BRNO UNIVERSITY OF TECHNOLOGY

VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

FACULTY OF ELECTRICAL ENGINEERING AND COMMUNICATION

FAKULTA ELEKTROTECHNIKY
A KOMUNIKAČNÍCH TECHNOLOGIÍ

DEPARTMENT OF RADIO ELECTRONICS

ÚSTAV RADIOELEKTRONIKY

VIDEO AND DATA SERVICES QUALITY IN THE FUTURE BROADBAND MULTIMEDIA SYSTEMS AND NETWORKS

KVALITA OBRAZU A SLUŽEB V ŠIROKOPÁSMOVÝCH MULTIMEDIÁLNÍCH SÍTÍCH A SYSTÉMECH BUDOUCNOSTI

DOCTORAL THESIS
DIZERTAČNÍ PRÁCE

AUTHOR Ing. Jan Kufa
AUTOR PRÁCE

SUPERVISOR prof. Ing. Tomáš Kratochvíl, Ph.D.
ŠKOLITEL

BRNO 2018

ABSTRAKT

Téma doktorské práce je zaměřeno na analýzu zpracování signálů v širokopásmových multimediálních sítích a systémech budoucnosti, kde se předpokládají systémy s ultra vysokým rozlišením (UHDTV), vysokým snímkovým kmitočtem (HFR) a stereoskopické systémy (3D). Tyto systémy budou umožňovat vysoce účinnou zdrojovou kompresi obrazu, zvuku a dat a také jejich vysoce účinný přenos, a to jak při volném vysílání (např. DVB-T2), tak ve službách placené televize (např. IPTV). Cílem práce je analýza a vyhodnocení kvality obrazu a služeb v těchto systémech na základě objektivních metrik a subjektivních testů. Práce se dále zaměřuje na analýzu vnímané kvality u stereoskopické televize, kódovací účinnost moderních stereoskopických enkoderů a vlivu sekvencí na uživatelský komfort.

KLÍČOVÁ SLOVA

HEVC, VP9, HFR, Kvalita videa, PSNR, SSIM, VQM, Ultra HD, Kódování, 3DTV.

ABSTRACT

The doctoral thesis is focused on the analysis of signal processing in future broadband multimedia networks and systems, where ultra-high definition televisions (UHDTV), high frame rate (HFR) videos and stereoscopic systems (3D) are expected. These systems will enable very high source coding and compression of video, audio and data, and also very effective transmission, even in free-to-air television broadcasting (e.g. DVB-T2) or on-demand television systems (e.g. IPTV). The aim of the work is to analyse and evaluate video quality based on objective metrics and subjective tests. The work also focuses on the analysis of perceived quality in stereoscopic televisions, the coding efficiency of modern stereoscopic encoders and the influence of stereoscopic sequences on user comfort.

KEYWORDS

HEVC, VP9, HFR, Video Quality, PSNR, SSIM, VQM, Ultra HD, Encoding, 3DTV.

KUFA, J. Kvalita obrazu a služeb v širokopásmových multimediálních sítích a systémech budoucnosti. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. 84 s. Vedoucí dizertační práce prof. Ing. Tomáš Kratochvíl, Ph.D..

DECLARATION

I declare that I have written my doctoral thesis on the topic of "Video and Data Services Quality in the Future Broadband Multimedia Systems and Networks" independently, under the guidance of the treatise on doctoral thesis supervisor and using technical literature and other sources of information which are all quoted in the treatise and detailed in the list of literature at the end of the treatise.

As the author of the doctoral thesis, I furthermore declare that regarding the creation of this treatise on a doctoral thesis, I have not infringed any copyright. In particular, I have not unlawfully encroached on anyone's personal and/or ownership rights and I am fully aware of the consequences in the case of breaking Regulation § 11 and the following of the Copyright Act No 121/200 Sb., and of the rights related to intellectual property right and changes in some Acts (Intellectual Property Act) and formulated in later regulations, inclusive of the possible consequences resulting from the provisions of Criminal Act No 400/2009 Sb., Section, Head VI, Part 4.

Brno ....

(author's signature)

ACKNOWLEDGMENT

I would like to thank my supervisor prof. Ing. Tomáš Kratochvíl, Ph.D. for his support and encouragement during my study and work.

Brno …………   ………………………….

(author's signature)

SIX research centre
sensor, information and communication systems

Faculty of Electrical Engineering
and Communication

Brno University of Technology
Technicka 12, CZ-61600 Brno, Czech Republic

http://www.six.feec.vutbr.cz

Experimentální část této disertační práce byla realizována na výzkumné infrastruktuře vybudované v rámci projektu CZ.1.05/2.1.00/03.0072
Centrum senzorických, informačních a komunikačních systémů (SIX)
operačního programu Výzkum a vývoj pro inovace.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

3D	Three dimensional
ACR	Absolute Category Rating
AVC	Advanced Video Coding
AVX	Advanced Vector Extensions
CABAC	Content Adaptive Binary Arithmetic Coding
CPU	Central Processor Unit
CTU	Coding tree unit
DS	Double Stimulus
CI	Confidence Interval
DTT	Digital Terrestrial Television
DVB	Digital Video Broadcasting
DVQL	Digital Video Quality Level
FPS	Frames Per Second
FUP	Fair User Policy
GOP	Group of Picture
GPU	Graphics Processor Unit
HD	High Definition
HDTV	High Definition Television
HEVC	High Efficiency Video Coding
HFR	High Frame Rate
IPTV	Internet Protocol Television
ITU	International Telecommunication Union
LNA	Low Noise Amplifier
MOS	Mean Opinion Score
MV-HEVC	Multiview High Efficiency Video Coding
MVC	Multiview coding
NVENC	Nvidia Encoder
P2P	Peer to Peer
PCC	Pearson Correlation Coefficient
PSNR	Peak Signal to Noise Ratio
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase-Shift Keying
QOE	Quality of Experience
SI	Spatial Information
SROCC	Spearman’s Rank Order Correlation Coefficient
SSIM	Structural Similarity
SS	Single Stimulus
TI	Temporal Information
UHD	Ultra HD
VIFp	Visual Information Fidelity
VOD	Video on Demand
VQM	Video Quality Metrics

1 INTRODUCTION

Modern technology allows creating more video content than in the past and high bandwidths system at high frequencies allow transferring better quality for us. If the 20th century can be described as a period of analog television, then the 21th century is a period of digital multimedia systems. Currently, there is a big interest in multimedia systems, not just in digital televisions. Now there are three major systems to transfer multimedia content: DVB (Digital Video Broadcasting) [1], IPTV (Internet Protocol Television) [2] and unicasting [3]. The term unicast is contrasted with the concept of broadcasting. Broadcasting means transmitting the same data to all devices. However, unicast is the sending of unique data to a single device identified by a unique address.

DVB is the standard for digital broadcasting. DVB has three basic platforms, characterized by three modes of distribution: DVB-T (Terrestrial), DVB-S (Satellite) and DVB-C (Cable). All these standards have their second generations (DVB-T2/S2/C2).

DVB-T is the standard for broadcasting digital terrestrial television, first published in 1998. The DVB-T2 standard was ratified in 2008. DVB-T/T2 uses Coded Orthogonal Frequency-Division Multiplexing (COFDM) [1]. This modulation is useful in environments with strong interference and multipath propagation. The specialty of the Czech Republic, unlike most European countries, is that the most widespread platform is DVB-T.

The DVB-S system for digital satellite broadcasting was standardized in 1993. Due to the low level of the received signal, however, without reflections, Quadrature Phase-Shift Keying (QPSK) is used. Nowadays, most of the largest satellite television providers are using the DVB-S2 standard with 8PSK modulation. Thanks to this change, the useful channel capacity has increased [1].

The DVB-C system for digital cable networks was developed in 1994. In 2008, the latest version DVB-C2 was announced, which improved possible data rate. DVB-C uses various types of Quadrature Amplitude Modulation (QAM). Thanks to the high-quality transmission channel and much greater level of signal spectrally efficient modulation, QAM may be used in the cable platforms DVB-C/C2 [1].

IPTV and Internet streaming are services which are called Video On Demand (VOD). YouTube and Netflix [4] are the most famous representatives. Viewers can browse a catalog of videos which is not related to the TV program. IPTV is a system whereby a television signal is delivered by using the internet protocol to the TV viewer, instead of being delivered through traditional terrestrial, satellite or cable television systems. IPTV offers the ability to play the data before the entire file has been transmitted. Video can be transmitted to the terminal device using wired or wireless technologies as different sub-standards in DVB.

Nowadays, the trend is to deliver still higher quality of video services to the customers. In contrast to the analog television, all these digital systems need some compression of the video sequences from the camera. Even though modern transmission systems support high bandwidths, they are still insufficient for the transmission of uncompressed video. Video encoding is a process in which there is a reduction of relevance and redundancy of the source video. Video applications on the internet are more expanded day by day and they are consuming much more of available bandwidth than in the past. In the future, it is also expected that the trend will move from Full HDTV (High Definition Television) to Ultra HDTV. Hence, efficient video compression with flexible parameters is a key factor for UHDTV video quality. Nowadays, there is consideration of introducing a modern encoder HEVC or VP9.

Codecs H.265 known as HEVC (High Efficiency Video Coding) [5] and VP9 [6] are successors of the most popular codecs H.264 also known as MPEG-4 AVC (Advanced Video Coding) and VP8. H.264 is nowadays used in most multimedia applications (e.g. HDTV broadcasting in DVB-T). VP8 was used for videos on YouTube, but it was replaced by the latest VP9. The HEVC encoder is used for broadcasting Ultra HDTV for IPTV Netflix and will be used in the upcoming DVB-T2. We can expect a new technology that will come and improve the user experience, for example, 3DTV (stereoscopic TV) or HFR (High Frame Rate) video.

Three-dimensional (3D) video is one of the next generation TV services with specific features from the video image quality point of view. Naturally, efficient compression tools enabling significant bitrate reduction to ease 3D video distribution are an integral part of any practical 3D video system implementation. Consequently, it is necessary to find appropriate video codecs whose performance will be sufficient to encode stereoscopic view, one for each eye. Therefore, each modern encoder has a 3D extension for its 2D encoder, which can reduce redundancy of the second view. Assessing the impact of the compression tool for stereoscopic video is very important for emerging video services.

In the following text, the possibilities and features of modern multimedia systems are discussed. The second chapter briefly describes the state of the art in multimedia systems such as capabilities of video broadcasting, video encoding, objective and subjective video metrics, as well as quality in stereoscopic TV. The definition of the aims of the dissertation follows as chapter three. The next chapters summarize the reached results.

2 STATE OF THE ART IN BROADBAND MULTIMEDIA SYSTEMS

A viewer usually wants to watch their favorite movies in the best quality. This is not just about resolution. The user continuously wants something new. Or at least the sellers and producers of hardware are trying to push their products. Modern multimedia systems have the huge complexity of the whole chain from capture to display. This includes capturing video by camera, and efficient video encoding before broadcasting. Transmitting systems with a high bandwidth are needed to transfer all multimedia contents. At the end of the system, there must be a video player that can decode the multimedia content and a monitor which can display it.

2.1 Multimedia Systems and Networks

Nowadays, research and development in multimedia systems can be divided into these areas: video broadcasting and unicasting. For broadcasting, DVB-T/S/C systems are used and now their second generations have already been introduced. For unicast and IPTV we can use classic IP internet networks, peer to peer networks (P2P) [7], or cellular networks (LTE).

A description of DVB-T and DVB-T2 performance in fixed terrestrial TV channels can be found in [8]. This paper deals with the DVB-T and DVB-T2 standards and analysis and simulation of the performance of both standards. Classic Ricean and Rayleigh multipath fading channels with 20 independent paths were used. On the other hand, SISO/MISO performances in DVB-T2 and fixed TV channels are in paper [9]. The main goal of this paper is to explore DVB-T2 performance in fixed reception scenarios when services are broadcasted by SISO/MISO techniques. Moreover, the benefits of the rotated constellation technique are investigated too. Real performance analysis of implementing the DVB-T2 network in Ulaanbaatar is described in paper [10]. Article [11] is focused on the measurement of DVB-S and DVB-S2 parameters. To give cost-effective and profitable broadband services, next generations of satellite systems will have to apply new technologies in satellite equipment and communication payload designs. Article [12] describes the second generation of the DVB-S2 transmission standard and its possibility to achieve greater performance and quality of the signal in presence of interference and high-level noise using constant, variable and adaptive encoding and modulation modes.

State of the art in IPTV is described in [2]. The most famous IPTV Netflix and other providers are described in [4]. In this paper an extensive measurement study was performed to uncover their architectures and service strategies. Characterizing bandwidth consumption by Netflix is in article [13]. Netflix accounts for 29.7 % of the peak downstream traffic in the US in 2011. A functional design of Broadcast Multicast Service Centre (BM-SC) to support mobile IPTV in LTE network can be found in [14]. The main purpose of this paper is the functional design of BM-SC to support IPTV. BM-SC is newly defined by 3GPP/LTE to support delivering IP Multimedia Subsystem (IMS) over a wireless broadband network. The research in this area also continues today. At present, there is a boom of watching online videos, but watching traditional TV is not on the decline. As can be seen in Table 1, the time of watching TV has increased in the last 5 years by 18 minutes [15].

Table 1: Average time of watching television per day.

Year	Time of watching TV
1/2017	4:20 / per day
1/2016	4:17 / per day
1/2012	4:02 / per day

Penetration of distribution platforms in households in the Czech Republic is presented in Figure 1 [15]. Digital Terrestrial Television (DTT), which consists of DVB-T and DVB-T2, has the highest penetration. Satellite television (SAT) DVB-S/S2 and television over cable (CATV) or Internet Protocol television (IPTV) has similar penetration. Analog television (ATV) was disabled at the end of 2012.

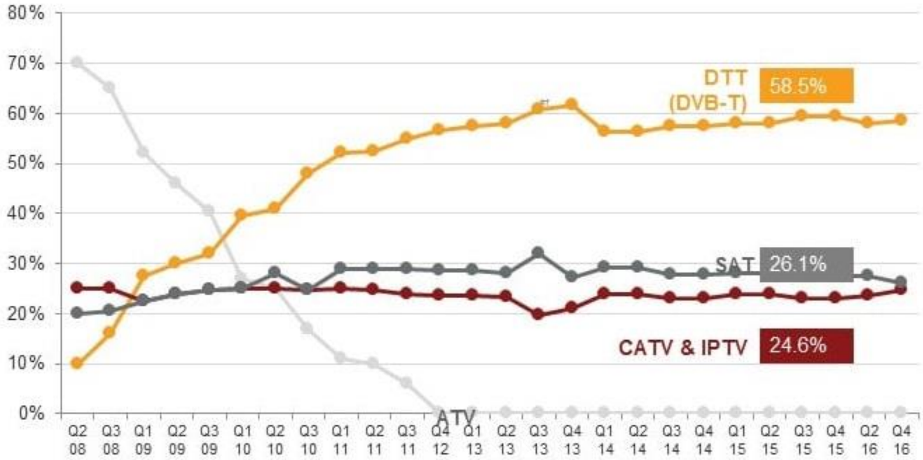


Figure 1: Penetration of distribution platforms on TV receivers in Czech households.

Providing TV content in Ultra HD (UHD) resolution is definitely the next step in the evolution of TV broadcasting. UHDTV, now marked as 4K (2160p), has four times the number of pixels than a Full HDTV screen (1080p). But this is not the end. The next step will be 8K UHDTV (4320p) which means 7680 pixels in the horizontal axis and 4320 pixels in the vertical axis (33 megapixels), which is sixteen times as many pixels compared to the current 1080p HDTV. Paper [16] describes subjective quality assessment comparing UHD and Full HD resolution in HEVC transmission chains. Comparison of upscaling algorithms from Full HD to UHD is in [17]. Paired Comparison methodology was used in the subjective experiment to evaluate their performances.

2.2 The Transmission Capacity of a Standard DVB (Terrestrial and Satellite)

At the beginning, it is necessary to determine how high bit rates are available in the Czech Republic for the receiving of television by different DVB standards. Within the research for satellite television providers were measured average bitrate for the TV program ČT24 continuously for 24 hours. Multiplex throughput is determined by various parameters of the physical layer. On the other hand, the provider adjusts a single stream (TV program) by themselves. Each program in the multiplex has a different bitrate, according to the requirements of the provider. He can decide to provide more programs in poorer quality or fewer programs in better quality. Table 2 shows the parameters of the individual standards which were measured.

Table 2: Measured parameters of the individual DVB standards and program ČT24.

Standard	DVB-T	DVB-S	DVB-S2
Modulation scheme	OFDM 8K	QPSK	8PSK
Transmission capacity of the multiplex [Mbps]	19.9	38.0	45.0
The average bitrate of the ČT24 program [Mbps]	3	2	4

2.2.1 Experimental Setup

The connection of measuring equipment is shown in Figure 2. In the first stage, the DVB-S signal, amplified by LNA (Low Noise Amplifier), is received. Thereafter, the signal is transferred in the first satellite inter-frequency to the DVB-S receiver "Katrein MSK-33". In the next stage, the signal is transferred via TS parallel to the Digital Video Measurement System "R&S DVM 400". The DVB-T signal is received by a second "R&S DVM 400". In the final stage of the diagram, there is "R&S DVQ", which measures the video quality of the decoded video signal. Description of the video metrics used by "R&S DVQ" is in chapter 2.2.2. The measurement was carried out for all platforms successively. It means three consecutive days. The reason was that we have available only one high-quality analyser [18].

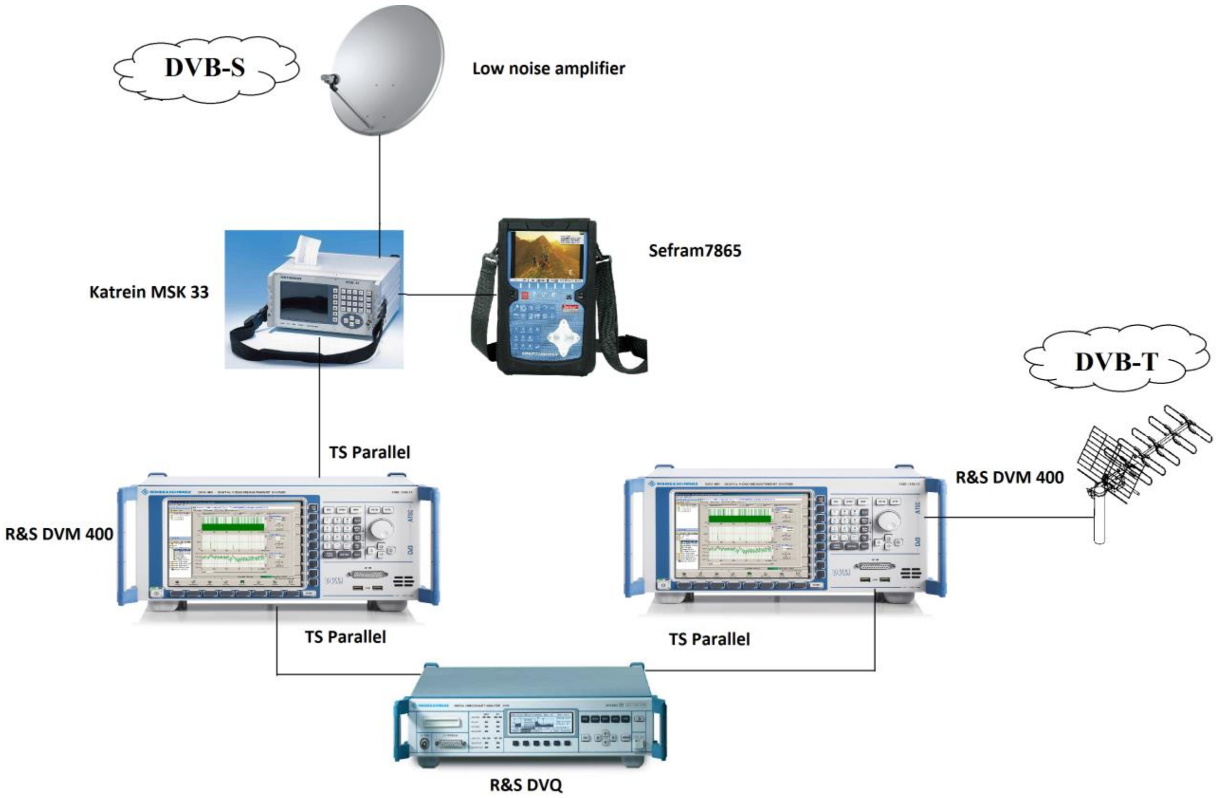


Figure 2: Block diagram of measuring instruments.

The physical layer of DVB-S and DVB-S2, illustrated by constellation diagrams, is shown in Figure 3. As can be seen in this picture, DVB-S uses QPSK modulation. The current standard, DVB-S2, uses 8PSK modulation. These constellation diagrams are from the device “Sefram 7865”.

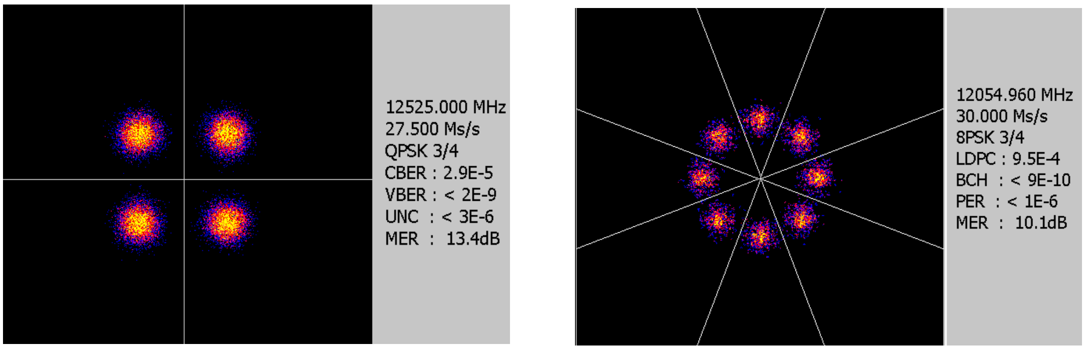


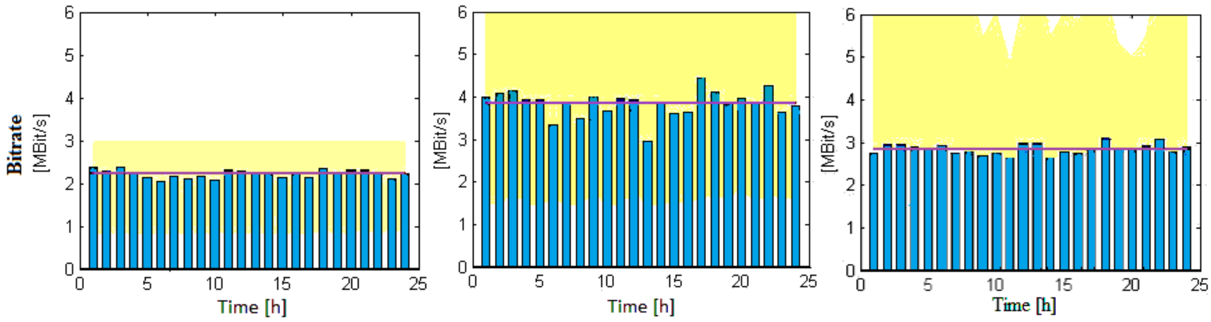
Figure 3: Constellation diagrams of DVB-S and DVB-S2.

2.2.2 Used Video Quality Metric

The basic metric of a Digital Video Quality Analyser for calculating the quality of coded video sequences is the high-quality DVQL-U (Digital Video Quality Level - Unweighted) [19]. DVQL-U is used as the absolute value for the existence of blocking type interference patterns within an original frame. In contrast to DVQL-W (Digital Video Quality Level - Weighted), DVQL-U is a direct measure of these blocking types of interference. Depending on the original frame, however, the test value does not always correlate with the impression of quality of a subjective test. To bring the objective quality closer to the subjective quality, other quantities in the video must also be taken into consideration. These are Spatial Information (SI) and Temporal Information (TI) [20]. This is because SI, and TI can make blocking structures invisible; they can mask them. These artifacts are then not seen by the human eye. The DVQL-W metric was chosen because it corresponds best to a subjective test.

2.2.3 Experimental Results

The measurements were carried out the entire day, i.e. 24 hours. The measurement was performed on the program ČT 24. It is unencrypted and video quality can be measured. Figure 4 shows the data flow of bitrate for the TV program ČT 24 during twenty-four hours. The violet line is the average value of bitrate for 24 hours. The blue line indicates average values per hour. The yellow bar graph shows the maximum and minimum of the bitrate. The same description is for the video quality in Figure 6-8. As can be seen in Figure 4, the bitrate of one video stream is variable. The difference is about 10 %. The maximum bitrate is approximately 80 % higher than the average value of bitrate. The maximum bitrate is constant for DVB-S. In DVB-T, it changes over time.



a) DVB-S b) DVB-S2 c) DVB-T

Figure 4: Average bitrate for program ČT 24 in standards: a) DVB-S, b) DVB-S2, c) DVB-T.

As can be seen in Figure 5, it is possible to transmit several video streams in one multiplex. These diagrams are from the device “R&S DVM 400”. For DVB-T, four to five programs are in SD resolution. For the satellite version of DVB, there are more than ten programs. The number of streams would be roughly half if programs in Full HD resolution were used. The number of programs in one multiplex depends on the provider. It must be mentioned that the sum of all data flows must be less than or equal to the total throughput of the multiplex.

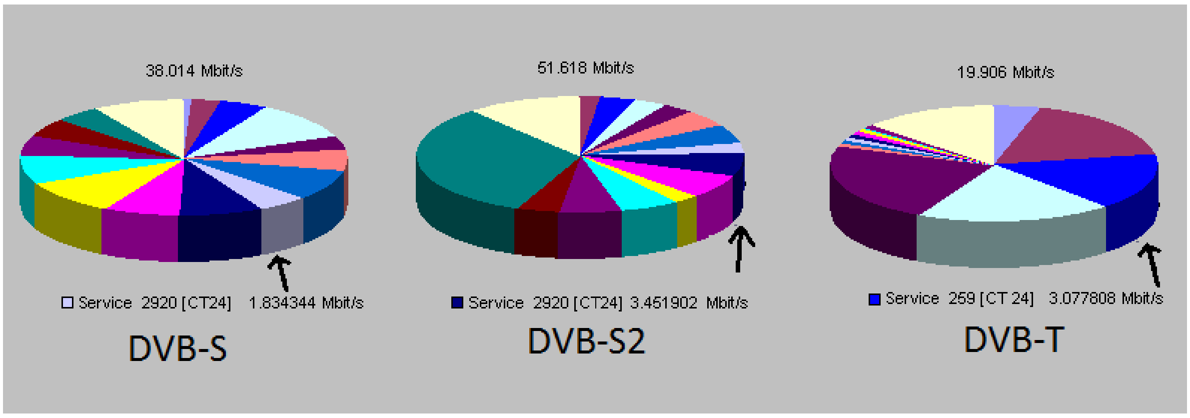


Figure 5: Allocation of bitrates for program ČT24 in standards DVB-S, DVB-S2 and DVB-T.

As can be seen from Figures 6-8, video quality is changing over time. When Figures 4a and 6a are considered, it can be concluded that quality depends on bitrate. Distribution of DVQL-W quality over the duration is in Figures 6b-8b. From the histograms of the time of occurrence of the video quality, it is obvious that the most common value of quality is 89 % for DVB-S2, 84 % in the case of DVB-T and 80 % for DVB-S.

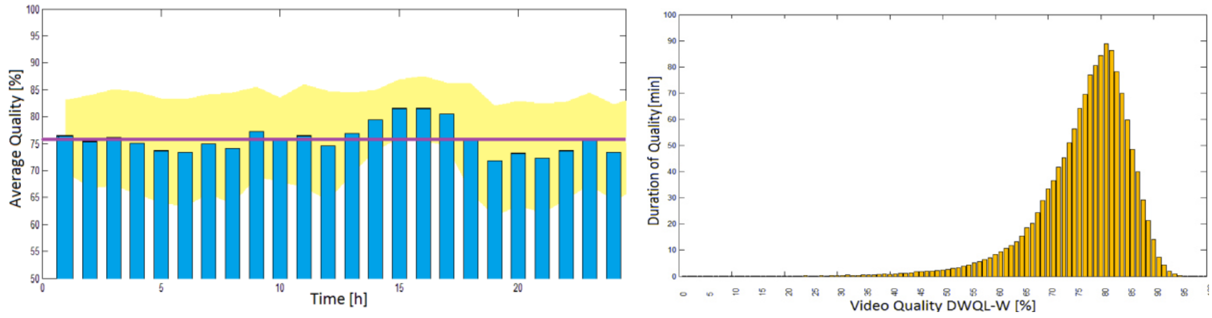


Figure 6: Standard DVB-S a) Average quality in 24 hours b) Duration of video quality.

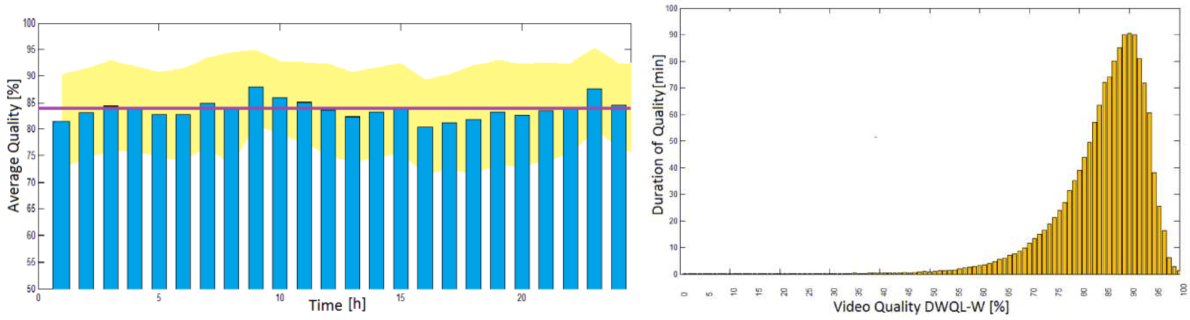


Figure 7: Standard DVB-S2 a) Average quality in 24 hours b) Duration of video quality.

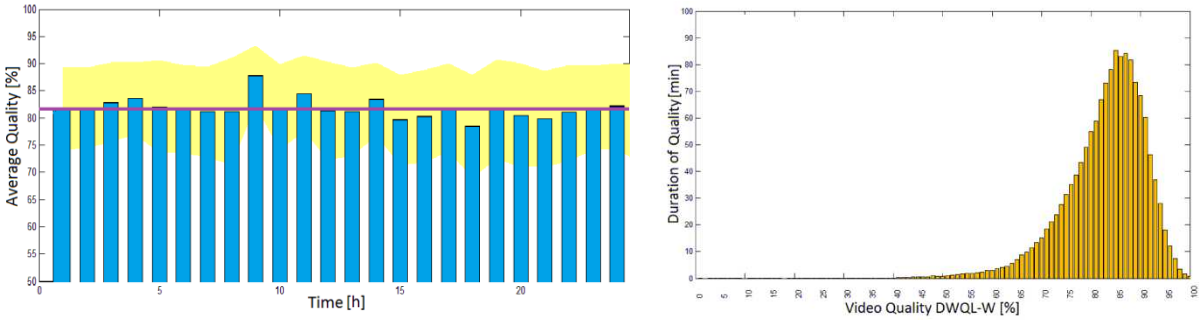


Figure 8: Standard DVB-T a) Average quality in 24 hours b) Duration of video quality.

From the results, it is possible to determine that the highest transmission rate is in DVB-S2. This corresponds to the highest video quality in this system. The provider has allocated a certain bandwidth that is used according to their requirements. The provider may have a lot of programs in poor quality, or a few programs with higher quality.

2.3 Modern Encoding Algorithms

Providing content in UHD resolution is the future of broadcasting. This resolution has four times the number of pixels than Full HD resolution (1920x1080). However, its main drawback is in higher hardware complexity and increased data volume. In such case, efficient video compression tools with flexible system parameters play a key role. Nowadays, HEVC and VP9 are the most promising compression algorithms to reduce high bandwidth requirements of UHD.

HEVC [5] is the next generation video compression tool, developed by the Joint Collaborative Team on Video Coding (JCT-VC). It was designed to offer about double data compression with the same quality or improve video quality at the same bitrate than the established and still used H.264 compression algorithm [21]. This is achieved by adaptive motion vector prediction and by the flexibility to encode predicted blocks in different block sizes. The disadvantage of HEVC is in higher CPU requirements and in its licensing situation.

