PROCEEDINGS

of the

2018 Symposium on Information Theory and Signal Processing in the Benelux

May 31-1 June, 2018, University of Twente, Enschede, The Netherlands
https://www.utwente.nl/en/eemcs/sitb2018/

Luuk Spreeuwers & Jasper Goseling (Editors)

The symposium is organized under the auspices of

Werkgemeenschap Informatie- en Communicatietheorie (WIC)

& IEEE Benelux Signal Processing Chapter

and supported by
Gauss Foundation (sponsoring best student paper award)
IEEE Benelux Information Theory Chapter
IEEE Benelux Signal Processing Chapter
Werkgemeenschap Informatie- en Communicatietheorie (WIC)

UNIVERSITY OF TWENTE.

Video Quality Assessment in Video Streaming Services: Encoder Performance Comparison

Rufat Alizada

University of Twente

Dept. EEMCS, Group TE

Drienerlolaan 5, 7522 NB Enschede, The Netherlands

`r.alizada@student.utwente.nl`

Abstract

Video streaming services over networks have increased significantly in the past decade. Maximizing end viewers quality of experience (QoE) became more crucial requirement than the quality of service (QoS) awareness when deploying new broadcast platforms for the provisioning of high-quality streaming services. Since the majority of current multi-user video streaming services are encoded and transmitted through bandwidth-limited networks, proper encoder settings are required to satisfy the total channel rate constraint. In this work a QoE-driven High Efficiency Video Coding (HEVC) encoder adaptation scheme is proposed, aiming to measure overall acceptability of video content as perceived subjectively and maximize QoE of all the users for broadcasting networks. First, the influence of HEVC encoder on video streaming is investigated. The encoding and compression efficiency of the new standard in comparison to its predecessor H.264/AVC is given. Afterwards, a QoE-maximized encoder adaptation framework is formulated based on the obtained encoder parameter model. It turned out that the proper deployment of such framework on HFC (hybrid fiber-coaxial) network may result in average bit-rate reduction of 44% and average cost savings of 20%-60%.

1 Introduction

Video Streaming has become one of the most popular applications over next-generation networks. Interest in multimedia content and services on demand has been increased significantly, creating the need for the high quality content provision and effective compression techniques. Well accustomed to a variety of multimedia devices, consumers want a flexible digital lifestyle in which high-quality multimedia content follows them wherever they go and on whatever device they use. To meet this industry requirement in a way that interoperability is reassured, standardization activities have been taken place for the various video encoding techniques. Along with the rapid development of video compression standards and network transmission technologies, video streaming application has faced to a situation where the end-user expects a high quality of experience (QoE) available. QoE, defined by ITU-T [1] as a measure of the overall acceptability of an application or service, as perceived subjectively by the end-user, is the ultimate measure of user satisfaction to be maximized in a multimedia application. Due to the bandwidth-limited nature of wireless channels, it is essential to develop efficient video compression and transmission schemes to maximize QoE of real-time video streaming applications.

High efficiency video coding (HEVC) [2], also known as H.265/MPEG-H Part 2 is being rapidly adopted and is the last generation of video coding standard finalized in January 2013 by The Joint Collaborative Team on Video Coding (JCT-VC). The HEVC is developed within the constraints imposed by real-time processing and also,

will further reduce by 50% the bit rate required for high-quality video encoding compared to the previous related work H.264/AVC standard [2]. This increase is achieved due to new and improved tools implemented in the HEVC standard, such as the highly flexible and efficient block partitioning structure, larger prediction blocks, and precise inter/intra predictions. The new block named the in-loop sample adaptive offset (SAO) filter and the algorithm of entropy encoder called Context Adaptive Binary Arithmetic Coder (CABAC).

In this paper, a QoE-driven HEVC encoder adaptation framework is formulated based on the obtained encoder parameters and objective QoE model. Within proposed framework, the assessment of the HEVC encoders implemented on hybrid fiber coaxial is performed in terms of video quality and coding performance. For this reason, a set of video signals is used as input to the reference encoders. The content of the experimental set covers different spatiotemporal activity levels, making the assessment framework of this paper to examine performance, especially in the cases where two different codecs are benchmarked.

The paper organized as follows: Section 2 presents an overview of the HEVC encoder. Section 3 contains the description of the test methodology and assessment setup. Then, the detailed experimental results are presented in Section 4, and this paper is concluded in Section 6.

2 Overview of HEVC encoder

Today, H.264/MPEG-4 AVC is the dominant video coding technology used worldwide. As a rough estimate, about half the bits sent on communication networks worldwide are for coded video using AVC, and the percentage is still growing [2]. However, the emerging use of HEVC is likely to be the inflection point that will soon cause that growth to cease as the next generation rises toward dominance.

Due to the popularity of HD video and the growing interest in Ultra HD (UHD) formats [3] with resolutions of, for example, 3840x2160 or even 7680x4320 luma samples, HEVC [4] has been designed with a focus on high-resolution video. Even though the coding of HD and UHD video was one important aspect in the HEVC development, the standard has been designed to provide an improved coding efficiency relative to its predecessor AVC for all existing video coding applications.

HEVC standard is designed along the successful principle of block-based hybrid video coding. Following this principle, a picture is first partitioned into blocks and then each block is predicted by using either intra-picture or inter-picture prediction. While the former prediction method uses only decoded samples within the same picture as a reference, the latter uses displaced blocks of already decoded pictures as a reference. Since inter-picture prediction typically compensates for the motion of real-world objects between pictures of a video sequence, it is also referred to as motion-compensated prediction. While intra-picture prediction exploits the spatial redundancy between neighboring blocks inside a picture, motion-compensated prediction utilizes a large amount of temporal redundancy between pictures. In either case, the resulting prediction error, which is formed by taking the difference between the original block and its prediction, is transmitted using transform coding, which exploits the spatial redundancy inside a block and consists of a decorrelating linear transform, scalar quantization of the transform coefficients and entropy coding of the resulting transform coefficient levels.

Each picture in HEVC is subdivided into disjunct square blocks of the same size, each of which serves as the root of a first block partitioning quadtree structure, the coding tree, and which are therefore referred to as coding tree blocks (CTBs). The CTBs can be further subdivided along the coding tree structure into coding blocks (CBs), which are the entities for which an encoder has to decide between intra-picture and motion-compensated prediction. While increasing the size of the largest supported
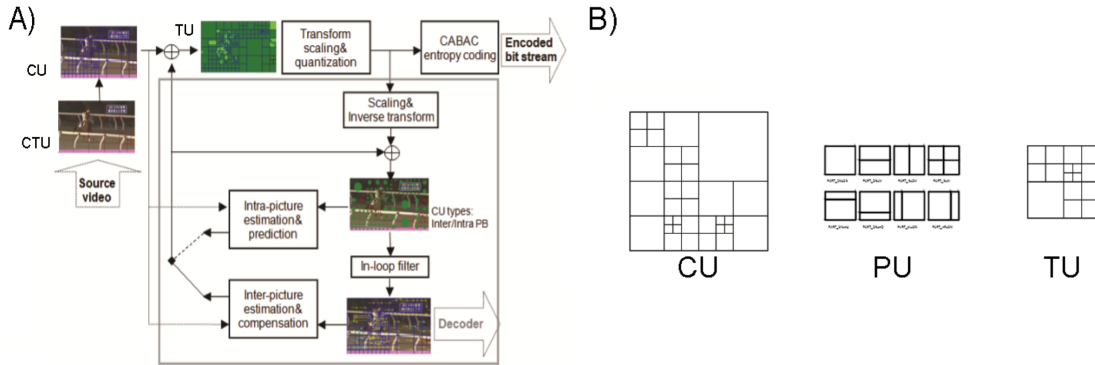
Figure 1: Picture partitioning example of coding quadtree CTU into CU (CB), partition modes for PU (PB), and transform quadtree within CU (TB).

