

IMPACT OF GoP ON THE VIDEO QUALITY OF VP9 COMPRESSION STANDARD FOR FULL HD RESOLUTION

Miroslav UHRINA, Juraj BIENIK, Martin VACULIK

Department of Telecommunications and Multimedia, Faculty of Electrical Engineering, University of Zilina, Univerzitna 8215/1, 010 07 Zilina, Slovakia

miroslav.uhrina@fel.uniza.sk, jura.j.bienik@fel.uniza.sk, martin.vaculik@fel.uniza.sk

DOI: 10.15598/aeec.v14i4.1866

Abstract. *In the last years, the interest on multimedia services has significantly increased. This leads to requirements for quality assessment, especially in video domain. Compression together with the transmission link imperfection are two main factors that influence the quality. This paper deals with the assessment of the Group of Pictures (GoP) impact on the video quality of VP9 compression standard. The evaluation was done using selected objective and subjective methods for two types of Full HD sequences depending on content. These results are part of a new model that is still being created and will be used for predicting the video quality in networks based on IP.*

Keywords

GoP, objective assessment, subjective assessment, video quality, VP9.

1. Introduction

Interest on new multimedia services has significantly raised in the last years. This increase goes hand in hand with demand for higher TV resolutions and bandwidth which leads to need to develop new compression standards. Nowadays, many new codecs have become available as VP9 or H.265/HEVC and other are being developed as DAALA or VP10. It is well known that compression together with transmission link imperfection are two major factors that influence the video quality. Because of that fact the video quality assessment still plays an important role of the research. This paper deals with the assessment of the Group of Pictures (GoP) impact on the video quality of VP9 compression standard using Full HD resolution. The rest of the paper is divided as follows. In the next part, the state

of art is written. The third part shortly describes the VP9 compression standard. In the fourth and fifth part the objective and subjective methods are described. The sixth part deals with the measurements and the seventh part with the results obtained from these measurements.

2. State of the Art

Recently, many studies and publications deal with exploring the video quality affected by the VP9 codec. Some of them explore [1], [2] and [3] the quality of multimedia services, others focus on objective testing [4], [5], [6], [7], [8] and [9] as well as on subjective tests [10] and [11], but not many deal with the comparison of the quality between sequences using GoP and without GoP. This paper focuses on the video quality evaluation of one of the newest compression standard - VP9 in terms of use GoP. The testing is done for two sequences depending on content for Full HD resolution.

3. VP9 Compression Standard

VP9 is one of the newest video compression standards. It has been developed by Google and has become available in June 2013. VP9 is a successor to VP8. The aim for VP9 includes reducing the bit rate by 50 % compared to VP8 while maintaining the same video quality. VP9 has many design improvements compared to VP8. It supports the use of superblocks of 64×64 pixels and a quadtree coding structure could be used with the superblocks. Some web browsers as Chromium, Chrome, Firefox, and Opera support playing VP9 video format in the HTML5 video tag. Its own successor, VP10, is being developed [12].

4. Objective Video Quality Assessment

Objective video quality assessment is a type of measurement where the evaluation using computational methods called "metrics" is done which produce values that score the video quality. They mostly measure the physical characteristics of a video signal. They are used very often because of its repeatability and simplicity of the calculation. Many objective metrics exist but the well-known and mostly used are Peak Signal-to-Noise Ratio (PSNR), Video Quality Metric (VQM) and Structural Similarity Index (SSIM). PSNR is the oldest but still very used metric. It is very fast and easy to compute [13]. The SSIM metric measures three parameters - the luminance similarity, the contrast similarity and the structural similarity and merges them into one value, which determines the quality. This final value is in the range from 0 to 1 where 0 stands for the worst and 1 for the best quality [14]. The VQM metric computes the visibility of artifacts expressed in the DCT domain. The final value of the VQM metric designates the amount of video distortion - for no impairment the value equals to zero and for increasing amount of impairment the output value rises, too [15] and [16]. All mentioned metrics can be included in so-called Full Reference (FR) metrics, which means that for ability to compute the video quality the reference sequence needs to be known.

5. Subjective Video Quality Assessment

The subjective assessment is a type of measurement where people are used to score the video quality. It is the most reliable and fundamental way how to determine the video quality called Quality of Experience. It involves visual psychological tests where human evaluators are subjected to a video stimulus and evaluate its quality based on their own subjective judgment. This type of assessment has one drawback - it is very time-consuming method and for proper assessment many people are needed. The well-known and mostly used subjective methods described in [17] and [18] are Double Stimulus Impairment Scale (DSIS) also known as Degradation Category Rating (DCR), Double Stimulus Continuous Quality Scale (DSCQS), Single Stimulus Continuous Quality Evaluation (SSCQE), Absolute Category Rating (ACR) also known as Single Stimulus (SS), Simultaneous Double Stimulus for Continuous Evaluation (SDSCE).

