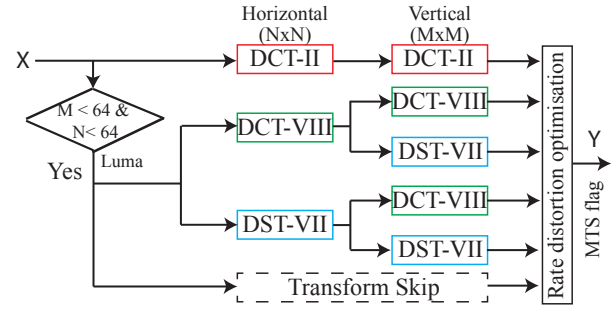


Fig. 3. The concept of 2D separable transforms selection in VVC.  $X$  is the input block of residuals,  $Y$  is the output transformed block and MTS flag is the index of the selected set of transforms.

#### C. Intra & Inter Prediction

In VVC, 65 intra directional intra prediction modes are used, instead of 33 in HEVC, in order to capture the arbitrary edge directions presented in a natural video. Moreover, rectangular blocks are used, compared to only square blocks in HEVC. The Intra prediction modes are coded in a list of 6 Most Probable Modes (MPM) instead of a list of 3 MPMs in HEVC. In the VTM-5.0, the results of intra prediction of planar mode are further modified by a position dependent intra prediction combination (PDPC) method. PDPC is an intra prediction method which invokes a combination of the un-filtered boundary reference samples and HEVC style intra prediction, with filtered boundary reference samples [6], [8]. Moreover, the concept of Multiple reference Line (MRL) intra prediction is introduced in VVC. MRL uses more reference lines for intra prediction while HEVC uses only the nearest reference line. The index of the selected reference line is signalled to the decoder in the bitstream.

For the Inter Prediction, the VVC includes several new and refined inter prediction coding tools, including Subblock-based Temporal Motion Vector Prediction (SbTMVP), Affine Motion Model (AFF), Bi-Directional Optical Flow (BDOF), more advanced Motion Vector (MV) prediction (inherit more information from reference, combine temporal and spatial prediction) [9].

#### D. Entropy Coding

Compared to HEVC, where the transform coefficients of a coding block are coded using non-overlapped Coefficient Groups (CGs) each of them contains the coefficients of a  $4 \times 4$  block of a coding block, VTM-5.0 uses various CGs ( $1 \times 16$ ,  $2 \times 8$ ,  $8 \times 2$ ,  $2 \times 4$ ,  $4 \times 2$  and  $16 \times 1$ ). In addition, the core of the Context Adaptive Binary Arithmetic Coding (CABAC) engine has some important changes in the VVC compared to the design in HEVC. The CABAC engine in HEVC uses a table-based probability transition process between 64 different representative probability states. In the VVC, a decode decision uses a 2-state model with variable probability updating window sizes [7].

TABLE I  
PERFORMANCE OF THE MAIN TOOLS INCLUDED IN THE VTM-5.0  
SOFTWARE [6]

Module	Tool description	BD-BR
Frame partitioning	Triangular partition mode	0.35%
	Chroma separate tree	0.14%
Transforms	Multiple Transform Selection	0.33%
	Low frequency non separable transform	0.79%
	Sub-block transform	0.41%
Inter Prediction	Subblock-based temp. merging candidates	0.43%
	Affine motion model	2.53%
	Merge with MVD	0.58%
	Bi-directional optical flow	0.78%
	Temporal motion vector predictor	1.19%
Intra Prediction	Multi-reference line prediction	0.20%
	Intra sub-partitioning	0.13%
	Matrix based intra prediction	0.27%
In-Loop Filters	Adaptive Loop Filter	4.91%
Quantization	Dependent Quantization	1.71%

### E. In-Loop Filtering

In VVC, besides deblocking filter and Sample Adaptive Offset (SAO), an Adaptive Loop Filter (ALF) with block-based filter adaptation is applied. Particularly, for the luma component, one among 25 filters is selected for each 4x4 block. This selection is mainly based on the direction and activity of local gradients. In addition, two diamond filter shapes are used; 7x7 diamond shape for luma component and 5x5 diamond shape for chroma components.

