Influence of Bit Depth on Subjective Video Quality Assessment for High Resolutions

Juraj BIENIK, Miroslav UHRINA, Peter KORTIS

Department of Multimedia and Information-Communication Technologies, Faculty of Electrical Engineering, University of Zilina, Univerzitna 1, 01026, Zilina, Slovak Republic

juraj.bienik@fel.uniza.sk, miroslav.uhrina@fel.uniza.sk, peter.kortis@fel.uniza.sk

DOI: 10.15598/aeee.v15i4.2407

Abstract. *This paper deals with the influence of bit depth on the subjective video quality assessment. To achieve this goal, eight video sequences, each representing a different content prototype, were analysed. Subjective evaluation was performed using the ACR method. The analysed video sequences were encoded to 8 and 10-bit bit depth. Two most used compression standards H.264 and H.265 were evaluated with 1, 3, 5, 10 and 15* $Mb\cdot s^{-1}$ *bitrate in Full HD and UHD resolution. Finally, the perceived quality of both compression standards using the subjective tests with emphasis on bit-depth was compared. From the results, we can state that the practical application of 10-bit bit depth is not appropriate for Full HD resolution in the range of bitrate from 1 to 15* $Mb\cdot s^{-1}$*; for Ultra HD resolution, it is appropriate only for videos encoded by H.265/HEVC compression standard.*

Keywords

ACR, bit depth, H.264, H.265, subjective quality assessment.

1. Introduction

The video quality evaluation has become a hot and often discussed topic in recent years, not only in academic sphere but also in the commercial area of the multimedia industry. The research and development teams working with the video quality are changing their focus from quality assessment based only on the Quality of Service (QoS) to the Quality of Experience (QoE). While QoS is strictly oriented on networking aspects such as the transmission network parameters and the parameters of multimedia stream, the QoE takes into the account the final recipient of multimedia information, which is the user. This is the reason why the perceived quality is still in the forefront of video quality research and why there is still increasing demand for precise video quality estimation quantification.

2. State of the Art

In multiple publications, the video quality of sequences with HDR is assessed. The evaluation based on subjective and objective methods is used in the papers [1] and [2], but assessment for H.264 compression standard is missing. The paper [1] presents the results based on four test sequences in Ultra HD resolution, while in the paper [2], six sequences in Full HD resolution with different content are used. Content oriented research is presented in the paper [3], where nine test sequences are analysed, but the evaluation is done only for the H.265/HEVC compression standard. In the paper [4], the efficiency of H.264 and H.265 was tested, but used test sequences were in low resolution and only four sequences were analysed. Other authors present the results only for H.265/HEVC codec in the paper [5], while only three sequences in Full HD resolution with 12, 10 and 8-bit bit depth are objectively evaluated. These mentioned lacks were motivation to create this paper.

3. Video Compression and Video Compression Standards

The development of video compression standards is closely related to the development of the hardware on which the coding and decoding is realized, namely on the performance of these devices. Currently, the most advanced and most used standard for video compression is the Advanced Video Coding (AVC) [6], also known as MPEG-4 part 10. This video codec, approved in 2003 by the ITU standardization organization, has been successfully established at the multimedia market. It is used by many multimedia applications, from mobile phone video call services to digital TV and on-demand video services. Its coding efficiency has been sufficient for many years and therefore its development has been completed only recently. The multimedia market introduction of Ultra HD resolution and HDR video services has shown its imperfections and weak adaptability to this type of video content. Therefore, the need for development of a new compression standard has risen.

The imperfections of H.264 have been tackled by two standardization organizations - the Video Coding Experts Group (VCEG) and the ISO / IEC Moving Picture Experts Group (MPEG), whose collaboration is known as Joint Collaborative Team on Video Coding (JCT-VC). The result of this collaboration is the new compression standard that was approved in January 2013 and its development is still in progress. The compression standard H.265 [7] is a direct follower of H.264 and is called the High Efficiency Video Coding, also known as H.265/HEVC. The kernel and the structure of this compression standard stayed unchanged, but several changes have been done to increase the video encoding efficiency. Because many modern multimedia players and smart TVs have integrated H.265 codec, we can claim the new standard is gradually replacing its predecessor H.264.