The methods can be divided by two aspects (Tab. 1):

Tab. 1: Subjective assessment methods.

	Assessment during the presentation	Assessment after the presentation
Full Reference method	SDSCE	DSIS, DSCQS
No Reference method	SSCQE	ACR

- whether the observers assess the quality during or after presentation,
- whether the reference sequence is hidden (No Reference methods) or not (Full Reference methods).

According to [17], minimum 15 observers should be used in an assessment to achieve valid results. Of course, the number of the observers needed for the tests depends upon the sensitivity and the reliability of the test procedure adopted and upon the anticipated size of the effect sought. The whole presentation structure of each test, which should not exceed 30 minutes, is shown in the Fig. 1. Before the test session, assessors should be introduced to many factors, as for instance the method of assessment, the types of impairments, the grading scale, the sequence, the timing (the reference, the test sequence time duration, the time duration for voting) and so on.

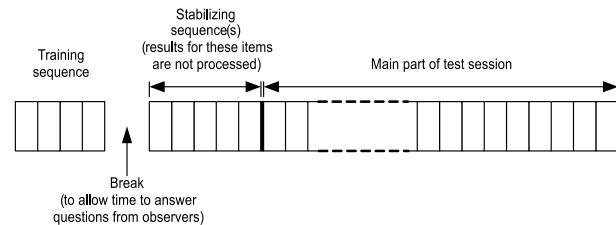


Fig. 1: The presentation structure of the subjective test session.

After the test session, the calculation of the mean score (MOS) using this formula is done:

$$\bar{u}_{jkr} = \frac{1}{N} \sum_{i=1}^N u_{ijk_r}, \quad (1)$$

where u_{ijk_r} is the score of observer i for test condition j , sequence k , repetition r and N stands for number of observers. Also, the 95 % confidence interval, which is derived from the standard deviation, and size of each sample is calculated. It is given by [17] and [18]:

$$\delta_{jkr} = 1.96 \frac{S_{jkr}}{\sqrt{N}}, \quad (2)$$

where:

$$\delta_{jkr} = \sqrt{\sum_{i=1}^N \frac{(u_{ijk_r} - \bar{u}_{jkr})^2}{(N-1)}}. \quad (3)$$

Due to the assessment of short sequences (10 sec.) after the presentation in our testing, DSIS, DSCQS and ACR methods were used.

6. Measurements

In our measurements two types of assessment were done:

- objective assessment using PSNR, SSIM and VQM metrics,
- subjective assessment using DSIS, DSCQS and ACR methods.

6.1. Source Signal

In our testing two types of test Source Sequences (SRCs) depending on content were used:

- one with dynamic scene called "Basketball" (Fig. 2(a)),
- one with slow motion called "Cactus" (Fig. 2(b)).



(a)



(b)

Fig. 2: The test sequences.

Both sequences were downloaded from [19] in the uncompressed format (*.yuv) and used as the reference ones. The basic parameters of these sequences are shown in the Tab. 2.

Tab. 2: Basic parameters of the test sequences.

Resolution	Aspect ratio	Framerate (fps)	Length (frames)	Length (s)
1920×1080	16:09	50	500	10

Since the compression difficulty is directly related to the spatial and temporal information of a sequence, regarding to [18], the Spatial (SI) and the Temporal Information (TI) of both sequences using the Mitsu tool [20] were calculated. The results are shown in the Tab. 3.

Tab. 3: Spatial (SI) and Temporal (TI) Indexes of test sequences.

	Basketball	Cactus
SI	71.08	74.86
TI	19.86	13.12

6.2. Coding

Both test sequences were encoded to the VP9 compression standard with two different GoP setting:

- without GoP setting ($N = 250$), which means the distance between two successive I frames was 250 frames (249 P frames between two successive I frames were used),
- the GoP was set to 12 ($N = 12$), which means the distance between two successive I frames was 12 frames (11 P frames between two successive I frames were used).

Since the VP9 compression standards does not use B frames, only P frames between two I frames were used.