### III. VVC CODING PERFORMANCE: SUBJECTIVE STUDY

In order to evaluate, subjectively, the coding performance of VTM-5.0 compared to HEVC HM-16.20, we have performed a set of subjective quality assessment tests. We present in this section the global environment and the implementation process to perform a such experience.

#### A. Experimental environment

The subjective study has been conducted in the INSA/IETR PAVIM Lab, which is a platform for video quality monitoring actively involved in the emerging video contents. This platform includes a psycho-visual testing room, complying with the ITU-R BT.500-13 Recommendation. A display screen Ultra High Definition (4K) of 55 inches Loewe Bild 7.55 was used to visualise the video sequences. 44 observers, 30 men and 14 women aged from 20 to 55 years, have participated in this experiment. All the subjects were screened for color blindness and visual acuity using Ishihara and Snellen charts, respectively, and have a visual acuity of 10/10 in both eyes with or without correction. Finally, all participants have been gratified. Viewers are placed at distances of 1.5 and 3 times the height of the screen for UHD and HD resolutions, respectively.

#### B. Test Video Sequences

In this experiment, a set of higher resolution video sequences, from various categories (music, sport, gaming, etc.) has been selected from several datasets (Huawei, SVT, b<>com) as well as 4EVER<sup>1</sup> database. The target resolutions

<sup>1</sup>For Enhanced Video ExpeRience 2 project, [www.4ever-2.com](http://www.4ever-2.com)

for this test are HD (i.e. 1920x1080) and UHD (3840x2160), in a Standard Dynamic Range (SDR). Firstly, 13 videos sequences in 4K resolution have been selected and down-sampled using the Scalable SHVC down sampling filters. The choice of these video is mainly based on the video encoding complexity in terms of colour, movement, texture and homogeneous content. All these videos were encoded using both HEVC (HM-16.20) and VVC (VTM-5.0), at different bitrates, in HD and UHD format in RA configurations. This first selection is done in order to retain those representing a good balance of content variety and video coding artefacts. After this initial selection, only seven scenes have been retained for the experiment, as given in “Table II”. In total 77 (7 scenes  $\times$  5 bitrates  $\times$  2 codecs + 7 original scenes) video sequences are used in this study. An example snapshots of used video sequences is shown in “Fig. 4”.

TABLE II  
TEST VIDEO SEQUENCES

Sequence	HD	UHD	Fps	Bit Depth
AerialCrowd2	1920 $\times$ 1080	3840 $\times$ 2160	30	10
CatRobot1	1920 $\times$ 1080	3840 $\times$ 2160	60	10
CrowdRun	1920 $\times$ 1080	3840 $\times$ 2160	50	8
DaylightRoad	1920 $\times$ 1080	3840 $\times$ 2160	60	10
Drums2	1920 $\times$ 1080	3840 $\times$ 2160	50	10
HorseJumping	1920 $\times$ 1080	3840 $\times$ 2160	50	10
Sedof	1920 $\times$ 1080	3840 $\times$ 2160	60	8

#### C. Evaluation Procedure

In this quality assessment experiment, the Subjective Assessment Methodology for Video Quality (SAMVIQ) method was used [11]. This method has specifically been designed for multimedia content. It takes into account a range of codec types, image formats, bitrates, temporal resolutions, etc. It has been recommended by ITU-R 6Q in 2004 [12]. Each scene (video sequence) is presented with the following conditions: an explicit reference, a hidden reference and 10 processed video sequences (PVSs). 4 categories of PVS are tested (HEVC-HD, VVC-HD, HEVC-UHD and VVC-UHD). The button with label REF clearly identifies the explicit reference sequence. The hidden reference is identical to the explicit reference but it is not readily accessible to the subject and it is “hidden” among other stimuli. For each scene, participants were asked to evaluate the processed video sequences, given by buttons with letter labels A to K (including the hidden reference), as indicated by the protocol SAMVIQ [10]. The conducted experiment is divided into two parts: HD and UHD. For an optimal visual comfort, these two parts have been done separately but using the same participants. Moreover, to prevent from visual fatigue, each part (HD and UHD) of the test is carried-out in two sessions. Before each experiment, participants receive clear and deep explanations about the evaluation procedures and the used interface. Finally, all viewers scores have been collected using a dedicated Graphical User Interface (GUI), developed in compliance with the SAMVIQ recommendation.