block size is advantageous for high-resolution video, it may have a negative impact on coding efficiency for low-resolution video, in particular if low-complexity encoder implementations are used that are not capable of evaluating all supported sub-partitioning modes. For this reason, HEVC includes a flexible mechanism for partitioning video pictures into basic processing units of variable sizes.

A schematic description of the whole HEVC encoder block diagram is given in Figure 1 [5]. It receives generally an input YUV frame type and generates encoded bitstream data on its output. The HEVC encoder is based on different coding tools with high computational complexity compared to its previous standard.

2.1 Block-Based Coding

The HEVC continues to implement the block-based hybrid video coding framework, with the exception of the increased macroblock size (up to 64x64) compared to AVC. Three novel block concepts are introduced, namely: the Coding Unit (CU), the Prediction Unit (PU) and the Transform Unit (TU). CU is the basic coding unit similar to the H.264/AVCs macroblock and can have various sizes but is restricted to be square shaped. PU is the basic unit for prediction, where the largest allowed PU size is equal to the CU size. Other allowed PU sizes depend on prediction type, where asymmetric splitting options for inter-prediction is also considered. Finally, TU is the basic unit for transform and quantization, which may exceed the size of PU, but not that of the CU.

The general outline of the coding structure is formed by various sizes of CUs, PUs, and TUs in a recursive manner, once the size of the Largest Coding Unit (LCU) and the hierarchical depth of CU are defined. Given the size and the hierarchical depth of LCU, CU can be expressed as a recursive quadtree representation as it is depicted in Figure 3B, where the leaf nodes of CUs can be further split into PUs or TUs.

2.1.1 Intra-Prediction in HEVC

Intra prediction in HEVC is designed to efficiently model different directional structures typically present in video and image content. The set of available prediction directions has been selected to provide a good trade-off between encoding complexity and coding efficiency for typical video material. The sample prediction process itself is designed to have low computational requirements and to be consistent across different block sizes and prediction directions. This has been found especially important as the number of block sizes and prediction directions supported by HEVC intra coding far exceeds

3

those of previous video codecs, such as H.264/AVC. In H.264 standard, nine modes of prediction exist in a 4x4 block for intra prediction within a given frame and nine modes of prediction exist at the 8x8 level. It is even fewer at the 16x16 block level, dropping down to only four modes of prediction. Intra prediction attempts to estimate the state of adjacent blocks in a direction that minimizes the error of the estimate. In HEVC, a similar technique exists but the number of possible modes is 35-in line with the additional complexity of the codec (Figure 2). This creates a dramatically higher number of decision points involved in the analysis, as there are nearly two times the number of spatial intra-prediction sizes in HEVC as compared to H.264 and nearly four times the number of spatial intra-prediction directions.
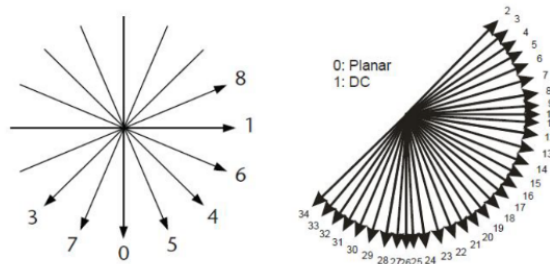

Figure 2: AVC vs HEVC Intra Prediction Modes.

2.1.2   Inter-Prediction in HEVC

The inter prediction in HEVC uses the frames stored in a reference frame buffer (with a display order independent prediction, as in AVC), which allows multiple bidirectional frame reference. A reference picture index and a motion vector displacement are needed in order to select reference area. The merging of adjacent PUs is possible, by the motion vector, not necessarily of rectangular shape as their parent CUs. In order to achieve encoding efficiency, skip and direct modes similar to the AVC ones are defined, and motion vector derivation or a new scheme named motion vector competition is performed on adjacent PUs. Motion compensation is performed with a quarter-sample motion vector precision. At TU level, an integer spatial transform (with the range from 4x4 to 64x64) is used, similar in concept to the DCT transform. In addition, a rotational transform can be used for block sizes larger than 8x8, and apply only to lower frequency components.
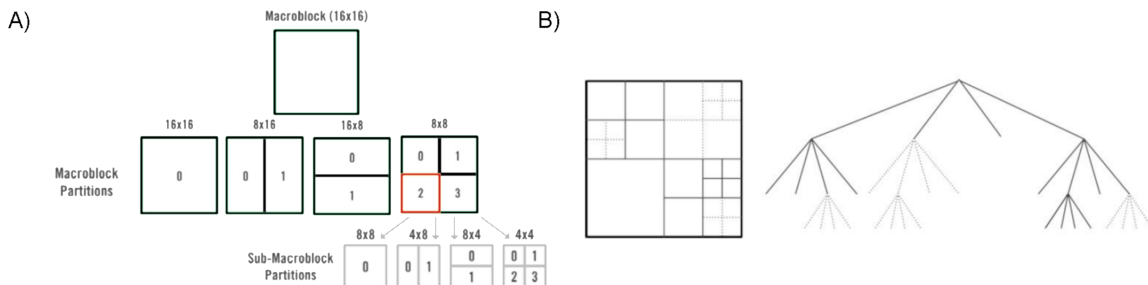

Figure 3: Inter Prediction and Coding: a) AVC Macroblock Partitions for inter prediction b) HEVC Quadtree Coding Structure for inter prediction.

4

2.1.3 Entropy Prediction in HEVC

Entropy coding is a lossless compression scheme performed at the last stage of video encoding (and first stage of video decoding), after the video signal has been reduced to a series of syntax elements. Syntax elements describe how the video signal can be reconstructed at the decoder. This includes the method of prediction (e.g., spatial or temporal prediction) along with its associated prediction parameters as well as the prediction error signal, also referred to as the residual signal. These syntax elements describe properties of the the aforementioned CU, PU and TU and loop filter (LF) of a coded block of pixels. The LF syntax elements are sent once per largest coding unit (LCU), and describe the type (edge or band) and offset for sample adaptive offset in-loop filtering.