The coding process was done using the FFmpeg tool [21]. The command line settings of this tool for the VP9 compression standard is shown in the Tab. 4. The bitrates were in the range from 1 to 10 Mbps with a step of 1 Mbps, which means 20 Hypothetical Reference Circuits (HRCs) were used - for each SRC ten HRCs restricted by maximum bitrate. It is important to mention that for the subjective assessment only five HRCs for each SRC were used 1, 3, 5, 7, 9 Mbps. If all HRCs were used, it would be too difficult for observers to recognize the video quality between two successive sequences. It was also necessary to take into account the maximum duration of test session, which should not last more than 30 minutes. The selected sequences were viewed by the observers in the random order.

Tab. 4: Command line settings of the FFmpeg tool for the VP9 compression standard.

Command line options	FFMPEG command line settings of VP9 Compression Standard
Input options	<i>i</i> Input Test Sequence.yuv video_resolution 1920x1080 pix_fmt yuv420p framerate 50
GoP options	g 12 / -
Bitrate options	<i>b</i> bitrate in Mbps <i>minrate</i> bitrate in Mbps <i>maxrate</i> bitrate in Mbps <i>bufsize</i> bufsize
Output options	Output Test Sequence.mkv

Afterwards, both sequences using the same FFmpeg tool back to the format *.yuv were decoded.

6.3. Evaluation

Finally, the video quality was evaluated.

- For the objective assessment the MSU Measuring Tool Pro version 3.0 was used [22]. PSNR, SSIM and VQM objective metrics for the measurements were used.
- For the subjective assessment the observers - people, who watched the sequences and assessed the video quality, were used. The DSIS, DSCQS and ACR methods were used.

In our experiments, 30 assessors (19 men and 11 women) in the range from 20 to 26 years were used. The average age was 22 years. Most of them were students of our department.

The whole process of the measurement and evaluation is shown in the Fig. 3.

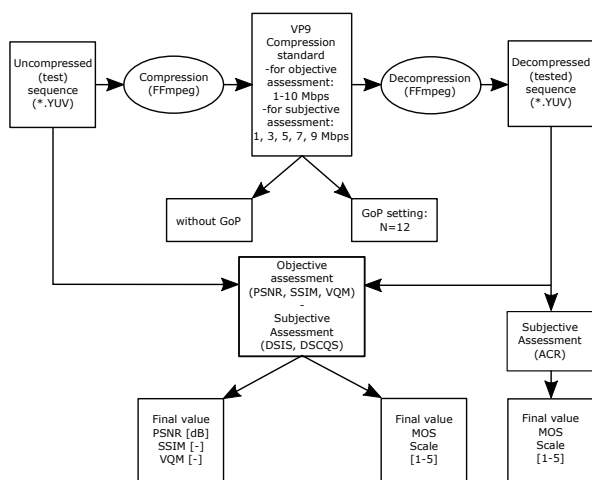
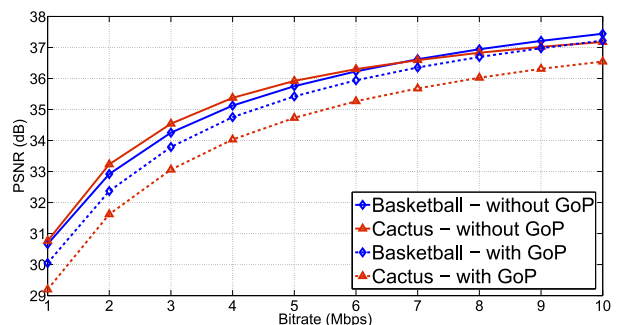


Fig. 3: The process of measuring and evaluating the impact of GoP of the VP9 compression standard on the video quality.

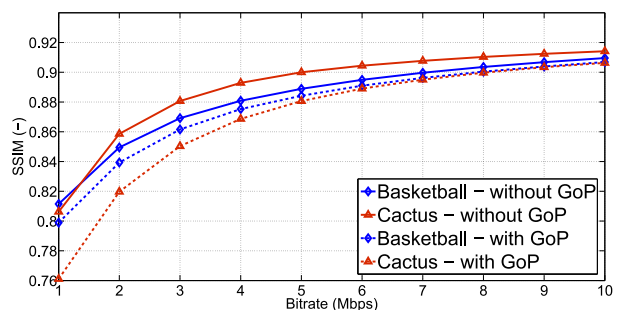
7. Experimental Results

Figure 4 shows the relationship between the video quality assessed by the objective metrics and the bitrate. The curves represent the test sequences with and without GoP. In this figure, three graphs are inset - depending on used objective metric Fig. 4(a), Fig. 4(b) and Fig. 4(c).