Fig. 4. An example frames of the used video sequences.

#### IV. EXPERIMENTAL RESULTS

##### A. Objective Evaluation Results

Based on the Peak Signal-to-Noise Ratio (PSNR) metric, Table III summarises the Bjøntegaard measurement (BD-BR) for HD and UHD contents. On average, the VTM-5.0 codec enables bit rate savings of about 31% and 34% for HD and UHD (4K) video sequences, respectively. “Fig. 5” illustrates the average performance in terms of PSNR metric of the whole used dataset. According to this figure, PSNR values increase significantly when using VTM-5.0 coding tools, and consequently VTM enables higher video quality than the HM codec. Moreover, using VMAF and SSIM, the same behaviours are noticed and a significant bitrate gains is obtained. These gains are summarised in Table V.

TABLE III  
BD-BR (PSNR) OF VTM-5 COMPARED TO THE ANCHOR HM VIDEO CODEC.

Sequence	BD-BR (HD)	BD-BR (UHD)
AerialCrowd2	-24,5%	-29,9%
CatRobot1	-39,8%	-41,8%
CrowdRun	-27,2%	-30,3%
DaylightRoad	-38,9%	-42,4%
Drums2	-29,5%	-32%
HorseJumping	-31,1%	-32,5%
Sedof	-27,7%	-31,9%
<b>Average</b>	<b>-31.24%</b>	<b>-34.4%</b>

##### B. Subjective Assessment Results

First, the standard outlier detection was used to the score screening, according to the ITU-R BT.1788 recommendation [12]. Based on this outlier detection 8 subjects (6 in HD and 2 in UHD test) were discarded. Consequently, only 36 viewers scores are retained. Table IV summarises the BD-Rate gains obtained by VVC relative to HEVC considering Mean Opinion Score (MOS). As shown in this table, VTM-5.0 outperforms the HEVC for the whole used sequences. “Fig. 6 and 7” present the MOS scores in HD and UHD

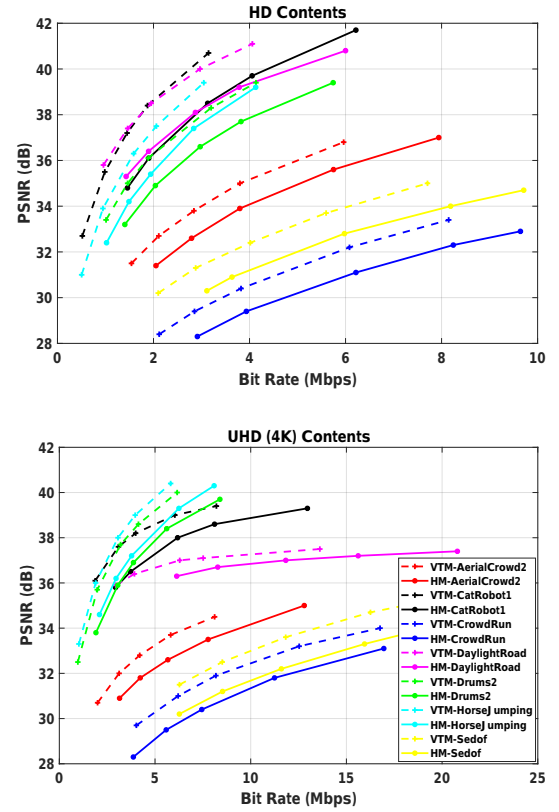


Fig. 5. Objective Quality Comparison, using PSNR, in both HD (top) and UHD (bottom) resolutions.