4. Video Quality Assessment

The quality assessment methods can be split into two basic groups - the objective and the subjective one.

Objective video quality assessment methods are based on the measurement of the physical properties and parameters of the video signal. Objective video quality evaluation is not time-consuming and is relatively fast. The algorithms for the objective quality testing can be implemented in image quality measurement equipment or in set-top boxes that can send the quality reports of received video signal to TV service provider. The objective methods are also known as metrics. Depending on the principle of objective methods, we can split them into pixel based metrics (e.g., MSE and PSNR) and metrics based on the Human Vision System (HVS) model, e.g. SSIM and VQM.

Subjective video quality methods, unlike the objective ones, are based on assessment of the video quality performed by respondents - people who subjectively classify the perceived image quality. This type of quality assessment is more credible (compared to objective metrics) and cannot be replaced by objective measurement.

From the aspect of count of stimulus (reference and impaired video) the subjective methods can be divided to:

- single-stimulus methods (e.g. ACR, SSCQE),
- double-stimulus methods (e.g. DSIS, DSCQS).

The subjective methods can be also differentiated depending on the stage when the quality measurement is performed:

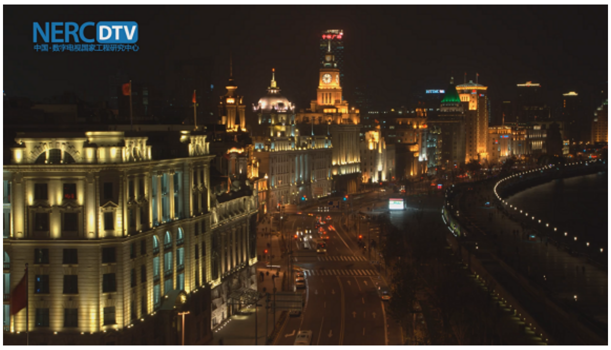
- methods in which the quality is evaluated after the presentation of both sequences from the sequence pair, respectively after the presentation of the test sequence only (e.g., DSIS, ACR),
- methods in which quality assessment takes place during the test sequence - continuous quality assessment (e.g. SSCQE, SDSCE).

The procedure and conditions of subjective testing are described in the recommendations ITU-T P.910 [8] and ITU-R BT.500-13 [9]. These recommendations define the requirements for source signals (reference sequences), assessors of quality, initial instructions to be presented before the start of testing, the steps of the test session, the way the results are documented, how the results should be analyzed, and the factors that may have negative effect on the subjective test. The assessment group must consist of minimum of 15 evaluators per test session.

Another very popular method is Crowdsourcing. Crowdsourcing is practically a specific type of subjective method. The quality of the video is rated by the evaluators (people), but the rating is realized in observer's comfort zone, e.g. at home on PC, laptop, tablet, or smartphone. The greatest advantage of this method is a vast number of evaluators (Crowd); the disadvantage is that each user evaluates the quality of different hardware too-different hardware performance.

Absolute Category Rating-ACR (also known as the Single Stimulus Method) is a subjective method based on evaluation of quality by respondents, which can quantify the quality of impaired sequences only (no reference video). Quality is defined by five-grade MOS scale (Mean Opinion Score), while only one discrete value is available for the sequence and no repetition of the sequence is allowed. The biggest benefit of this method is the speed of realization-not so much time-consuming like e.g. DSIS or DSCQS.

Five-grade scale of ACR: 1 = Bad, 2 = Poor, 3 = Fair, 4 = Good, 5 = Excellent (Quality).



(a) Bund Nightscape.

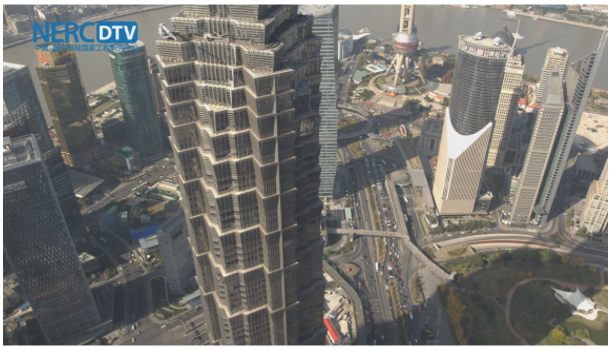
(b) Campfire Party.