VP9 is a freely available video compression tool, developed by Google. It is the successor to VP8 with numerous improvements. Like HEVC, VP9 also supports sub-pixel interpolation and adaptive in-loop deblocking filtering (its type is adjustable depending on encoding parameters). VP9 supports using YUV format 4:2:0 which is appropriate for online multimedia services (e.g. video streaming, and YouTube video). It is an open compression tool with limited system settings [22].

Nowadays, many studies deal with exploring HEVC and VP9 compression tools and with their mutual comparison. A comparison of HEVC and VP9 compression efficiency for UHD and Full HD video sequences by a subjective test was presented in [23] and [24], respectively. Researchers in works [25]-[29] focused on the performance comparison of H.264, HEVC and VP9 encoders, using objective metrics. Results from all of these works confirmed the dominance of the HEVC encoder over both H.264 and VP9 encoders. In other works [30]-[33], performances of HEVC and VP9 compression tools are tested for videos with either Full HD or 4K resolution, using both objective metrics and subjective tests. Results show a correlation between the results obtained by objective and subjective evaluations. On the other hand, many times only Peak-to-Signal-to-Noise Ratio (PSNR) [34] and the Structural Similarity Index (SSIM) [35] objective metrics are used.

The quality of the encoder is not only determined by the encoding efficiency, but also the encoding speed. An encoder that achieved excellent video quality but would take weeks to encode a one-hour film is unacceptable for real use. Even more stringent requirements are for real-time video encoding. The newest codecs are more complex and need more time to encode video. Real-time encoding is necessary for some systems like DVB and is not so important for other services like YouTube for example. Therefore, it is important to examine the speed of the encoder. We have some options how to change the speed of encoding: using hardware accelerated encoding, i.e. a special circuit which can encode video independently. Or we can use a different implementation of the encoder or buy a more powerful computer. The last possibility is to change predefined quality profiles in the encoder. Predefined quality profiles can change the balance between the speed of encoding and the quality. And why to examine predefined settings? If the compression of the video is five hours instead of one hour, it would be good to be sure that it is profitable and image quality is better and the time spent on the compression is not wasted. A comparison of different implementations of the HEVC codec was already done in [36]. In this comparison, only pure software encoding implementations were used. The implementations of standard HEVC are still under development. Authors of [37] deal with transferring the load of some elements of the encoder (transformation and quantization) to the GPU. Their results show approximately a 10 times higher increase in encoding speed compared with the conventional encoding method. In [38] authors achieved about 950 times faster inter-prediction compared with the reference software. Improved algorithms were proposed to offload part of the encoder to the GPU [38].

2.4 Objective and Subjective Video Quality Metrics

Video quality can be evaluated by two metrics - objective (standardized computer metric) or subjective (quality is evaluated by independent viewers).

2.4.1 Objective Metric:

The basic division of objective metrics is to metrics with reference and no-reference metrics. The advantage of metrics with reference is that they can be applied to a broader range of encoders and are more accurate, but they need raw uncompressed video. The next division can be into two parts using a Human Visual System (HVS) [39]. The first part is without knowing HVS characteristics and the second part takes advantage of knowing the HVS characteristics. Below are listed and described objective metrics used in the dissertation:

- **PSNR** (Peak Signal-to-Noise Ratio): very simple metric based on differences of the corresponding pixel values; higher dB value means higher video quality and it is usually between 20 and 50 dB [40].
- **PSNR-HVS:** Peak Signal-to-Noise Ratio considering Contrast Sensitivity Function (CSF) of HVS [40].
- **PSNR-HVS-M**: Peak Signal-to-Noise Ratio taking into account Contrast Sensitivity Function (CSF) of HVS and between-coefficient contrast masking of discrete cosine transform (DCT) basis functions [40].
- **SSIM** (Structural Similarity): computes the structural differences in the pictures reflecting basic properties of the human visual system (HVS). A higher index value means higher quality. SSIM can have values from 0 to 1 [35].
- **MS-SSIM** (Multi-Scale Structural Similarity): uses the same basic algorithm except that it operates over scales. The reference and coded images are iteratively driven through a low pass filter and down-sampled by a factor of two. The resulting image pairs are processed with the SSIM algorithm and multiplied together [41].
- **VIFp** (Visual Information Fidelity) pixel domain version: this metric has been developed using HVS models and natural scene statistics [40].
- **VQM** (Video Quality Metric): video quality metric that closely predicts the subjective quality ratings. VQM scores can be produced by the General Model developed by National Telecommunications and Information Administration (NTIA). NTIA General Model is designed to be a general-purpose quality model. It compares the original characteristics with the processed characteristics of video sequences and then it produces VQM scores. The range can be from 0 (no perceived deterioration) to approximately 1 (maximum perceived deterioration). The final score is calculated by using Equation (1) which consists of seven parameters [42]. Details of VQM metrics, including an explanation of Equation (1), are described in article [43].

$$VQM = -\,0{,}2097 \cdot si_{loss} + 0{,}5969 \cdot hv_{loss} + 0{,}2483 \cdot hv_{gain} + 0{,}0192 \cdot chroma_{spread} - 2{,}3416 \cdot si_{gain} + 0{,}0431 \cdot ct_{atigain} + 0{,}0076 \cdot chroma_{extreme} \quad (1)$$

In contrast with PSNR and SSIM metrics, in which a higher value means higher quality, in VQM a lower value indicates higher quality.

2.4.2 Subjective Metrics:

To compare the quality of encoded video sequences by subjective metrics, the Single Stimulus (SS) or the Double Stimulus (DS) metrics [20], [44] can be used.

- **Single-Stimulus:**
  - **ACR** (Absolute Category Rating): each sequence is rated individually.
  - **ACR-HR** (Absolute Category Rating with Hidden Reference).
  - **SSCQE** (Single Stimulus Continuous Quality Rating): a long sequence is rated continuously over time using a slider device.
- **Double-Stimulus:**
  - **DSCQS** (Double Stimulus Continuous Quality Scale): participants see an unimpaired reference and the impaired sequence in random order.
  - **DSIS** (Double Stimulus Impairment Scale): participants see an unimpaired reference video, then the same video impaired, and after that, they vote on the change of quality.
  - **PC** (Pair Comparison): different impairment types of quality are compared.

Subjective quality of video sequence is almost always represented as a MOS value (Mean Opinion Score). This means a quality score of a sequence is estimated by a group of participants [45]. General viewing conditions for subjective assessment of the quality of SDTV and HDTV television pictures on flat panel displays are described in detail in ITU recommendation ITU-R BT.2022 [46].

2.5 High Frame Rate Video

In the future, we can expect an expansion of video content with HFR. In recent years, considerable attention has been on resolution and stereoscopic videos. Now, it is time to focus on smooth and fluent video. This can be achieved by HFR. The frame rate is the frequency at which a frame in a video sequence is displayed. The frame rate is usually expressed in frames per second (FPS). It is supposed that HFR can make fast action scenes look smoother. Other techniques about frame rate and speed of playback are “Slow motion” and “Time-lapse”. Slow motion is achieved when each film frame is captured at a frame rate much faster than it will be played. When we replay this video at normal speed, movements look slower. On the other hand, time-lapse is a technique whereby the speed, at which film frames are captured, is much lower than the speed used to play the video.

Nowadays, there is no problem to find HFR videos. In 2014 the biggest video server YouTube already had the ability to record and play video at frame rates 48 and 60 fps. In 2012, the film "The Hobbit" was released which was filmed in Ultra HD resolution and 48 fps. This is double frame rate compared to the normal film standard. In article [47], the subjective quality of HFR video is evaluated, but only in Full HD resolution and encoded by using H.264. The impact of frame rate on objective metrics is in article [48]. The encoder and resolution of sequences are the same as in the article before. How to model HFR video transmission over wireless networks is described in [49].

2.6 Visual Quality in Stereoscopic TV

In the near future, we can expect the application of new technologies that will improve the user Quality of Experience (QoE). One of them can be Stereoscopic TV. There are several possibilities how to encode 3D video content. Each view of the stereo pair can be encoded separately as an independent video sequence using common video coding algorithms for 2D video sequences. Another possibility is to use video coding algorithms specifically designed to support multiple views. These algorithms usually consider the similarity of both views which can lead to significant bitrate savings. Also, specialized video coding algorithms for 3D exist which can take advantage of depth maps if present. The following paragraphs relate to previously published works about video coding of 3D content for multimedia purposes and related Quality of Experience.

Many recent studies have focused on exploring 3D video quality, processed with different video codecs and methods. The impact of asymmetric stereoscopic video encoding on the perceived quality was studied in [50]. Video plus depth compression, using MPEG-4 AVC and MVC (Multiview coding) for 3D content in mobile scenarios was explored in [51]. It was confirmed that the bitrate needed for compressed video depends on the properties of 3D content, on the quality of 3D depth maps and on the view synthesis approach. MVC [52] and its applicability for stereoscopic videos are investigated in [53] - [55]. Authors of [56] proposed efficient encoding tools for stereoscopic video compression with depth modeling modes based on HEVC [57] while in [58] an overview of the 3D-HEVC video encoding standard is presented. Its features are compared with the MVC standard. Results of software evaluations suggest that it is possible to achieve about 52 % coding efficiency gain on average when using 3D-HEVC compared to standard MVC. A special case is described in [59], where an extension of 3D-HEVC considering a circular camera arrangement is proposed.

Possibilities of using common 2D objective metrics for stereoscopic video were examined in [60] and [61]. In the first paper, the impact of encoding artifacts in stereoscopic video quality has been evaluated with three 2D objective metrics. The evaluation was done using PSNR, SSIM and VIFp. The results show that only the VIFp results were highly correlated with subjective data among selected objective metrics. In [62], the use of 2D objective metrics for 3D quality assessment has been explored. Two objective metrics, VQM and Perceptual Quality Metric (PQM), have been investigated and their alignment to MOS has been analysed. In that research, unlike ours, the video sequences were encoded only by using the AVC encoder. Based on the statistical Pearson Correlation Coefficient (PCC) analysis, PQM correlates better with MOS than VQM, 0.78 versus 0.97. Results also indicated that the correlation is strongly content dependent. In another work, Han et al. [63] proposed an extended no-reference objective 3D Video Quality Metric (eNVQM) for 3D video quality assessment. Performance of eNVQM was studied in comparison with two 2D objective video quality metrics, SSIM and VQM. The PCC analysis showed that eNVQM has better accuracy, PCC equal to 0.944, in terms of human perception for 3D video, compared to two current common assessment methods. Pearson correlation for SSIM was 0.911 and 0.932 for VQM.

The authors of [64] analysed the use of ACR for stereoscopic content. A study of the subjective quality of monoscopic and stereoscopic video in adaptive streaming in [65] presents a comparative analysis of different bitrate adaptation strategies in adaptive streaming in 2D and 3D scenarios. We can observe that if the experiment was done on monoscopic video content, then no statistical differences were found when changing the bitrate in an abrupt or a gradual way. Also, high quality oscillations were hardly perceptible if there is not a large coding bitrate difference. Tests on stereoscopic video confirm that switching from 3D to 2D could be the best possibility to reduce the bitrate, while the inverse behavior does not provide a significant improvement to QoE. Paper [66] studied the response of the human visual system (HVS) to compressed stereoscopic sequences and compared the visibility of artefacts in 3D and 2D views (individually left and right eye views) over a different range of bitrates. The 2D and 3D MOS from the test showed that there is a bitrate threshold above which compression artefacts tend to be suppressed in the 3D view when compared to the classic 2D view. Based on the above-mentioned article, we can conclude that the correlation between objective metrics and subjective tests is highly dependent on the test sequence. Therefore, it is not possible to compare the metrics to each other unambiguously.

3D TV technology expanded very slowly. There are many reasons for it. One of them is the maintenance of quality spatial effects watching 3D video sequences. The requirement on the research of quality of experience is obvious. New aspects of the quality of experience must be considered for a 3D video [67]. A survey of QoE in 3D can be found in literature [68]. The stereoscopic perception is influenced by many factors: the content of the video sequence, processing and coding, viewing conditions including the type of used display and finally by the psychological state of the observer [69]. All phases of the 3D video processing can affect the final visual experience. The QoE of 3D content can be evaluated from three viewpoints: quality of the image, quality of spatial percept and visual fatigue. The quality of the image is consistent with the QoE in a classic 2D video. It is influenced especially during image capturing [70], data coding [71], and data transmission [72]. The possible errors are the following: blur, noise, and crosstalk. The discrepancy between accommodative and convergence stimuli contained in the image can cause fatigue in observers and their discomfort [73], [74]. In paper [75], the author presents the first comprehensive review of available image processing methods for reducing discomfort in stereoscopic images and videos. Many researches have dealt with the influence of coding [76] or used 3D displays [77], [78], [79]. Nowadays, many databases are available online. Some of the databases have general content [80]-[83]. There are many databases for special applications and purposes, e.g. for analysing a traffic situation [84] or face recognition [85] and other.

3 DISSERTATION AIMS

In the previous chapter, the state of the art of the multimedia systems and its future development was described. There were also mentioned some of the future improvements of the video quality and their evaluation in multimedia systems. Other sections describe systems and methods which are needed for transferring content to the terminal equipment. Goals in the dissertation thesis are summarized in the following paragraphs:

- **Modern Encoding Algorithms and Video Quality.**

The first aim is focused on the analysis of current and future modern encoding standards. The reason is that there is a big boom in the number of videos, but the bandwidths of transmission networks are limited. Another sub-goal is a comparison of available video sequences in Full HD and Ultra HD resolution. This is followed by analyses and assessments of spatial and temporal information of the new video sequences. Also, to create a database of video content suitable for future multimedia systems (high resolution, high frame rate). This is because there is no possibility of further development without high-quality videos with detailed defined parameters. Analyses of common and advanced objective metrics and their comparison with subjective metrics are also highly required. Since a significant increase has been seen with high frame rate video, it is necessary to use subjective tests to decide if the viewers will ever appreciate such videos. And if so, what are the parameters of videos in which the viewer will require a lower frame rate.

- **Options of Video Encoding Acceleration.**

In the previous goal, we have only focused on video quality. However, video quality is not the only parameter which must be addressed during coding. The encoding speed is also important. For this reason, it is necessary to examine the possibilities of improvements which can be used for speeding up encoding. This can be achieved by various methods, including: Predefined quality profiles. Changing processor architecture and the version of the encoder. Using the graphics card (GPU) for hardware accelerated video encoding. As well as the impact of video bitrate on the speed of encoding. The effect of the above-mentioned options to the video quality and speed of encoding needs to be explored.

- **Visual Quality in Stereoscopic TV.**

The task of this aim is also the quality of videos but not in monoscopic TV, like in previous goals, but in stereoscopic TV. One of the outcomes is creating a database of stereoscopic videos where various parameters are defined. This helps us estimate how good the QoE for the users will be. The aim of the next subchapter should be to explore the performance of recent compression algorithms for 3D stereoscopic video. This should be done by using objective and subjective methods of video quality. An important aspect should also be a comparison of the suitability of 2D objective metrics for stereoscopic video. In the case where no metric is able to achieve excellent correlations, developing an objective model for better modeling of stereoscopic subjective quality is necessary. Authentication of the created model on independent sequences is desirable due to validation and verification.

4 MODERN ENCODING STANDARDS AND EVALUATION OF VIDEO QUALITY

4.1 Comparison of Encoders HEVC and VP9

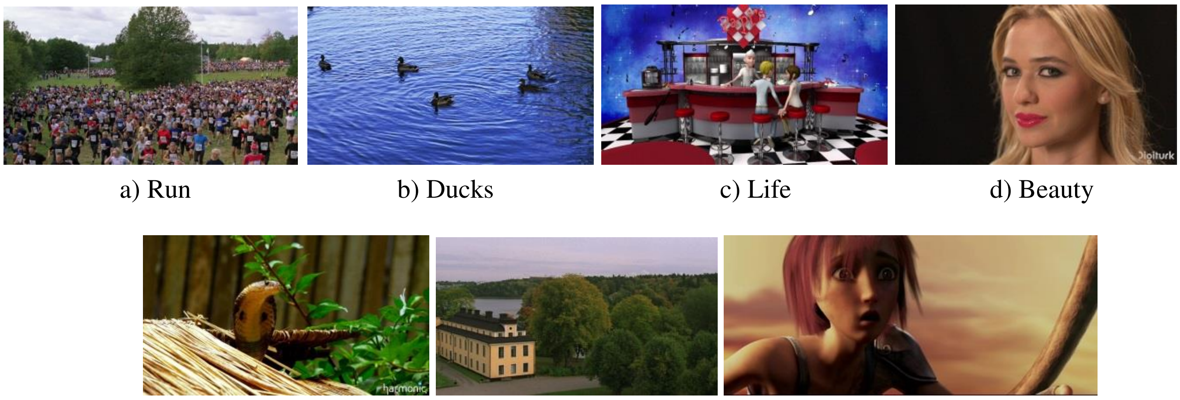
Nowadays, interest in excellent video quality is rapidly increasing. Such interest is closely related to the provided video services in HD formats and in the future in UHD. It is evident that flexible and highly efficient video coding algorithms are very important for video distribution in such formats and in a required quality. As an example, we can state the scenario where we are very limited by transmission speed like in Wi-Fi networks. Alternatively, we are limited by the maximum amount of transferred data, so-called Fair User Policy (FUP), in mobile networks [86]. The aim of this section is to explore the encoding efficiency of HEVC and VP9 compression tools for video content [87], [27]. This thesis will answer following questions: What is the difference between HEVC and VP9 compression performances for video content in Full HD and UHD resolutions? Is there any significant difference in QoE for Full HD and UHD video content? Which objective metrics correlate well with users MOS for HEVC and VP9 Full HD and UHD videos?

4.1.1 Defining a Set of Video Sequences

First, it is essential to select the videos at which the test will be conducted. For a set of tested videos, it is necessary to choose different kinds of videos which are available in Full HD and also in UHD simultaneously (see Figure 9). There were chosen videos with different spatial and temporal information's, different frames per second for outdoor, indoor, and also synthetic video. Table 3 presents an overview of the videos used in the test [89], [90]. The research was performed on uncompressed video clips in Y4M format. This format is defined as RAW, that means that each frame is stored as a sequence of pixels encoded in the YUV color space. If the quality of video samples should be comparable, the samples must have the same value of bitrate. The difference between the set and real bitrate was between -2 and 3 %. This is enough small error for our measurement. There were defined only bitrate and profile without any other modification and without tune options.

Table 3: Parameters of videos used in the comparison.

Name	Frames per second [Fps]	Frames [-]	Time [s]
Run	50	500	10
Ducks	50	500	10
Life	30	825	27,5
Sintel	24	1253	52,2
Beauty	60	600	5
Tree	50	500	10
Cobra	30	352	12



a) Run b) Ducks c) Life d) Beauty

f) Cobra g) Tree e) Sintel

Figure 9**:** Thumbnails of video sequences.

Video sequences have various TI and SI complexity (some of them can be seen in Figure 10). Both values were calculated according to ITU-T P.910 [20].

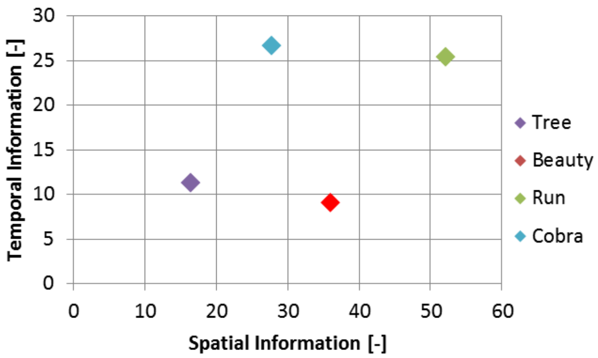


Figure 10: Temporal and spatial information of used video sequences.

4.1.2 Parameters of Encoders

For encoding video, HEVC reference codec HM, was not chosen because it is extremely slow in the processing. It seems that the better solution is to use x.265 implementation of HEVC codec because this encoder has a much higher speed of encoding and impact on quality is negligible. This implementation is still in the research and the new releases of this software are available [5]. The 64-bit version of this encoder was chosen. Parameters like a bitrate, output name and quality were set by MATLAB script used for effective and automated encoding. For UHD resolution four time's higher bitrate was set than for Full HD. For the older H.264 or VP8 codecs, small bitrate was set for good video quality, but bitrates are suitable for the newest encoders which need 50% of required bandwidth for the same quality [23]. Used bitrates are in Table 4.

Table 4: Bitrates of videos used in comparison

Video	Bitrate [Mbps]						
Full HD	0.5	1	1.5	2	2.5	3	6
UHD	2	4	6	8	10	20	

For encoding the video by VP9 codec, the program FFmpeg [91] was chosen which has in latest builds included support for VP9 encoding. Implementation of this codec is still in the research and very often there are compiled new builds of this software. New versions are focused on speed improvements because encoding speed is poor in comparison with x.265. Parameters like a bitrate, output name and quality were set by batch files used to automate encoding again. The 64-bit version of this encoder was chosen for this comparison [91]. Used parameters for coders are in Table 5.

Table 5: Parameters of encoders HEVC and VP9.

Encoder	HEVC	VP9
Build date	30.10.2014	1.11.2014
Implementation	x.265	Vpx
Encoder version	1.3+844	Lavf56.11.101
Preset Options	Medium	-
WPP streams	17	-
Frame threads	3	-
Pool	8	-
CTU size	61	-
RQT depth inter	1	-
RQT depth intra	1	-
Range	57	-
Quality	-	good
Cpu	-	0
Threads	-	8
Qmin	-	0
Qmax	-	63

4.1.3 Objective Video Metrics Used in Comparison

In research were used only full reference objective metrics to evaluate videos. The basic PSNR related metrics are appealing because they are simple to calculate and have easily understandable results. But they are not very well matched to perceived visual quality. The disadvantage of PSNR and SSIM evaluation is that they are computed independently from each frame and at the end the average value of sub-results is made. The video could be of very good quality, but for example quality of 10 frames is terrible and human brain evaluates this video as a bad quality. But in average value, from PSNR metric, this bad part is masked by another good frames and result is not so bad. The value of SSIM for each frame for video "Tree" with bitrate 3 Mbps is in Figure 11. VQM based metrics are more like subjective tests because they are designed to respect the characteristics of the human eye and HVS. Metrics based on PSNR and SSIM were computed only for luminance Y components from source uncompressed RAW files and decoded RAW files. The PSNR and SSIM were computed also with encoder and results were similar to results by Video Quality Measurement Tool [40]. Values from measurement tool were used as results in the evaluation. More information about metrics can be found in section 2.4.1.

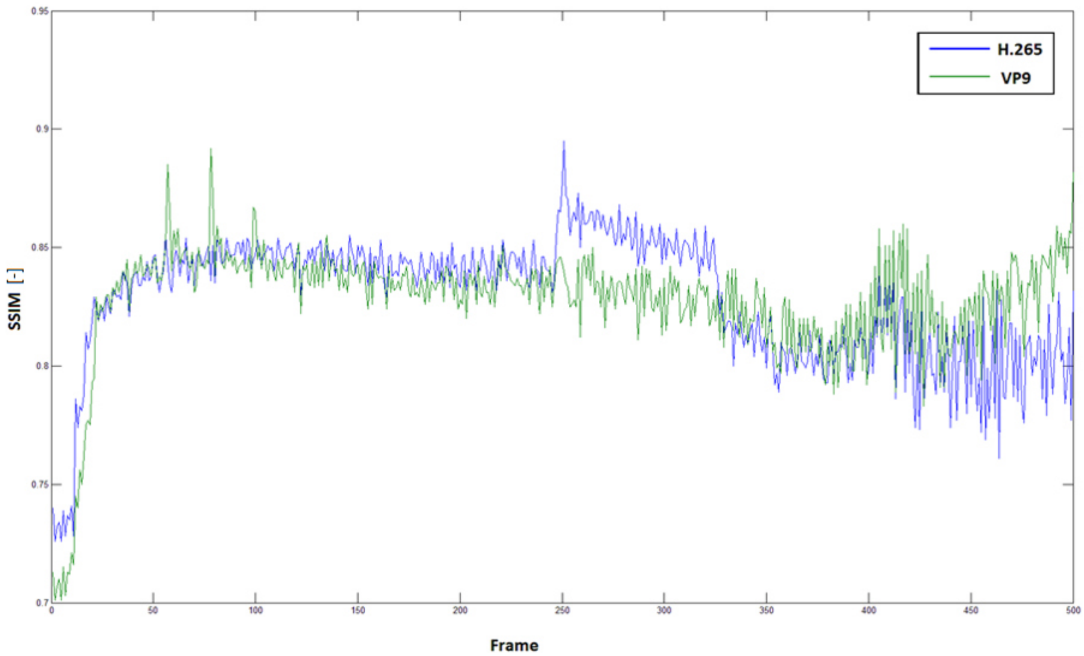
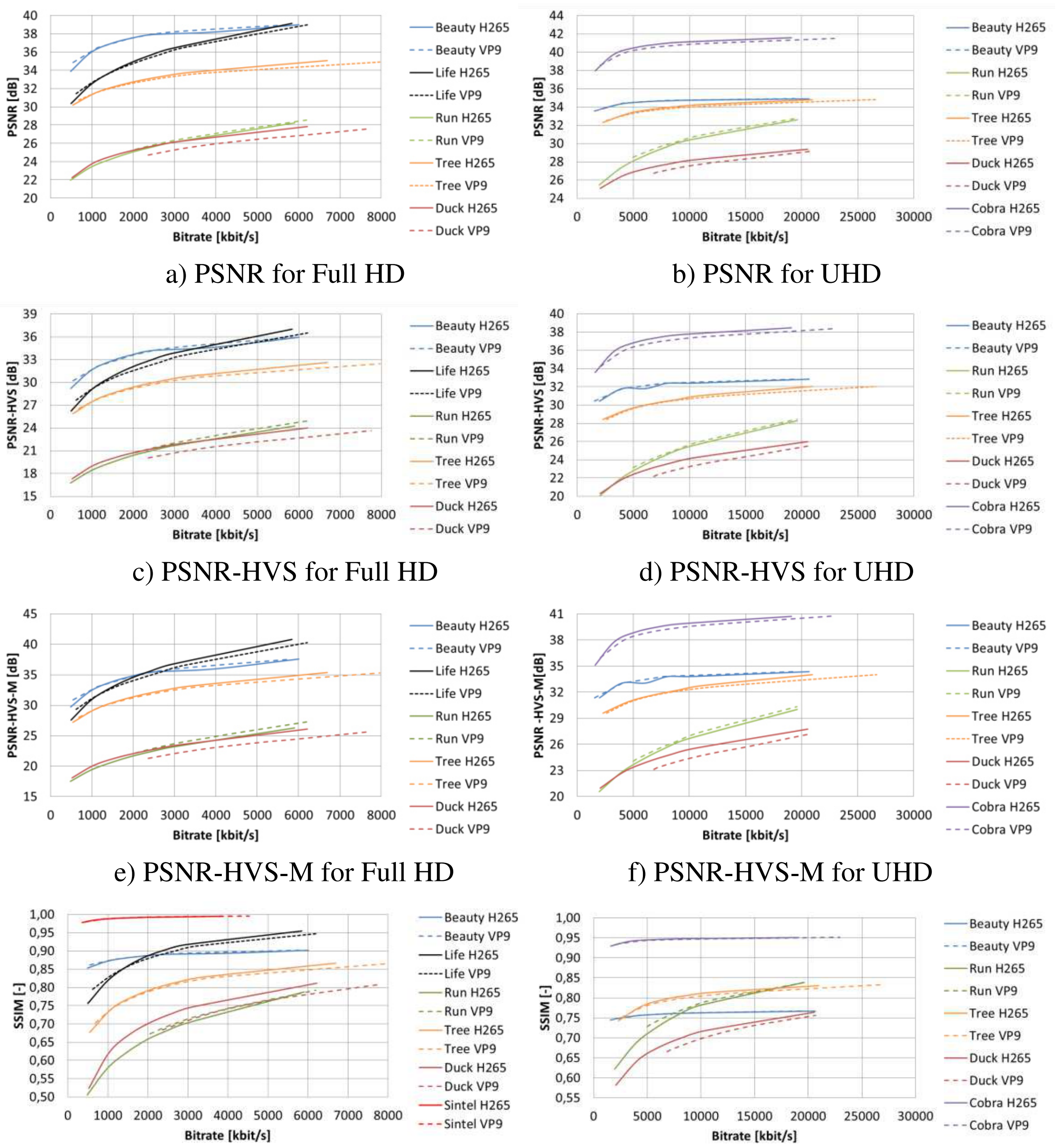


Figure 11: Results of SSIM for each frame of video "Tree".

4.1.4 Quality Evaluation by Objective Metrics

In Figure 13 first three graphs show the several types of PSNR curves corresponding to each video with different bitrates. The waveform is very similar but with different values of PSNR. If the bitrate for Full HD and HEVC is set to 5 Mbps the encoder VP9 needs for the same quality bitrate 6.2 Mbps. This means that encoder HEVC has higher compression efficiency by approx. 25 %. The SSIM for Sintel video is very close to 1 and the difference of encoding efficiency is negligible. The slope of curves for SSIM and PSNR is sharper for SSIM with lower bitrates. The increase of quality with higher bitrates and for PSNR metrics is linear. Results of VIFp, in Figure 13, show that video Beauty has worse quality in comparison with other videos and the other metrics. Encoder VP9 in VQM has better results for Sintel and Run videos than HEVC. Generally, the values of VQM are very similar in SSIM and PSNR values. The SSIM for UHD shows that video Beauty has an only minor increase of video quality depending on bitrate.

For Full HD video "Tree", which has quality measured by PSNR equal to 34 dB, it is necessary bitrate 4 Mbps. For the same quality of UHD video it is needed 10 Mbps. Video "Duck", which should have PSNR equal to 27 dB, needs 4.5 Mbps for Full HD and 7.5 Mbps for UHD. For "Tree" video with SSIM quality evaluation equal to 0.8 it is needed bitrate 2.1 Mbps for Full HD and 7.5 Mbps for UHD. For "Duck" video with SSIM equal to 0.75 3.2 Mbps bitrate is needed for Full HD and 18 Mbps for UHD. This shows that we are not able exactly to decide how many times the bitrate of UHD must be higher in comparison with Full HD. It is between 150 and 550 %. Results also show that codec VP9 cannot allow low bitrate for videos which are difficult to encode and automatically increase the bitrate. The lowest PSNR of encoded video for VP9 was about 25 dB. HEVC encoder sets very precisely specified values of bitrate. The curves show very similar results of encoding efficiency for Full HD and UHD.



a) PSNR for Full HD b) PSNR for UHD

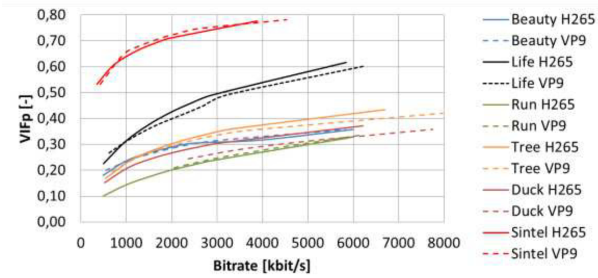
c) PSNR-HVS for Full HD d) PSNR-HVS for UHD

e) PSNR-HVS-M for Full HD f) PSNR-HVS-M for UHD

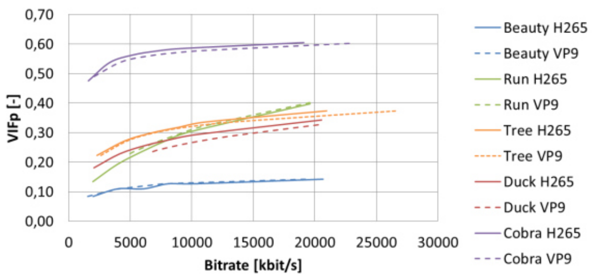
g) SSIM for Full HD h) SSIM for UHD

Figure 12: Results of objective metrics for Full HD and UHD videos.

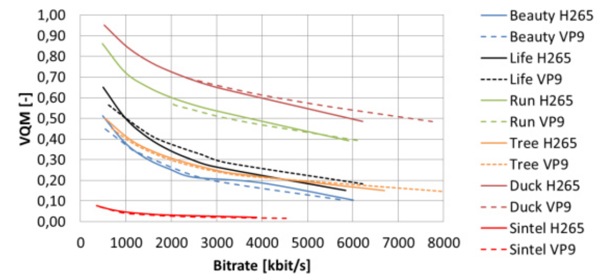
We have also faced some technical problems during testing: Curves of PSNR from Sintel are not shown in graphs because standard deviation was about 30 % and measurement was inaccurate. Our version of VQM metrics cannot be run on UHD sequences because the software reports that it is unsupported image resolution (image rows>=1260). A similar problem was found with Full HD resolution with MS-SSIM metrics because the number of pixels in horizontal and vertical resolution must be divisible by sixteen. The height of Full HD resolution it is not a multiple of 16.