(4K) formats, respectively. In these figures only some example video sequences as presented, but the behaviour is similar for the rest of the used dataset. As shown in these figures, VTM-5.0 codec enables a higher MOS score and consequently a higher video quality, compared to HEVC reference software. In addition, for some sequences, a bitrate reduction of 50% is obtained with the same quality level (*Drums2* in HD and *CrowdRun* in UHD format).

As a conclusion, for HD resolution, a bitrate saving of 31% and 35% can be achieved with the VTM-5.0 in terms of PSNR and VMAF metrics, respectively. This gain exceed 40% for the UHD resolution, using VMAF metric. For the subjective comparison, the obtained gains is ranging between 37% and 40% for HD and UHD resolutions, respectively, as summarised in Table V.

TABLE IV  
BD-BR (MOS) OF VTM-5.0 COMPARED TO THE ANCHOR HM-16.2 VIDEO CODEC.

Sequence	BD-BR (HD)	BD-BR (UHD)
AerialCrowd2	-33%	-41%
CatRobot1	-48%	-47%
CrowdRun	-35%	-35%
DaylightRoad	-49%	-48%
Drums2	-35%	-39%
HorseJumping	-38%	-45%
Sedof	-19%	-21%
<b>Average</b>	<b>-37%</b>	<b>-40%</b>

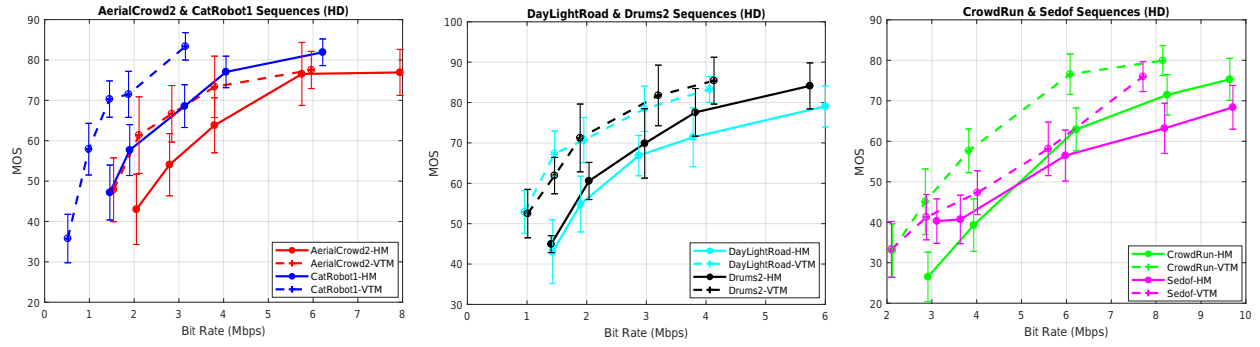


Fig. 6. MOS-based comparison, with associated 95% confidence intervals, for six example video sequences in HD format.

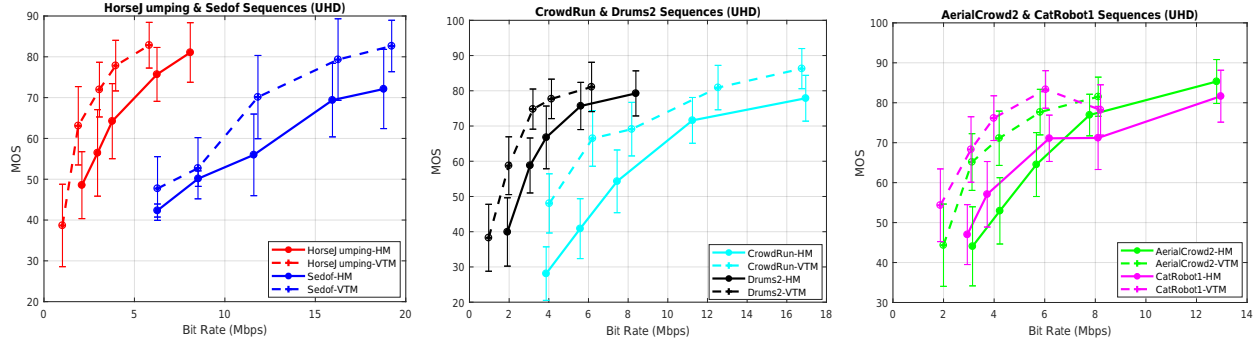


Fig. 7. MOS-based comparison, with associated 95% confidence intervals, for six example video sequences in UHD (2160p) format.