(c) Construction Field.

(d) Fountains.

(e) Marathon.

(f) Runners.



(g) Tall Buildings.

(h) Wood.

Fig. 1: Test video sequences.

Tab. 1: Parameters of source video sequences.

Resolution	Chroma subsampling	Bit depth	Aspect Ratio	Framerate (fps)	Length (seconds)
3840×2160 (UHD)	4:4:4	10 bit per channel	16:9	30	10

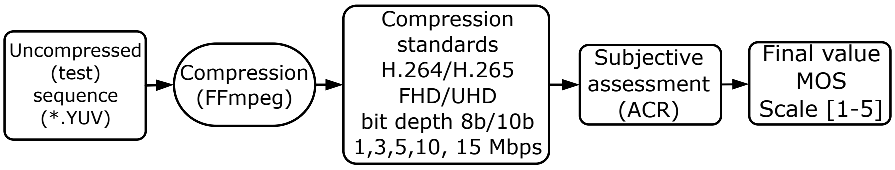


Fig. 2: Scheme of measurement process.

5. Measurement

5.1. Test Sequences

The quality evaluation process was done for eight test video sequences (Fig. 1(a), Fig. 1(b), Fig. 1(c), Fig. 1(d), Fig. 1(e), Fig. 1(f), Fig. 1(g) and Fig. 1(h)) which were downloaded from [10] in uncompressed format. The parameters of source video sequences are shown in the Tab. 1. Content information of used sequences (spatial and temporal information) are described by SI and TI diagram (Fig. 3) [8]. The calculations of SI and TI values were realized by Mitsu tool [11].

1) Characteristics of Used Test Sequences

- Bund Nightscape - time lapsed night city scene captured by static camera. The static parts are represented by buildings and the horizon, walking people and driving cars are the only dynamic objects in the scene (Fig. 1(a)).
- Campfire Party - night scene of the fire in the front of the image and group of people in the background. The flaming bonfire is changing quickly (the fast change of temporal and luminance information). Group of people in the background is moving slowly. At the end of the sequence, the camera zooms on the group of people (Fig. 1(b)).
- Construction Field - slow-motion scene of the building site with the static background. The only dynamic objects are excavator and walking workers. The scene is captured by static camera (Fig. 1(c)).
- Fountains - view on the city fountain. The spurting water in the foreground (a lot of edges in the image). The background is static and consists of trees and buildings. The camera is static, scene with minimum of motion (Fig. 1(d)).

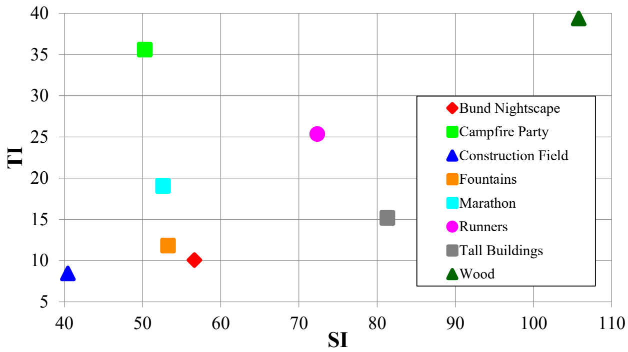


Fig. 3: Information about content of used sequences - SI and TI diagram.

- Marathon - marathon competition captured from the static point of view. Runners represent moving objects; the background is static street (Fig. 1(e)).
- Runners - relatively dynamic scene of running competition, but unlike to "marathon" scene, there are fewer runners. The camera is static and the runners are running closer to the camera. The camera is angled to the side (higher spatial information) (Fig. 1(f)).
- Tall buildings - birds eye view of modern city. The static objects are the skyscrapers, river and city infrastructure. The slow-motion objects are cars. The camera is panning slowly (Fig. 1(g)).
- Wood - shot of the trees in the forest. The camera is moving from the left to the right side and the speed of moving is slightly increasing. Relatively high value of the temporal and spatial information (Fig. 1(h)).