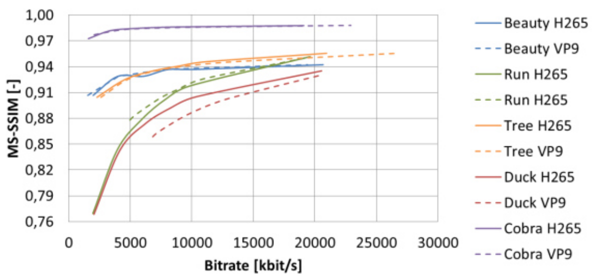
i) VIFp for Full HD



j) VIFp for UHD



k) VQM for Full HD



l) MS-SSIM for UHD

Figure 13: Results of objective metrics for Full HD and UHD videos.

4.1.5 Subjective Test

After objective quality tests were also performed subjective tests on respondents and mutual comparison of results. Because of the time demands on respondents, the test was made only at four sequences and four bitrates. Chosen videos were “Tree”, “Run”, “Beauty” and “Cobra” [89],[90]. More information about the used videos can be found in Table 3.

To ensure reproducibility of the experiment, subjective tests were realized in a laboratory (see Figure 14) which was set up according to ITU BT.500-13 [44]. A workplace in a room with controlled lighting was created. The display device was Samsung UE50JU6900, a UHD 50-inch television. The distance of the participants from the TV display should be approximately 1.6 times the height of the TV screen [46]. In our case, the viewing distance was 1.5 meters.

Figure 14: Testing room setup.

Bitrate for decompressed UHD video with 50 fps, color depth 8 bit and color model 4:2:0 is approximately 5 Gbps. The TV was fed by a personal computer (PC), equipped with a fast SSD connected via PCI-Express with a read speed up 11 Gbps. The reading speed of the disk must exceed the data stream of movies. The PC is equipped with NVIDIA GTX 960 that supports HDMI 2.0. Older version HDMI 1.4 is also capable of transmitting an image in UHD resolution, however only in 30 fps. For the considered subjective tests is necessary to use a newer version HDMI 2.0 that can handle images in UHD resolution in 60 fps.

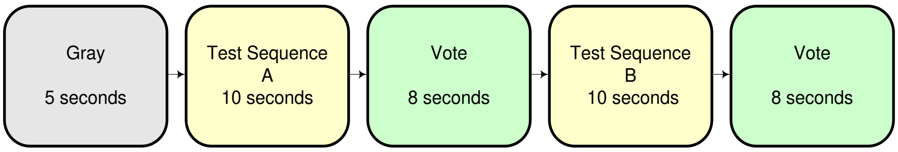


Figure 15: Single Stimulus subjective method – time pattern.

As a subjective method, ACR was chosen as a representative of a group SS test. This method was chosen due to its simplicity and needs of the shortest time to evaluate video sequences by participants. For such subjective method, test video sequences are presented once and rated independently on a category rate. The time pattern for the subjective test is illustrated in Figure 15. If a constant voting time is used, then the length of such time should be less or equal to 10 sec. The time to vote was reduced to 8 sec. (to increase the speed of the evaluation based on pretest users). Sequences in UHD and Full HD were played randomly, and observers did not know what resolution is currently presented. The used scale was chosen on five points scale from the lowest quality 1 (Bad) to the best quality 5 (Excellent). The whole time for the subjective test was 16 minutes. To collect scores, participants fill their QoE into the paper sheet. The scores were then converted into an electronic database by using Optical Character Recognition (OCR).

4.1.6 Results of Subjective Test

The performance of the compressed video sequences in Full HD and UHD resolution, evaluated by objective metrics and subjective method ACR are discussed in this section. The raw subjective scores were processed to obtain the Mean Opinion Score (MOS). Discarding the outliers, the MOS has been evaluated together with the 95% confidence interval. Objectively and subjectively measured quality of the encoded video sequences in Full HD and UHD resolution are shown in Figure 16 and Figure 17, respectively. In each row, the first three columns of the graphs show the results of objective metrics PSNR, SSIM, VQM [43] whereas the column in the same row shows the MOS scores representing the opinion of observers. The results of the HEVC encoder are green, the results of the VP9 encoder are orange.

Full HD Resolution:

- **Tree**

Performance of HEVC and VP9 codecs is very similar. The video quality, evaluated by PSNR and SSIM, is increasing almost linearly with the increasing bitrate. On the other hand, HEVC at lower bitrates has slightly less performance than VP9. Such a difference at higher bitrates is not visible. This phenomenon was proved by the MOS scores from subjective tests. The MOS scores of the "Tree" sequence corresponds well with the scores from SSIM and VQM objective metrics.

- **Beauty**

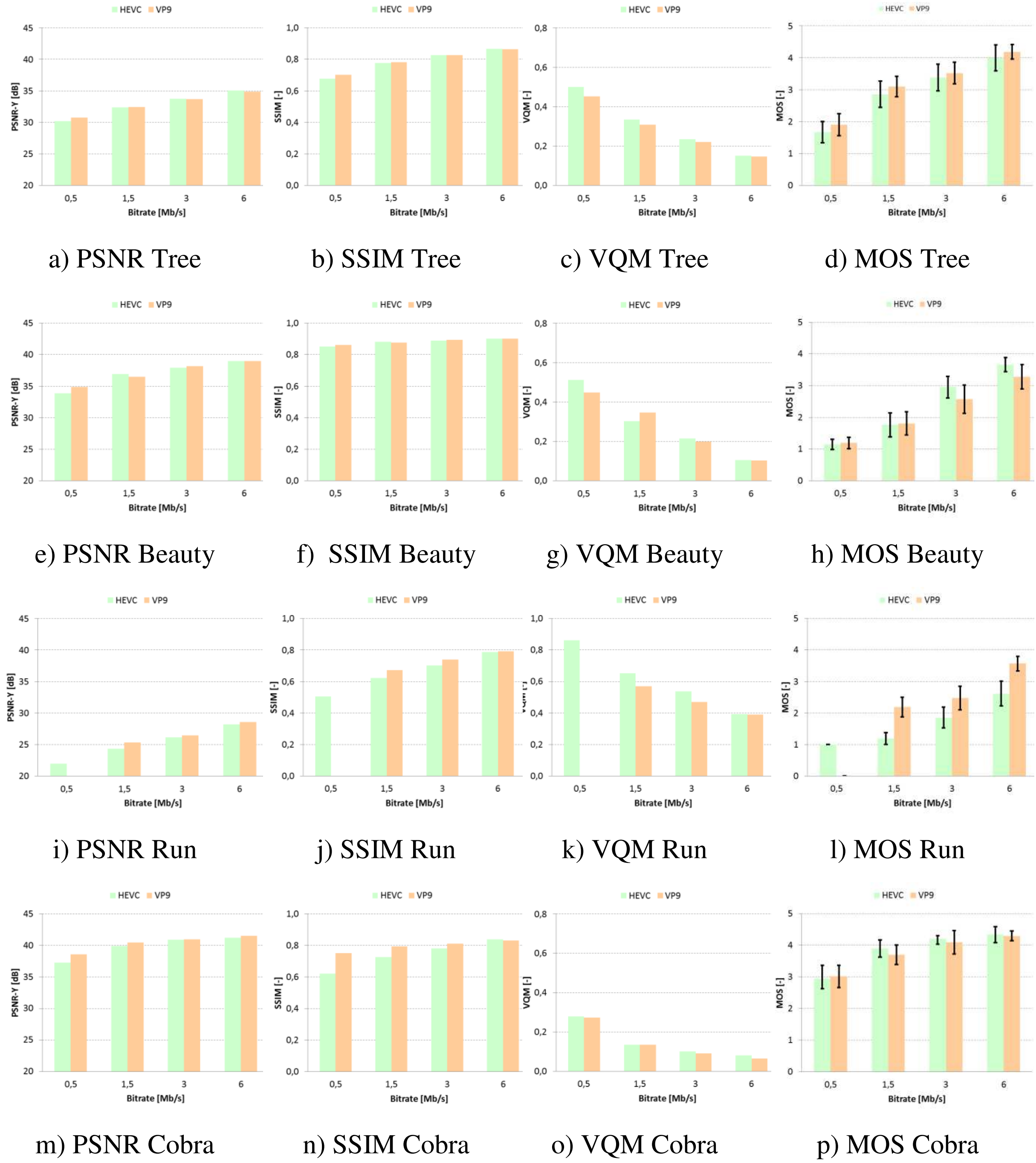
From the point of objective scores, a sequence, compressed by HEVC and VP9, with similar performances at all considered bitrates. Furthermore, objectives scores, in comparison with scores for the "Tree" sequences, are better (mainly for SSIM and VQM). However, MOS scores, in comparison with MOS scores for "Tree" video sequences, are less almost by one level. The MOS scores for HEVC and VP9 are similar at lower bitrates, but at higher bitrates (higher than 3 Mbps) HEVC overcome VP9. It is probably caused by the features of the video sequence "Beauty" – the viewer´s attention is more focused on the face of the woman than on the background of the whole video content.

- **Run**

While for HEVC this sequence is not a problem, VP9 cannot encode it at the bitrate less than 1.5 Mbps (missing bar graphs in Figure 16 from i) to l)). Furthermore, all objective metrics reflect less video quality overall bitrates for both HEVC and VP9 compression tools. A similar effect is visible in the MOS scores. In general, the MOS scores for bitrates from 0.5 Mbps to 3 Mbps are in the interval from 1 (Bad) to 2 (Poor). At the highest bitrate (6 Mbps) the QoE is not better than "Fair".

- **Cobra**

An exceptional sequence. While objective and subjective scores for previous sequences at all bitrates were different, the QoE for the "Cobra" sequence is very similar trough all considered bitrates. HEVC and VP9 performances, evaluated by PSNR and VQM objective metrics, are the same. Even slightly differences between both compression tools are obvious for SSIM metric, also favor in VP9. However, the MOS indicate slightly better quality.
performances for HEVC codec. The lowest and highest MOS score for both HEVC and VP9 are 3 (Fair) and 4 (Good), respectively. Such results can be caused by the features of the video "Cobra" – blurred background of the video at lower bitrates has less viewer´s attention than the head of the cobra with detailed textures in front.



a) PSNR Tree b) SSIM Tree c) VQM Tree d) MOS Tree

e) PSNR Beauty f) SSIM Beauty g) VQM Beauty h) MOS Beauty

i) PSNR Run j) SSIM Run k) VQM Run l) MOS Run

m) PSNR Cobra n) SSIM Cobra o) VQM Cobra p) MOS Cobra

Figure 16: Results of quality evaluated by PSNR, SSIM, VQM and MOS with 95 % confidence interval for Full HD resolution.

Ultra HD Resolution:

- **Tree**

While differences between the HEVC and VP9 objective performances for full HD resolution are very small, for UHD resolution the VP9 slightly overcome HEVC, evaluated by SSIM and VQM metrics. A similar effect can be found between the MOS values, where the HEVC compressed ”Tree” video is evaluated better than VP9 only at 6 Mbps. Moreover, the MOS values at all bitrates are between 3 (Fair) and 5 (Excellent). In comparison with Full HD resolution, the QoE is not increasing linearly with the increasing bitrates.

- **Beauty**

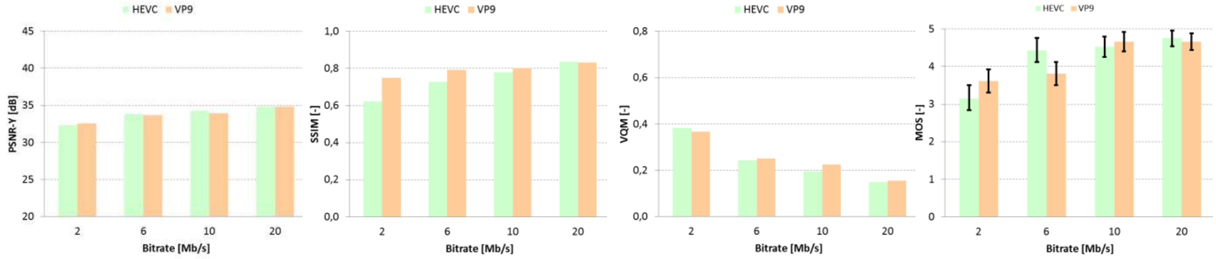
Scores from PSNR and SSIM objective metrics for both HEVC and VP9 codecs are practically the same, only values obtained by VQM metric show slightly higher performance for VP9. Users QoE (see Figure 17 h)) also verified such tendency. There is no difference in encoding efficiency of both codes from 2 Mbps to 10 Mbps. However, at 20 Mbps the VP9 outperforms HEVC. In the case of Full HD resolution, the HEVC has slightly higher performance than VP9.

- **Run**

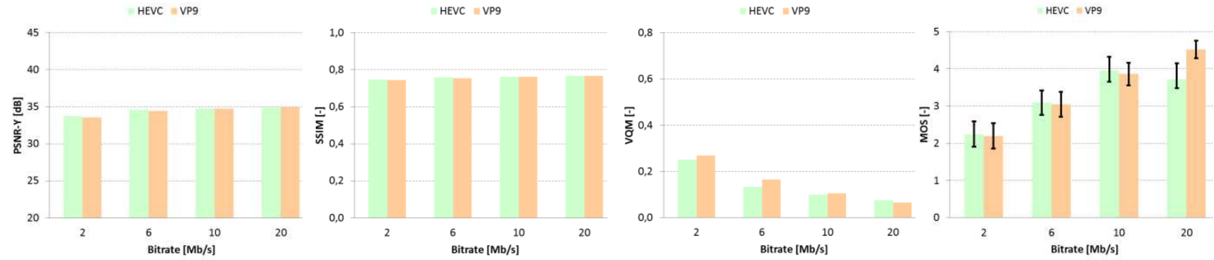
Objective and subjective scores for the “Run” video sequences in UHD resolution show the more comparable performance of HEVC and VP9 as it was in the case of Full HD resolution video. However, the objective video quality for Full HD and UHD resolution at lower bitrates has the same low performances. Once again, QoE for both HEVC and VP9 compressed sequences highly depends on the considered bitrate. At higher bitrates, HEVC is better than VP9. Moreover, at 20 Mbps, almost all participants rated the HEVC compressed sequences with the score “Excellent”.

- **Cobra**

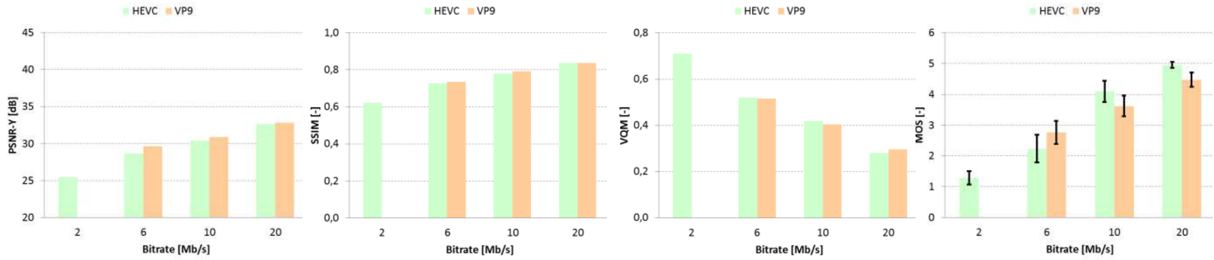
How it can be seen, the video “Cobra” has very similar objective and subjective scores in Full HD and UHD resolution at different bitrates. While SSIM and VQM metrics practically show the same performances for both HEVC and VP9 codecs, slight differences are indicated in QoE for HEVC and VP9 (see Figure 17 p)). Interestingly, at the lowest (2 Mbps) and highest (20 Mbps) bitrates, the MOS scores are better for VP9 than for HEVC. Once again, such results can be caused by the features of the video “Cobra” (see again Table 3).



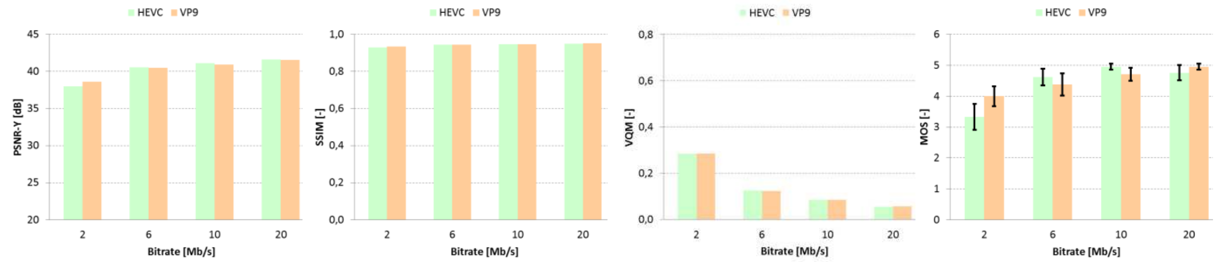
a)PSNR Tree b) SSIM Tree c) VQM Tree d) MOS Tree



e) PSNR Beauty f) SSIM Beauty g) VQM Beauty h) MOS Beauty



i) PSNR Run j) SSIM Run k) VQM Run l) MOS Run



m) PSNR Cobra n) SSIM Cobra o) VQM Cobra p) MOS Cobra

Figure 17: Results of quality evaluated by PSNR, SSIM, VQM and MOS with 95 % confidence interval for Ultra HD resolution.

4.1.7 Conclusion of modern encoders

The test results show that HEVC encoder offers better encoding compression efficiency than video encoded by VP9. The dominance of HEVC is not in all movies, the best one is in video "Duck". This video has a high level and step change in temporal information between video frames. Waves on water are very difficult to encode. In most cases, the results are the same on lower bitrates, but HEVC is better from bitrate higher than 1.5 Mbps for Full HD. Only Run video is the exception, where the VP9 encoder is dominant for all bitrates. This video has a stationary background but runners in the forefront have large temporal and spatial information. The encoding speed of HEVC codec is an advantage. The time needed to encode 10 s of Full HD video by HEVC was about 70 seconds. Time needed for VP9 was hundredfold longer.

Regarding the facts and obtained objective and subjective scores, answers to the questions from the introduction of section 4.1 are as follows:

1) From a broader point of view, objective and subjective scores confirmed our general assumption that currently there is no significant difference between Full HD and UHD videos compressed by HEVC and VP9. The quality of using MOS for all videos encoded by HEVC is 2.72 and 2.57 for VP9.

2) Spearman's Rank Order Correlation Coefficient (SROCC) was applied to the results [87]. This analysis is used to determine the correlation between objective and subjective metrics. A higher number indicates a greater correlation between the metrics. The correlation analysis was applied to each sequence separately and then an average value for all sequences was calculated. The difference in correlations between the resolutions is insignificant. Results can be found in Table 6. The correlation of the objective metrics has comparable results as in [92].

Table 6: Correlation coefficient of objective metrics.

Correlation	**Resolution**	**Objective metric**		
		PSNR	**SSIM**	**VQM**
SROCC	Full HD	0.71	0.78	0.91
	UHD	0.68	0.80	0.89

After a thorough comparison of all objective and subjective scores, it can be concluded that, in our case, the VQM objective metric best reflects the user's QoE for HEVC/VP9 compressed Full HD and UHD videos.

3) In the case of 6 Mbps bitrate, the MOS scores for Full HD and UHD video sequences sometimes show higher QoE in favor of Full HD, independently of the considered video codec. See Table 7. An analogous situation can be observed in Full HD and UHD video with 1.5 Mbps and 2 Mbps, respectively. Such a phenomenon is probably caused by different features of Full HD/UHD resolutions – the human vision can more tolerate an image with fuzzy edges than an image with distortions, due to lower bitrates.

Table 7: MOS for Full HD and Ultra HD videos for bitrate 6 Mbps.

	Resolution	
Video	**Full HD (MOS)**	**UHD (MOS)**
Tree	4.00	**4.42**
Beauty	**3.66**	3.09
Run	**2.66**	2.23
Cobra	4.35	**4.76**

Overall, the encoding efficiency of both HEVC and VP9 compression tools are very similar. Such similarity is also shown by PSNR objective score where a difference between HEVC and VP9 compressed videos is not higher than 2 dB. Furthermore, the SSIM objective metric also indicates similar performances of HEVC and VP9 codecs. The MOS scores at the highest bitrates (6 Mbps) for Full HD are between 3 (“Fair”) and 4 (“Good”) whereas, for UHD videos with 20 Mbps bitrate the scores are very close to 5 (“Excellent”).

Encoder VP9 was not able to encode video sequence with extra poor image quality, in our study sequence “Run”, for VQM worse than 0.6. Such a situation has not happened for HEVC compression tool, but QoE for this scenario was always the lowest one (“Bad”). However, this feature of VP9 has not a significant impact on the overall QoE.

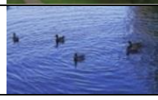
4.2 Standard and Advanced Video Quality Metrics

From the viewpoint of the assessment of the video quality in multimedia services, using of different objective metrics and subjective methods is essential. Objective metrics, based on mathematical approaches, can obtain information about the video quality in a relatively short time. Scores from subjective methods in comparison with scores from objective metrics are more adequate. However, they are expensive and not so effective from time consumption. Therefore, there is a big effort to develop advanced objective metrics with scores having a high correlation with subjective scores. The main aim of this chapter is to compare the performances of three established objective metrics with three advanced objective metrics for the evaluation of the UHD video quality, encoded by HEVC compression technology. The comparison is completed by the subjective scores and with corresponding correlation analysis [93].

4.2.1 Video Sequences

Three short uncompressed raw video sequences were used, downloaded from [89], [90]. More information, including spatial and temporal information, can be found in Table 8. Video sequences were compressed by HEVC video encoding algorithm and the bitrates were 1, 2, 4, 8 and 12 Mbps.

Table 8: Parameters of the used video sequences.

Name	Description	FPS	Resolution	Frames	SI [-]	TI [-]	Time	Thumbnail
Bospor	Floating boat on the river	30	3840x2160	300	15.9	4.4	10s	
Tree	Trees in the park	50	3840x2160	500	35.9	11.3	10s	
Duck	Ducks on the lake	50	3840x2160	500	73.6	15.7	10s	

4.2.2 Used Standard and Advanced Objective Metrics

Six objective metrics (3 established and 3 advanced) are considered altogether. The PSNR and SSIM metrics were calculated by the VQMT software. The metrics ST-MAD, VSNR and NQM were calculated in program MeTrix MuX [94]. The BVQM software was used to calculate the VQM metric [95].

- ***Peak Signal-to-Noise Ratio (PSNR)***
  It is one of the simplest and most widely used pixel-oriented metrics for image and video quality evaluation purposes [40]. The higher is the PSNR value, the higher is the image quality.
- ***Structural Similarity (SSIM)***
  It is a full reference objective metric, based on the Human Visual System (HVS), which measures the similarity between two images [35]. The function is based on the fact that human eyesight is more sensitive to relative changes in brightness than to absolute changes. Value of this metrics is between 0 (low) and 1 (high) similarity.
- ***Video Quality Metric (VQM)***
  The VQM metric is fully based on the HVS system and compares the compressed video with the uncompressed video [95]. The index can be between 1 (low quality) and 0 (high quality).
- ***Spatiotemporal Most Apparent Distortion (ST-MAD)***
  This advanced algorithm using Most Apparent Distortion (MAD) method to estimate spatial and motion-based distortions in the video. Its lower value indicates a higher quality of the compressed video sequence [94].
- ***Visual Signal-to-Noise Ratio (VSNR)***
  The VSNR video metric quantifies the visual fidelity of distorted images. It calculates the contrast thresholds, defined as a disturbing against the reference picture with using HVS masking. If the disturbing is evaluated as an over-threshold, then the analysis continues with the perception of the low-level contrast [95]. Lower VSNR value means lower video quality.
- ***Noise Quality Measurement (NQM)***
  The NQM metric is based on the phenomenon that the psycho-visual effects of filtering and noise are separate. Its value is calculated from the measure of the frequency distortion and additive [94] noise. The higher is the NQM value, the higher is the video quality.

The quality of the HEVC encoded UHD videos, evaluated by the above considered objective metrics, are plotted in Figure 18 a) – g). According to the theoretical assumptions, the higher is the bitrate, the higher is the objectives score. It is also visible that the performance of objective metrics depending on the features of the video (see Table 8). The performance of VSNR and NQM metrics is similar. Such a similarity is visible between PSNR vs. SSIM and VQM vs. ST-MAD.

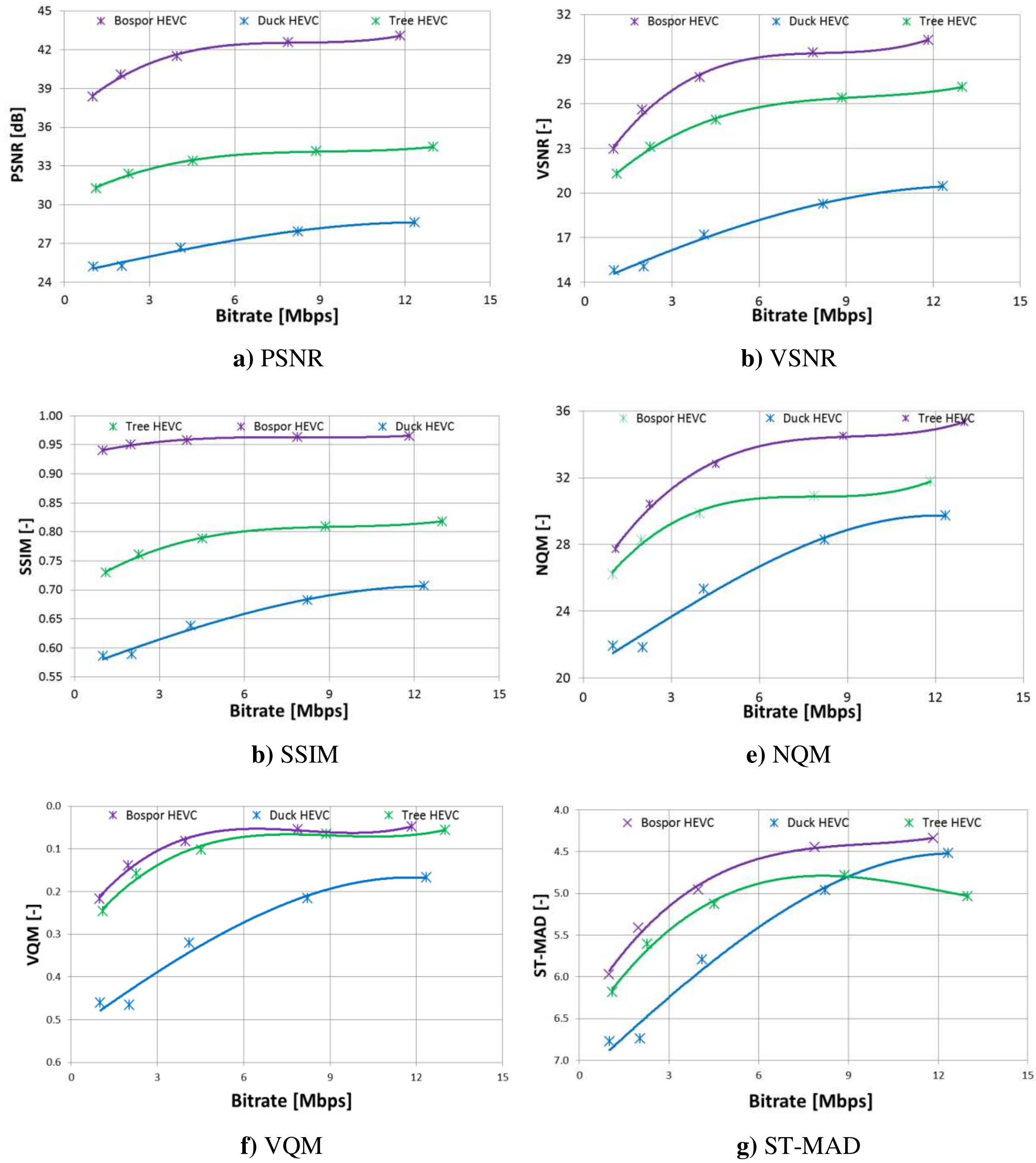


Figure 18: Dependence of objective quality on the bitrate.

4.2.3 Subjective Test

Results from the objective metrics were complemented with the scores from subjective tests. In such tests, overall 21 people were participated, after the testing of their visual acuity. All these tests were realized in controlled laboratory conditions. The presentation of video sequences was done on a computer connected to the TV. The display device was 49" LG 49UF8527. The viewing distance for all participants was equal to 104 cm (screen height 65 cm multiplied by 1.6). The ACR was adopted for the subjective tests due to the lowest time duration. ACR time pattern is shown in Figure 15. The test sequences were randomly presented. At the end of every single sequence, the participant rated the quality of the video using the simple 5-point continuous scale. The range was from 0 (Bad) to 5 (Excellent).

The subjective results from ACR method were processed and the MOS together with 95 % Confidence Interval (CI) were obtained (see Figure 19). In general, the MOS scores correspond to objectively measured quality. For the video “Duck”, the obtained MOS curve is almost linear. It is an interesting fact that video “Tree” from bitrate 4 Mbps has slightly higher MOS scores than video “Bospor”, which is probably caused by the properties of the video sequence.

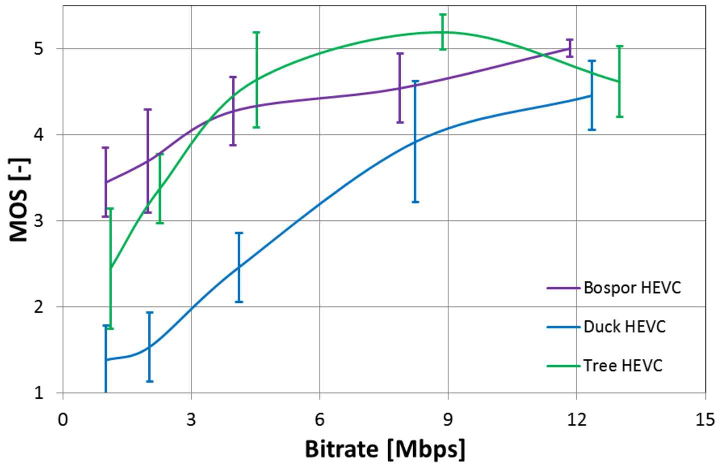


Figure 19: MOS values and 95 % CI intervals for UHD videos.

4.2.4 Correlation of Objective Metrics and Subjective Test

To evaluate the correlation between the objective and subjective scores, the PCC was computed [96]. The outputs of such a correlation analysis are clearly presented in Table 9. The PCC scores are between +1 and -1, where -1 and +1 mean total positive and negative linear correlation respectively, and 0 denotes no linear correlation. The VQM and ST-MAD objective metrics have negative values because their lower score indicates higher video quality. From the obtained results is visible that objective and subjective scores correlate well. More precisely, the VQM, VSNR, NQM and ST-MAD metrics have the highest correlation with the subjective scores (bold values in the Table 9). The ST-MAD metric has the highest average correlation across all videos.

Table 9: Correlation between subjective and objective metrics.

Video	Objective metric					
	PSNR	**SSIM**	**VQM**	**VSNR**	**NQM**	**ST-MAD**
Bospor	0.849	0.918	-0.933	0.972	**0.973**	-0.969
Tree	0.824	0.902	-0.962	0.935	0.945	**-0.995**
Duck	0.875	0.965	-0.989	**0.996**	0.993	-0.995
Average	0.849	0.932	-0.961	0.968	0.970	**-0.986**

4.2.5 PSNR Metric in Encoder and Professional Tool

Nowadays available video encoders can calculate the PSNR values during the encoding of the video sequences. Compared to professional video quality measurement tools, such a calculation can be less accurate because the encoder tries to display a better value to look more efficient. Figure 20 shows a comparison of PSNR values obtained from the HEVC encoder and VQMT tool [95]. The obtained PSNR versus bitrate curves show that the difference between the PSNR values is not higher than 2 dB. Hence, the PSNR values estimated by the HEVC encoder are relevant.