## V. CONCLUSION

In this paper, we present a subjective-based comparison, as well as an objective measurements, between the newest VVC, through the VTM-5.0 coding tools, and the well-known HEVC, using the HM-16.2 version. As results, the VTM coding tools enable a significant quality improvement compared to the HM reference software, for different videos sequences used in the experiment. Using PSNR and VMAF, as quality metrics, the obtained gain is ranging between 31% to 40% depending on video format. Subjectively, the VTM codec outperforms, in a significant manner, the HEVC reference software, especially for the low bitrate. Moreover, in some cases, we can notice that VTM enables the same perceived visual quality as HEVC with a bitrate reduction of 50%.

TABLE V  
BITRATE SAVING OF VTM-5.0 OVER HEVC STANDARD.

Resolution	PSNR	VMAF	MOS
HD	-31.24 %	-35.18 %	-37%
UHD	-34.42%	-40.44%	-40%

## REFERENCES

- [1] J-S. Gary, J-R. Ohm, W-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard" IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 16491668, Dec. 2012.
- [2] J-R. Ohm, G. Sullivan, J. Schwarz, T-K. Tan, and T. Wiegand, "Comparison of the Coding Efficiency of Video Coding Standards - Including High Efficiency Video Coding (HEVC)," IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 16691684, 2012.

- [3] N. Sidaty, W. Hamidouche, O. Deforges and P. Philippe, "Compression efficiency of the emerging video coding tools," 2017 IEEE International Conference on Image Processing (ICIP), Beijing, 2017, pp. 2996-3000.
- [4] J. Chen, E. Alshina, G. J. Sullivan, J-R. Ohm, and J. Boyce, "Algorithm description of Joint Exploration Test Model 3 (JEM3), in Document JVET-C1001 3rd JVET Meeting, Geneva, CH, May 2016.
- [5] N. Sidaty, P-L. Cabarat, W. Hamidouche, D. Menard and O. Deforges "Performance and Computational Complexity of the Future Video Coding", IEEE Workshop on Signal Processing Systems (SIPS), 2018. .
- [6] W.-J. Chien, J. Boyce, Y.-W. Chen, R. Chernyak, K. Choi, R. Hashimoto, Y.-W. Huang, H. Jang, S. Liu, and D. Luo, "JVET AHG report: Tool reporting procedure (AHG13), in JVET document O0013 (JVET-O0013), 15th JVET Meeting: Gothenburg, SE, July 2019.
- [7] B. Bross, J. Chen, S. Liu, "Versatile Video Coding (Draft 5)," document JVET-N1001, 14th JVET meeting: Geneva, MA, 1927 Mar. 2019.
- [8] F. Bossen, J. Boyce, X. Li, and V. Seregin, K. Shring, "JVET common test conditions and software reference configurations for SDR video," document JVET-N1010, 14th JVET meeting: Geneva, MA, 1927 Mar. 2019.
- [9] J. Chen Y. Ye and S-H. Kim "Algorithm description for Versatile Video Coding and Test Model 5 (VTM 5)," Joint Video Experts Team (JVET) document, 14th Meeting: Geneva, CH, 1927 Mar. 2019.
- [10] EBU BPN 056: "SAMVIQ Subjective Assessment Methodology for Video Quality," Report by EBU Project Group B/VIM (Video In Multimedia), May 2003.
- [11] F. Kozamernik, V. Steinman, P. Sunna, E. Wyckens "SAMVIQ A New EBU Methodology for Video Quality Evaluations in Multimedia," IBC 2004, Amsterdam, pp. 191 - 202.
- [12] ITU-R BT.1788 "Methodology for the Subjective Assessment of Video Quality in Multimedia Applications," 2007.