5.2. Coding Process

The assessed sequences were encoded in Full HD and Ultra HD resolutions using the FFmpeg tool [12] into the H.264 and H.265 compression standards with bitrates 1, 3, 5, 10 and 15 Mb·s^{-1}. The GoP was set to

Tab. 2: Structure of assessment group.

	Assessed codec	Resolution	Count of women	Count of men	Average age (years)
Group 1	H.264	UHD	13	17	approx. 22
Group 2	H.265	UHD	9	21	approx. 22
Group 3	H.264 and H.265	FHD	5	25	approx. 24

the half of the framerate, i.e. M = 3, N = 15. The main test criteria were: the perception of different bit depth: 8-bit and 10-bit bit depth per channel, and the exploration of their effect on subjective quality evaluation (Fig. 2). The preprocessing of test sequences is shown in the Fig. 4. The command line settings of this tool for both compression standards are shown in the Tab. 3. In this step, 160 sequences of each resolution used for the assessment were created.

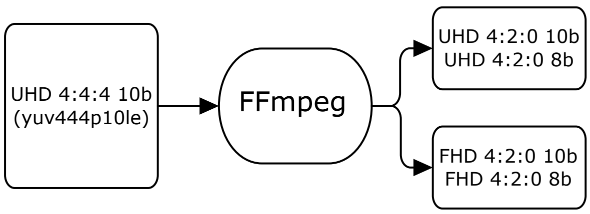


Fig. 4: Preprocessing of test sequences.

Tab. 3: Command line settings of the FFmpeg tool.

Options settings	FFMPEG command line settings
Input options	-i Input,Test Sequence.yuv, -video_resolution 3840×2160, -pix_fmt yuv444p10le, -framerate 30
Codec option	-vcodec libx264 (libx265)
GoP options	-keyint=15, -minkeyint=15, -bframes=3, -b-adapt=1
Bitrate options	-bitrate=bitrate in Mb·s^{-1}, -vbv-maxrate=max bitrate in Mb·s^{-1}, -vbv-bufsize= max bufsize in Mb·s^{-1}
Output options	Output,Test Sequence.mp4

5.3. Subjective Assessment

For the assessment, the ACR method was chosen. Evaluation groups consisted of 30 assessors, their structure is shown in the Tab. 2 Video quality in Ultra HD was assessed by Group 1 and 2. By Group 3, sequences in Full HD resolution were evaluated.

6. Experimental Results

The Fig. 5(a), Fig. 5(b), Fig. 5(c) and Fig. 5(d) show the results for subjective evaluation for Full HD resolution of each scenes in MOS scale. Figure 5(a) and Fig. 5(b) were created for H.264 (8-bit and 10-bit bit depth) and Fig. 5(b) and Fig. 5(d) for H.265 compression standard.

From the Fig. 5(a), Fig. 5(b), Fig. 5(c) and Fig. 5(d), we can state that the best results indicate slow motion scenes (low TI value) like "Construction field" and "Bund nightscape". Surprising results from the evaluation are values of the scene Wood (highest values of SI and TI), which trendline is close to middle of quality curves. This fact is possible to substantiate that human brain is not perceiving camera moving the same as moving of object on the static background.

Also, we can state that with higher bitrates the differences in MOS between scenes are smaller (lower scattering of values).

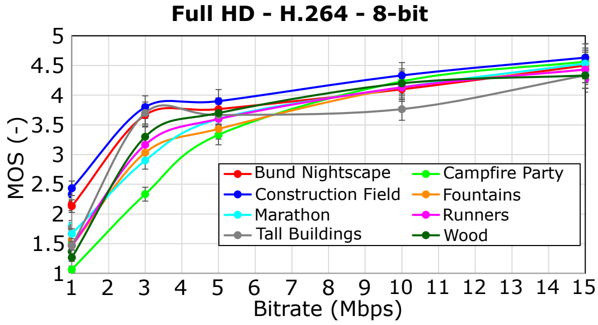
The Fig. 6(a), Fig. 6(b), Fig. 6(c) and Fig. 6(d) show the results from tests of perceived quality for Ultra HD resolution, while Fig. 6(a) and Fig. 6(b) show results for H.264 (8-bit and 10-bit bit depth) and Fig. 6(c) and Fig. 6(d) for H.265 compression standard.