Context-Based Adaptive Binary Arithmetic Coding (CABAC) [6] is a form of entropy coding used in H.264/AVC [7] and also in HEVC [4]. In H.264/AVC, CABAC provides a 9% to 14% improvement over the Huffman-based CAVLC [8]. CABAC involves three main functions: binarization, context modeling, and arithmetic coding. Binarization maps the syntax elements to binary symbols (bins). Context modeling estimates the probability of the bins. Finally, arithmetic coding compresses the bins to bits based on the estimated probability.

3   Test Methodology

This paper assesses the efficiency of recent implementations of the video encoders along two dimensions: the video quality obtained when a video signal is decoded at the receiver, and the computational complexity of the encoding and decoding processes. For performing the detailed performance analysis and in order to be as fair as possible due to the significant difference in the capabilities of the individual encoders, very similar settings for all tested encoders were used.

Below, the test methodology and the evaluation setup are explained in detail. Particularly, in Sub-Section 3.1, the VQ assessment is discussed, followed by the discussion of compression efficiency assessment, in Sub-Section 3.2. Finally, Sub-Section 3.3 gives an overview of the testing platform performed in HFC network.

3.1   Video Quality Assessment

In this section, firstly, the most popular current test methodologies for video quality assessment are reviewed and classified. Currently, there are many image and video quality assessment methods, each meeting different purposes. These methods can be classified in different ways depending on the criteria set adopted, as illustrated in Figure 4.

Currently, QoE under the subject of video quality assessments is roughly divided into two section that none other than Objective and Subjective methods. Most used subjective video quality assessment methods are described in the ITU recommendations ITU-R BT-500 [9] and ITU-T P.910 [10]. Aforementioned methods are focused on multimedia services. These tests are generally conducted under laboratory conditions, in which the supervisor explains the test instructions to the assessors. Later, assessors watch a test video; they grant an adjective score using 5-point MOS scale described in the Absolute Category Rating (ACR) method standardized in ITU-T Recommendation P.910. Although running subjective tests for video quality evaluation is the main way of evaluating VQ, it is rather an expensive, time consuming and tedious procedure. All this makes applying objective video metrics the best option for evaluating VQ for real-time applications.

Objective video quality assessment methods can be classified by using several considerations. Depending on the type of application service, objective methods are di-
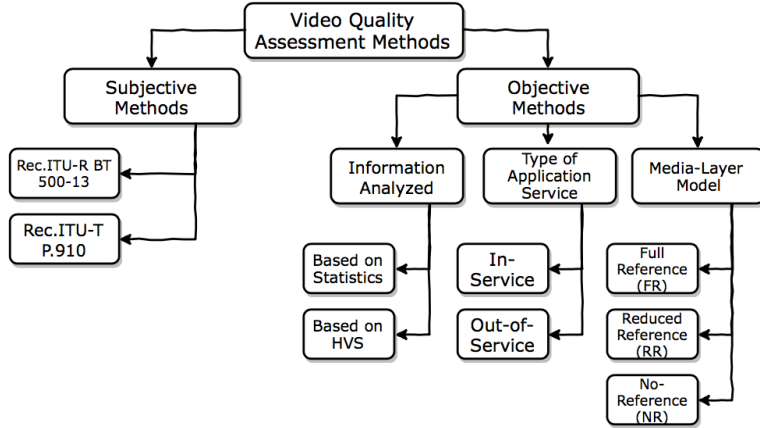
<div align="center">5</div>

Figure 4: Classification of video quality assessment methods using different criterions.

vided into two categories [11]: (1) In-service methods: Real time VQ assessment applications with time constraints [12], [13]. (2) Out-of-service methods: Do not have time restrictions and are used in different tasks, such as video codec performance evaluation and video streaming services [14], [15]. In addition, as illustrated in Figure 5, the media-layer objective quality assessment methods can be further categorized as full-reference, reduced-reference, and no-reference [13] depending on whether a reference, partial information about a reference, or no reference is used in assessing the quality, respectively.
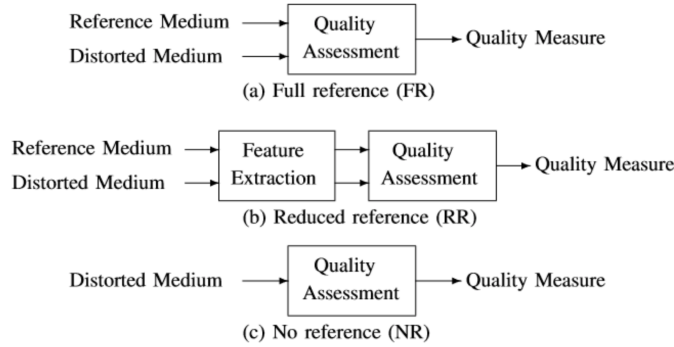

Figure 5: Overview of media layer models.

As illustrated in Figure 4, the category of information analyzed is sub-classified according to the information analyzed. The well known objective metrics, Mean Squared Error (MSE) and Peak Signal-to-Noise Ratio (PSNR) are calculated based on statistical analysis of the pixels information. In turn, the metrics Structural Similarity (SSIM) [16], Video Quality Metric (VQM) [17] and algorithms based on Region of Interest (RoI) [18] or attentions maps [19], [20] are based on the Human Visual System. The overview of the objective video quality metrics is included in the Section 3.3.

For the evaluation of the VQ, four reference video clips were chosen out of recently designed LIVE-Netflix Video QoE Database, with content that represents various levels of spatial and temporal activity. A representative snapshot of each signal is depicted on Figure 6. The test signals have spatial resolutions of 1920x1080 (HD), 3840x2160 (QFHD) and 4096x2160 (UHD) and frame rates of 30 to 60 fps. For the experimental need of this paper, test signals were encoded from their original uncompressed YUV

format to AVC and to HEVC profiles. In order to maintain and achieve ideal comparison between the various profiles, it is necessary all the profile configurations have identical or very similar parameter values.

a) Factory 1920x1080p (HD/Animated)   b) Ducks 3840x2160p (QFHD)

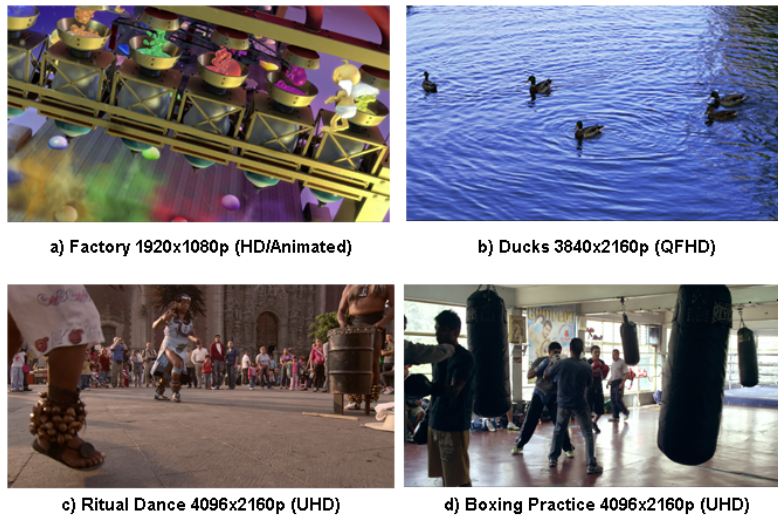c) Ritual Dance 4096x2160p (UHD)   d) Boxing Practice 4096x2160p (UHD)

Figure 6: Representative frames taken from the test signals.