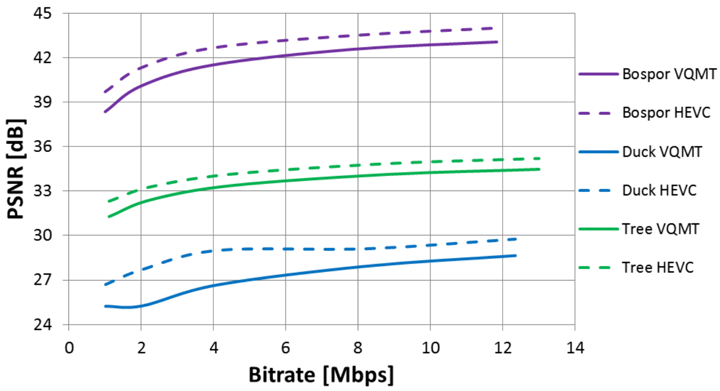


Figure 20: Comparison of PSNR metrics obtained by HEVC encoder and VQMT program.

4.2.6 Video Quality Based on PSNR versus Frames

The video quality is changing during the time due to different temporal and spatial information in the video. Of course, this cause different objective values for each video frame. Such a phenomenon for the video “Duck” and “Bospor”, evaluated as PSNR vs. frame, is depicted in Figure 21. For both scenarios, 6 Mbps bitrate is considered. The video “Duck” starts with frames, where spatial and temporal information’s are low. These information’s gradually increasing between the frames 90 and 200, which means lower estimated PSNR values for encoded videos. After that, thanks to lower spatial and temporal information’s, the PSNR is increasing. In the case of the video “Bospor”, almost constant PSNR values indicate significant spatial and temporal information’s. Similar PSNR values reflect that the encoder has enough information about the motion vectors.

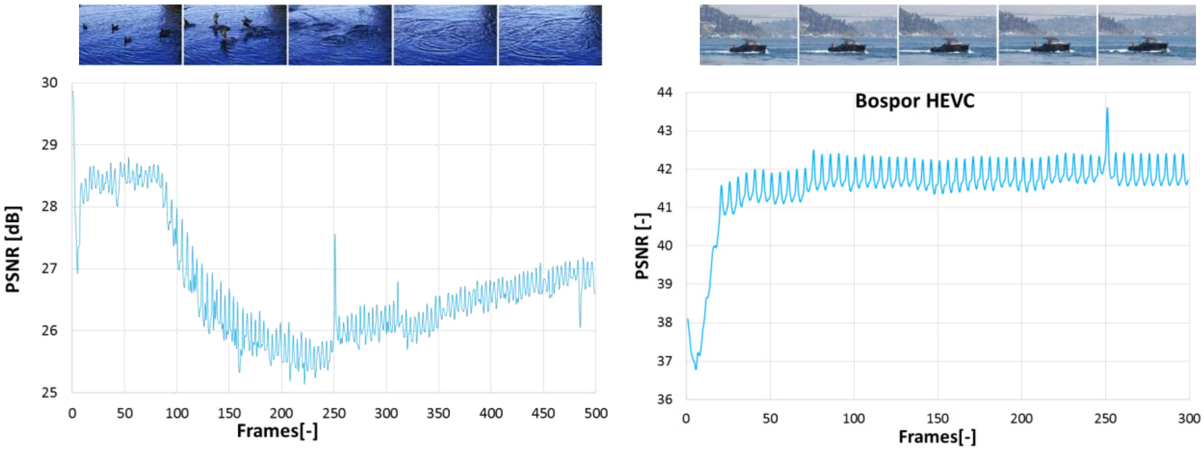


Figure 21: PSNR versus Frames for the video “Duck” (left) and video “Bospor” (right).

4.2.7 Conclusion of standard and advanced objective metrics

Performances of different objective metrics were explored to estimate the quality of HEVC encoded UHD videos. Objectives scores were extended with scores from subjective tests. The obtained results, according to the PCC computation, show that VSNR, VQM, NQM and ST-MAD objective scores have a good correlation with the subjective scores. The ST-MAD metric has the highest average correlation across all videos. The PSNR metric has the worst correlation. This metric was developed to compare image quality. The obtained MOS for the video “Duck” was almost linearly increasing with set higher bitrate. In the case of videos “Bospor” and "Tree" were not so high increase in quality at higher bitrates.

The PSNR values estimated by the HEVC encoder and calculated by a professional tool were compared in next part. This comparison revealed minor differences (lower than 2 dB) between the PSNR values. The difference is mostly constant for different bitrates and sequences. We can conclude that the metrics built in encoder can be used for accurate comparison of video quality, but only within an HEVC encoder. The built-in metric must be corrected (normalized) for use to compare the coding efficiency of different encoders.

Finally, dependences of PSNR values on the video frames were studied. Two representative examples show that PSNR values for each frame can be different due to different spatial and temporal information’s. Moreover, such a behavior is also depending on the used compression tool. A dominant part of the objective video metrics calculates the average PSNR value based on the video quality in each frame. In general, the observer gives a worse evaluation for the video, in which the quality has a steep drop. On the other hand, videos with the same average quality with no drop in the quality obtain a better subjective score. This is the biggest drawback of all objective metrics.

4.3 High Frame Rate Video

With the upcoming Ultra HD resolution, we have already reached a stage where higher resolution does not have so high importance because the human eye is not enough sensitive to take benefit from an even higher resolution. In the future, we can expect an expansion of video content with HFR. The aim of this chapter is to examine if the viewers appreciate the Ultra HD video with a higher frame rate encoded by HEVC. We would like to answer the question of whether viewers appreciate the impact of higher frame rates on the usual Ultra HDTV. The method of transfer HFR Ultra HD video to TV is also mentioned in this chapter [96].

4.3.1 Videos and Encoder Used for the HFR Comparison

Three videos were chosen for a set of test sequences (see Figure 23). Description of videos is in Table 10. Calculated SI and TI of the video are in Figure 22. As a source sequences, uncompressed video clips in the RAW format were selected. Nowadays it is hard to find the uncompressed video with Ultra HD resolution and 60 fps. A video database of sequences from Netflix [90] can be used as a useful source. Currently, there is a boom in using action outdoor cameras. However, at present, exist just a few action cameras, which can record in Ultra HD resolution at 60 fps. Mostly is just possible to select Ultra HD video at 30 fps or 60 fps video in Full HD.

Table 10: Parameters of sequences used in the test.

Color	Name	Time [s]	Description
(blue)	Park Joy	10	Moving footage of the people running in the park.
(yellow)	Aerial	10	View of the landscape by using a drone.
(green)	Toddler Fountain	10	A child running inside the fountain.

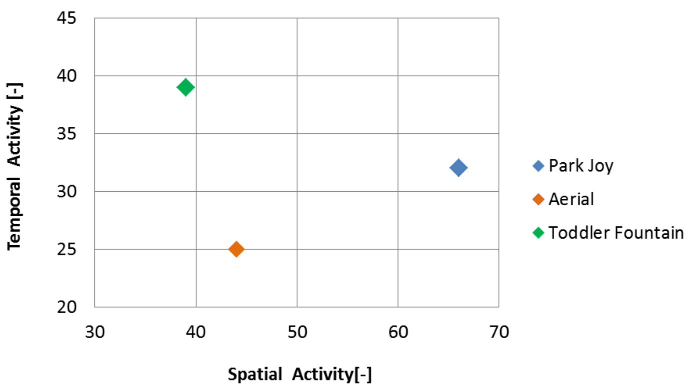


Figure 22: Temporal and Spatial information of used HFR sequences.

a) Park Joy b) Aerial c) Toddler Fountain

Figure 23: Thumbnails of used HFR videos.

Source sequences were encoded by codec HEVC. This codec can encode video up to 8k resolution and frame rate 120 fps. The most widespread implementation of this codec, x265, was selected. Parameters for encoder are in Table 11. Level 5.1 was set in the encoder because Level 5.0 does not support frame rate 60 fps with Ultra HD video. The group of pictures (GOP) was set to 8. Other parameters of encoder were selected according to the recommendations in document [5].

Table 11 Parameters of a used encoder for HFR.

Implementation	x265
Built date	24.4.2016
Encoder version	1.9
Preset	Medium
Profile	Main

Two quality levels were chosen, and each level is defined by its bitrate, see Table 12. Low bitrate 5 Mbps represents streaming with very limited bandwidth. On the other hand, bitrate 50 Mbps represents almost limitless bandwidth, for example, Ultra Blu-ray disk. We also focus on the fact if viewers appreciate the HFR video provided that are not limited by bandwidth. What is the difference if they are very limited by the bandwidth? In the case where we have the same bitrate, then at an HFR, the size of one frame is smaller than the size of the frame at low frame rates. That means whether people appreciate more fluency than image quality.

Table 12: Used framerates and bitrates for HFR.

Parameter	**Value**		
Frame rate [FPS]	24	30	60
Bitrate [Mbps]	5	50	

Used frame rates are also in Table 12. The frame rate of 24 fps represents film standard, almost all films that can be downloaded from the internet have this frame rate. Generally mobile phones and other consumer electronics record in a frame rate of 30 fps. Video with 60 fps represents HFR playback.

4.3.2 Testing Environment and Subjective Test Setup

To ensure reproducibility of results, the subjective experiment was realized in a laboratory of video and multimedia technology. Laboratory conditions were set up according to ITU BT.500-13[44]. A workplace in a room with controlled lighting was created. Ultra HD television 49-inches LG 49UF8527 was used to display video sequences. The distance of the participants from the Ultra HD display was 1.5 meters.

The bitrate for decompressed Ultra HD video with 60 fps, color depth 8 bit and color model 4:2:0 is approximately 6 Gbps. The TV was fed by a Personal Computer (PC), it was equipped with a fast SSD connected via PCI-Express with a read speed up 11 Gbps. The reading speed of the disk must exceed the data stream of movies. The PC is equipped with NVIDIA GTX 960 that supports HDMI 2.0. Older version HDMI 1.4 is also capable of transmitting an image in Ultra HD resolution, however only in 30 fps. For the considered subjective test, it is necessary to use a newer version HDMI 2.0 which can handle image in Ultra HD resolution in 60 fps [97]. DisplayPort (DP) in version 1.2 is an additional interface, which is able to transfer the video in Ultra HD resolution in frame rate 60 fps to display. This interface, however it is not often used in television, is mainly used in monitors [98].

As a subjective method, pair comparison was chosen. This method was chosen because of its suitability for comparing two similar videos. The time to vote was 8 second. Sequences were played randomly, and observers did not know what frame rate is currently presented. Total groups of 22 people were joined into the subjective test, half recruited from university students and the second half from the public. An average age was 25 years.

4.3.3 Results and Evaluation of HFR

The impact of frame rate is more subjective aspect than evaluation of image quality. From this point of view, it is difficult to evaluate the exact perceptual improvement by switching from normal to the high frame rate. Participants compared a pair of two videos on a scale from -2 to +2. The limit value -2 means that the viewer conclusively prefers a smaller frame rate, while the value +2 indicates that the viewer prefers video with higher frame rate. A detail description of the scale for comparison between 24 and 30 fps is in Table 13. The same meaning of scale is in the comparison between 30 and 60 fps.

Table 13: Description of the scale used in graphs.

Scale	Description
-2	Frame rate 24 fps definitely better than 30 fps
-1	Frame rate 24 fps slightly better than 30 fps
0	The same quality for both frame rate
+1	Frame rate 30 fps slightly better than 24 fps
+2	Frame rate 30 fps definitely better than 24 fps

Results are not represented by the average values for a better understanding of the measured data. In the graphs, there are plotted cumulative boxplot for each value of QoE. In this representation of results is better seen how many viewers see differentness of QoE in various frame rates and how significant. The effect of video on QoE can be also better observed. Each video in the graph is represented by a different color, as can be seen in Table 10. Video "Park Joy" is represented by blue color, video "Aerial" by yellow color and green is video "Toddler Fountain".

From the result, in Figure 24, it can be determined that viewers did not observe a significant difference between frame rate 24 and 30 fps if low bitrate 5 Mbps is set. Insignificant dominance can be seen for sequences that have a frame rate of 30 fps. About 95 % of respondents rated QoE for video "Park Joy" equal or better if the video has a frame rate of 30 fps.

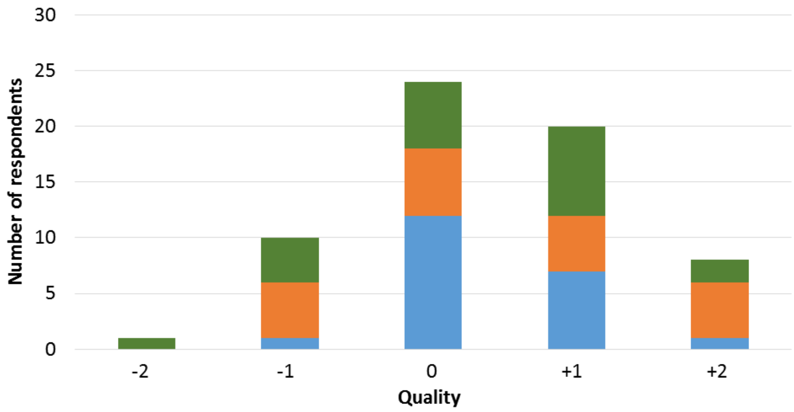


Figure 24: Spread of QoE for frame rate 24 and 30 fps with bitrate of 5 Mbps.

In the case, which is shown in Figure 25, participants predominantly rated higher frame rate as better. A quarter of respondents did not see the difference in QoE between a frame rate of 30 and 24 fps.

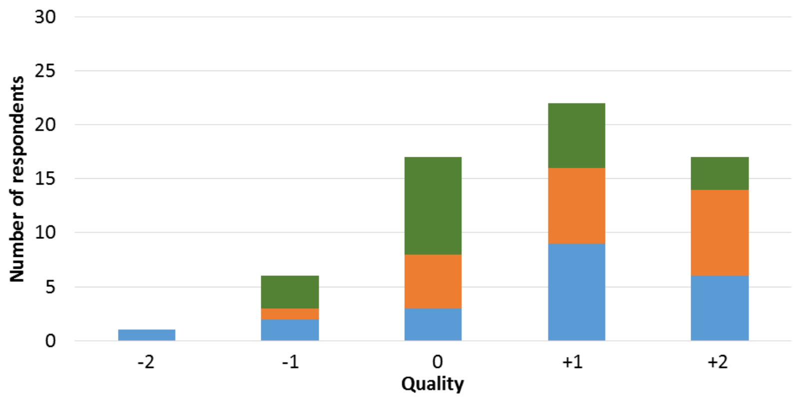


Figure 25: Spread of QoE for frame rate 24 and 30 fps with bitrate of 50 Mbps.

The result of QoE for low bitrate of 5 Mbps at a high frame rate of 60 fps is ambiguous, see Figure 26. The QoE highly depends on the used video. In video "Aerial" viewers clearly preferred frame rate of 60 fps. In contrast to this fact, the preferred frame rate in sequences "Park Joy" and "Toddler fountain" was 30 fps. Preference of video "Aerial" in HFR version can be caused by the properties of this sequence. The video is captured by flight drone and the whole scene is moving.

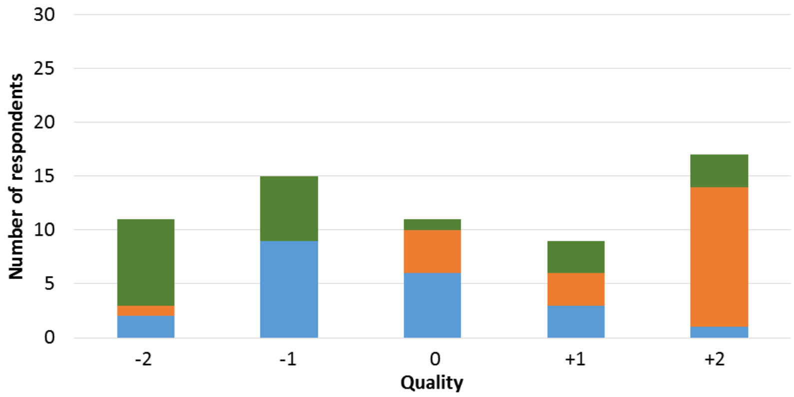


Figure 26: Spread of QoE for frame rate 30 and 60 fps with bitrate of 5 Mbps.

Unambiguous results are for QoE for frame rate 60 fps and bitrate 50 Mbps, as can be seen in Figure 27. More than 75 % of viewers prefer HFR video. This applies to all sequences. About 20 % of viewers do not see the difference between frame rate 30 and 60 fps.

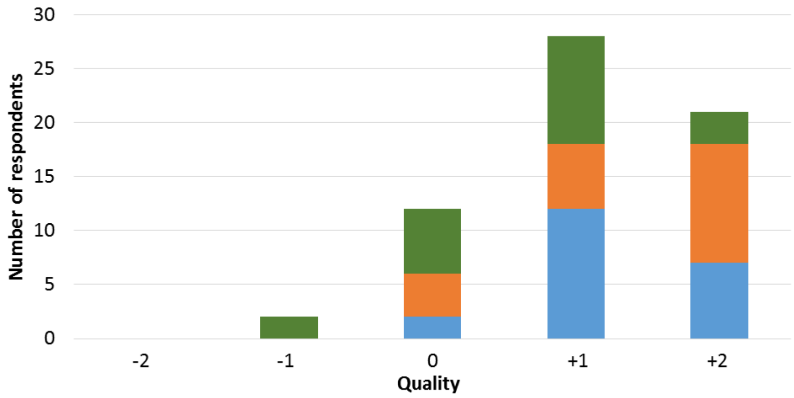


Figure 27: Spread of QoE for frame rate 30 and 60 fps with bitrate 50 Mbps.

4.3.4 Conclusion of HFR videos

In the case, if it is possible to have big data throughput, then viewers clearly prefer videos with higher frame rate. At lower bitrate results are not already conclusive. Observers almost did not see the difference in the QoE if frame rate 24 and 30 fps, with bitrates 5 Mbps, was set. On the other hand, with the option of 30 and 60 fps, evaluation is extremely dependent on the type of sequence. Generally, it can be said that QoE strongly depends on the video sequence. In the optimal case, it would be perfect to have the option to switch the frame rate for each video. From the results, it is also clear that viewers can see the difference between the video with a classical frame rate and the HFR even though they have classic Ultra HDTV. Connecting TV to the source of HFR Ultra HD content is a complex task. The modern interface must be used, especially HDMI 2.0.

5 HEVC - SPEED OF ENCODING AND HARDWARE ACCELERATED ENCODING

5.1 The speed of Encoding of HEVC encoder

Newest codecs are more complex and need more time to encode video. Real-time encoding is necessary for some systems like DVB-T and is not important for another service like YouTube for example. Ways to accelerate the encoding are listed at the end of section 2.3. We have concentrated in this paragraph on the three options: predefined quality profiles, the newer architecture of PC and the newest version of the encoder [99], [100]. Predefined quality profiles can change the balance between the speed of encoding and the quality. Unanswered questions are: How much can have predefined quality profiles change the encoding speed? What is the performance of the new version of the encoder? Can the new processor architecture improve the speed of video encoding? Answers for the aforementioned questions are in the next paragraphs.

5.1.1 Codec HEVC and Predefined Quality Profiles

Used implementation x265 is an HEVC video encoder application library, designed to encode video or images into an HEVC encoded bit stream. x265 is the most widespread implementation of HEVC encoder. Versions of used implementations are in Table 14. Parameters of predefined quality profiles of used encoder HEVC are in Table 15. The encoder is able to measure the time needed to encode video by itself. This is a big advantage for our measurement. There were defined only bitrates and profiles without any other parameter modification and without any tune of the encoder. Dispersion of the set bitrate was about 3 %. Such dispersion is sufficiently small.

Table 14: Parameters of used encoders HEVC.

Implementation	x265	x265
Built date	24.1.2015	24.1.2014
Compiler	GCC 4.6.3	GCC 4.6.3
Encoder version	1.4	0.6

Table 15: Parameters of quality profiles for encoder HEVC.

Profile	1	2	3	4	5	6	7	8	9	10
	Ultrafast	Superfast	Veryfast	Faster	Fast	Medium	Slow	Slower	Veryslow	Placebo
ctu	32	32	32	64	64	64	64	64	64	64
bframes	4	4	4	4	4	4	4	8	8	8
b-adapt	0	0	0	0	2	2	2	2	2	2
rc-lookahead	10	10	15	15	15	20	25	30	40	60
scenecut	0	40	40	40	40	40	40	40	40	40
refs	1	1	1	1	3	3	3	3	5	5
me	dia	hex	hex	hex	hex	hex	star	star	star	star
merange	25	44	57	57	57	57	57	57	57	92
subme	0	1	1	2	2	2	3	3	4	5
rect	0	0	0	0	0	0	1	1	1	1
amp	0	0	0	0	0	0	0	1	1	1
max-merge	2	2	2	2	2	2	3	3	4	5
early-skip	1	1	1	1	0	0	0	0	0	0
fast-intra	1	1	1	1	1	0	0	0	0	0
b-intra	0	0	0	0	0	0	0	1	1	1
sao	0	0	1	1	1	1	1	1	1	1
signhide	0	1	1	1	1	1	1	1	1	1
weightp	0	0	1	1	1	1	1	1	1	1
weightb	0	0	0	0	0	0	0	1	1	1
aq-mode	0	0	1	1	1	1	1	1	1	1
cuTree	0	0	0	0	1	1	1	1	1	1
rdLevel	2	2	2	2	2	3	4	6	6	6
tu-intra	1	1	1	1	1	1	1	2	3	4
tu-inter	1	1	1	1	1	1	1	2	3	4

5.1.2 Used Computers, Videos and Objective Quality Metrics

Hardware parameters of used computers are in Table 16**.** Computer C2D is an old PC appropriate only for office work. Computers i5 and i7 are up to date average powerful PCs. The performance of all PC configurations was calculated by parameter FLOPs. FLOPS is an acronym for floating-point operations per second and it is a benchmark of computer performance. Single precision flops were measured. The performance of the C2D computer is five times weaker than the performance of i5 and i7. Computer i5 is newer than i7 but i5 is for a common user while i7 has high-end performances.

Table 16: Parameters of used computers.

Name	CPU	Architecture	Core/Threat	Frequency	RAM	SP Flops
C2D	C2D-E6700	Conroe	2/2	2.66 GHz	4GB 800MHz	42,4 GFlops
i5	i5-3550	Ivy Bridge	4/4	3.3 GHz	8 GB 1600 MHz	220,7 GFlops
i7	i7-2600	Sandy Bridge	4/8	3.4 GHz	16GB 1333MHz	219,7 GFlops

For a set of tested videos were chosen only two videos “Life” and “Tree” but with different kinds of videos (see Table 3). Only Full HD resolution of video sequences was chosen. Consumption of time by encoding with profile “Placebo” is enormous and therefore was chosen only these two videos.

In research, there were used only full reference objective metrics to evaluate videos. It was chosen only a few most used metrics, PSNR, SSIM and VQM.

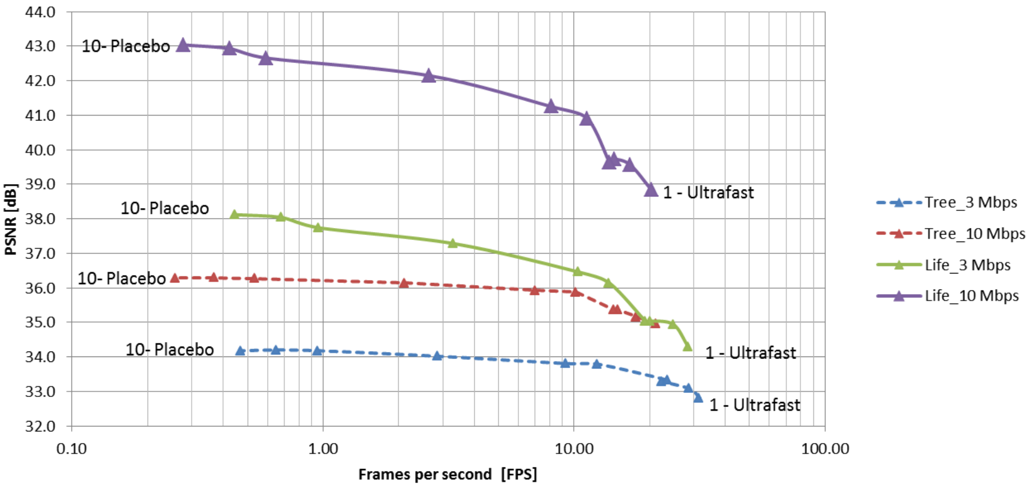
5.1.3 Results of Predefined Quality Profiles

In this chapter, was focus on how affect the predefined quality profiles to encoding speed and video quality can have. In this section, computer i5 is used only. On the horizontal axis in Figure 29 is the value of frames per second because it is independent on the length of the video. In the left side of graphs starts the best quality profiles and gradually to the right side are the faster profiles. The lower the bitrate, the more important is the choice of the better quality profile used. In other words, the lower bitrate which was set, the higher profile should be used for compression. The increase of image quality is negligible for profile better than “Medium” for video “Tree” with bitrate of 10 Mbps. Faster profiles have small problems with utilizing all cores in Intel Core i5, speed can be also reduced by other components of the computer because encoding speed is very fast. The “Medium” profile was chosen by the x265 creators as the default profile. In my own opinion, it could be said that this is not the best choice.

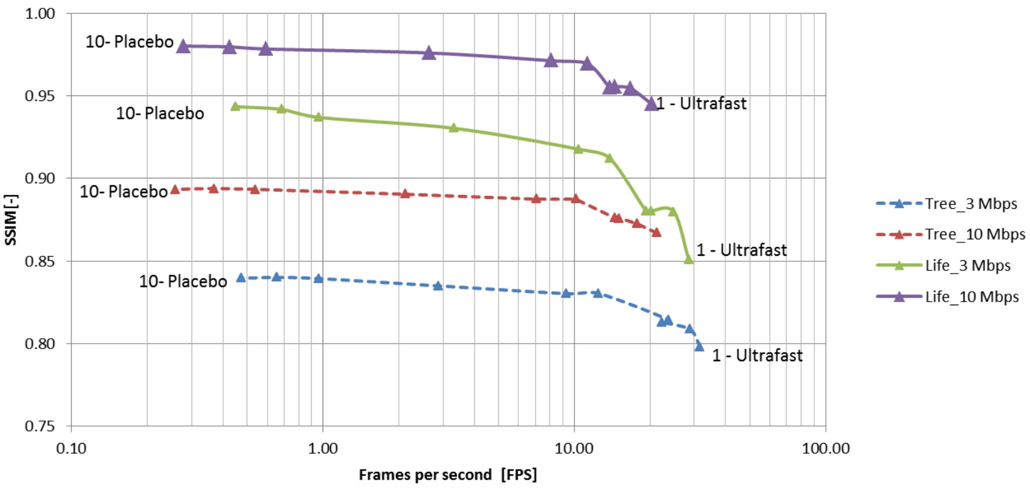
The usage of “Slow” profile provides a considerable increase of quality at the expense of compression slow down measured with all the objective metrics. Therefore, I recommend using this profile as the default because it is the best compromise between the image quality and the time needed for compression. The difference in image quality between the profiles “Ultrafast” and “Placebo” is shown in Figure 28.

b) Profile Placebo b) Profile Ultrafast

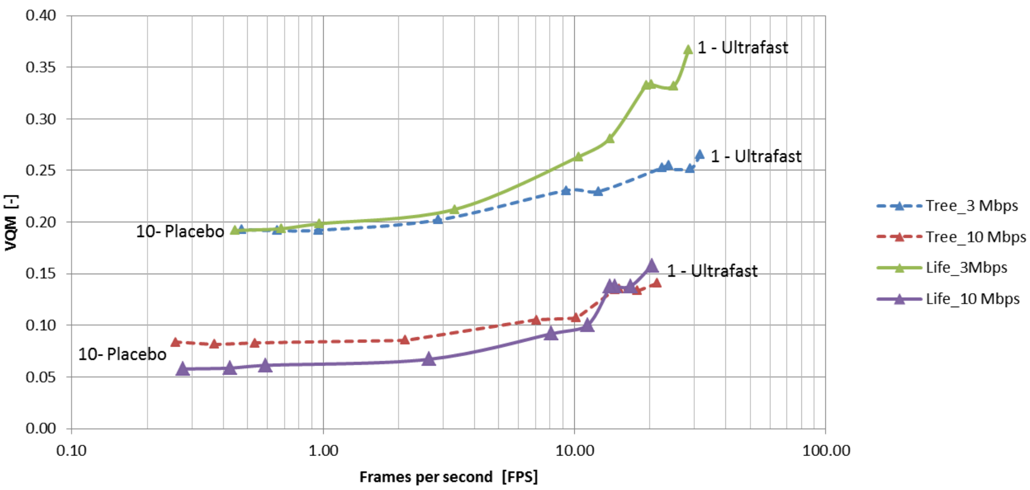
Figure 28: Video Life with bitrate of 3 Mbps and zoom 400 % with different profiles.



a) PSNR



b) SSIM

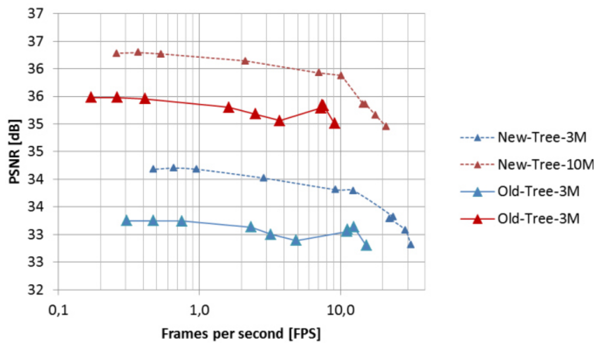


c) VQM

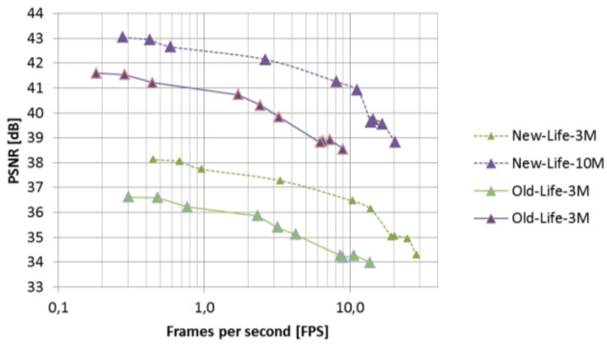
Figure 29: Results of the speed of encoding and image quality by using objective metrics.

5.1.4 Results of Versions of HEVC Encoder

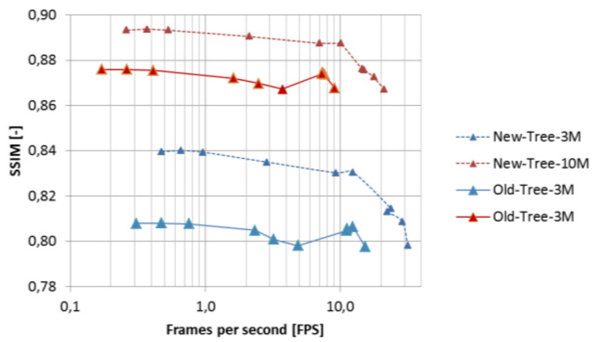
The following paragraph deals with the different versions of the encoder HEVC. Two differently old versions of the same HEVC encoder implementation were chosen. The time difference between both versions is exactly one year (see in Table 14). The dependency of encoding efficiency and the speed of encoding on the version of the encoder is shown in Figure 30. The best quality profiles (e.g. "Placebo") start on the left side of graphs and gradually, results for faster profiles (e.g. "UltraFast") are indicated on the right side. Name of dependencies (Legend) in figures is presented as follows: „version of the video codec_name of the video_bitrate".