From the Fig. 6(a), Fig. 6(b), Fig. 6(c) and Fig. 6(d), we can generally state that the best evaluation scores gained the slow-motion scenes again, MOS curves of sequences have a bigger scattering compared to the results for Full HD resolution. Trendlines of the sequence "Tall Buildings" indicate quite interesting results-the quality by H.264 codec with 10-bit bit depth raises, vice versa, by H.265 codec the quality decreases.

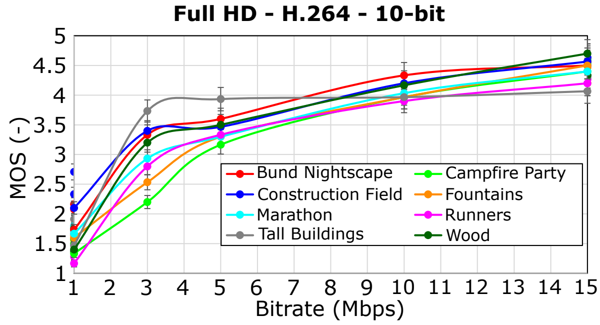
The Fig. 7(a) and Fig. 7(b) and the Tab. 4 show complex comparison of both used codecs and bit depths. The Fig. 7(a) shows average MOS values for Full HD resolution and the Fig. 7(b) average MOS values for Ultra HD resolution.

From the Fig. 7(a) and Fig. 7(b) and from the Tab. 2, we can state following:

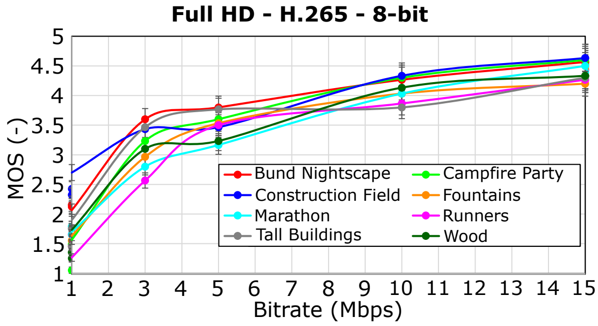
- The quality of all analyzed sequences rises logarithmically.
- The differences in MOS are bigger with growing resolution.



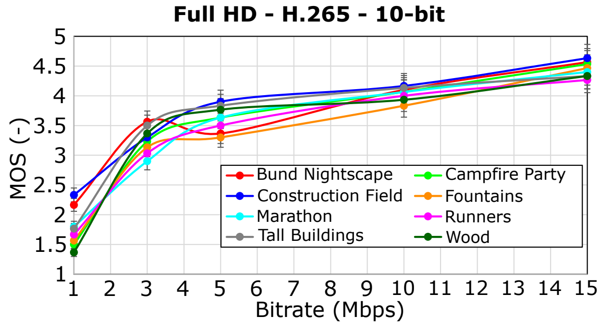
(a) MOS values for H.264, 8-bit bit depth.



(b) MOS values for H.264, 10-bit bit depth.

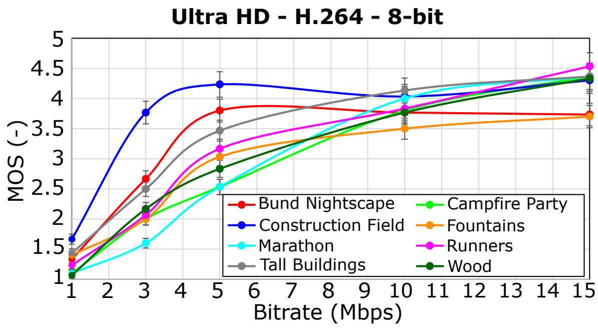


(c) MOS values for H.265, 8-bit bit depth.

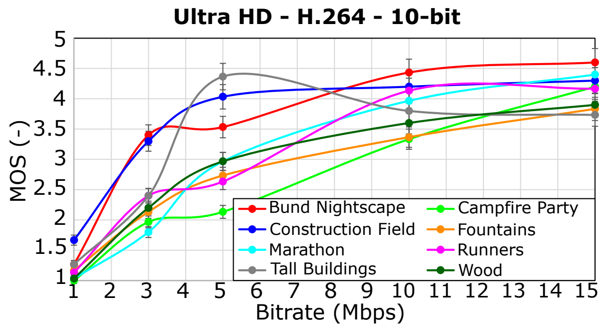


(d) MOS values for H.265, 10-bit bit depth.