Main objective VQ metric used in this research is a reduced-reference video QoE measure, SSIMPLUS, that provides real-time prediction of the perceptual quality of a video based on human visual system behaviors, video content characteristics (such as spatial and temporal complexity, and video resolution), display device properties (such as screen size, resolution, and brightness), and viewing conditions (such as viewing distance and angle) [21]. SSIMPLUS assessment model combines the most significant HVS features, which include spatial frequency sensitivity, luminance masking, texture masking, temporal frequency sensitivity, and short-term memory effect. Moreover, it has a simple and clear structure and is easy for software implementation.

3.2   Compression Assessment

Compression efficiency is the most fundamental driving force behind the adoption of modern digital video compression technology, and HEVC is exceptionally strong in that area. However, it is also important to remember that the standard only provides encoders with the ability to compress video efficiently, it does not guarantee any particular level of quality since it does not govern whether or not encoders will take full advantage of the capability of the syntax design. As it was already mentioned, AVC and HEVC codecs follow the "block-based hybrid" coding approach. This type of coding exploits the spatial and temporal redundancy of the video frames. Frames are divided into three types: (1) I-frames (Intra Coded Picture) serve as an anchor for other frames to be decoded; (2) P-frames (Predictive Coded Picture) are predicted in a temporal manner only from previous frames (P-or I-frames); (3) B-frames (Bidirectional Coded Picture) are predicted from predicted from previous and following frames (I- P- or even B-frames) and achieve highest compression rate. The frames contained between two consecutive I-frames are called Group of Picture (GOP) [22]. The visual quality usually decreases with the GOP size. Thus, investigating the effect of encoder settings, such as input sequence resolution, frame rate, the length of GOP and quantization parameter (QP), on HEVC encoded video quality is necessary and challenging.
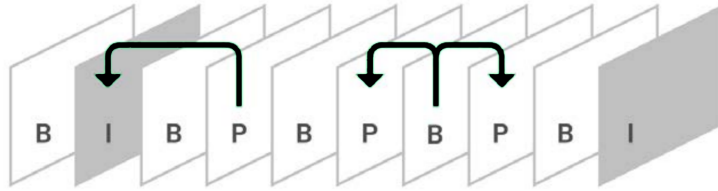
7

Figure 7: A typical sequence with I-, B- and P-Frames.

For the purpose of this analysis, multiple video sequences were selected and evaluated. All videos contain a non-static picture in a duration of ten seconds. One of the test sequences, named *Ducks* contains scenes of multiple ducks taking off from the surface of the water. This sample contains a large amount of changes and effects to stress various aspects of processing and test the stability of the encoder. Next two samples named *Ritual Dance* and *Boxing Practice* contain fast action artifacts and with the introduction of different amount scene changes and effects. The reason for these three different video sequences being evaluated is to determine if scene structure has some influence on the compression efficiency evaluation. In order to achieve accurate comparison, the GOP structure for all the encoding profiles and between two encoding methods consisted of the same I, P and B sequence, ensuring accurate benchmarking of both Intra and Inter-coding methods of AVC and HEVC profiles. The assessment of compression efficiency of HEVC is realized through objective measurement software Elecard StreamEye Tool by Elecard [23]. StreamEye is practical objective visual distortion measurement model for digital video compression. The primary purpose of StreamEye is to evaluate the video coding algorithms for the compression and visual quality through the comparison with the reference raw data.

3.3 Specification of testing platform

To evaluate how encoders perform, number of test sequences are prepared and the objective quality scores are collected. In contrast with previous studies, the sequences under test included ultra high definition footage which had been transcoded and compressed to varying degrees. The efficiency of the encoders is examined to extend to which the various objective metrics are compared and assessed. This section describes sequence preparation, gives an overview of used objective quality metrics and provides specifications of testing platform.

Table 1: Characteristics of test material used in the evaluation

File Name	Video Codec	Video Profile	Resolutions	fps	Bit Depth
Factory	AVC/HEVC	High@L5.1	1920x1080	30	8
Ducks	AVC/HEVC	High@L5.1	3840x2160	50	8
Ritual Dance	HEVC	Main@L5.1@Main	4096x2160	60	8
Boxing Practice	HEVC	Main@L5.1@Main	4096x2160	60	8

Thirty test video sequences are generated by subjecting four different original undistorted HD and UHD video sequences (*Factory, Ducks, Ritual Dance* and *Boxing Practice*) to AVC/HEVC coding schemes with different bitrates (2 to 5 Mbps for HD and 10 to 30 Mbps for UHD content). Selected bit rates are chosen to represent different real-life HDTV consumer and broadcasting applications from IPTV at the lower end to UHD on the upper end of the bitrate scale. Some target bit rates were rather aggressive

8

in order to be able to evaluate encoders at a point where they were stressed to process the content.

Evaluation of all test scenarios include five common objective video quality metrics. Chosen metrics are described below. Those metrics were chosen to represent a number of different approaches to the quality assessment problem. Aforementioned statistical metrics like PSNR and SSIM are included to provide a baseline against which the other metrics can be compared.

1. *Peak Signal-to-Noise Ratio (PSNR):* Traditional image quality assessment method. It is the ratio between the maximum power of the signal and the power of the difference signal between the reference and test images.

2. *Structural Similarity (SSIM)[16]:* Measures the structural similarity between the reference and test images based on the assumption that HVS is adapted for extracting structural information from a scene.

3. *Visual Information Fidelity (VIF)[25]:* Approaches quality evaluation by attempting to quantify distortion-induced differences in source information which can be usefully processed by the HVS. Scores range from 0 (worst) to 1 (best).

4. *Video Multi-Method Assessment Fusion (vmaf)[25]:* Objective FR VQ metric developed by Netflix. The metric is fusion of VIF and Detail Loss Metric (DLM), an image quality metric based on rationale of separately measuring the loss of details which affects the content visibility, and the redundant impairments which distracts viewer attention.

5. *Perceptual Fidelity (PF)[21]:* Novel SSIMPLUS QoE measure that provides straightforward predictions on what an average consumer says about the quality of the video content being delivered on a scale of 0-100 and also categories the quality as either bad, poor, fair, good or excellent.