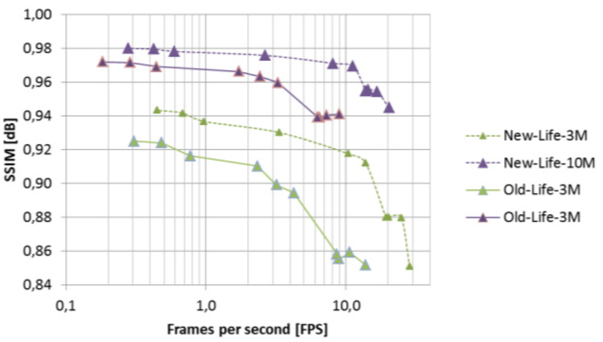
a) PSNR for video Tree



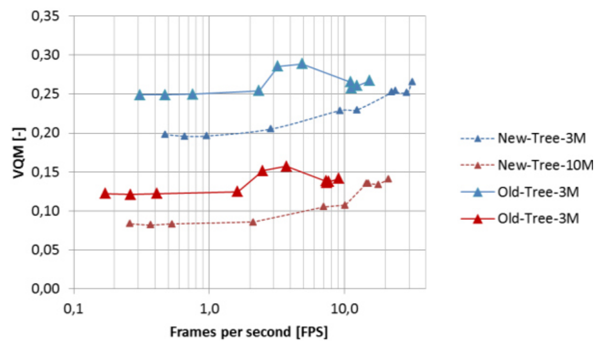
b) PSNR for video Life



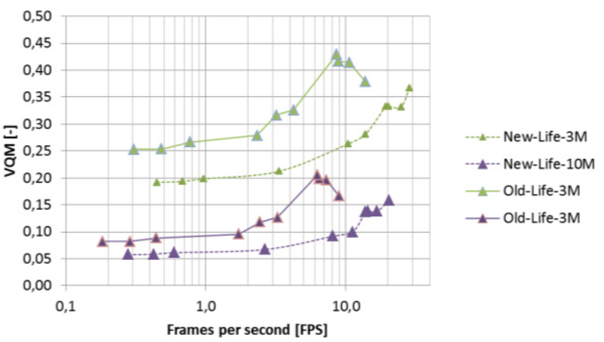
c) SSIM for video Tree



d) SSIM for video Life



e) VQM for video Tree



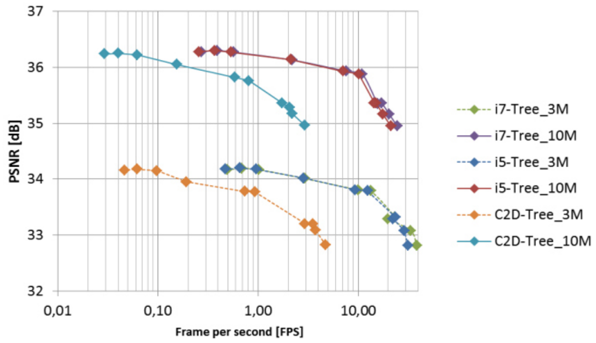
f) VQM for video Life

Figure 30: Impact of the version of the encoder on the speed of encoding and video quality.

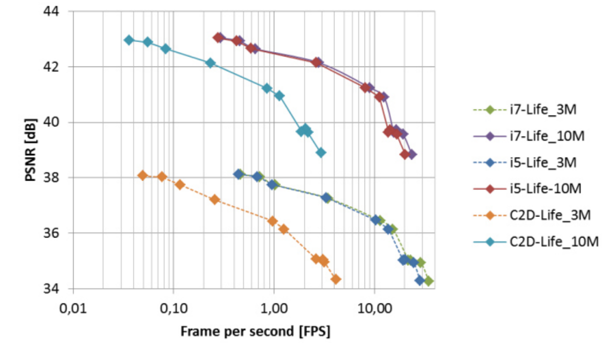
Results show that different versions of the same encoder have a different influence on the speed of encoding and on the final video quality. For example, the speed of the encoding at older encoder with profile “Medium” for video “Tree” at bitrate 10 Mbps was 2.5 fps whereas, with the new version of the encoder the speed of encoding was 7 fps (see Figure 30 a)). The PSNR is increased from 35.0 dB (old version of the encoder) to 35.8 dB (new version of the encoder). The performance of the old version of the encoder is not uniform. Especially, in the case of “Ultrafast” profile, the speed of encoding is increased from 9 fps to 22 fps for video “Tree” while the changes in video quality are negligible.

5.1.5 Results of Impact of Processor Architecture

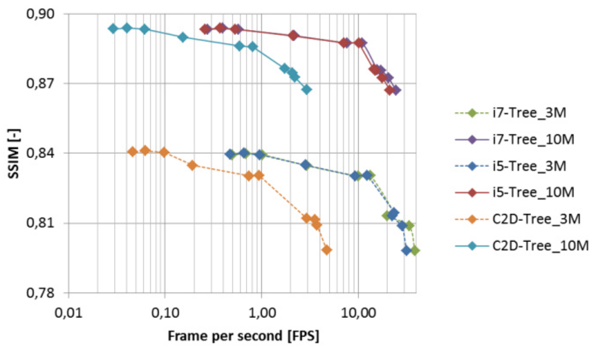
In this paragraph, we focus on the study of the influence of processor architecture from the point of encoding speed. Furthermore, we explore the impact of the newer instruction set of the processor on the video quality.



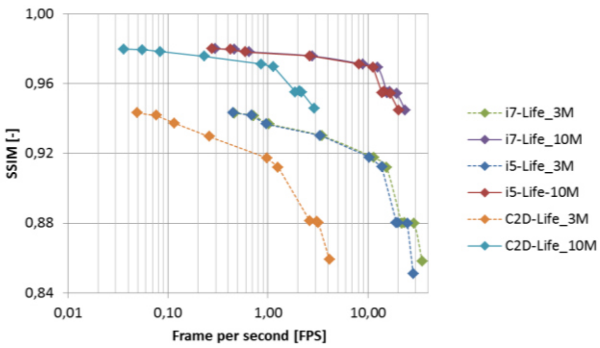
a) PSNR video Tree



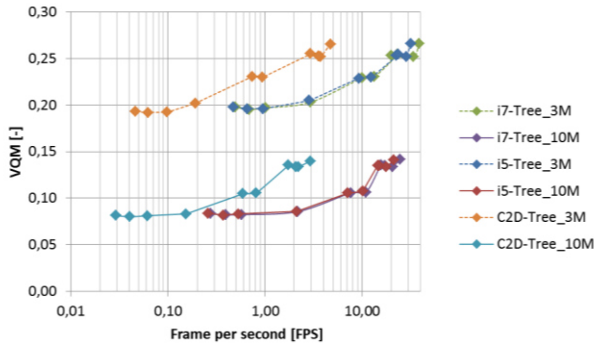
b) PSNR video Life



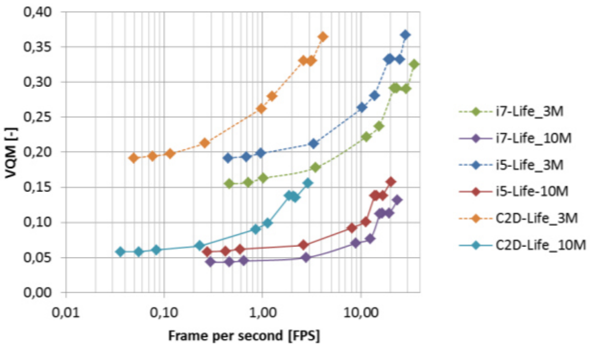
c) SSIM video Tree



d) SSIM video Life



e) VQM video Tree



f) VQM video Life

Figure 31: Impact of the architecture of processor on the speed of encoding and video quality.

Advanced Vector Extensions (AVX) and Streaming SIMD Extensions 4 (SSE4) are new extensions of the instruction set which are supported only by the PC with processor i5 and i7. These extensions can be helpful from the point of video encoding speed. The speed of encoding as the dependence of fps vs. objective metric on the processor architecture is shown in Figure 31. Name of dependencies (Legend) in figures is presented as follows: "the name of the PC_name of the video_bitrate". In this study, only the version 1.4 of HEVC encoder was used. We know that, theoretically, the processor with C2D has approximately five times weaker computing performance than other two processors (see Table 16). What effect have this on the encoding performance? The impact of processor cores for some implementations is insignificant, for example for the default HM implementation. This implementation can handle only with one core.

For example, the speed of the encoding on C2D with quality profile "Medium" for video "Tree" at bitrate 10 Mbps was just 0.6 fps whereas, with the processor i5 and i7 it is 7 fps and 7.5 fps, respectively (see Figure 31 a)). When we compare the speed of encoding on C2D and on i5, where i5 has a clock frequency higher about 24 %, the encoding speed of i5 is faster by 1200 %. It is an extreme speed up of the encoding. In the profile "Fast" the difference between i5 and i7 is only a few percents. In faster profiles, the difference between new processors is about 15 %. Processor i7 can fully take advantage of the Hyper-Threading. In these profiles, C2D is eight times slower than i5 and i7 processors.

5.1.6 Conclusion of Speed of Encoding

As can be seen from the results, the predefined quality profiles fundamentally change the balance between the video quality and speed of encoding. Placebo quality profile, as one might expect, did not provide quality improvement and taking into account the extremely long encoding time it should be used only by people who do not take care for a time. Bearing in consideration all the results "Slow" profile is the best one in my own opinion. In the case of a higher quality profile, the quality increase is not too high to compensate for the expenditure of time. The bitrate increase of the output compressed videos prolongs the time of compression. Three times higher bitrate needs about twice long time to encode video. If the fast speed of encoding is a priority I recommend choosing "Superfast" profile. The profile "Very Fast" and "Faster" provide very similar image quality like a "Superfast" but the speed of encoding is slower. "Ultrafast" profile is a special case. This profile has extremely poor image quality but needs the shortest time to encode. This profile is recommended only for some specialized case like a real-time encoding on some low powerful devices where hardware encoding is not available. For example, mobile phones which have not HEVC hardware encoder integrated directly to the chip. Next results show that a new version of the HEVC encoder has a higher influence on the video quality and speed of the encoding is also higher. It was proved that PC with advanced architecture (but with the same clock frequency and with the same number of cores) could grow the speed of encoding by five times in comparison with old PC. Differences between the results obtained by various objective quality metrics are negligible.

5.2 Software and Hardware HEVC Encoding in Full HD and Ultra HD

In comparison with older standards, HEVC significantly improves coding efficiency. At the same time, it increases computational complexity of coding and therefore encoding takes a longer time. In this chapter, the usage of different implementations of HEVC is proposed where some of them can take advantage of a multi-core Central Processing Unit (CPU). The others are accelerated by using a Video Engine (VE) in the Graphics Processing Unit (GPU). The different predefined quality profiles are also used. Another aspect was to compare power consumption and utilization of components in a computer depending on different HEVC implementations. Research has been carried out for both resolutions, Full HD and Ultra HD. It can be used software or hardware accelerated encoding [101].

5.2.1 Types of Software Encoders

The CPU performs encoding. The codec can be modified but encoding still works. The video quality for a given implementation is always the same, independent on the type of CPU. This does not apply to the encoding speed. It widely varies on the type of architecture and performance of the CPU.

- HM (HEVC Test Model)

The reference software for HEVC is called HM. It is made to provide a reference implementation of the HEVC standard. One of the main goals of the reference software is to provide a basis for experiments that determine which coding tools provide the best coding performance. It is not a particularly efficient implementation. It is unsuitable for daily use. The reference implementation does not take advantage of parallelization on a multicore processor. The encoding rate is extremely slow [101].

- x265 free implementation

The x265 is a free H.265 / HEVC video encoder for encoding video streams. It is released under the terms of the GNU GPL. Its main goals are to offer the highest possible quality at a given bitrate and to be the fastest and most efficient HEVC encoder. The encoder has ten predefined encoder profiles. See again Table 15. In our test, the fastest preset "Ultrafast" and the preset with the best quality "Placebo" was used [5].

- Turing codec

Turing is an HEVC software video encoder for efficient video compression with extra low memory consumption. The implementation is optimized for fast encoding of Ultra HD video content and at the same time preserving a high quality of generated compressed videos. The software is free of charge available under the GPLv2 license. The Turing codec currently supports three preset speeds: slow, medium and fast. The presets control the predefined values of all options and tools in the encoder. A description of used presets "slow" and "fast" is in Table 17 [102].

Table 17: Used predefined profiles in the Turing encoder.

Option / Tool	Speed preset	
	Slow	Fast
Search range for ME	64	32
Search range for bi-prediction	5	1
Number of merge candidates	5	2
Number of intra modes to test with RD search	8	4
Deblocking filter	enabled	enabled
CTU size	64	64

5.2.2 Types of Hardware Encoders

Hardware encoding uses a dedicated media processor, in our case, the GPU. It allows the CPU or graphics card to complete other tasks. This provides higher PC performance. The hardware is designed for a defined set of codecs. The major manufacturers of graphics cards already have a built-in ASIC IP block on the latest GPU that performs video encoding. In the case of Intel, it is Intel Quick Sync Video (QSV). Nvidia has its own NVidia ENCoder (NVENC). Unlike video encoding on the CPU or a generic GPU, Quick Sync or NVENC is a dedicated hardware core on the die. This allows much more power efficient video encoding.

- Nvidia NVENC

From the Kepler architecture, NVIDIA GPUs contain a hardware-based encoder, which provides fully accelerated video encoding. It is independent of the graphics performance of the GPU. With complete encoding offloaded to NVENC, the graphics engine and the CPU are free for other tasks. The biggest handicap of HEVC NVENC is the fact that it does not support B frames (bi-predictive coded pictures). These apply to the second generation of Maxwell GPUs.

The older generation GPUs, Kepler and the first Maxwell generation, only had one NVENC engine. The second generation of Maxwell and Pascal generation has two NVENC engines. That enables to support a larger number of parallel encoding streams. Table 18 [103] provides an overview of NVENC performance of Kepler, Maxwell and Pascal GPUs for video with resolution 1920x1080, color format YUV4:2:0 and 8 bit color depth [103].

Table 18: The performance of NVENC in Different GPU and architectures.

H.264 [FPS]				HEVC [FPS]	
Kepler	First generation Maxwell	Second generation Maxwell	Pascal	Second generation Maxwell	Pascal
220	345	432	631	200	395

- Intel Quick Sync Video

The QSV uses the dedicated media processor to make video processing and conversion fast with sufficient quality. The name "Quick Sync" refers to quickly recoding, for example, a video from Blu-ray to a suitable format for a smartphone. If the 2nd~5th generation of Intel Core processors is used, it is possible to use the QSV H.264 encoder. In the case that the Intel Skylake processor is used, then there is an opportunity to use the QSV HEVC encoder. In the encoder, there are built-in modes to change the balance between the speed of encoding and video quality. The modes are: best, higher, high, balanced, fast, faster, fastest. In our test, the fastest preset and preset with the best quality was used [104].

5.2.3 Used Videos and Computers

The research was performed on uncompressed video clips in RAW format in Ultra HD and Full HD resolution. Full HD content was created by down-sampling using FFmpeg [91]. We used three 20-second long video sequences. Video sequences and their symbols in graphs can be found in Table 19 [89], [90].

Table 19: Overview of used test sequences.

Name	Screenshot	Symbol	Description
Tree		×	Camera moving slowly which shows the chateau park. Then focuses on the branches with leaves on the trees.
Duck		✱	A view of ducks in the lake. Ducks take off from the lake. Then there are waves on the water surface.
Cobra		◆	Static background with a slowly moving snake. The texture of the cobra's head is in detail and background is blurred.

Two quality levels were chosen, and each level has been defined by its bitrate. A low bitrate of 5 Mbps represents streaming with very limited bandwidth. On the other hand, a bitrate of 20 Mbps represents a high bandwidth system, for example, Ultra Blu-ray disk. The Group of Pictures was set to 8. Tolerance between the set bitrate and the real bitrate is in our test less than 3 %. This is an acceptable small error. Parameters of used implementations are in Table 20.

Table 20: Parameters of used implementations.

Codec	Version	Preset	Symbol in graphs
HM	15.0	/	A
X265	2.1	Placebo	B
		Ultrafast	C
Turing	1.01	Slow	D
		Fast	E
NVENC	2.57	H.265	H
		H.264	I
QSV	3.00	Best	F
		Fastest	G

For the realization of the tests, a personal computer was used where software encoding and encoding by NVENC was performed (in Table 21 labeled as "PC1"). A laptop was used for QSV encoding (in Table 21 labeled as "PC2"). The reason for using a laptop was that we did not have a PC with the latest Skylake architecture that would support QSV HEVC encoding. Therefore, it is not possible to directly compare QSV with other encoders. This gives us at least an approximate preview of the quality and speed of encoding. A personal computer was equipped with a fast SSD connected via PCI-Express. A laptop was equipped with a fast SSD connected via SATA interface.

Table 21: Parameters of used computers.

Name	PC1	PC2
CPU	i7-2700	i5-6200U
Architecture of CPU	Sandy bridge	Skylake
Frequency [GHz]	3.4	2.8
Core/Threat	4/8	2/4
GPU	Nvidia GTX 960	Intel HD Graphics 520
Architecture of GPU	Maxwell gen.2	Intel Gen. 9
RAM	16 GB 1333 MHz	8 GB 1600 MHz
HDD	SSD Kingston Predator 480GB	SSD SanDisk SD7SN6S-256G
HDD Speed R/W [Mbps]	1400/1000	520/470
CPU Flops [GFlops]	220.8	172.5
GPU Flops [GFlops]	2785.0	362.2

5.2.4 Objective Metrics and Realization of the Experiment

In research, there were used only full reference objective metrics to evaluate videos. Objective video quality metrics were calculated by VQMT [40] software. The objective metrics PSNR, SSIM VQM were calculated.

Measurement of encoding speed: All implementations, except for the reference HM, already have a built-in measuring algorithm of the time needed for encoding. For this reason and the inability to objectively assess the accuracy of the built-in measuring algorithm, an in-house program created in MATLAB was used. This program measured the time which was required to carry out the entire command of coding. It includes a load input uncompressed sequence and subsequent encoding. The difference between this time and the time measured by the encoder was 200 ms. This means that loading of video sequences, in our case, has no significant impact on the speed of encoding. This is mainly caused by very fast data storage.

Measurement of power consumption: Important aspects of encoding are not only video quality and speed of encoding. Power consumption is also a very important aspect. The power consumption of the whole PC consumed from the electrical network was measured. Energy consumption, which is listed in Table 22, represents only increase in power consumption against its idle state. The inaccuracy of various computers is partly eliminated by this method. In the idle state, PC1 consumes 61 W and PC2 only 5 W.

5.2.5 Experimental Results of Power Consumption

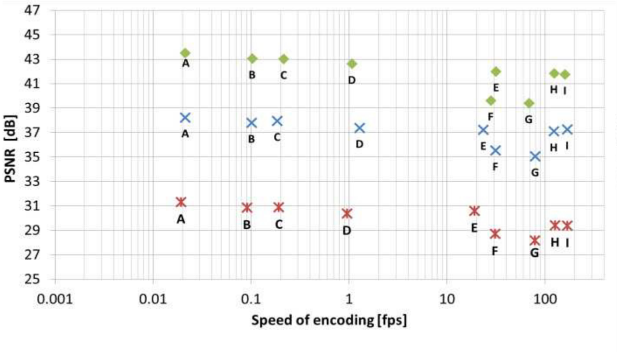
The result of power consumption suggests that implementation HM is not able to fully use the possibilities of the CPU. From this reason, power consumption is lower than in other software implementations. Utilization of the computer and power consumption are similar for Turing and x265 encoders. Utilization of RAM (Random-Access Memory) by the encoder is the lowest in the Turing encoder. On modern PCs, RAM consumption is acceptable in all implementations. The NVENC hardware encoder, in comparison with the non-reference implementation, has energy consumption about 20 % lower. The QSV encoder has the lowest power consumption. This is not directly comparable with the other encoders. As already mentioned, the QSV encoder uses a different computer.

Table 22: Power and sources consumption.

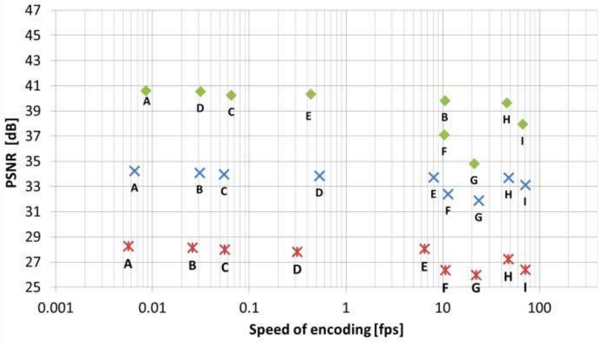
Encoder	Power consumption [W]	CPU Load [%]	GPU Load [%]	VE Load [%]	Used RAM [MB]
HM	32	17	0	0	1275
X265	79	99	0	0	2786
Turing	81	92	0	0	180
NVENCH.264	67	14	18	84	620
NVENCH.265	65	11	12	94	840
QSV	17	45	36	68	706

5.2.6 Results of Speed of Encoding and Video Quality

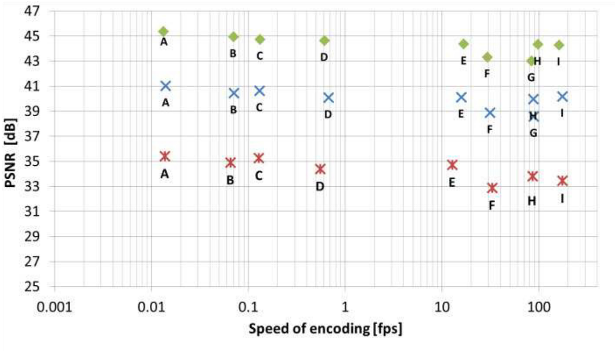
The results obtained from the objective metrics suggest that the speed of encoding depends a lot on the kind of encoder. As can be seen in Figure 32 and Figure 33, the x265 encoder with preset "Placebo" and Turing with preset "Slow" have similar encoding speeds. The Turing encoder with preset "Fast" is ten times faster. The x265 implementation with "Placebo" preset is two hundred times faster. Five times slower than the x265 "Placebo", is the reference implementation HM. Any software implementation is not able to achieve real-time encoding. Encoding in Full HD is approximately three times faster than encoding in Ultra HD. This applies to the software and hardware encoding. The PSNR quality of the HM encoder is better about 0.5 dB than the x265 implementation with the predefined profile “Placebo”. As it was mentioned in chapter 5.2.2 encoder NVENC H.265 does not support B frames. For this reason, it can be expected a lower video quality performance in comparison with other encoders. Despite this fact, the overall quality of NVENC is comparable with software implementations. The NVENC H.265 encoder is capable of encoding four Full HD streams parallelly in real-time. In the case of Ultra HD, it is just one stream. The quality of the QSV encoder is one class lower, however, the encoding speed is still high. QSV with preset "best" is five times slower than the QSV preset "fastest". The HM encoder has about 3 dB higher PSNR quality than QSV "Fast". To obtain four times higher bitrate, approximately 30 % longer time is needed to encode video. In the case of hardware encoding, the time is around 10 % longer. Encoding speed depends just slightly on the type of the video sequence. Non-reference software implementation can fully utilize the potential of a powerful PC. This results in significantly faster video encoding with just a small impact on image quality.



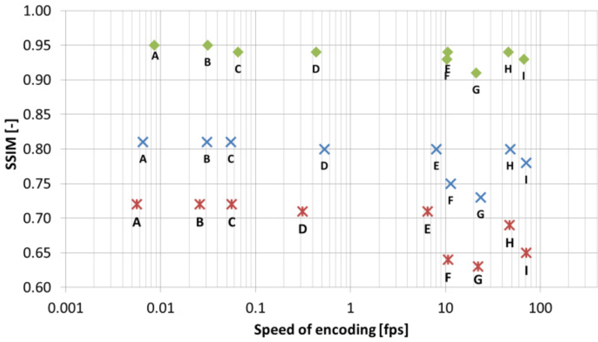
a) PSNR Full HD



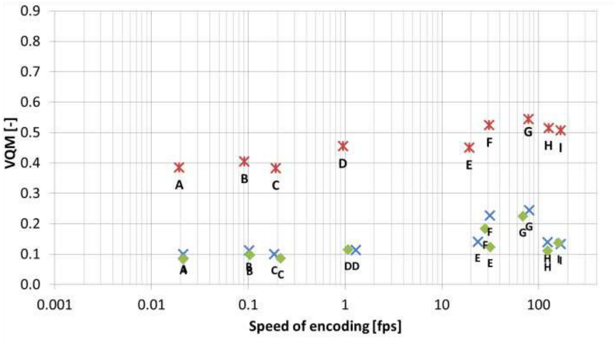
b) PSNR Ultra HD



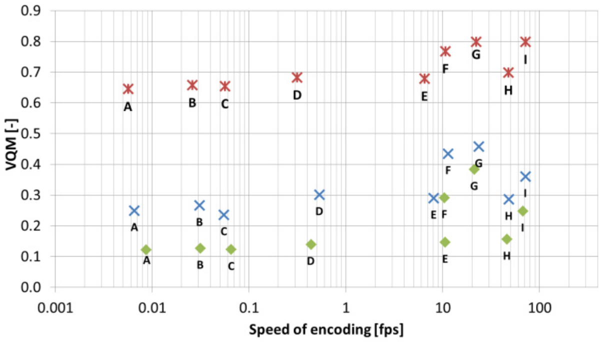
c) SSIM Full HD



d) SSIM Ultra HD

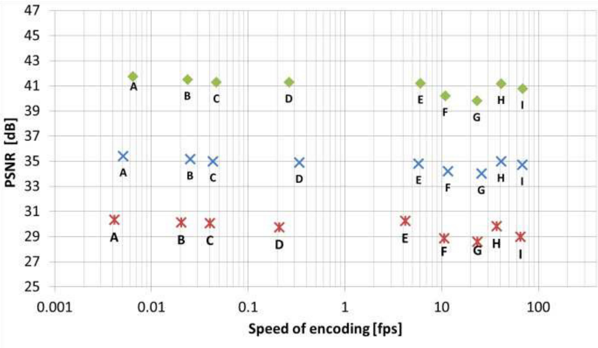


e) VQM Full HD

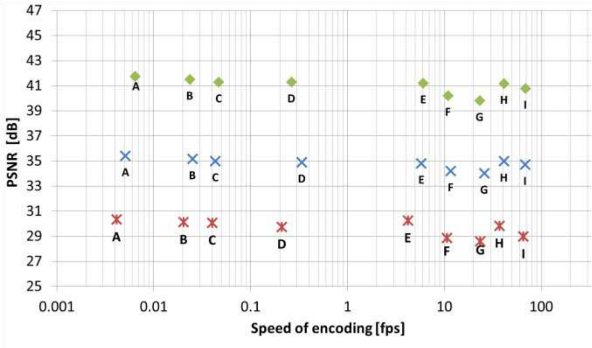


f) VQM Ultra HD

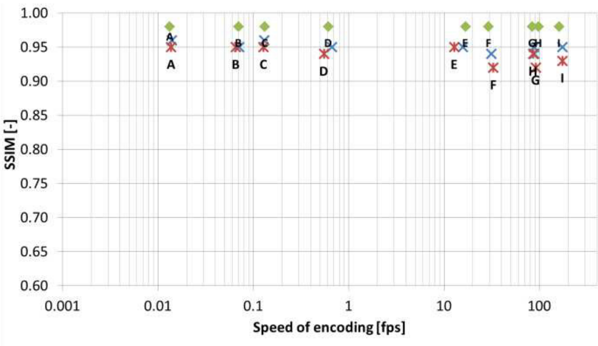
Figure 32: Results of quality and speed of encoding for videos with bitrate of 5 Mbps.



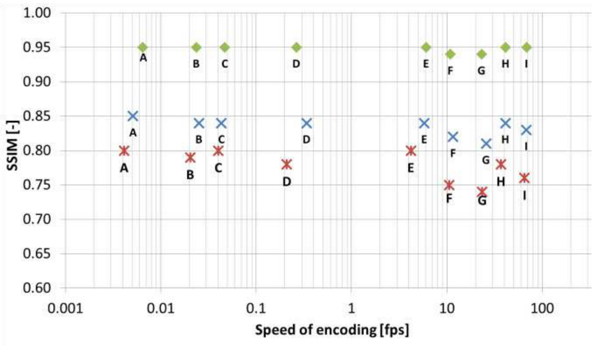
a) PSNR Full HD



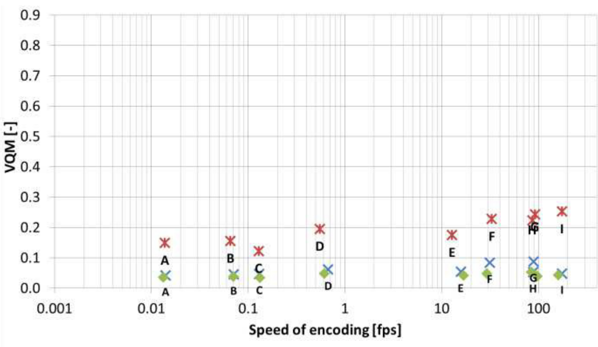
a) PSNR Ultra HD



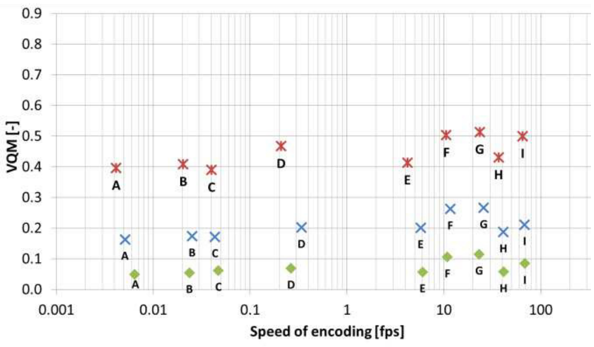
b) SSIM Full HD



b) SSIM Ultra HD



c) VQM Full HD



c) VQM Ultra HD

Figure 33: Results of quality and speed of encoding for videos with bitrate 20 Mbps.

5.2.7 Conclusion for Software and Hardware HEVC Encoding

The difference between the video quality of the reference implementation, Turing and x265 encoders, with the highest quality preset, is insignificant. The hardware NVENC encoder is about six thousand times faster than the reference HM with no enormous difference in quality. The encoder NVENC H.265 does not support B frames. In the experiment, there could be disabled B frames at all encoders, but it would be an unnecessary disadvantage for encoders that support the B frames. We can assume that, if the new graphics cards will support B frames in NVENC, then there will be an increase in video quality. The high speed of encoding, sufficient quality, and low power consumption can be achieved by the NVENC H.265 encoder. This encoder can be claimed as a “good encoding tool”.

6 VISUAL QUALITY IN STEREOSCOPIC TV

6.1 Database of Stereoscopic Sequences with Known Parameters

The perception of the spatial effect of the 3D video is influenced by several factors. One of the most interesting and crucial factors is the content of the observed video sequence. This factor is little investigated. The content of the scene can be described by particular parameters. The parameters have to describe spatial, temporal, depth and brightness distribution in sequences. Was created a database of the stereo video sequences complete with their quantified parameters, which express their contents. The sequences should serve for execution of subjective tests dealing with the impact of video sequence parameters on the observers' perception. Was executed a subjective test with this database [105].

6.1.1 Videos Sequences in the New Database

Was created short uncompressed sequences with various properties. In particular, the distance (depth) of captured objects, temporal dynamics (movement) and brightness of scenes are various. We assume these parameters play an important role in a stereoscopic perception. The influence of the depth of the scene was dealt with in articles [105] - [107], however we deal with a various parameter of the sequences.

The database used for investigation of the influence of scene content should include a large amount of sequences. For the investigation and detection of the influence of impact of one of the previously mentioned parameters, the database should contain groups of sequences with similar parameters. Various groups should mutually differ always in one property. This is important to assess the impact of each property.

The sequences in the new database are divided into two groups according to the method of their acquirement. To cover a variety of different source formats, we used two different sources of video sequences. The process of obtaining short uncompressed video sequences is shown in the right part of Figure 34. The program FFmpeg [91] was used for cutting and treatment of the video sequences. The first group contains sequences acquired from Blu-ray disc. The second group is a Camcorder recording. This group of sequences was created by using a consumer-level stereoscopic camcorder. The used camera has a stereo base of 2.5 cm and focal length (eq. 35 mm) 58 mm. Most of the sequences are without substantial cuts because it is necessary calculated average parameters for sequences. However, the database contains video sequences with cuts for examination of the impact of extremely high temporal information on a visual fatigue. The video sequences are uncompressed due to the elimination of the influence of coding. The main parameters of the video sequences are the following: resolution is Full HD and framerate is 25. The thumbnails of the sequences used in the subjective test are shown in the left part of Figure 34.

Figure 34: The thumbnails of used stereoscopic sequences and flowchart of the creating database.

6.1.2 Parameters of Stereoscopic Videos

This database is created primarily for a purpose of investigation of the influence of scene parameters to spatial percept. Therefore, the parameters of sequences and their calculation are important. The sequences can be described by many parameters. In order to preserve the clarity and acceptable amount of sequence groups, we have selected four base parameters, which determine a division into groups and three subsidiary parameters. Basic parameters are SI, TI, Depth Map Range (DR) and Brightness Median (BM) of the sequences. The subsidiary parameters are SSIM between left and right images, the Percentile of Brightness (BP) and amount of details in the sequence (ID). The mean value and standard deviation were calculated for each parameter. The values of the parameters for each image were obtained by a simple script created in the software MATLAB. Subsequently, the objectives will be shortly described. All parameters except SSIM and DR are calculated only from the left image. The basic parameters of the sequences used in the subjective test are shown in Figure 35. In the graph, there are plotted cumulative boxplot for each property. The sequences are sort by the division to groups.