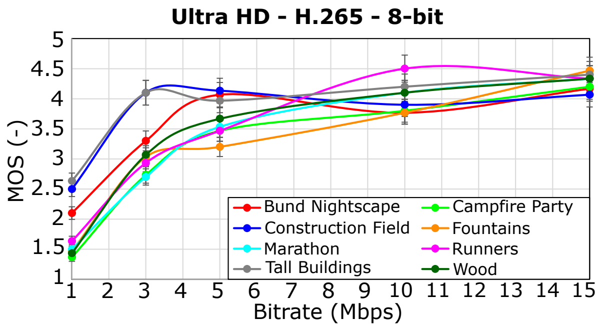
Fig. 5: MOS values for Full HD.



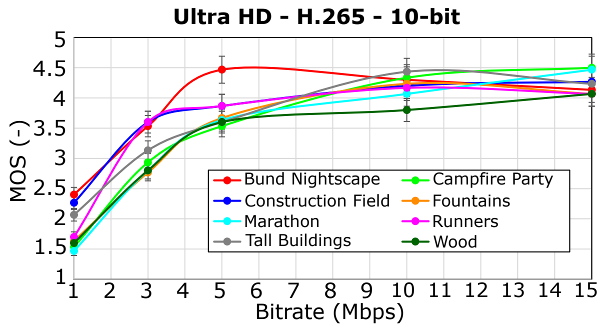
(a) MOS values for H.264, 8-bit bit depth.



(b) MOS values for H.264, 10-bit bit depth.



(c) MOS values for H.265, 8-bit bit depth.

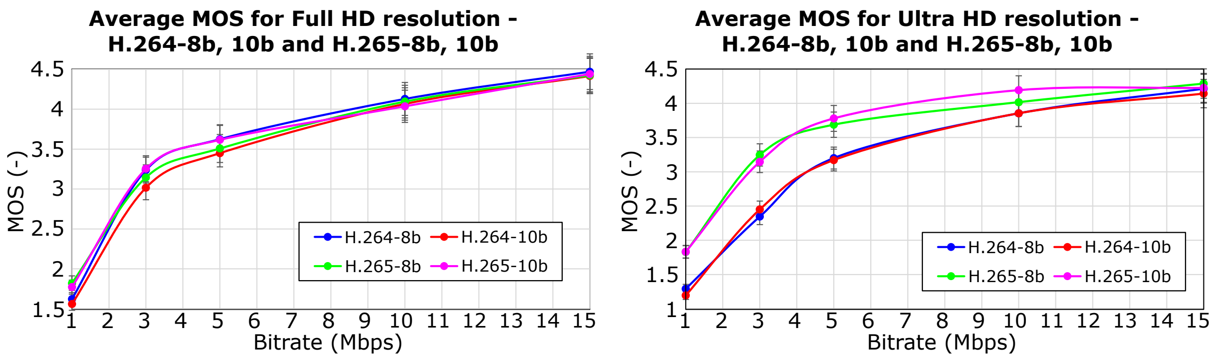


(d) MOS values for H.265, 10-bit bit depth.

Fig. 6: MOS values for Ultra HD.

Tab. 4: Complex efficiency comparison for average MOS values.