First test scenario considers the direct output of the encoders. The compression efficiency evaluation of all test sequences is performed over the first test scenario. Second test scenario considers a video streaming service over hybrid fiber-coaxial (HFC) access network. In HFC network, the content is sent from the cable system's distribution facility to local communities through optical fiber subscriber lines. The tests were performed over this network in order to replicate real-life television broadcasting and simulate various cases of network conditions, that may cause "damage" of video stream, in particular, due to delay and packet losses in the network. Third test scenario considers the evaluation of the content by utilizing the device-adaptation feature of SSIMPLUS software [26]. This feature takes into consideration the fact that human quality assessment of the same video content can be significantly different when it is displayed on different viewing devices, such as HDTV, digital TV, projectors, PCs, and smartphones, and many more. The outcome of the third test scenario could be used to adapt video QoE analysis to any display device and viewing conditions.

In case of the first scenario, the reference software was used for both AVC and HEVC coding. For this work, the FFmpeg library libx264 which is the x264 H.264/MPEG-4 AVC encoder wrapper and libx265 which is the x265 H.265/HEVC encoder wrapper were used to encode the content. FFmpeg is an easy to use open source software capable of performing a wide range of multimedia operations including transcoding, encoding and conversion of audio/video content [27]. In this work, FFmpeg version 3.4.1, built with Apple LLVM version 9.0 was used for encoding the videos. Encoding was performed on a MacBook Pro laptop with 8 GB RAM, Intel Iris and Intel Dual Core i5@2.70GHz running 64-bit macOS High Sierra v.10.1.3. Rest of the evaluation was carried out on a AWS Elemental Server rack, also equipped with Intel multi-core technology. This Intel platform is composed of two Intel Core i7-6820HQ processors,

running at 2.7 GHz, and 16GB of DDR4 memory. AWS Elemental Live software v.4.0 is implemented over the server and provides support for all recent audio/video coding schemes.

4 Evaluation

4.1 Video Quality of HEVC

For obtaining experimental results, most of the test sequences were selected according to the encoder test conditions, as presented in Table 1. All sequences include different texture and motion characteristics to bring up a reasonable relationship between bit rate, PSNR, and encoding settings.

Figure 8 illustrates VQ curves of HEVC and x264 encoders for two typical examples of tested sequences. As it is clearly seen from Figure 8, the HEVC encoders provides significant gains in terms of perceived quality compared to its predecessor AVC.
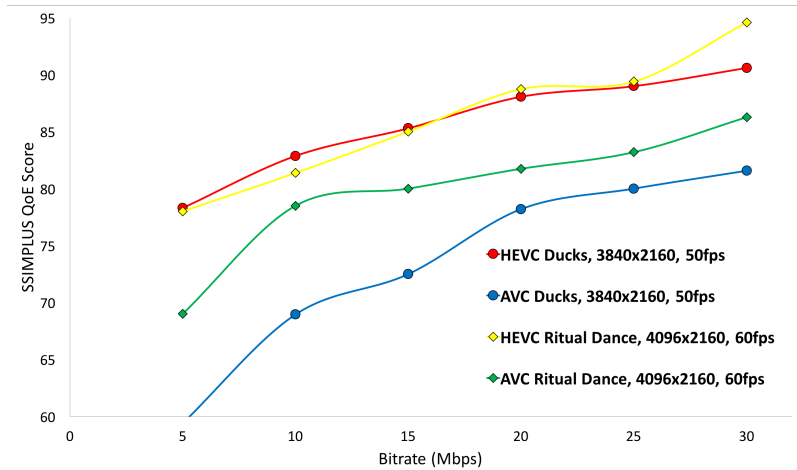

Figure 8: Video quality assessment for several typical examples of tested sequences.

Table 2 illustrates the results comparing all possible conditions for the 1080p and 4k content, respectively. Comparing HEVC and AVC at similar bit rates, HEVC always provides statistically better visual quality when compared to AVC for *Ducks*, *Ritual Dance* and *Boxing Practice*. The table also includes "Motion" metric, a simple measure of the temporal difference between adjacent frames. The score typically ranges from 0 (static) to 20 (high-motion). This simple feature illustrates that content under assessment contains a large amount of changes and effects to stress various aspects of the processing of encoder. For the animated content *Factory*, there is not sufficient statistical evidence to show that HEVC outperforms AVC, especially at high bit rates.

4.1.1 QoE driven bandwidth optimization

Once the quality of content is evaluated, many benefits come as a natural next step. One of such benefits is bandwidth optimization. Although a significant number of solutions have been proposed in the industry for saving bandwidth, talking about bandwidth reductions without maintaining the right level of visual QoE makes little sense. Due to the lack of proper QoE assessment tools, existing bandwidth saving approaches, whether it is applied to encoding/transcoding or streaming optimization, result in unstable results. To perform bandwidth optimization properly, the first step

10

Table 2: Video quality assessment of 1080p and UHD content encoded at 10, 20 and 30 Mbps. Including the comparison of the HEVC and AVC encoding schemes.

Content	VQM	HEVC			AVC		
		10Mpbs	20Mpbs	30Mpbs	10Mpbs	20Mpbs	30Mpbs
QFHD 3840x2160 Ducks 500 frames Motion: 4.84	PSNR	27.7	28.8	29.3	25.59	26.79	27.4
	SSIM	0.67	0.71	0.73	0.59	0.64	0.66
	vmaf	48.36	58.57	64.42	33.69	42.25	46.8
	PF	82.87	88.07	90.6	68.95	78.2	81.57
	VIF	0.68	0.72	0.74	0.61	0.64	0.67
UHD 4096x2160 Ritual Dance 600 frames Motion: 16.59	PSNR	27.4	28.7	29	26.1	26.4	27.2
	SSIM	0.64	0.69	0.71	0.63	0.65	0.67
	vmaf	46.45	59.66	62.51	43.11	47.2	54.68
	PF	81.38	88.74	94.61	78.49	81.74	86.27
	VIF	0.67	0.72	0.8	0.64	0.67	0.71
UHD 4096x2160 Boxing Practise 600 frames Motion:12.13	PSNR	27.4	28.3	28.9	27.9	28	28.1
	SSIM	0.67	0.7	0.72	0.64	0.66	0.69
	vmaf	48.39	58.98	61.41	41.08	48.53	55.42
	PF	84.78	87.26	92.95	74.81	84.06	87.43
	VIF	0.69	0.7	0.76	0.62	0.68	0.7

has to be adopting a trusted QoE metric with powerful functionalities, e.g., accurate, meaningful and consistent quality assessment cross resolutions, frame rates, dynamic ranges, viewing device and video content. An illustrative example using SSIMPLUS as the example QoE metric is given to demonstrate how large bandwidth savings can be achieved in live and file-based operations by making use of such a QoE metric.