The **spatial information** informs about the frequency of changes of the intensity. The SI is calculated as a mean change between the adjoining pixels on the vertical/ horizontal direction by the following relation in one frame in the left image

$$SI = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}|I_L(x_{i-1},y_j)-I_L(x_i,y_j)|+|I_L(x_i,y_{j-1})-I_L(x_i,y_j)|}{mn}, \quad (2)$$

where I_L is a pixel intensity in the left image, x_i, y_j represent a position in the image, m/n in number of pixels in the horizontal/vertical direction [20]. The first subtraction represents a difference in horizontal direction and the second one in the vertical direction.

The **temporal information** describes the time change of the scene. The difference between the brightness of two consecutive frames is calculated. Subsequently, the average value for whole sequences is determined. The following relation was used for calculation TI

$$TI = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}|I_L(x_i,y_j)-I_R(x_i,y_j)|}{mn}, \quad (3)$$

where symbols I_L, m, n, x_i, y_j have the same meaning as in relation (2), I_R is a pixel intensity in the right image [20].

The **median brightness and 80th percentile of brightness** inform better about the brightness of video sequences than the average value of brightness. The median and 80th percentile for individual frames are calculated using MATLAB function *prctile* and *median*.

The **depth map range** informs about a range of the depth in sequences. This parameter is determined for one frame as a difference between disparities of nearest and furthermost objects in the particular image. Subsequently, the average value for whole sequences is calculated (see the following relation)

$$DR = \frac{d_{\max(f)}-d_{\min(f)}}{n_f}, \quad (4)$$

where d_{max} and d_{min} are a maximal respective minimal disparity in frame f of the sequences and n_f is a number of frames in sequences. The value is given in a percentage of the image width. This parameter was obtained by a use of a program Stereophoto Maker [108].

SSIM informs about the similarity between left and right image of sequences. This parameter is mildly dependent on the depth map range. SSIM was calculated by use of the MATLAB function *ssim*.

The **details in the image** give similar information as SI, therefore we use it as a subsidiary parameter. This parameter is obtained as a number of edge pixels in the image. The edge representation was calculated by using a canny detector. The MATLAB function is *BW=(I,'Canny',threshold)*. The average value for a video sequence is calculated.

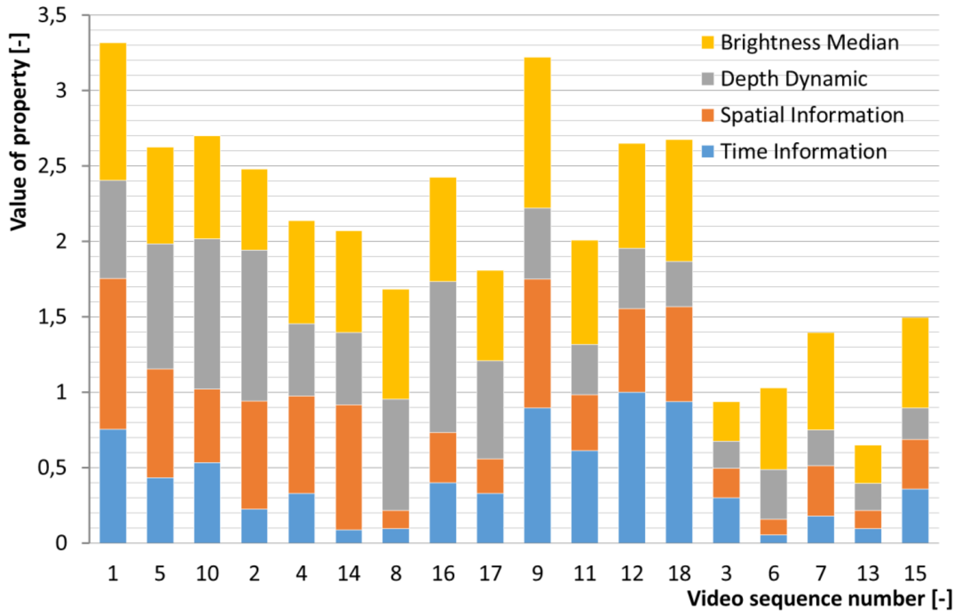


Figure 35: Basics parameters of sequences used in the subjective test.

6.1.3 Subjective Test with Database

For the purpose of showing the usefulness of a created database, was executed a simple subjective test. Only four groups of sequences are used in the executed test. The groups were selected to determine the influence of spatial information, temporal information, and depth map range. The average values of the important basic parameters in each group are in Table 23.

Table 23: Basic parameters of the used groups.

Group	*SI*	*TI*	*DR*
1	7.75	7.18	73.75
2	2.07	9.16	79.67
3	6.65	2.38	65.33
4	4.72	9.51	34.33

The subjective test was focused on the investigation of the impact of the video sequences parameters on the spatial perception. In this order, we strive to eliminate all other aspects influencing a spatial perception. Therefore, the video sequences are played uncompressed. The test methodology complies with the standard R BT.2021 [109]. For the best result, the test was executed in two steps. In the first step, respondents evaluated each video sequence on the discrete scale 1-5 point. Therefore, the single stimulus method is used in the first step. Eighteen sequences were used in the first step. In the second step, the respondents observed a reduced playlist of sequences. Only sequences, which were evaluated by 4 points in the first round, are in this new playlist. The second step is executed to distinguish the evaluation of the similarly evaluated sequences. The stimulus comparison is used in the second step.

The respondents attended a Randot test and a color perception test for verification that they are able to participate in the test. Twenty students of the BUT Brno have participated in the test. Stereoscopic Ultra HDTV with a diagonal 123 cm was used in the test. The display used the passive 3D system. Schematic layout of the measuring workplace is shown in Figure 36.

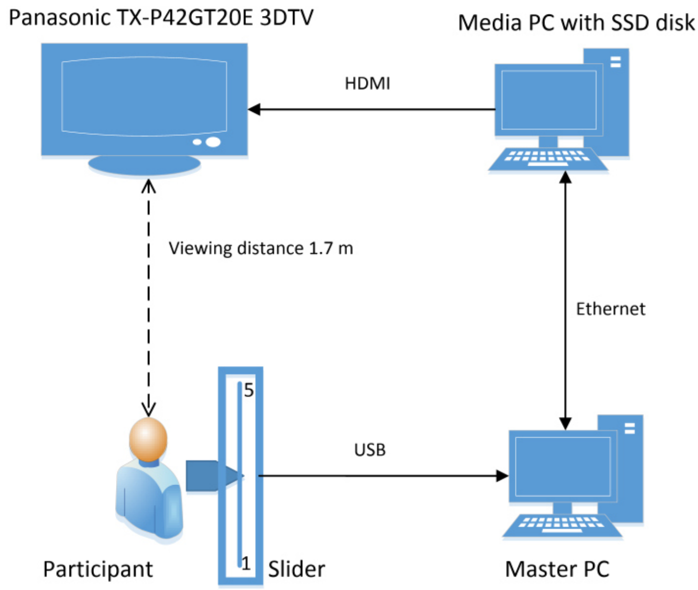


Figure 36: Schema of the testing place.

The optimal distance for observed display of this size was determined as 1.7 m. The respondents performed the test one by one. The respondents sat in an ideal viewing position. The respondents were asked two questions:

- How do you like this scene (**attractiveness**)?
- How comfortable do you feel (**comfortability**)?

The respondents were informed that the first question is focused on their enjoyment, so they do not evaluate the potential errors in sequences. The second question is focused on the unpleasant feelings originating from the properties of a nowadays used stereoscopic system.

6.1.4 K parameter of video sequences

The aim of the evaluation was finding of dependency of the evaluation by the respondents on the parameters of video sequences. The dependency of the MOS of attractiveness on the basic parameters SI, TI, DR and BM are shown in Figure 37. Obviously, some dependencies exist. However, these dependencies are not strong. Therefore, we empirically determined coefficient K which is given by following formula

$$K = \frac{SI + TI + DR \cdot 5}{BM}. \tag{5}$$

From the results, we can see that DR has the greatest impact to the attractiveness, therefore DR has greater weight in relation. The attractiveness rises with increase of all parameters except BM, there for BM is in the denominator. The coefficient K determined with good probability evaluation of the feeling of respondents. The dependency of the attractiveness on the parameter K is shown in Figure 37.

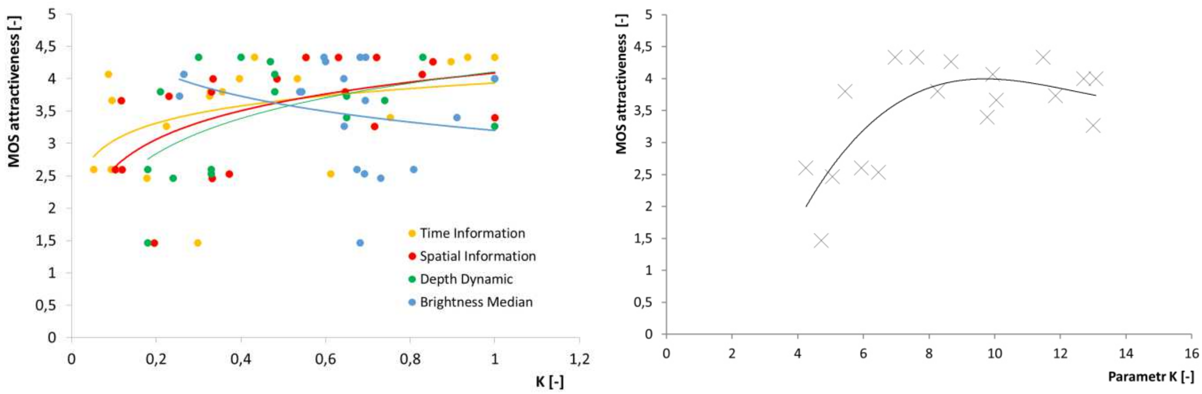


Figure 37: Dependencies of the MOS of attractiveness on the parameters of sequences and on the proposed parameter K.

The dependency of the comfortable on the basic parameters SI, TI, DR and BM are shown in left part of Figure 38. From the result, the comfortability increases with the decrease of the parameters SI, TI, DR and BM. The dependency of the comfortability on the parameter K is shown in Figure 38.

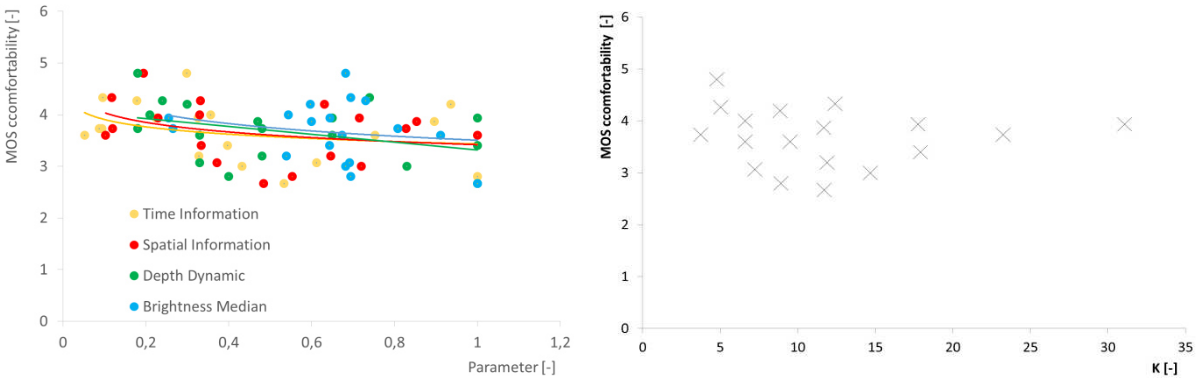


Figure 38: Dependencies of the MOS of comfortability on the basic parameters of sequences and on the proposed parameter K.

6.1.5 Conclusion of new stereoscopic database

This chapter deals with the creation of a new database of 3D video sequences with quantifying parameters of visual content designed to investigate the influence of video parameters on the stereoscopic viewer perception assessed by a subjective evaluation of observers. The aim of these evaluations of video sequences with quantified video content should be an objective metric based on the parameters of the scene. The database contains a large amount (50) of video sequences divided into eight groups depending on their properties. The following properties were calculated for each video sequence: Spatial Information, Temporal Information, Depth map Range, Median of the Brightness on the scene and its percentile, SSIM between left and right partial frame, and the number of details in the scene. The created database was used in the subjective test. The test used only a part of the database (18 video sequences). The results of the test with quantified video content, that allow examining the influence of the spatial information, temporal information and depth map rang, were obtained. The attractiveness of the sequences has the strongest dependency on the depth map range. In accordance with the test, the parameter K was proposed. From the dependency of the evaluation of the video sequences on the parameter K, it is obvious that the probability of good attractiveness of sequences rises if the parameter K reaches a certain size.

6.2 Models for Comparison of Compression Algorithms for 3DTV

This chapter focuses on a study of different compression algorithms for stereoscopic videos. The well-established H.264 AVC and upcoming HEVC standards as well as their multiview extensions H.264 Annex H - MVC and H.265 Annex G – Multiview HEVC are considered. Objective video quality metrics are used to analyse the compression efficiency of the considered codecs, extended with results from subjective tests. The correlation between the objective and subjective scores for stereoscopic videos are statistically analysed. The last important outcome is the creation of objective models for the best correlation with subjective stereoscopic video quality. Some parts of this chapter are a joint work with Ondrej Kaller [110]. We try to answer the following points:

1) Which 2D objective metrics best correlate with the users mean opinion score for 3D videos, based on additional statistical analyses?

2) Can any 2D objective metrics be optimized for better 3D accuracy? Alternatively, can a model be created that has a better correlation with 3D video sequences?

6.2.1 Video Sequences and Encoding Parameters

As the source of 3D video sequences, was used four samples which are available in databases [111] and [80], to make our research have a wider range of uses. All these sequences were in Full HD resolution for each view and had a frame rate of 25 frames per second. The length of each video sequence before encoding was adjusted to 10 seconds. This is a typical length used in subjective video quality studies. The selected video sequences cover a wide variety of contents as can be seen from the SI and TI in Table 24. The table also contains one frame of each corresponding sequence. Both parameters SI and TI were calculated according to ITU-T P.910 [20]. The average value of depth for 5, 50 and 95 % for each video sequence [108] was calculated, as can be seen in Table 25. The average value of the depth of the videos varies considerably.

Table 24: Description of used stereoscopic videos.

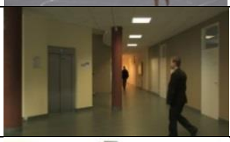
Name	Thumbnail	Description	SI [-]	TI [-]
Basketball		Moving, dolly and panning wide shot. Fast and unpredictable movement. Moderate 3D effect.	35.4	18.0
PoznanHall		Static view into the hall with the slightly moving camera, walking man. Significant 3D effect.	14.8	9.4
Train		Moving subject. Fine details and textures. Moderate 3D effect.	61.6	4.4
WishingWell		Static close-up shot. Major warping and highlights. The strong 3D effect, emphasis on warping.	41.7	26.9

Table 25: Average depth of video sequence.

	Video			
	Basketball	**PoznanHall**	**Train**	**WishingWell**
Depth d05	-10.59	-40.95	8.17	-24.06
Depth d50	-3.96	-24.25	12.40	-11.63
Depth d95	3.66	-12.54	15.27	18.56

As input to encoders, only bitrates were defined together with searching motion range without any other system parameter modification and without any tuning of the encoders. The quality profile was set to the highest quality because we were focusing only on the quality of encoding, not the encoding speed. Encoders selected other parameters automatically by itself. A summary of video encoders settings used in encoding is provided in Table 26. Target bitrates were adjusted between 0.5 Mbps and 4 Mbps. The target bitrate applies to one view only. Let us give an example for the 1 Mbps bitrate: For the 2D encoder, the total bitrate 1 + 1 Mbps was set. For multiview encoders, the stream of both views was 2 Mbps. This means that the total data rate is the same. The searching range for HEVC-based encoders was set to 64 pixels to take full advantage of these modern encoders.

Table 26: Parameters of used encoders.

Codec	AVC	MVC	HEVC	MV-HEVC
Encoder	x264 r2597	FRIM x64 1.25	HM 15	HTM 15.1
Profile	High	High	Main	Main
Level	5.1	4.0	5.1	-
Preset	Very Slow	1 (quality)	-	-
Search Range	32	32	64	64
GOP Size	8	25	8	8
Deblocking filter/SAO	Yes	No	Yes	Yes
Entropy coding	CABAC	CABAC	CABAC	CABAC

6.2.2 Objective and Subjective Stereoscopic Video Quality

For objective evaluating the quality of the encoded video sequences, we used three full-reference objective quality metrics, namely PSNR, SSIM and VQM.

A subjective test was then performed. All subjective video quality assessment was conducted in a special test room. Laboratory conditions were set up according to ITU-R BT.500-13 [44] including a room with controlled lighting. A plasma stereoscopic television (Panasonic TX-P42GT20E) was used to display video content. The television's active shutter 3D system was used. The biggest advantage of this method of 3D view is that it does not reduce the resolution, in contrast to polarization 3D systems. The peak luminance of the display was adjusted to 200 cd/m^2. The viewing distance of the participants from the display, according to

ITU-R BT.2022 [46], is the height of the picture multiplied by 3.2. In our case, the optimal viewing distance is 1.7 meters (see Figure 35). In the subjective test, only one participant was in front of the television to eliminate the effect of different observation positions.

The ACR subjective method was adopted. The used measuring scale was chosen on a five-point scale from the lowest quality 1 (Bad) to the best quality 5 (Excellent). As the pretest 3D sequences in three qualities were played. These sequences were different from the sequences used during the test. Observers had an overview of how the 3D movie could look. Sequences were played for each group of participants randomly and participants did not know what encoder or bitrate was presented. Participants rated the quality on sliders which were connected to the master computer (see Figure 35). This computer also controls the media computer from which the video sequences were played.

A total group of 37 observers participated in the 3D subjective test. Two of them were female. University students and employees were recruited with an average age of 24. Color blindness - Ishihara test (See Figure 39 [77]) of all participants was tested as well as their ability of stereoscopic vision (Randot stereo test) [77]. Three people who did not pass the tests were not included in the final evaluation.

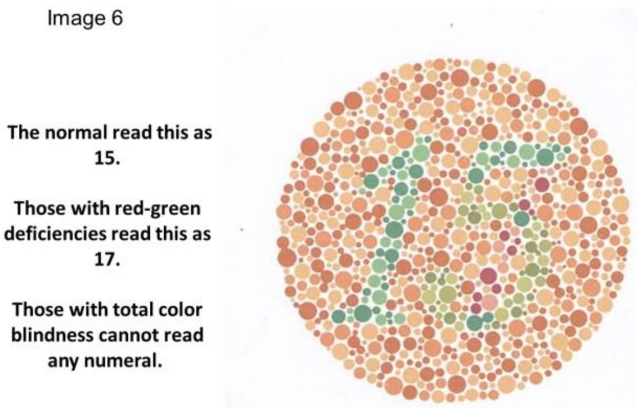


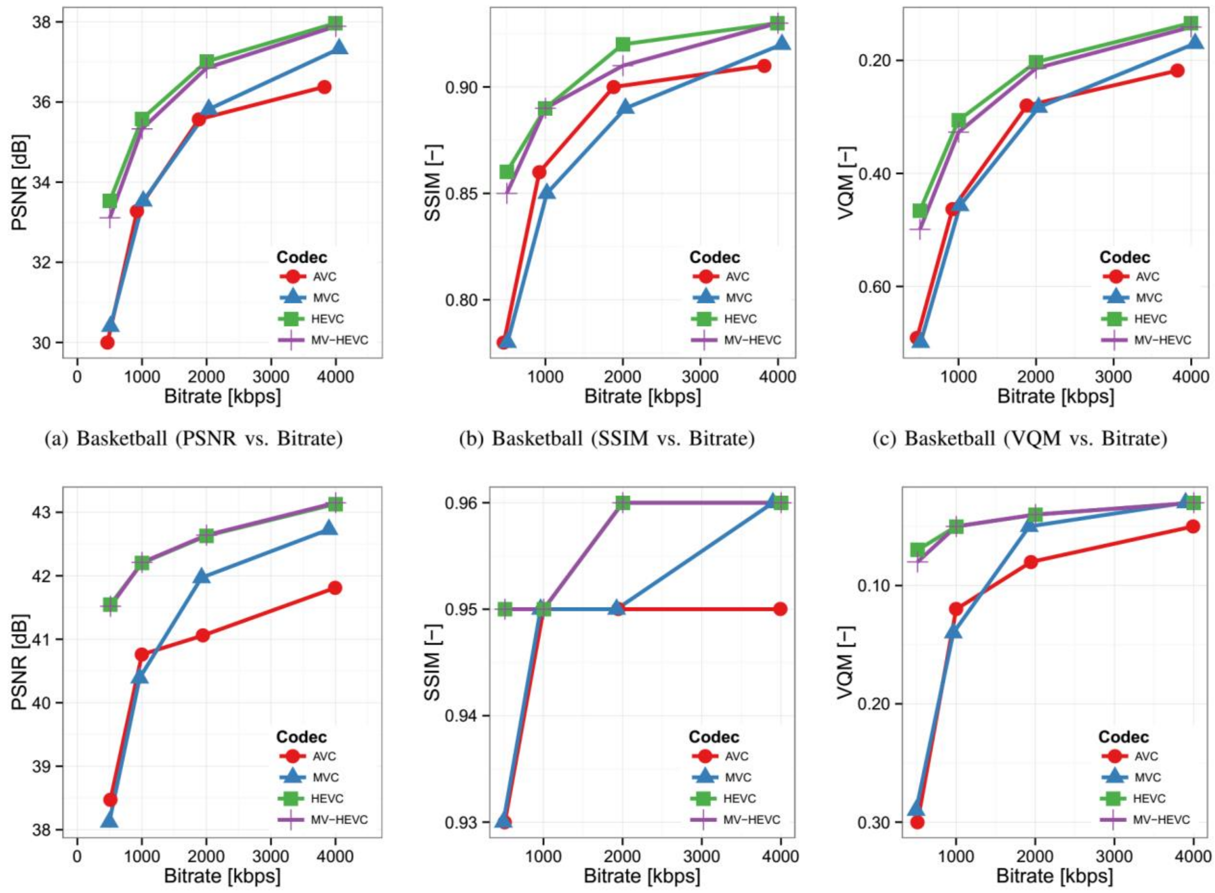
Figure 39: Thumbnail of Ishihara test for testing of color blindness.

6.2.3 Coding Efficiency of Modern Stereoscopic Encoders

The results obtained from the objective metrics and subjective test are evaluated, compared, and discussed in this section.

The results of objective metrics show that the performance of standard codec and the mutual comparison is highly content dependent. For several content types, the multiview coding gain is significant, while for other contents the multiview coding only brings undesired overhead with no performance improvement. Objectively measured quality of the encoded sequences can be seen in Figure 40 and Figure 41. The first row of the graphs shows the results of metrics PSNR, SSIM and VQM for content Basketball, respectively. As we are applying a 2D video quality metric on 3D video, the metrics were computed for both views separately. Considering the results for left and right views differed only slightly, the values for the left view only are used [71].

Overall, the codecs belonging to the same standard exhibit very similar performance with differences in PSNR in the order of 1 to 3 dB [112]. There are, however, interesting exceptions and the behavior changes significantly content-wise:

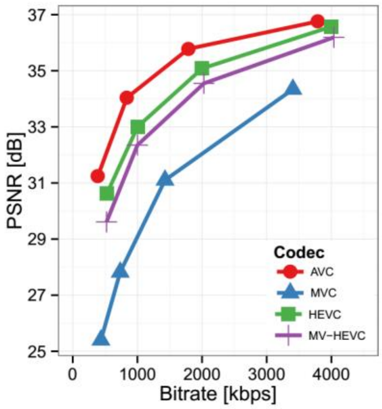


(a) Basketball (PSNR vs. Bitrate) (b) Basketball (SSIM vs. Bitrate) (c) Basketball (VQM vs. Bitrate)

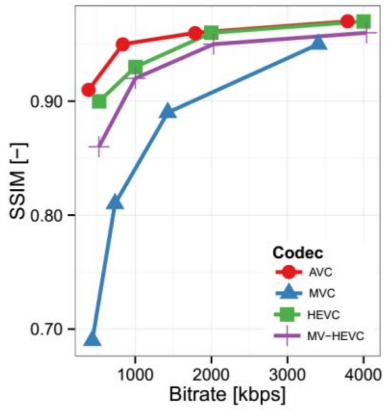
(d) Poznan-Hall (PSNR vs. Bitrate) (e) Poznan-Hall (SSIM vs. Bitrate) (f) Poznan-Hall (VQM vs. Bitrate)

Figure 40: Results of objective metrics for Basketball and PoznanHall videos.

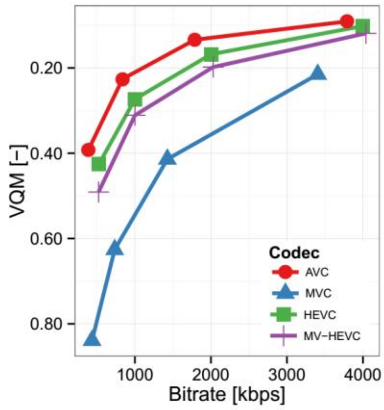
- **Basketball:** The performance of MPEG-4 AVC and MVC is very similar, as well as the performance of HEVC and MV-HEVC. Thus, for this particular sequence, the encoding gain of multiview encoding is negligible.
- **PoznanHall:** While for HEVC and MV-HEVC the rate distortion efficiency is almost identical, visible difference rise between MPEG-4 AVC and MVC at bitrates above 2 Mbps, in favor of MVC. The scene contains significant global motion which appears to be efficiently described by the multiview extension.
- **Train:** An exceptional sequence. While the HEVC-based codecs still prove similar performance, obviously HEVC performs better than MV-HEVC, i.e., the encoding gain of the multiview codec is negative. Even larger differences are obvious between MPEG-4 AVC and MVC, also in favor of MPEG-4 AVC.
- **WishingWell:** While MVC performs significantly better than MPEG-4 AVC, the situation is opposite for HEVC having worse performance than MV-HEVC. We have identified a sequence in which MVC is really able to exploit multiview encoding potential.



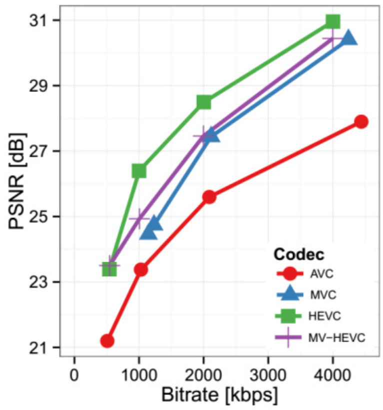
(g) Train (PSNR vs. Bitrate)



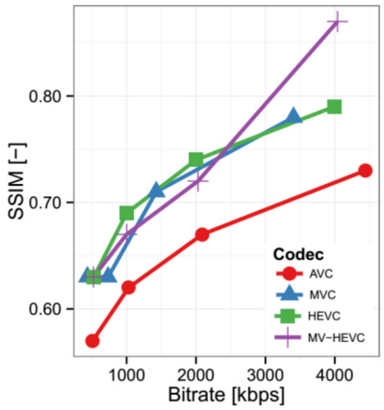
(h) Train (SSIM vs. Bitrate)



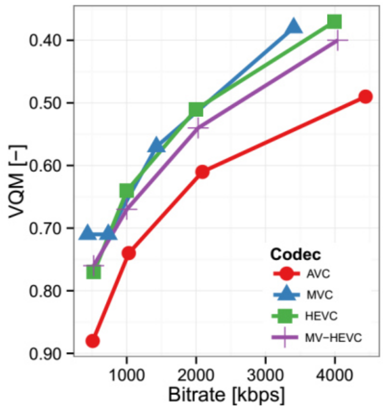
(i) Train (VQM vs. Bitrate)



(j) Wishing-well (PSNR vs. Bitrate)



(k) Wishing-well (SSIM vs. Bitrate)



(l) Wishing-well (VQM vs. Bitrate)

Figure 41: Results of objective metrics for Train and WishingWell.

Now let us focus on the subjective test. Results of subjective tests for all sequences and codecs are presented in Figure 43. A legend in the figure is presented as follows: "First is the numbering and after that is the name of the video sequence, used encoder and target bitrate. The last column presents the MOS". For example, in the first row, the third line is the sequence "Basketball" encoded by AVC with bitrate of 2 Mbps. The central red mark in MOS is the median, the edges of the blue box are the 25th and 75th percentiles. The most extreme data points, without outliers, are the black whiskers. Outliers (Red Cross) are plotted individually. The following lines describe the subjective test results.

- **Basketball:** The performance of AVC and MVC is very similar. In addition to the highest bitrate, there MVC is better. For HEVC and MV-HEVC, the quality is the same for all bitrates. There is no increase in quality between the bitrates 2 and 4 Mbps. The results of the subjective tests correspond approximately to the objective metrics.

- **PoznanHall:** In the case of the HEVC codec, there is a gradual increase in quality with higher bitrates. On the other hand, with MV-HEVC, the quality was similar for all bitrates.

- **Train:** The coding efficiency of HEVC and MV-HEVC is similar. Bitrate higher than 1 Mbps does not cause predicted improvement in the QoE. In the case where bitrate is higher than 2 Mbps, then the coding efficiency is similar for all codecs. There is no coding gain of the multiview codecs.

- **WishingWell:** The performance of MVC is significantly better than AVC. It is a situation in which the codec is able to exploit multiview coding potential. Comparable results were obtained for codecs HEVC and MV-HEVC.

The results show that the scattering of the test subject's evaluation in the subjective test is large. For this reason, it was necessary to evaluate the participants who acted as outliers. We have used the whisker method for outlier values detection. Whisker extends the interval of quartiles (Q25, Q75) by w (itself) on both sides. In our case, w is equal to 1.5, which would correspond to 99.3 percentiles coverage of values, in case of normal distribution [113].

There are two hypotheses about outlier rate. First, there is a difference in variance of QoE evaluation among the sequences. Second, the variance is larger at the beginning and at the end of the testing session, due to disorientation and fatigue. A hypothesis was tested concerning the uniform distribution of outliers through sequences and time. The Hi-square goodness-of-fit tests against discrete uniform distribution, in the case of sequences and time, have rejected this proposition at a significance level of 0.05 [114]. We can prove that in our subjective test, after significantly lower outlier parts (first 8 video sequences) the rest of the evaluation time has uniform outlier rate. This hypothesis can be seen in Figure 42. In these graphs, the blue color indicates the results which were below the permissible deviation. The results marked in yellow color are those that were above the error of the mean.

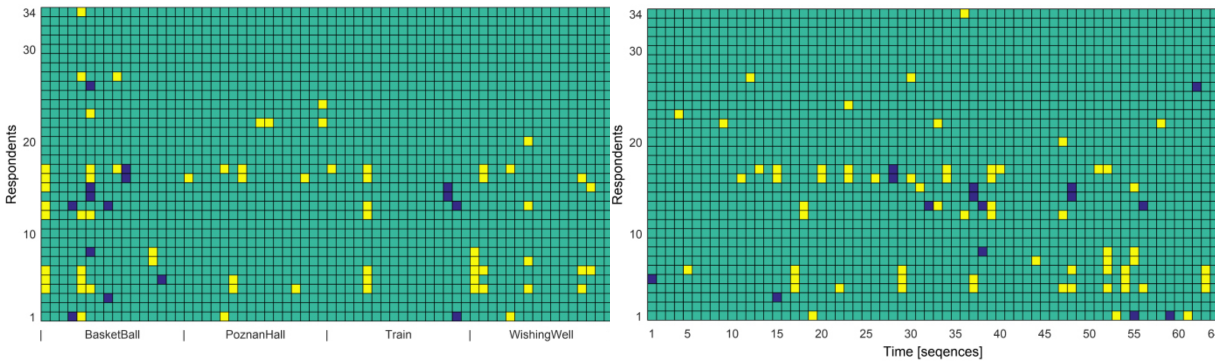


Figure 42: Dependence of the number of outliers on the sequence and on the playing time.