Full HD resolution								
Bitrate ($Mb{\cdot}s^{-1}$)	$H.264_{10b}$-$H.264_{8b}$		$H.265_{10b}$-$H.264_{8b}$		$H.265_{8b}$-$H.264_{8b}$		$H.265_{10b}$-$H.264_{10b}$	
	Diff.	**%**	**Diff.**	**%**	**Diff.**	**%**	**Diff.**	**%**
1	0.13	10.58	−0.05	−2.75	0.20	12.36	−0.03	−2.43
3	−0.10	−3.03	0.11	3.45	−0.09	−2.83	0.17	5.19
5	−0.20	−5.41	0.11	3.09	−0.12	−3.22	0.27	7.60
10	−0.03	−0.81	−0.06	−1.43	−0.03	−0.81	−0.23	−5.59
15	0.37	8.47	0.02	0.37	−0.05	−1.02	−0.37	−7.81
Ultra HD resolution								
Bitrate ($Mb{\cdot}s^{-1}$)	$H.264_{10b}$-$H.264_{8b}$		$H.265_{10b}$-$H.264_{8b}$		$H.265_{8b}$-$H.264_{8b}$		$H.265_{10b}$-$H.264_{10b}$	
	Diff.	**%**	**Diff.**	**%**	**Diff.**	**%**	**Diff.**	**%**
1	−0.10	−7.42	0.00	0.23	0.54	41.61	0.64	53.31
3	0.10	4.44	−0.10	−3.08	0.90	38.37	0.70	28.40
5	−0.03	−0.91	0.09	2.49	0.49	15.23	0.61	19.19
10	0.00	0.00	0.18	4.36	0.16	4.22	0.34	8.76
15	−0.07	−1.68	−0.06	−1.46	0.07	1.78	0.08	2.01



(a) Average MOS value for Full HD resolution.

(b) Average MOS value for Ultra HD resolution.

Fig. 7: Average MOS value for Full HD and Ultra HD resolution.

If we chose quality thresholds of MOS equal to 3 and 4 (3 = the lowest limit of acceptable perceived quality and 4 = the minimal value from which is quality evaluated as a good) considering the Fig. 7(a) and the Tab. 2, for Full HD resolution we can also state:

- The sequences encoded with 8-bit bit depth for H.264 in FHD resolution indicate better results than the sequences with 10-bit bit depth.
- For 10-bit bit depth videos encoded by H.265 codec, results are better only in the range from 2 to 8 $Mb{\cdot}s^{-1}$, but these differences are negligible.
- The results from comparison of both compression standard with 8-bit bit depth indicate, that H.264 outperformed H.265 (MOS equal to 3 is responsible to 2.7 $Mb{\cdot}s^{-1}$ for H.264 and 2.8 $Mb{\cdot}s^{-1}$ for H.265; MOS equal to 4 is responsible to 8.5 $Mb{\cdot}s^{-1}$ for H.264 and 9.0 $Mb{\cdot}s^{-1}$ for H.265).
- Vice versa, comparison of sequences with 10-bit bit depth indicate that the H.265 codec outperformed the H.264 codec (MOS equal to 3 respond to 2.6 $Mb{\cdot}s^{-1}$ for H.265 and 3.0 $Mb{\cdot}s^{-1}$ for H.264; MOS equal to 4 is the same for both codecs).

Generally, we can state that the commercial use of sequences with 10-bit bit depth for Full HD resolution is irrelevant because differences between 8-bit and 10-bit depth are minimal, regardless of used compression standard-negligible quality enhancement. Considering Fig. 7(b) and Tab. 2, for Ultra HD resolution, we can state:

- The coding efficiency of H.265 codec is higher than by H.264-H.265 outperformed H.264 in both bit depths.
- The results for the sequences with 8-bit bit depth indicate MOS equal to 3 is respond to 2.5 $Mb{\cdot}s^{-1}$ for H.265 and 4.2 $Mb{\cdot}s^{-1}$ for H.264; MOS equal to 4 is responsible to 10 $Mb{\cdot}s^{-1}$ for H.265 and 12 $Mb{\cdot}s^{-1}$ for H.264.
- The results for the sequences with 10-bit bit depth indicate MOS equal to 3 is responsible to 2.8 $Mb{\cdot}s^{-1}$ for H.265 and 4.2 $Mb{\cdot}s^{-1}$ for H.264; MOS equal to 4 is responsible to 7 $Mb{\cdot}s^{-1}$ for H.265 and 12 $Mb{\cdot}s^{-1}$ for H.264.

Commonly, we can state that using of 10-bit bit depth is appropriate for video in Ultra HD resolution, encoded by the H.265 compression standard, streamed with bitrates higher than 4 $Mb\cdot s^{-1}$; for the H.264 codec 10-bit bit depth, the relevance of usage is ambiguous.