Significant bandwidth savings can be obtained by adopting a QoE measure that produces consistent QoE assessment across content, resolution, and user device, each of which could lead to significant gain. Firstly, because of the difference in encoding difficulty of different content (Sample 3 and Sample 4 shown in the Fig.9), to reach a guaranteed QoE quality level (SSIMPLUS = 90), using a fixed bandwidth to encode all videos may be a waste, depending on video content, e.g., using a fixed 30 Mbits when only 27 Mbits is necessary for Sample 4.
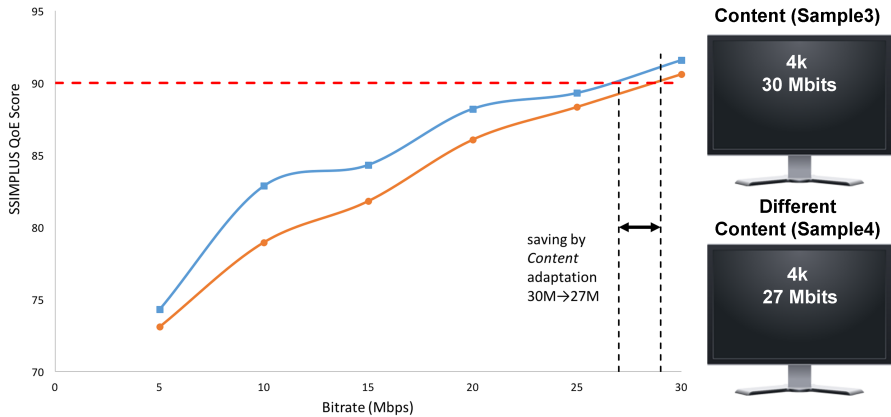

Figure 9: Illustration of how bandwidth savings is achieved by using a QoE metric that is able to adapt to video content.

Second, when the same content is encoded to two or more spatial/temporal resolutions, the capability of picking the most cost-effective spatial/temporal resolution to achieve the guaranteed quality level can also help save large bandwidth, e.g., a bandwidth reduction from 3.4 Mbits to 2.3 Mbits is obtained by switching from 1080p to 720p resolutions, as shown in the Figure 10.
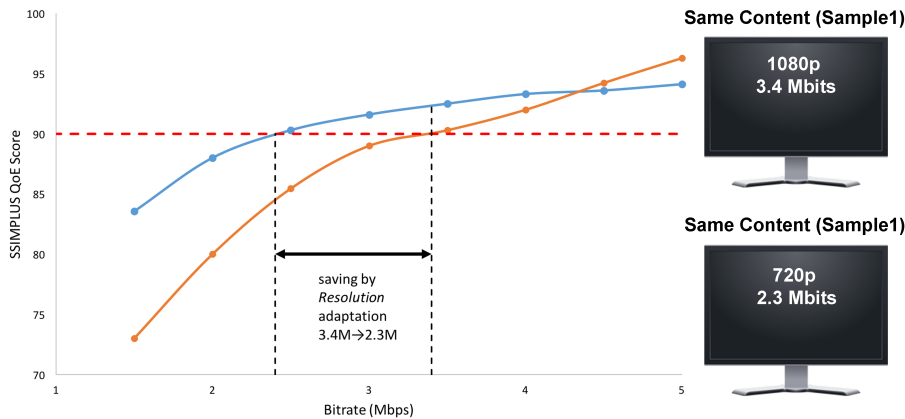
11

Figure 10: Illustration of how bandwidth savings is achieved by using a QoE metric that is able to adapt to video resolution.

Finally, the perceptual QoE varies significantly on different viewing devices. This is illustrated in the quality-bitrate curve in the Figure 11, which shows that when the user is known to use a smartphone rather than a TV to watch the video, a bandwidth of 17 Mbits is sufficient to achieve the same target quality level (SSIMPLUS = 90). With all three factors combined, a total of 44% bandwidth savings may be obtained (from 30 Mbits to 17 Mbits).
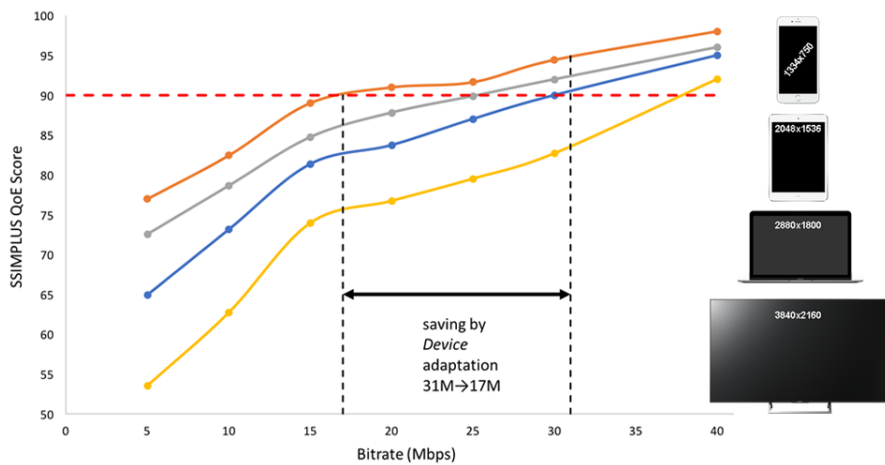

Figure 11: Illustration of how bandwidth savings is achieved by using a QoE metric that is able to adapt to user viewing device.

Although given examples are here for illustration purposes only, and in practice users may be constrained to explore all three factors for maximum cost-savings, conducted research suggests that for most video content and the most common usage profiles, an average cost saving of 20%-60% is typically achieved by properly adopting this QoE metric-driven bandwidth optimization technology. Such bandwidth savings can be implemented by adaptive operation of video encoders/transcoders, and may also be incorporated into adaptive streaming frameworks to achieve similar goals in a dynamic way.

12

5   Coding Efficiency of HEVC

For illustration, Figure 12 shows the partitioning of a picture with 3860x2160 luma samples into 16x16 macroblocks and 64x64 CTUs. It can be seen that 16x16 macroblock covers only a very small area of a picture, much smaller than the regions that can typically be described by the same motion parameters. Taking into account that some of the CTUs will be subdivided for assigning different prediction modes and parameters, the partitioning into 64x64 CTUs provides a more suitable description.
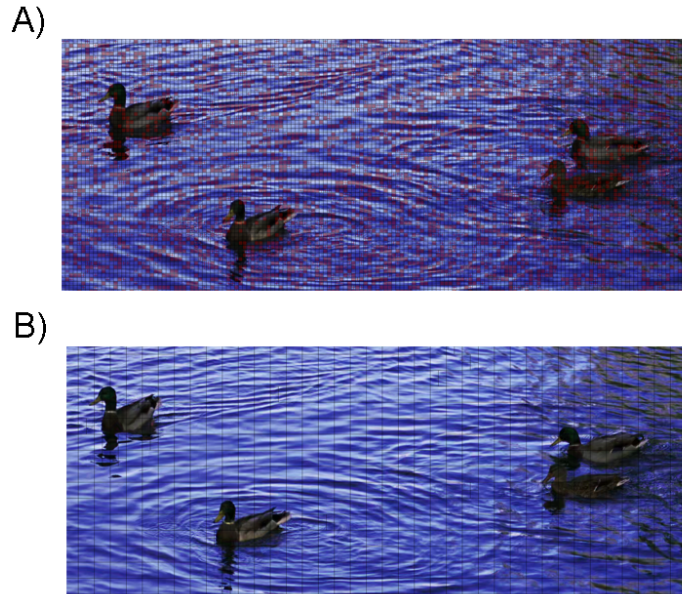

Figure 12: Illustration of the partitioning of a picture with 3840x2160 luma samples into macroblocks and coding tree units: (a) Partitioning of the picture into 16x16 macroblocks as found in all prior video coding standards of the ITU-T and ISO/IEC; (b) Partitioning of the picture into 64x64 coding tree units, the largest coding tree unit size supported in the Main profile of HEVC.