The data from participants, which has more than 10 % of outlier evaluations, were excluded. The number of participants not included in the final evaluation is four, which amounts to 11.8 % of the test base.

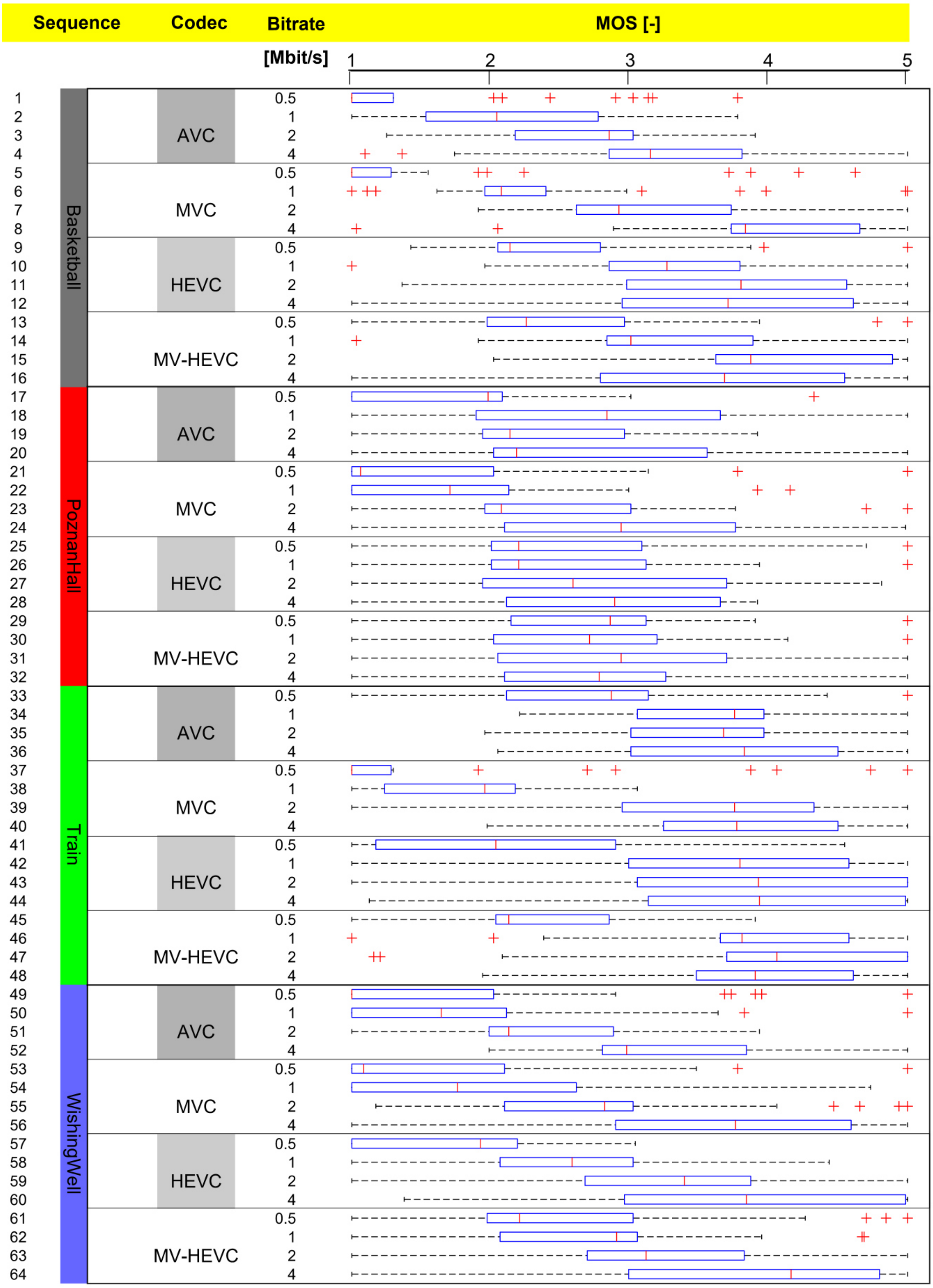


Figure 43: Results of the subjective stereoscopic test.

Table 27: Results of Wilcoxon signed-rank test.

Hypothesis about MOS	Rejection of the hypothesis	p [-]
AVC ≈ MVC	0	0.6483
HEVC ≈ MV-HEVC	0	0.4616
Codecs H.264 vs. H.265 for stereoscopic compress efficiency		
Same compress efficiency	1	0.0024
H.265 has double compress efficiency	0	0.3128

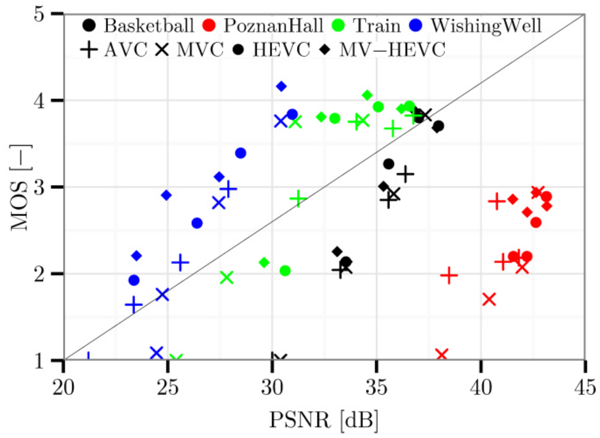
The results of the Wilcoxon signed-rank test [115] are presented in Table 27. This test did not reject the hypothesis that AVC and MVC have similar coding efficiency. The same result is also obtained for HEVC and MV-HEVC. The H.265 standard was designed to produce a 50 % less bitrate compared to the H.264 standard for the same image quality [101]. The hypothesis that H.265 generation needs half the bitrate to compare to H.264, for the same quality also in the stereoscopic video, has been proved.

6.2.4 Correlation of Objective Metrics and Subjective Test

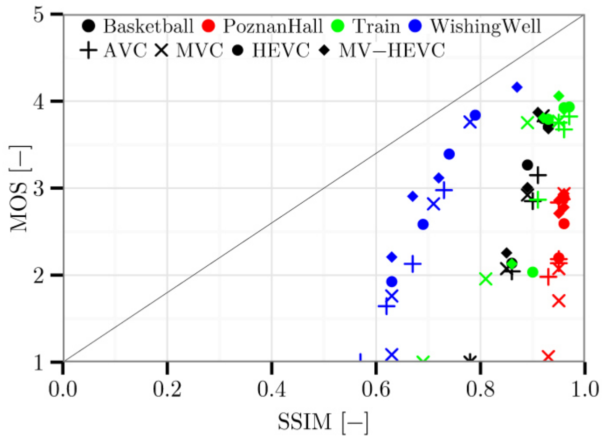
After evaluating the coding efficiency, it is necessary to determine which 2D metric has the greatest correlation with the subjective 3D test. For these purposes, SROCC and PCC were applied to the results. These analyses are used to determine the correlation between objective and subjective metrics. The correlation scores are between +1 and -1, where -1 and +1 mean total positive and negative linear correlation respectively, and 0 denotes no linear correlation. The VQM objective metrics have negative values in correlation because their lower score indicates higher video quality. The correlation analysis was firstly applied to each sequence separately (see Table 28). Due to the fact that we need a universal metric, the correlation value was then calculated across all sequences. The results show that the correlation depends on the video content. The video "PoznanHall", which is from another video database, has a different correlation than other videos. It may also be due to the fact that the video has a large stereoscopic parallax (see Table 25). For some viewers, it could be distracting and therefore the video has a non-standard rating. For this reason, in the last row of Table 28, the "PoznanHall" sequence is omitted and the resulting score, just in this row, is calculated without it.

Table 28: Pearson correlation coefficient and Spearman’s rank order correlation coefficient.

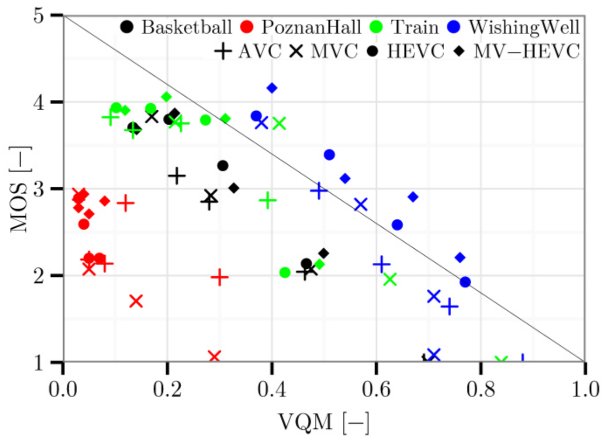
Video Sequence	PSNR		SSIM		VQM	
	PCC	*SROCC*	*PCC*	*SROCC*	*PCC*	*SROCC*
Basket	0.980	0.919	0.965	0.899	-0.974	-0.892
PoznanHall	0.261	0.442	0.293	0.472	-0.502	-0.484
Train	0.606	0.648	0.848	0.696	-0.802	-0.648
WishingWell	0.872	0.939	0.888	0.965	-0.877	-0.905
∑	0.261	0.226	0.462	0.415	-0.511	-0.407
Seq.- 1.,3.,4.	0.608	0.625	0.672	0.702	-0.815	-0.792



a) PSNR metric and the subjective test



b) SSIM metric and the subjective test.



c) Correlation between the VQM metric and the subjective test.

Figure 44: Dependence of objective metrics on the subjective test.

The correlation between objective and subjective methods is plotted in Figure 44. The black markers represent the video "Basketball", whereas red, green and blue colors indicate videos "PoznanHall", "Train" and "WishingWell", respectively. After a thorough comparison of all objective and subjective scores, it can be concluded that in our case the VQM objective metric best reflects the user's QoE for compressed 3D videos.

6.2.5 Generic Creation of Models for 3D Video

Although the VQM metric has the highest correlation, it is still not ideal for evaluating 3D videos. We thus propose our own model, which better models our subjective test results. Such a model should provide sufficient general predictions at least for content with similar parameters as the used video sequences. This section describes the model proposal, validation and verification of the models, and a description of each proposed model is provided at the end.

Table 29: Available objective parameters of the sequence.

2D parameters	2D metrics	Depth description	VQM coefficients
SI	PSNR	d05	si_loss
TI	SSIM	d50	hv_loss
	VQM	d95	hv_gain
		dDR	color1
			si_gain
			contati
			color2

We have several objective parameters specifying Source Referent Contents (SRCs) as SI, TI and disparity. Other parameters describe our interventions -- Hypothetical Reference Conditions (HRCs) as PSNR, SSIM or VQM coefficients. All the available sequence parameters (potential regressors) are summed up in Table 29. The column titled "Depth description" contains four parameters related to content depth. The first three are the quantiles (d05, d50, d95) of disparity distribution. The fourth parameter is disparity dynamic range, defined as d95 - d05. The disparity is calculated for a sufficient amount of significant corresponding pixels by the Speeded-Up Robust Features (SURF) algorithm [116]. The last column contains seven coefficients whose linear combination forms the VQM value.

We have only 64 samples of MOS, which is the response variable. To avoid over-parameterization, it is necessary to reduce the number of regressors. A good model needs about a hundred observations to one regressor. According to [117], to detect reasonable size effects with reasonable power, 10 - 20 samples per parameter are needed. The disproportion between the number of potential model parameters and "training" data is also the main reason of that why we focused on linear modeling.

Now we are focusing on model estimation methods. The simplest and very common model estimation method for the General Linear Model (GLM) is Ordinary Least Squares (OLS). The OLS method minimizes the sum of squared residuals, which are the differences between the observed values and the estimated values of the quantity of interest. In our case, these values are the median of subjectively estimated quality (MOS) and the modeled MOS value. As we cannot exclude the correlation of regressors, Generalized Least Squares (GLS) has been utilized as the model estimation method [117], [118]. There are two criteria on which regressors have been selected into our models: Akaike information criterion and coefficient of determination [119].

The Akaike Information Criterion (AIC) is a measure of the relative quality of statistical models for a given set of data. AIC is based on minimizing the relative information lost when a given model is used to represent the process that generated the data. It sets the proportion between the goodness of fit of the model and the complexity of the model. This level of parsimony is a function of input data sample relevance in a population. The AIC coefficient does not keep any absolute information about model quality, but the model with the lowest AIC is relatively best from the tested set.

The coefficient of determination R^2 is the proportion of variance explained by the model to the variance of explained (modeled) variable. In the case of linear regression with statistically independent regressors, R^2 is the square of the coefficient of multiple correlations between model output and independent (explanatory) variables. The coefficient of determination is increasing with the number of regressors, even if they do not bring other additional information. To choose a model with the optimal number of parameters, the adjusted R^2 is used. The adjusted R^2 ($\approx \bar{R}^2$) is the best estimate of the degree of relationship in the basic population. The coefficient $\bar{R}^2$ determines how our linear model would describe the population if we had ideal data samples.

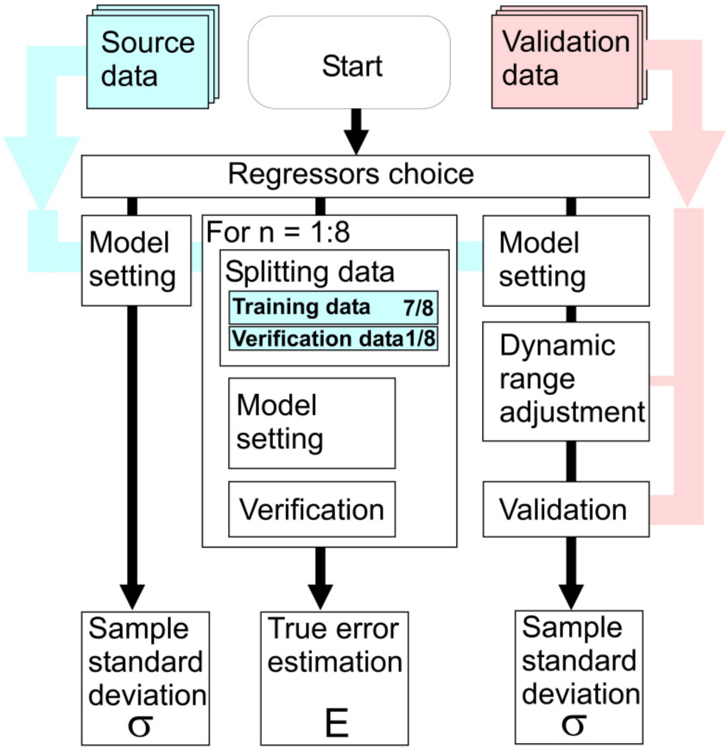


Figure 45: Model setting - verification and validation process.

The flowchart in Figure 45 shows the process of setting models (left column), their verification (middle one) and validation (right column). First, the regressors are chosen from Table 29 at the base of the criteria mentioned in previous paragraphs. Secondly, the model is set by the GLS model estimation method. The standard deviation per sample (σ), sometimes in literature called as Root Mean Square Error (RMSE), is calculated. Here, RMSE is the ideal point estimation of σ. In the case of model verification, the dataset is randomly divided into 8 parts (literature recommends 5-10, a divisor of 64 was chosen). The model is set to training data (7/8 subpart of original data) and σ_1 is calculated from verification data (1/8 subpart). After 8 repetitions, the arithmetic mean value of σ_1 - σ_8 is calculated, called true error estimation (*E*). Figure 46 shows a residual plot, the scatter plot of verification samples deviations. It demonstrates how the observed values differ from the point of best fit. We can obtain a good overview about model bias and homoscedasticity.

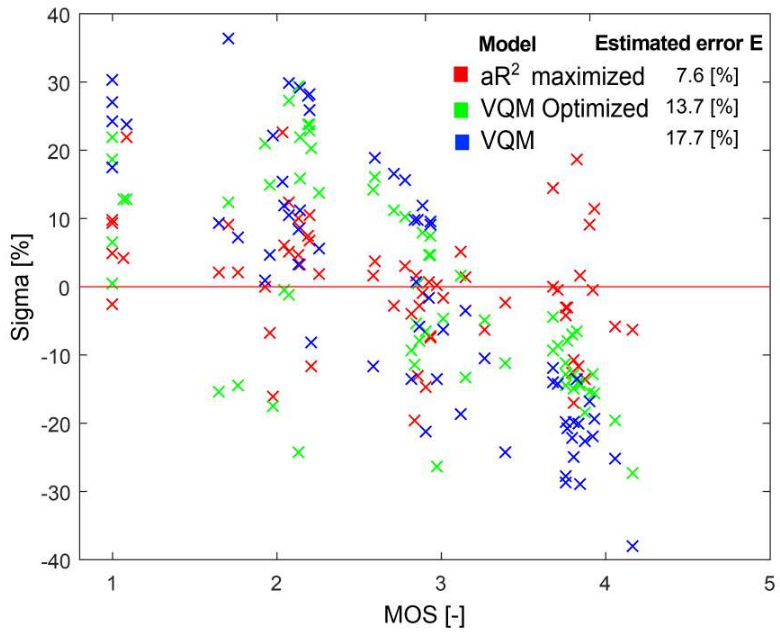


Figure 46: Scatter diagram - the residual plot of relative error of MOS.

The next step is validation. The right column of the flowchart in Figure 45 describes the process of model validation. For this purpose, a validation dataset has been added -- other video sequences than those used in the subjective test. The validation videos (SRC 1-8) come from RMIT3DV - an uncompressed stereoscopic 3D HD video library. This database has been provided by RMIT University in Melbourne [80]. From the sequences, those whose (potential regressors) parameter values are within the range of the original data values have been chosen. As validation data (SRC 1-8), the sequences 3D_01, 3D_03, 3D_05, 3D_16, 3D_17, 3D_29, 3D_42, 3D_48 were used. The HRC applied on selected sequences was HEVC with four levels of compression ratio (2x [250, 500, 750, 1000] kbps). The validation data is fully independent. The subjective tests have been done with other respondents. Once again, the ACR subjective method was used. Furthermore, the same display technology and test environment have been used.

The dynamic range adjustment is the second step done with the set of model. The full-reference objective video quality metrics as SSIM, VQM, Moving Pictures Quality Metric (MPQM) [120], Noise Quality Measure (NQM) [120] tend to be global QoE models. The generality of the metrics goes against accuracy, even in very complex models. Our goal was to make the most accurate model with limiting data amounts. Although the respondents are trained to set their quality dynamic range, they tend to utilize the full range of the QoE scale. This is the reason, why we decided to adjust the dynamic range of our model to validate the data optimally [40], [121], [57].

Validation is done by calculating the standard deviation of the model results and MOS values. The bar graph in Figure 47 shows MOS values and a gray box containing 50 % of voted quality values. Three various point estimations of MOS, as the three corresponding linear models results, are plotted as color cross marks. The colored background refers to the validation content SRCs 1-8. Each colored surface contains the MOS values of four HRCs applied in one sequence.

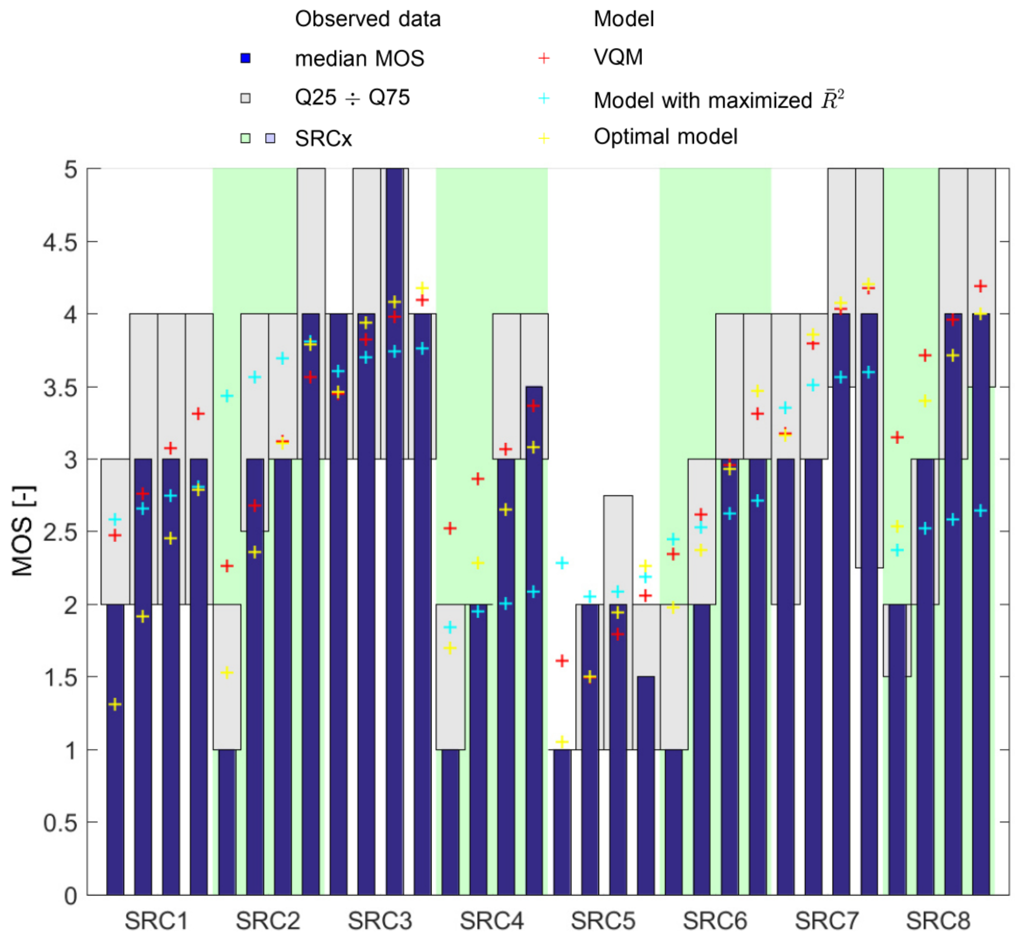


Figure 47: Validation chart - The MOS values of validation data and their model's predictions.

6.2.6 Innovative Models for 3D Video Content

The details of the models are described in this subsection. Each block of text describes an individual model including its properties and differences from others. Table 30 sums up the model's accuracy and verifications. The correlation coefficient was calculated from original and validation data, therefore, they did not correspond to results from Table 28. The σ denotes the standard deviation per sample. It is calculated for the original (training) data, verification data and through both datasets (designated ∑). The standard deviation has not been calculated for the PSNR metric. This metric does not have defined range. Unlike, for example SSIM metric, where the value ranges from 0 to 1. The standard deviation for Model VI. cannot be calculated separately for original and validation data because both datasets are used as the input data of the model.

Table 30: Models description, their standard deviation and Pearson Correlation Coefficient.

Model	Description	Regressors	Adjusted R^2 [-]	σ/sample [%]			PCC [-]
				Original data	Validation data	∑	
PSNR	Generic PSNR	PSNR	0.376	-	-	-	0.411
SSIM	Generic SSIM	SSIM	0.489	46.54	43.09	45.40	0.547
VQM	Generic VQM	VQM	0.592	32.06	15.65	26.59	-0.567
I.	VQM optimized	si_loss, hv_loss, hv_gain, color1, si_gain, contati, color2	0.577	13.80	25.51	17.70	0.631
II.	AIC minimized	PSNR, SI, TI, color2, d50	0.842	8.57	26.37	14.50	0.713
III.	aR^2 maximized	VQM, TI, si_gain, d95	0.847	8.51	21.00	12.67	0.811
IV.	VQM adjusted	VQM	0.248	19.36	13.94	17.55	0.671
V.	Crossoptimized	TI, hv_gain, color1	0.276	19.65	12.71	17.34	0.675
VI.	Full-data model	SI, si_loss, si_gain, contati, color2, d05, d50, dDR	0.772	-		10.91	0.889

Model VQM: is a classic VQM, according to the recommendation. It serves us as the reference for other models, due to the fact that it had the highest precision from the general metrics. The deviation per sample is more than two times higher for our original data than for the validation dataset. This indicates that our original dataset is very heterogeneous, which it really is (original sequences are comprised of two different databases).

Model I: is a linear combination of VQM coefficients, optimized for the original data. There are two aspects to demonstrate this model. First, VQM could be improved by training on particular data. The original coefficients of the VQM metric have been established for general 2D video sequences and their analog/digital distortions caused by transmission/broadcasting. There was improve of the VQM model by training it on a specific type of data (3D stereoscopic sequences in HD resolution). Secondly, there was a lack of data samples to do this VQM optimization properly. The model is over-parameterized and loses its generality, which is manifested by the increase of deviation on the validation dataset.

Model II: it has been estimated for our original data. The regressors have been chosen for minimum AIC of the model. More precisely, the AICc coefficient has been minimized, which is preferable in the case of small amounts of data samples (less than 40 samples per 1 coefficient of the model). The coefficients of a linear model for the original data are [7.72, 3.47, 2.46, 1129, -32.10]. They are ranked in the same order as the regressors in Table 30. This model is quite well-fitted to our data which is indicated by the adjusted coefficient of determination. However, the validation of this model shows an estimated error value of 26 %, which is more than in the case of the optimized VQM model. Validation with an independent dataset confirms the concerns that this model is not general enough.

Model III: it includes the regressors which have been chosen based on criteria of the maximum adjusted coefficient of determination. Validation of this model brings better results than the previous one. A scatter plot also indicates good homoscedasticity, even if one of the model coefficients is VQM. The coefficients of regressors for the original data are [-99.162, 0.844, 470.542, 35.254]. The result of validation of this model is not cogent (the sample standard deviation is 21 %). The author's opinion is that it may be due to the difference between the datasets.

Model IV: it is the VQM value, whose dynamic range is adjusted separately to both datasets. The same process of dynamic range adjustment is done for validation data of models I.-V. The dynamic range of the modeled data (MOS values) is the additive information of the model. So, the model's results should be compared with adjusted MOS values to get relevant information about the model's selective accuracy.

Model V: it has been established to improve the validation dataset results. On the other hand, is not desirable to lose the benefit of two independent datasets which provide information about the generality of the model. The cross-optimization process has been done. From all the models (all the regressor combinations) this has been set to the original data, the algorithm chose the one which has the lowest sample standard deviation for validation data. The algorithm of cross optimization chose the regressors, which have the most similar influence on the modeled MOS value over both datasets. The resulting model is not optimal for any of the datasets, neither for the conjugate dataset, but we do not lose the possibility of validation.

Model VI: it is the optimal model for the conjugate dataset (which includes both original and validation data). The model has been set to a minimum of sample standard deviation and 8 regressors. Although 8 regressors is a reasonable large value for 96 sample of the conjugate dataset, we can deduce nothing about the generality of this model.

6.2.7 Conclusion of objective model for stereoscopic video

Regarding the above-mentioned facts and obtained objective and subjective scores, answers to the questions from begin of chapter 6.2 are as follows:

1) After a thorough comparison of all objective and subjective scores, it can be concluded that the VQM objective metric best reflects the user's QoE for 3D video content. However, even this metric does not reach a very high correlation with our subjective test. For more details see the results in Table 28 which shows a statistical comparison of objective metrics and subjective tests. From the point of view of outlier rate, it can be concluded that our assumption, that the variance of results is larger at the beginning and at the end of the testing session, due to disorientation and fatigue, was wrong. It was proved that in the subjective test, after a significantly better beginning part, the rest of the evaluation has uniform outlier rate. Dependence of the number of outliers on the sequence was not significant.

2) For better modeling of our results, six new models of objective metrics were created. These models have been validated and verified on other 3D video sequences and compared to the general VQM model. In general, we can state that the models that had the smallest error for our sequences were less accurate on other databases. On the other hand, models that were less accurate had a wider usable scope on other databases. Table 30 lists the most important data of our models, such as model descriptions, regressors, and their standard deviations. *Model III* has the highest correlation with MOS for our dataset. This model, compared to a general VQM metric that had 32 % deviation, had only 8 % standard deviation. When we consider our source data and validation data, then *Model VI* had the smallest standard deviation. On the other hand, the most regressors are included in this model and the model is trained on the original input video sequences as well as on validation sequences. The model, which is the most balanced in all areas, is *Model V*. The standard deviation for the original sequence is one third lower than for the general VQM. The deviation for the validation data is also slightly lower than for the classic VQM. Another benefit is that only three regressors enter the model calculation. *Model V*, for these reasons, can be determined as the most appropriate model due to its great versatility and sufficient accuracy.

7 CONCLUSION

The dissertation focuses on visual quality and multimedia services in broadband networks. Additional research was carried out to extend the state of the art. It dealt with the quality of services available in the Czech Republic. These were DVB-S, DVB-S2 and DVB-T services. Real transmission capacity of the multiplex was measured as well as the bitrate of individual TV programs. It can be observed how the bit rate and video quality changes over time, during the day. The research was published in paper [18]. Based on market research, it can be seen that the average time of watching television per day increases every year. The dissertation was focused on three main goals.

The first goal was to evaluate the coding efficiency of modern encoding standards, including the comparison of objective and subjective metrics, as well as the impact of sequences with high frame rate. The test results show that there is not such a significant difference between HEVC and VP9 encoders. In most cases, the results are similar on lower bitrates, but HEVC is better at higher bitrates. The encoding speed of the HEVC codec is an advantage. The time needed to encode 10 s of Full HD video using HEVC was about 70 seconds. The time needed for VP9 was hundredfold longer. This may change in the future because these codecs are especially improving by optimizing encoding speed. As an interesting fact, it may seem that in some cases, the MOS scores for Full HD and UHD video sequences sometimes show higher QoE for Full HD content. Such a phenomenon is probably caused by different features of Full HD/UHD resolutions. The human vision can more tolerate an image with fuzzy edges than an image with distortions, due to lower bitrates. Similar findings can be seen in the HFR videos. Based on the SROCC analysis, for standard objective metrics, it was determined that the VQM objective metric best reflects the user's QoE for Full HD and UHD videos. The correlation of the objective metrics has comparable results as in [92]. In the case of advanced objective metrics, according to PCC, results indicate that VSNR, VQM, NQM and ST-MAD metrics have a good correlation with the subjective scores. The ST-MAD metric has the highest average correlation across all videos. The PSNR values estimated by the HEVC encoder and calculated by a professional tool revealed slight differences (lower than 2 dB). The difference is mostly constant for different bitrates and sequences. Dependence of PSNR and SSIM values on the video frames show that values for each frame of the sequence are different due to different spatial and temporal information. Such behavior also depends on the used compression tool. A dominant part of the objective video metrics calculates the average value based on the image quality in each frame. On the other hand, the observer provides a worse evaluation for a video in which the quality has a steep drop. Videos with the same average quality with no drop in quality obtain a better subjective score. This is the biggest drawback of all objective metrics. The end of the chapter focused on videos with a high frame rate. In the case of HFR videos, it can be concluded that if it is possible to have high bitrate, then viewers clearly prefer videos with higher frame rates. Observers almost always did not see the difference in the QoE if a frame rate of 24 and 30 fps was set. On the other hand, with the option of 30 and 60 fps, evaluation is extremely dependent on the type of sequence. Viewers clearly preferred a frame rate of 60 fps in dynamic videos. In contrast to this fact, the preferred frame rate in static sequences was 30 fps. Preference of dynamic videos in the HFR version can be caused by the properties of this sequence. The viewer appreciates greater fluency over the detailed texture of the image. Due to this reason, it would be perfect to have the possibility to switch the frame rate for each video. These issues were published in papers [27], [87], [93] and [96].

The second goal was to find out how to get faster video encoding as well as what software and hardware aspects have the most impact on it. Speeding-up encoding can be done by using predefined quality profiles which change the balance between the speed of encoding and the video quality. It is important, from the results, to notice the following. A “Placebo” quality profile did not provide quality improvement and due to the extremely long encoding time, it should be used only by people who do not care about time. Bearing in consideration all the results, the “Slow” profile is the best one in my own opinion. In the case of a higher quality profile, the quality increase is not high enough to compensate for the expenditure of time. If fast encoding speed is a priority, I recommend choosing the “Superfast” profile. The “Ultrafast” profile is a special case. This profile has extremely poor image quality but needs the shortest time to encode. Subsequent results indicate that new version of the HEVC encoder has an influence on increasing both video quality and encoding speed. It was proven that a PC with an advanced architecture could increase the speed of encoding by five times in comparison with an old PC. By comparing software encoders to hardware encoders, it can be observed that the difference in quality is negligible, but the difference in encoding speed is huge. The difference between the video quality of the reference implementation, Turing and x265 encoders, with the highest quality preset, is insignificant. The hardware NVENC encoder is about six thousand times faster than the reference software HM with no dramatic difference in quality. The high speed of encoding, sufficient quality, and low power consumption can be achieved by the NVENC H.265 encoder. This encoder can be claimed as a “good encoding tool”. The main principles of these methods were published in the papers [99], [100] and [101].