7. Conclusion

This paper dealt with the influence of bit depth on the subjective video quality assessment. For quantifying perceived quality, the ACR method was used. Analyses of videos in Full HD and Ultra HD resolutions, H.264 and H.265 compression standards with 8-bit and 10-bit bit depth were done. Finally, subjective quality of both mentioned codecs was compared with emphasis on perceiving of bit depth. Generally, we can state that 10-bit bit depth can be implemented, but significant quality increment indicates only videos encoded by H.265 compression standard in Ultra HD resolution.

Acknowledgment

This work was supported by Centre of excellence for system and services of intelligent transport II, IMTS 2622120050 supported by the Research & Development Operational Programme funded by the ERDF.

References

[1] HERROU, G., W. HAMIDOUCHE and X. DUCLOUX. HDR video quality evaluation of HEVC and VP9 codecs. In: *Picture Coding Symposium*. Nuremberg: IEEE, 2016, pp. 1–5. ISBN 978-150905966-9. DOI: 10.1109/PCS.2016.7906332.

[2] MUKHERJEE, R., K. DEBATTISTA, T. BASHFORD-ROGERS, P. VANGORP, R. MANTIUK, M. BESSA, B. WATERFIELD and A. CHALMERS. Objective and subjective evaluation of High Dynamic Range video compression. *Signal Processing: Image Communication*. 2016, vol. 47, iss. 1, pp. 426–437. ISSN 0923-5965. DOI: 10.1016/j.image.2016.08.001.

[3] ZHANG, Y., M. NACCARI, D. AGRAFIOTIS, M. MRAK and D. R. BULL. High Dynamic Range Video Compression Exploiting Luminance Masking. *IEEE Transactions on Circuits and Systems for Video Technology*. 2016, vol. 26, iss. 5, pp. 950–964. ISSN 1051-8215. DOI: 10.1109/TCSVT.2015.2426552.

[4] THOMA, H., M. DE-FRUTOS-LOPEZ and J. AUER. Chroma subsampling for HDR video with improved subjective quality. In: *Picture Coding Symposium*. San Jose: IEEE, 2013, pp. 345–348. ISBN 978-147990294-1. DOI: 10.1109/PCS.2013.6737754.

[5] THOMA, H., M. DE-FRUTOS-LOPEZ and J. AUER. Chroma subsampling for HDR video with improved subjective quality. In: *Picture Coding Symposium*. San Jose: IEEE, 2013, pp. 345–348. ISBN 978-147990294-1. DOI: 10.1109/PCS.2013.6737754.

[6] ITU-T H.264. *Advanced video coding for generic audiovisual services*. Geneva: ITU, 2004.

[7] ITU-T H.265. *High efficiency video coding*. Geneva: ITU, 2017.

[8] ITU-T P.910. *Subjective video quality assessment methods for multimedia applications*. Geneva: ITU, 2009.

[9] ITU-R BT.500-13. *Methodology for the subjective assessment of the quality of television pictures*. Geneva: ITU, 2012.

[10] Test Sequence. In: *Test Sequence* [online]. 2014. Available at: `http://medialab.sjtu.edu.cn/web4k/index.html`

[11] Video Quality Indicators. In: *Video Quality* [online]. 2017. Available at: `http://vq.kt.agh.edu.pl/metrics.html`.

[12] FFmpeg tool. In: *FFmpeg* [online]. 2017. Available at: `https://www.ffmpeg.org/`.

About Authors

Juraj BIENIK was born in 1987 in Zilina, Slovakia. He received his M.Sc. in Telecommunications at the Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, at the University of Zilina in 2012. Nowadays he is a Ph.D. student at the same department. His research interests include audio and video signal processing, functionality and optimisation of networks and video quality assessment.

Miroslav UHRINA was born in 1984 in Zilina, Slovakia. He received his M.Sc. and Ph.D. degrees in Telecommunications at the Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, at the University of Zilina in 2008 and 2012, respectively. Nowadays he is an assistant professor at the same department. His research interests include audio and video compression, video quality assessment, TV broadcasting and IP networks.

Peter KORTIS was born in 1977 in Zilina, Slovakia. He received his M.Sc. and Ph.D. degrees in Telecommunications at the Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, at the University of Zilina in 2000 and 2007, respectively. Nowadays he is an assistant professor at the same department. His research interests include IP networks, VoIP and genetic algorithms.