5.1   Performance Comparison of HEVC vs AVC

This section presents evaluation of the compression efficiency of the HEVC algorithm in comparison to the AVC for the test signals under test, when same encoding parameters have been selected.

The first step in compressing of content is to segregate the data into different classes. Depending on the importance of the data it contains, each class is allocated a portion of the total bit budget, such that the compressed data has the minimum possible distortion. This procedure is called bit allocation. Based upon the demanded bit rate and the current fullness of the buffer, a target bit rate for the entire GOP is determined, together with the QP for the GOP's I-picture and first P-picture. In video coding, it is expected that the encoder could adaptively select the encoding parameters to optimize the bit allocation to different sources under the given constraints. Table 3 illustrates the bit allocation of the encoders under the assessment. The experimental results for the x264 reference encoder illustrate that some frames in the assessed content were encoded using 20.2 Mbits, even tough target bitrate was set to 10 Mbits. In addition, Table 3 illustrates that AWS Elemental encoder has the most accurate rate controller that

13

tries to allocate available bit budget equally between video units (macroblock, frame, GOP). Proper usage of bit allocation data could be used for efficient management of available bit budget and accurate performance comparison of the encoders.

Table 3: Bit allocation of encoders under comparison for *Ducks* encoded at 10 Mbps

	Bit Allocation (Mbits)		
Encoders	Maximum	Average	Minimum
x264 AVC	20.2	10.5	9.2
x256 HEVC	11.9	10.1	8.7
AWS Elemental HEVC	10.4	10.1	10

To evaluate performance even further, rate-distortion assessment of HEVC in terms of the bit budget is performed. The bit rate reduction of one codec over another for a similar quality is estimated using the Bjøntegaard Delta Rate (BD-Rate) [28]. The Bjøntegaard model relies on PSNR measurements to determine the average bit-rate difference for the same objective quality. Since evaluation is performed with *yuv420* content, seperate PSNR values are obtained for luma (Y) and chroma (U,V) components. Combined $PSNR_{YUV}$ value are calculated as a weighted sum of the PSNR values per each frame of each individual component [29], as shown in Eq.1.

$$PSNR_{YUV} = \frac{6 \times PSNR_Y + PSNR_U + PSNR_V}{8} \qquad (1)$$

Although a more realistic estimate of the performance efficiency can be obtained by considering subjective ratings instead of PSNR values. Using the combined $PSNR_{YUV}$ in rate-distortion assessment could be used to determine the trade-offs between luma and chroma component fidelity [29].
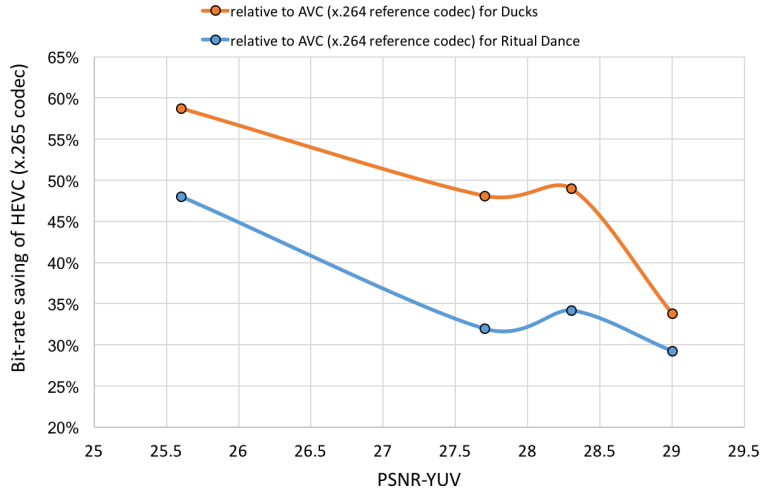

Figure 13: Bit-rate saving plots for several typical examples of tested sequences.

Figure 13 present rate-distortion curves of HEVC bitrate savings for two typical examples of tested sequences. As it is clearly seen, the HEVC provides significants gains in term of coding efficiency compared to H.264/AVC. Table 4 provides a summary of the bitrate reduction results, where negative BD-rate values indicate bitrate savings in contrast to positive values, which indicate the required overhead in bitrate to achieve the same $PSNR_{YUV}$ values.

14

Table 4: Summarized BD bit rate experimental results.

	BD-rate in %		
Encoders	x265 HEVC	x264 AVC	AWS Elemental HEVC
x256 HEVC		-47.3	19.2
x264 AVC	46.6		69.1
AWS Elemental HEVC	-21.4	-64.3	

As shown in Table 4, AWS Elemental HEVC outperforms both reference codecs. The average BD bit rate savings of AWS Elemental HEVC encoder relative to AVC and HEVC (x265) are 64.3% and 21.4%, respectively. As it is also observed from Table 4, the bit rate savings, on average, the HEVC (x265) encoder achieves an average gain of 47.3% in terms of bit-rate savings compared to AVC. Proper usage of BD bit rate values could be used for accurate performance comparison of the encoders.

5.2 Encoding Efficiency of HEVC

This section presents the experimental results of the comparison between HEVC and AVC encoded signals. Across the encoding process of both the HEVC and AVC profiles, all encoding parameters remained identical in order to quantitatively compare encoding efficiency of the encoder. Figure 14 illustrates the VQ evaluation per single frame, without focusing only on the average results. This facilitates the full-length content comparison together with opportunities to highlight the frames with higher complexity and discover anomalies in the transcoded files.
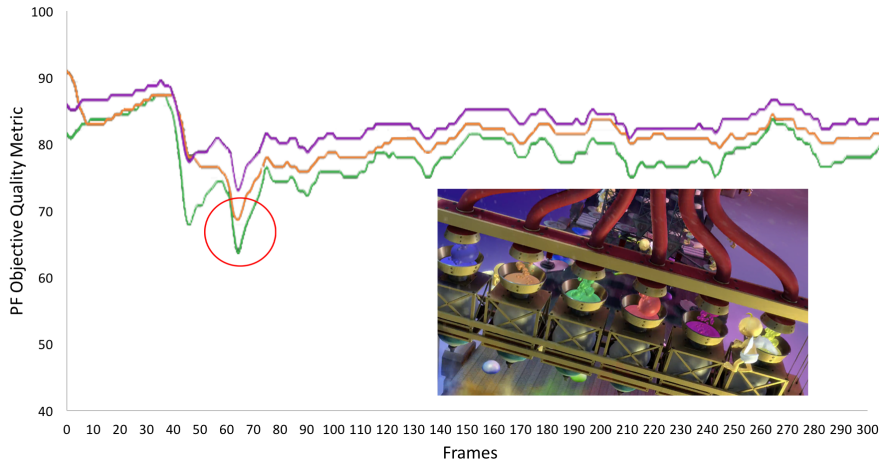

Figure 14: Per-frame VQA for several codecs.