The third goal was to examine visual quality in stereoscopic television systems. At the beginning, a custom video database was created. In this database, different parameters of each sequence were calculated, such as image depth, image activity, average image brightness, etc. The sequences were divided into groups and subjective assays were performed. The coefficient "K" was then determined which gives us an estimate of how the sequence will appeal to the user. From the dependency of the evaluation of the video sequences on the parameter K, it is obvious that the probability of good attractiveness of sequences rises if the parameter K reaches a certain size. The visual quality degradation by encoding is not considered in this section. The next part focuses on a study of different compression algorithms for stereoscopic videos. The well-established H.264 AVC and upcoming H.265 HEVC standards, as well as their multiview extensions H.264 MVC and H.265 – Multiview HEVC, are considered. Objective video quality metrics were used to analyse the compression efficiency of the considered codecs, extended with results from subjective tests. Overall, the codecs belonging to the same standard (AVC, MVC and HEVC, MV-HEVC) exhibit very similar performance with differences in PSNR in the order of 1 to 3 dB. The results of objective metrics show that the performance of a standard codec and the mutual comparison is highly content dependent. The results of the subjective test show that the scattering of the evaluation in the subjective test is large. For this reason, it was necessary to evaluate the participants who acted as outliers. After evaluating the coding efficiency, it was necessary to determine which 2D objective metric has the greatest correlation with the subjective 3D test. For these purposes, SROCC and PCC were applied to the results. Although the VQM metric has the highest correlation, it is still not ideal for evaluating 3D videos. We thus propose our own model, which better models our subjective test results. Six new models of objective metrics were created. These models have been validated and verified on other 3D video sequences and compared to the general PSNR, SSIM and VQM model. The best of the models had about half the error-rate than the classic VQM metric. The database creation was published in paper [105]. An objective test of stereoscopic encoders was published in [71], [112]. A subjective test, correlation coefficient, and proposed models were published in paper [110].

At this point, all the goals of the thesis can be considered as accomplished. In the future, I would like to focus on creating a program that produces a Hyperlapse video adapted to GPS parameters. Action cameras are experiencing a great boom. However, there is no program for automated speed-up of video playback. A video from an action camera, such as a bicycle ride, would be loaded into the program. The program would adjust Hyperlase speed based on GPS speed as well as temporal and spatial video activity. When the user slowly rides up a hill, the playback speed would be higher. On the other hand, if they go downhill, the playback speed would be lower. This would achieve an almost constant speed of Hyperlapse video. The output video would be in HFR format. In addition, the rider's speed, proper acceleration, based on an accelerometer, and altitude would be plotted in the video. Video encoding should be accelerated by using a graphic card.

REFERENCES

[1] U. H. Reimers, "DVB-The Family of International Standards for Digital Video Broadcasting," in Proceedings of the IEEE, vol. 94, no. 1, pp. 173-182, Jan. 2006. DOI: 10.1109/JPROC.2005.861004

[2] B. Veselinovska, M. Gusev and T. Janevski, "State of the art in IPTV," in Proceedings of the 37th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija (HR), 2014, pp. 479-484. DOI: 10.1109/MIPRO.2014.6859615

[3] T. Lohmar and U. Horn, "Hybrid Broadcast-Unicast distribution of Mobile TV over 3G Networks," in Proceedings. 31st IEEE Conference on Local Computer Networks, Tampa (FL), 2006, pp. 850-851. DOI: 10.1109/LCN.2006.322189

[4] V. K. Adhikari et al., "Measurement Study of Netflix, Hulu, and a Tale of Three CDNs," in IEEE/ACM Transactions on Networking, vol. 23, no. 6, pp. 1984-1997, Dec. 2015. DOI: 10.1109/TNET.2014.2354262

[5] x265 free library/codec, [cit. 2017-03-04]. Online: http://x265.org/hevc-h265/

[6] The WebM Project- VP9 Video Codec, [cit. 2017-03-04]. Online: http://www.webmproject.org/vp9/

[7] F. Liu and Z. Li, "A Measurement and Modeling Study of P2P IPTV Applications," in Proceedings of Computational Intelligence and Security, Suzhou, 2008, pp. 114-119. DOI: 10.1109/CIS.2008.112

[8] L. Polak and T. Kratochvil, "DVB-T and DVB-T2 performance in fixed terrestrial TV channels," in Proceedings of 35th International Conference on Telecommunications and Signal Processing (TSP), Prague (CZ), 2012, pp. 725-729. DOI: 10.1109/TSP.2012.6256392

[9] L. Polak, O. Kaller and T. Kratochvil, "SISO/MISO performances in DVB-T2 and fixed TV channels," in Proceedings of 38th International Conference on Telecommunications and Signal Processing (TSP), Prague (CZ), 2015, pp. 768-771. DOI: 10.1109/TSP.2015.7296369

[10] C. Naimannaran and E. Badarch, "Performance analysis of implementation Dvb-t2 network in Ulaanbaatar," in Proceedings of 8th International Forum on Strategic Technology (IFOST), Ulaanbaatar, 2013, pp. 182-185. DOI: 10.1109/IFOST.2013.6616884

[11] K. Malarić, I. Suć and I. Bačić, "Measurement of DVB-S and DVB-S2 parameters," in Proceedings of 23rd International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split (HR), 2015, pp. 160-164. DOI: 10.1109/SOFTCOM.2015.7314105

[12] G. Gvozden, M. Gosta and S. Grgic, "DVB-S2 coding modes for broadcasting and unicasting," in Proceedings of 47th International Symposium ELMAR, Zadar (HR), 2005, pp. 193-196. DOI: 10.1109/ELMAR.2005.193674

[13] J. Martin, Yunhui Fu, N. Wourms and T. Shaw, "Characterizing Netflix bandwidth consumption," in Proceedings of IEEE 10th Consumer Communications and Networking Conference (CCNC), Las Vegas (NV), 2013, pp. 230-235. DOI: 10.1109/CCNC.2013.6488451

[14] O. Bataa, O. Chuluun, T. Orosoo, E. Lamjav, Y. i. Kim and K. Gonchigsumlaa, "A functional design of BM-SC to support mobile IPTV in LTE network," in Proceedings of 7th International Forum on Strategic Technology (IFOST), Tomsk, 2012, pp. 1-5. DOI: 10.1109/IFOST.2012.6357671

[15] České Radiokomunikace, Start vysílání DVB-T2, Available from: http://www.digizone.cz/clanky/prezentace-ceskych-radiokomunikaci-ke-vcerejsimu-spusteni-prechodove-dvb-t2-site/

[16] K. Berger, Y. Koudota, M. Barkowsky and P. Le Callet, "Subjective quality assessment comparing UHD and HD resolution in HEVC transmission chains," in Proceedings of 2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX), Pylos-Nestoras (GR), 2015, pp. 1-6. DOI: 10.1109/QoMEX.2015.7148114

[17] J. Li, Y. Koudota, M. Barkowsky, H. Primon and P. Le Callet, "Comparing upscaling algorithms from HD to Ultra HD by evaluating preference of experience," in Proceedings of Sixth International Workshop on Quality of Multimedia Experience (QoMEX), Singapore, 2014, pp. 208-213. DOI:10.1109/QoMEX.2014.6982320

[18] J. Kufa. "Quality of Services in Digital Video Broadcasting in Czech Republic". In Proceedings of the 23nd Conference STUDENT EEICT 2017. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních, 2017. s. 390-394. ISBN: 9788021454965

[19] Rohde&Schwarz. Digital video quality analyser DVQ. Operating manual. Munich, 1999/05-20

[20] ITU-T Rec. P.910. "Subjective video quality assessment methods for multimedia applications". ITU, Geneva, (Switzerland), 2008.

[21] T. Wiegand, G. J. Sullivan, G. Bjontegaard and A. Luthra, "Overview of the H.264/AVC video coding standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560-576, July 2003 DOI: 10.1109/TCSVT.2003.815165.

[22] D. Mukherjee et al., "The latest open-source video codec VP9 - An overview and preliminary results," in Proceedings of Picture Coding Symposium (PCS), 2013, San Jose (CA), pp. 390-393. DOI: 10.1109/PCS.2013.6737765

[23] M. Rerabek and T. Ebrahimi "Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments" in Proc. SPIE 9217, Applications of Digital Image Processing XXXVII, 92170U, September 2014. DOI: 10.1117/12.2065561

[24] M. Rerabek, P. Hanhart, P. Korshunov and T. Ebrahimi, "Quality evaluation of HEVC and VP9 video compression in real-time applications," in Proceedings of Seventh International Workshop on Quality of Multimedia Experience (QoMEX), 2015, Pylos-Nestoras (Greek), pp. 1-6. DOI: 10.1109/qomex.2015.7148088

[25] D. Grois, et al, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," in Proceedings of Picture Coding Symposium (PCS), 2013, San Jose (CA), pp. 394-397. DOI: 10.1109/pcs.2013.6737766

[26] T. Nguyen and D. Marpe, "Objective Performance Evaluation of the HEVC Main Still Picture Profile," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 5, pp. 790-797, May 2015. DOI: 10.1109/tcsvt.2014.2358000

[27] J. Kufa and T. Kratochvil, "Comparison of H.265 and VP9 coding efficiency for full HDTV and ultra HDTV applications," in Proceedings of 25th International Conference Radioelektronika, Pardubice (CZ), pp. 168-171, 2015. ISBN: 978-1-4799-8117- 5. DOI: 10.1109/radioelek.2015.7128999

[28] Q. Huangyuan, et al., "Performance evaluation of H.265/MPEG-HEVC encoders for 4K video sequences," in Proceedings of Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific, Siem Reap, pp. 1-8. DOI: 10.1109/apsipa.2014.7041782

[29] D. Zegarra Rodriguez and G. Bressan, "Performance assessment of High Efficiency Video Coding - HEVC," in Proceedings of Global High Tech Congress on Electronics (GHTCE), 2013 IEEE, Shenzhen, pp. 110-111. DOI: 10.1109/ghtce.2013.6767252

[30] M. Uhrina, L. Sevcik, J. Frnda and M. Vaculik, "Impact of H.265 and VP9 compression standards on the video quality for 4K resolution," in Proceedings of 22nd Telecommunications Forum Telfor (TELFOR), 2014, Belgrade, pp. 905-908. DOI: 10.1109/telfor.2014.7034551

[31] N. Ramzan, Z. Pervez and A. Amira, "Quality of experience evaluation of H.265/MPEG-HEVC and VP9 comparison efficiency," in Proceedings of 26th International Conference on Microelectronics (ICM), 2014, Doha, Qatar, pp. 220-223. DOI: 10.1109/icm.2014.7071846

[32] R. Weerakkody, et al., "Verification testing of HEVC compression performance for UHD video," in Proceedings of Signal and Information Processing (GlobalSIP), 2014 IEEE Global Conference on, Atlanta, Georgie, pp. 1083-1087. DOI: 10.1109/globalsip.2014.7032288

[33] W.Song, Y.Xiao, D. Tjondronegoro, and Antonio Liotta. "QoE Modelling for VP9 and H.265 Videos on Mobile Devices". In Proceedings of the 23rd ACM international conference on Multimedia (MM '15). ACM, New York (USA), 2015, pp. 501-510. DOI: 10.1145/2733373.2806256

[34] O. Zach, M. Slanina. "A Matlab-based tool for video quality evaluation without reference" Radioengineering, vol.23, no.1, pp. 405-411, Apr. 2014.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Trans. on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004. DOI: 10.1109/tip.2003.819861

[36] O. Zach and M. Slanina, "A comparison of H.265/HEVC implementations," Proceedings ELMAR, Zadar, 2014, pp. 1-4. DOI: 10.1109/ELMAR.2014.6923337

[37] M. Grellert et al., "An adaptive workload management scheme for HEVC encoding," 2013 IEEE International Conference on Image Processing, Melbourne, VIC, 2013, pp. 1850-1854. DOI: 10.1109/ICIP.2013.6738381

[38] X. Jiang et al., "High efficiency video coding (HEVC) motion estimation parallel algorithms on GPU," 2014 IEEE International Conference on Consumer Electronics - Taiwan, Taipei, 2014, pp. 115-116. DOI: 10.1109/ICCE-TW.2014.6904012

[39] H.R. Sheikh and A.C. Bovik, "Image information and visual quality", IEEE Trans. on Image Processing, vol. 15, no. 2, pp. 430-444, February 2006. DOI: 10.1109/tip.2005.859378

[40] VQMT Software- Video Quality Measurement Tool (École polytechnique fédérale de Lausanne). [online].[cit. 2017-03-04]. Available from: http://mmspg.epfl.ch/vqmt

[41] W.S. Malpica, A.C. Bovik, "Range image quality assessment by Structural Similarity," in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Taipei (Taiwan), 2009, pp. 1149-1152. DOI: 10.1109/icassp.2009.4959792

[42] M. Pinson, S.Wolf, "A new standardized method for objectively measuring video quality," IEEE Transactions on Broadcasting, vol.50, no.3, pp. 312-322, Sept. 2004. DOI: 10.1109/tbc.2004.834028

[43] Software ITS. Video Quality Metric (VQM) Software. [online]. [cit. 2017-03-04]. Available from: http://www.its.bldrdoc.gov/ resources/video-quality-research/software.aspx

[44] ITU-R BT.500-13. "Methodology for the subjective assessment of the quality of television pictures. " ITU, January 2012.

[45] T. Alpert and J. P. Evain, "SSCQE (Single Stimulus Continuous Quality Evaluation): A New Subjective Assessment Method Introduced in ITU-R Recommendation 500-7: Presentation and Results," in Proceedings of 96th International Workshop, Los Angeles (USA), 1996, pp. 417-427. DOI:10.5594/M001273.

[46] ITU-R BT.2022. "General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays" ITU, August 2012.

[47] R. M. Nasiri, et al., "Perceptual quality assessment of high frame rate video," in Proceedings of IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), 2015, Xiame (PRC), pp. 1-6. DOI: 10.1109/MMSP.2015.7340831

[48] T. Zinner, O. Hohlfeld, O. Abboud and T. Hossfeld, "Impact of frame rate and resolution on objective QoE metrics," in Proceedings of Quality of Multimedia Experience (QoMEX) 2010, Trondheim, (Norway), pp. 29-34. DOI: 10.1109/QOMEX.2010.5518277

[49] J. Wu, C. Yuen, N. M. Cheung, J. Chen and C. W. Chen, "Modeling and Optimization of High Frame Rate Video Transmission Over Wireless Networks," in IEEE Transactions on Wireless Communications, vol. 15, no. 4, pp. 2713-2726, April 2016. DOI: 10.1109/TWC.2015.2509063

[50] G. Saygili, et al., "Evaluation of Asymmetric Stereo Video Coding and Rate Scaling for Adaptive 3D Video Streaming, " IEEE Trans. on Broadcast., vol. 57, no. 2, pp. 593-601, Jun. 2011. DOI: 10.1109/tbc.2011.2131450

[51] P. Merkle, et al., "Stereo Video Compression for Mobile 3D Services, " in Proceedings of 3rd Int. Conf. 3DTV, May 2009, pp. 1-4. DOI: 10.1109/3dtv.2009.5069649

[52] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard" in Proceedings of the IEEE, vol. 99, no. 4, pp. 626-642, Apr. 2011. DOI: 10.1109/jproc.2010.2098830

[53] Y. Chen, M. M. Hannuksela, T. Suzuki, and S. Hattori, "Overview of the MVC + D 3D Video Coding Standard" Journal of Vis. Commun. Image R., vol. 25, no. 4, pp. 679-688, May 2014. DOI: 10.1016/j.jvcir.2013.03.013

[54] A. Vosoughi, V. Testoni, P. C Cosman, and L. B. Milstein, “Multiview Coding and Error Correction Coding for 3D Video over Noisy Channels,” Signal Processing: Image Communication, vol. 30, pp. 107-120, Jan. 2015 DOI: 10.1016/j.image.2014.10.006

[55] K. Muller, P. Merkle, and T. Wiegand,”3D Video Representation Using Depth Maps,” in Proceedings of IEEE, vol. 99, no. 4, pp. 643-656, Apr. 2011. DOI: 10.1109/jproc.2010.2091090

[56] K. Muller, P. Merkle, G. Tech, and T. Wiegand, “3D Video Coding with Depth Modeling Modes and View Synthesis Optimization,” in Processing of International. Conf. APSIPA ASC, Dec. 2012, pp. 1-4.

[57] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” IEEE Trans. on Circuits and Systems for Video Technology. vol. 22, no. 12, p. 1649-1668, Dec. 2012. DOI: 10.1109/TCSVT.2012.2221191

[58] G. Balota, and et. al., ”Overview and Quality Analysis in 3D-HEVC Emergent Video Coding Standard,” in Proceedings of 5th Int. Conf. LASCAS, Feb. 2014, pp. 1-4. DOI: 10.1109/LASCAS.2014.6820260

[59] J. Stankowski, et al. ”3D-HEVC extension for circular camera arrangements" 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2015, Lisbon, 2015, pp. 1-4. DOI: 10.1109/3DTV.2015.7169371

[60] K. Wang., et al. ”Stereoscopic 3D video coding quality evaluation with 2D objective metrics”. In Proc. of SPIE 8648, Stereoscopic Displays and Applications XXIV, 86481L. DOI: 10.1117/12.2003664

[61] S.L.P., Yasakethu, et al., ”Quality Analysis for 3D Video Using 2D Video Quality Models” IEEE Transaction on Consumer Electronics, 2008, vol. 54, no. 4, pp. 1969–1976. ISSN: 0098-3063. DOI: 10.1109/TCE.2008.4711260

[62] P. Joveluro, et al., ”Perceptual Video Quality Metric for 3D video quality assessment”, 2010 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video, Tampere, 2010, pp. 1-4. DOI: 10.1109/3DTV.2010.5506331

[63] Y. Han, Z. Yuan, G. M. Muntean, ”Extended no reference objective Quality Metric for stereoscopic 3D vide”o, 2015 IEEE International Conference on Communication Workshop (ICCW), London, 2015, pp. 1729-1734. DOI: 10.1109/ICCW.2015.7247430

[64] J. Li, M. Barkowski, P. Le Callet, ”Validation of reliable 3DTV subjective assessment methodology - Establishing a Ground Truth Database”. VQEG eLetter, Video Quality Expert Group, 2014, Verification and Validation, 1 (2), pp 33-35. DOI: 10.1109/ivmspw.2013.6611917

[65] S. Tavakoli, et al., ”Subjective Quality Study of Adaptive Streaming of Monoscopic and Stereoscopic Video”, in IEEE Journal on Selected Areas in Communications, vol. 32, no. 4, pp. 684-692, April 2014. DOI: 10.1109/JSAC.2014.140402

[66] R. Palaniappan, N. Jayant. and P. Mane, ”Visual quality in stereoscopic 3DTV”, 2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR), Pacific Grove, CA, 2012, pp. 726-728. DOI: 10.1109/ACSSC.2012.6489107

[67] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New requirements of subjective video quality assessment methodologies for 3DTV," in Proc. Int. Workshop Video Process. Quality Metrics, Scottsdale, AZ, USA, Jan. 2010.

[68] A. Moorthy and A. Bovik, "A survey on 3D quality of experience and 3D quality assessment", Proc. SPIE 8651, Human Vision and Electronic Imaging XVIII, 86510M DOI: 10.1117/12.2008355

[69] P. Le Callet, S. Moller, and A. Perkis, Eds., "Qualinet white paper on definitions of quality of experience," Eur. Network on QoE in Multimedia Syst. Services (COST Action IC 1003), Ver. 1.1, Jun. 3, 2012.

[70] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in Proc. SPIE-IS&T Electron. Imaging 3D Image Process. Appl, Jan. 2010, vol. 7526, Art no 75260S. DOI: 10.1117/12.839438

[71] L. Polak et al., "Study of advanced compression tools for stereoscopic video by objective metrics," 26th International Conference Radioelektronika, Kosice, 2016, pp. 268-272. ISBN: 978-1-5090-1673- 0. DOI: 10.1109/radioelek.2016.7477357

[72] J. Gutiérrez, et al., "Subjective assessment of the impact of transmission errors in 3DTV compared to HDTV," in IEEE 3DTVConf., Antalya, Turkey, May 2011, pp. 1-4. DOI: 10.1109/3dtv.2011.5877209

[73] M. Vinnikov, R.S. Allison, and S. Fernandes. 2016. ”Impact of depth of field simulation on visual fatigue”. Int. J. Hum.-Comput. Stud. 91, C (July 2016), 37-51. DOI: 10.1016/j.ijhcs.2016.03.001

[74] J. Kuze and K. Ukai, "Subjective evaluation of visual fatigue caused by motion images," Displays, vol. 29, no. 2, pp. 159-166, Mar. 2008. DOI: 10.1016/j.displa.2007.09.007

[75] Terzić, K., & Hansard, M., “Methods for reducing visual discomfort in stereoscopic 3D: A review,” Signal Processing: Image Communication, 47, pp. 402-416. DOI: 10.1016/j.image.2016.08.002

[76] K. Wang, et al, "Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating," in Proc. SPIE-IS&T Stereosc. Displays and Appl, Feb.2011,vol.7863, Art.ID7863IT. DOI: 10.1117/12.872462

[77] M. Slanina, et. al, "Testing QoE in different 3D HDTV technologies," in Radioengineering, vol. 21, no. 1, pp. 445-454, Apr. 2012.

[78] J. Gutiérrez, F. Jaureguizar and N. Garcıa, "Subjective Comparison of Consumer Television Technologies for 3D Visualization," in Journal of Display Technology, vol. 11, no. 11, pp. 967-974, Nov. 2015. DOI: 10.1109/jdt.2015.2448758

[79] J. D. Yun, Y. Kwak, and S. Yang, "Evaluation of perceptual resolution and crosstalk in stereoscopic displays," Journal of Display Technology, vol. 9, no. 2, pp. 106-111, Feb. 2013. DOI: 10.1109/jdt.2012.2228252

[80] E. Cheng, et al, ”RMIT3DV: Pre-Announcement of a Creative Commons Uncompressed HD 3D Video Database,” 012 Fourth International Workshop on Quality of Multimedia Experience, Yarra Valley, VIC, 2012, pp. 212-217. DOI: 10.1109/qomex.2012.6263873

[81] M. Klíma, et al., "DEIMOS - an Open Source Image Database" Radioengineering 4(20):1016-1023. Dec 2011.

[82] M. Urvoy, et al., ”NAMA3DS1-cospad1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3d stereoscopic sequences,” In Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on (pp. 109-114). DOI: 10.1109/qomex.2012.6263847

[83] J. Berclaz, F. Fleuret, E. Turetken and P. Fua, "Multiple Object Tracking Using K-Shortest Paths Optimization," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 9, pp. 1806-1819, Sept. 2011. DOI: 10.1109/tpami.2011.21

[84] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015, pp. 3061-3070. DOI: 10.1109/cvpr.2015.7298925

[85] R. Fransens, C. Strecha, L. Van Gool "Parametric Stereo for Multi-Pose Face Recognition and 3D-Face Modeling," ICCV 2005 workshop "Analysis and Modeling of Faces and Gestures", 2005, pp.108-123. DOI: 10.1007/11564386_10

[86] A. Wang, et al., “QoE-Oriented Resource Allocation for DASH Based Video Transmission over LTE Systems”, In Proceedings of IEEE 85th Vehicular Technology Conference (VTC Spring), Sydney, NSW, 2017, pp. 1-5. DOI: 10.1109/VTCSpring.2017.8108535

[87] J. Kufa, et al., “HEVC/H.265 vs. VP9 for Full HD and UHD Video: Is There Any Difference in QoE?” In proceedings of the 58th International Conference Elmar, Zadar, 2016, pp. 51-55. DOI: 10.1109/elmar.2016.7731753

[88] J. Kufa, T. Kratochvil, “Comparison of H265 and VP9 Coding Efficiency for Full HDTV and Ultra HDTV Applications”. In Proceedings of 25th International Conference Radioelektronika 2015, Pardubice (Czech Republic), 2015. pp. 168-171. ISBN: 978-1-4799-8117- 5. DOI: 10.1109/radioelek.2015.7128999

[89] UltraVideo Group. Test sequence. [online]. [cit. 2017-03-04]. Available from: http://ultravideo.cs.tut.fi/#testsequences

[90] Xiph.org::Derf\'s Test Media Collection. Xiph.org Video Test Media. [online]. [cit. 2017-03-04]. Available from: https://media.xiph.org/ video/derf

[91] Zeranoe FFmpeg – Builds. Zeranoe FFmpeg. [online]. [cit. 2017-03-04]. Available from: http://ffmpeg.zeranoe.com/builds/

[92] K. Seshadrinathan and A. C. Bovik, "Motion Tuned Spatio-Temporal Quality Assessment of Natural Videos," IEEE Trans. on Image Processing, vol. 19, no. 2, pp. 335-350, Feb. 2010. DOI: 10.1109/tip.2009.2034992

[93] J. Kufa., "On the Performance of Objective Video Quality Metrics for UHD Videos". In Proceedings of the 24th Conference STUDENT EEICT 2018. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2018. s. 302-306. ISBN: 978-80-214-5614-3.

[94] M. Gaubatz, MeTrix MuX 1.1 [online]. Cornell University. cit. [2018-02-03]. Available from: http://foulard.ece.cornell.edu/gaubatz/metrix_mux/#modernization

[95] B. Bršel.: "Porovnání objektivních a subjektivních metrik kvality videa pro Ultra HDTV videosekvence". Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2016. 51s. Diplomová práce. Vedoucí práce Ing. Jan Kufa

[96] J. Kufa J. et al. "High Frame Rate Video: Is It Always Better?" In Sborník příspěvků studentské konference Blansko 2016. Brno: Vysoké učení technické v Brně, Fakulta elektrotechniky a komunikačních technologií, 2016. pp. 43-46. ISBN: 978-80-214-5389- 0.

[97] HDMI Forum's Technical Working Group. [cit. 2016-07-06]. Available from: http://www.hdmi.org/manufacturer/hdmi_2_0/

[98] C. Wiley, "DisplayPort Technical Overview" in Proceedings of IEEE International Conference on Consumer Electronics (ICCE), 2011, Las Vegas (USA). Available from: http://www.vesa.org/wp-content/uploads/2011/01/ICCE-Presentation-on-VESA-DisplayPort.pdf DOI: 10.1889/1.3621384

[99] J. Kufa, "Predefined Quality Profiles in different versions of HEVC encoder– Influence of PC architecture of processor". In Student EEICT Proceedings of the 22th Student Competition Conference. Brno: 2016. pp. 43-46. ISBN: 978-80-214-5389-0

[100] J. Kufa, T. Kratochvil, "Impact of Predefined Quality Profiles in H. 265 to Video Quality and Speed of Encoding" In Sborník příspěvků studentské konference Kohútka 2015. Brno, Czech Republic: VUT v Brně, pp. 32-34. ISBN: 978-80-214-5239- 8.

[101] J. Kufa J., T. Kratochvil, "Software and Hardware HEVC Encoding". In 2017 International Conference on Systems, Signals and Image Processing (IWSSIP). Poznaň: Poznań University of Technology, 2017. pp. 1-5. ISBN: 978-1-5090-6343-7. DOI: 10.1109/iwssip.2017.7965585

[102] S.G. Blasi et al., "The open-source turing codec: towards fast, flexible and parallel HEVC encoding", IET Conference Proceedings, DOI: 10.1049/ibc.2016.0038

[103] Nvidia Video codec. [cit. 02-02-2016]. Available from: https://developer.nvidia.com/nvidia-video-codec-sdk

[104] Intel Media Software. [cit. 02-02-2016]. Available from: https://software.intel.com/sites/default/files/mediasdk_release_notes.pdf

[105] L. Boleček., J. Kufa, V. "3D Video Database with Known Parametrs". In 2017 International Conference on Systems, Signals and Image Processing (IWSSIP). Poznan: 2017. s. 1-5. ISBN: 978-1-5090-6343-7. DOI: 10.1109/iwssip.2017.7965616

[106] J. P., López, et al., “Stereoscopic 3D video quality assessment based on depth maps and video motion,” EURASIP Journal on Image and Video Processing, 2013(1). DOI: 10.1186/1687-5281-2013-62

[107] J. P., López, et al., “Proposal for characterization of 3DTV video sequences describing parallax information,” In Consumer Electronics (ICCE), 2015 IEEE International Conference on, (pp. 353-355). DOI: 10.1109/icce.2015.7066442

[108] Stereophoto maker, Available: http://stereo.jpn.org/eng/stphmkr/

[109] ITU-R BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," Recommendation ITU-R BT.2021, Aug. 2012.

[110] J. Kufa, O. Kaller, O. Zach, L. Polak, M.Slanina, T. Kratochvil, “Objective Models for Performance Comparison of Compression Algorithms for 3DTV”. Radioengineering, 2018, In the second phase of review procedure.

[111] M. Domaski, et al. ”Poznan Multiview Video Test Sequences and Camera Parameters”, ISO/IEC JTC1/SC29/WG11 MPEG 2009/M17050, Xian, China, October 2009.

[112] D. Kukolj, J. Kufa, et al. (2019) “3D Content Acquisition and Coding.” In: Assunção P., Gotchev A. (eds) 3D Visual Content Creation, Coding and Delivery. Signals and Communication Technology. Springer, Cham, DOI: 10.1007/978-3-319-77842-6_3

[113] MATLAB Documentation, [Online] Cited 2017-03-04. Available from: http://www.mathworks.com/help/matlab/

[114] S. B. Lemeshko, “Distribution of Statistics of Chi-Square Goodness-of-Fit Tests for Small Samples”, 2006 8th International Conference on Actual Problems of Electronic Instrument Engineering, Novosibirsk, Russia, 2006, pp. 287-287. DOI:10.1109/APEIE.2006.4292559

[115] R. F Woolson. “Wilcoxon Signed-Rank Test”. Wiley Encyclopedia of Clinical Trials, Medical University of South Carolina, Charleston, SC, USA. DOI: 10.1002/9780471462422.eoct979

[116] H. Bay, et al. “SURF: Speeded Up Robust Features”, In Proceedings of Computer Vision - ECCV 2006:9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006. pp. 404–417, DOI: 10.1007/11744023_32

[117] J., Holcik, M. Komeda, a kol. “Matematicka biologie: elearningova ucebnice” [online]. 1. vydani. Brno: Masarykova univerzita, 2015. ISBN 978-80-210-8095-9.

[118] F., E. Harrell, “Regression Modeling Strategies with Applications to Linear Models”, Logistic and Ordinal Regression, and Survival Analysis, Nashville: Springer, 2001. 571 pages. ISBN 978-3-319-19424-0 DOI: 10.1007/978-3-319-19425-7

[119] H. Akaike, “A new look at the statistical model identification”, IEEE Transactions on Automatic Control, pp. 716-723, MR 0423716, (1974), DOI: 10.1109/TAC.1974.1100705.

[120] Y. Wang, “Survey of objective video quality measurements”. 2006. Available from: http://digitalcommons.wpi.edu/computersciencepubs/42, DOI:WPI-CS-TR-06-0212

[121] D. Mmilovanovic, et al. “MPEG video deployment in digital television: HEVC vs. AVC codec performance study”, 2013 11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS), Nis, 2013, pp.101-104. DOI: 10.1109/TELSKS.2013.670490

Curriculum Vitae

Personal

Name: Ing. Jan Kufa
Born: June 29th 1990 in Šumperk, Czech Republic
Contact: Jan.Kufa@vutbr.cz

Education

2014 – 18 Brno University of Technology / Department of Radio Electronics
Ph.D. study of Electronics and Communication
Dissertation thesis: Video and data services quality in the future broadband multimedia systems and networks.

2012 – 14 Brno University of Technology / Department of Radio Electronics
Master's study of Electronics and Communication
Diploma thesis: Shortwave power amplifier.

2009 – 12 Brno University of Technology / Department of Radio Electronics
Bachelor's study of Electronics and Communication
Bachelor's thesis: Accumulator charger for bicycles.

Courses

Training School on 3D Audiovisual Content Processing and Communications (3D-AVCom 2015) organized by 3D-ConTourNet, ICT COST Action IC1105, Portugal.

Participant in research projects

- CORTIFF 1/2015-6/2017
  LF14033 - Coexistence Of Radiofrequency Transmission In the Future.

- QOCIES 10/2015 – 05/2016
  LD15020 - Quality Optimized Coding and Transmission of Stereoscopic Sequences.

- FEWERCON 6/2018-
  Future Wireless Radio Communication Networks in Real Scenarios.

Languages

Czech, English, German