Measurement values illustrated in Figure 14 are different, but the behaviour is similar between multiple codecs. This means that the most of the peaks are on the same scene/frame of the asset. The red circle highlights the scene or frame(s) with the highest complexity to transcode. The screenshot of this particular frame is also illustrated in Figure 14. This particular frame of animated content contains relatively fast moving objects with the complex color background, which explains why tested encoder struggles to meet its bit budget.

As it can be observed from experimental results, the encoding efficiency of the HEVC encoded signals appear to be better compared to the AVC/H.264. This is in

15

line with the objective of the HEVC to maintain the encoding efficiency compared to its predecessor AVC, while almost doubling compression and performance efficiency of the bitstream, as shown in the previous section. Commenting further on the experimental results, although the average PSNR score of the HEVC is similar to the respective one of the AVC in some cases, PF results illustrate that the HEVC encoding performance appears to have greater variance with higher QoE, that outperforms instantly the AVC performance.

6 Conclusions

This paper presents a detailed description of the objective quality evaluation tests conducted to benchmark the performance of HEVC and AVC video codecs for real-time video applications. The evaluation was performed using various HD and UHD content encoded at various bit rates. High accuracy software assessment tools were used to accurately compare the performance of the investigated codecs. Evaluation of test results shows that HEVC offers improvements in compression performance when compared to AVC, if one considers a wide range of bit rates from low to high, corresponding to video with low to transparent quality. More specifically, objective based QoE measurements show that average bit-rate reduction of 43% and average cost saving of 20%-60% is typically achieved by properly adopting this QoE metric-driven bandwidth optimization technology.

Acknowledgments

References

[1] (ITU-T), *"ITU-T Recommendation G.1070 Opinion model for video telephony applications"*, Tech. Rep.,2012

[2] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," Circuits and Systems for Video Technology, IEEE Transactions on, vol. 22, no. 12, pp. 16491668, 2012.

[3] ITU-R Rec. BT.2020 (2012) Parameter values for ultra-high definition television systems for production and international programme exchange.

[4] ITU-T Rec. H.265 and ISO/IEC 23008-10 (2013) High efficiency video coding.

[5] B. Bross, W.-J. Han, J.-R. Ohm, G. J. Sullivan, and T. Wiegand, "High efficiency video coding (HEVC) text specification draft 8," JCTVC-J1003 July, 2012.

[6] Marpe D, Schwarz H, Wiegand T, "Context-based adaptive binary arithmetic coding in the H.264/AVC video compression standard," IEEE Trans CSVT 13(7):620636 (2003)

[7] ITU-T Rec. H.264 and ISO/IEC 14496-10 (2003) Advanced video coding

[8] Alshina E, Alshin A, "Multi-parameter probability up-date for CABAC, Joint Collaborative Team on Video Coding (JCT-VC)", Document JCTVC-F254, Torino, July 2011

[9] ITU-R Recommendation BT.500, Methodology for the Subjective Assessment of the Quality of Television Pictures, ITU-T, Geneva, Switzerland, Jan. 2012.

[10] ITU-T Recommendation P.910, Subjective Video Quality Assessment Methods for Multimedia Applications," ITU-T, Geneva, Switzerland, Apr. 2008.

[11] S. Chikkerur, V. Sundaram, M. Reisslein, and L. Karam, "Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison.," IEEE Trans. on Broadcasting, vol. 57, no.2, pp. 165-182, Jun. 2011.

[12] K. Yamagishi and T. Hayashi, "Parametric packet-layer model for monitoring video quality of IPTV services," in Proc. Int. Conference of Communications, pp. 110-114, Beijing, China, May 2008.

[13] M. Garcia and A. Raake, "Impairment-factor-based audio-visual quality model for IPTV," in Proc. International Workshop on Quality Multimedia Experience, pp. 1-6, California, US., Jul. 2009.

[14] M. Martines, M. Lopez, P. Pinol, M. Malumbres, and J. Oliver, "Study of Objective Quality Assessment Metrics for Video Codec Design and Evaluation," in Proc. IEEE International Symposium on Multimedia, pp. 517-524, California, US., Dec. 2006.

[15] B. Ciubotaru, G.-M. Muntean, and G. Ghinea, "Objective assessment of region of interest-aware adaptive multimedia streaming quality," IEEE Trans. on Broadcasting, vol. 55, no. 2, pp. 202-212, Jun. 2009.

[16] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," IEEE Trans. Image Processing 13, 600612 (Apr. 2004).

[17] M. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," IEEE Trans. Broadcast., vol. 50, no. 3, pp. 312322, Sep. 2004.

[18] H. Kwon, H. Han, S. Lee, W. Choi, and B. Kang, "New Video Enhancement Preprocessor Using the Region-Of-Interest for the Videoconferencing," IEEE Trans. Consumer Electron., vol. 56, no. 4, pp. 2644-2651, Nov. 2010.

[19] J. You, A. Perkis, M. Gabbouj, and M. M. Hannuksela, "Perceptual quality assessment based on visual attention analysis," in Proc. International Conference on Multimedia, pp. 561564, Beijing, China, May 2009.

[20] A. K. Noorthy and A. C. Bovik, "Visual importance pooling for image quality assessment," IEEE J. Select. Topics Signal Processing, vol. 3, no. 2, pp. 193201, Apr. 2009.

[21] SSIMPLUS: The most accurate video quality measure, `https://www.ssimwave.com/from-the-experts/ssimplus-the-most-accurate-video-quality-measure/`

[22] Sze V, Budagavi M, Sullivan G.J., "High Efficiency Video Coding (HEVC), Algorithms and Architectures," Springer; 2016.

17

[23] ELECARD StreamEye: Video analysis test software, `https://www.elecard.com/products/video-analysis/streameye`

[24] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Temporal effects on subjective video quality of experience," Transactions on Image Processing, under review.

[25] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric." `http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html`

[26] A. Rehman, K. Zeng and Z. Wang, "Display device-adapted video quality-of-experience assessment," IST/SPIE Electronic Imaging: Human Vision Electronic Imaging, Feb.2015.

[27] FFMPEG. FFmpeg and H.264 Encoding Guide. `https://trac.ffmpeg.org/wiki/Encode/H.264`

[28] G. Bjøntegaard, "Calculation of average PSNR differences between RD-curves", ITU-T Q.6/SG16 VCEG 13th Meeting, Document VCEG-M33, Austin, USA, Apr. 2001.

[29] J. Ohm, G.J. Sullivan, H. Schwarz, T.K. Tan, and T. Wiegand, "Comparison of the coding efficiency of video coding standardsincluding High Efficiency Video Coding (HEVC)," Circuits and Systems for Video Technology, IEEE Transactions on , vol. 22, no.12, pp.1669-1684, Dec. 2012.