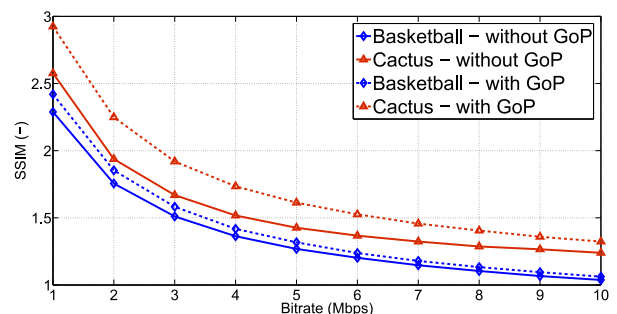
According to the graphs, the quality raises logarithmically with increasing bitrate. What is important for us is the difference of the quality between the sequences with and without GoP setting. As seen from the plots, the quality of the sequences without GoP setting reach better quality than the sequences with typically GoP setting. For better representation, the difference be-



(a) PSNR.



(b) SSIM.



(c) VQM.

Fig. 4: The relationship between the video quality measured by the objective metrics and the bitrate. The curves represent the test sequences with and without GoP.

tween the sequences with and without GoP setting was calculated. The values represent the quality difference expressed in percentage between the sequences with and without GoP setting, measured by the objective metrics, are shown in the Tab. 5. The same values are plotted in the Fig. 5. The table, as well as the figure, shows that the difference in quality between the two mentioned sequences is biggest in low bitrates and with increasing bitrate the difference decreases. Subsequently, the same measurements but using subjective methods were done.

Tab. 5: The quality difference (in percentage) between the sequences with and without GoP setting by the objective metrics.

Mbps	Basketball difference (%)			Cactus difference (%)		
	PSNR	SSIM	VQM	PSNR	SSIM	VQM
1	2.04	1.57	-5.43	5.37	5.93	-11.89
2	1.70	1.21	-5.28	5.09	4.75	-13.85
3	1.37	0.88	-4.54	4.48	3.57	-13.05
4	1.07	0.64	-3.85	3.95	2.79	-12.51
5	0.93	0.53	-3.72	3.43	2.20	-11.59
6	0.80	0.44	-2.86	2.94	1.73	-10.40
7	0.73	0.38	-2.74	2.56	1.42	-9.15
8	0.68	0.35	-2.57	2.24	1.18	-8.42
9	0.64	0.32	-2.64	1.95	1.00	-6.86
10	0.59	0.29	-2.32	1.73	0.86	-6.32

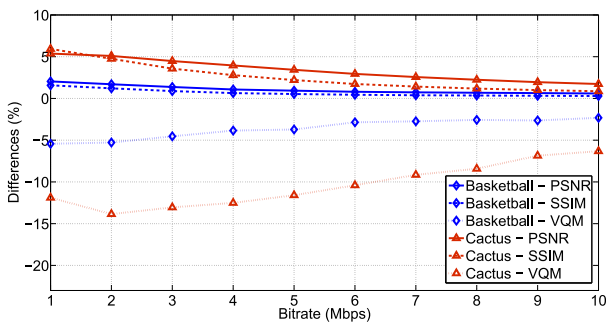
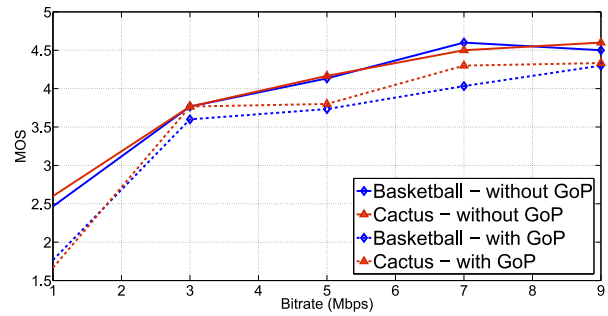


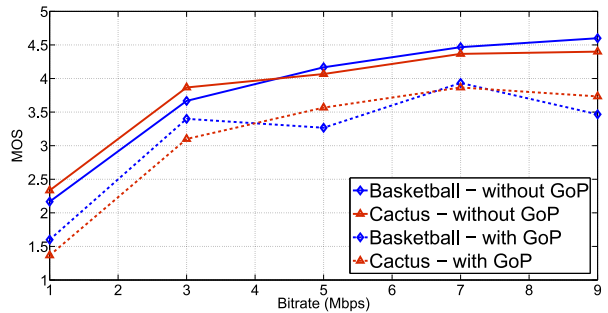
Fig. 5: The quality difference (in percentage) between the sequences with and without GoP setting measured by the objective metrics.

Figure 6 shows the relationship between the video quality assessed by the subjective methods and the bitrate. The curves represent the test sequences with and without GoP. In this figure also three graphs are inset - depending on used subjective methods Fig. 6(a), Fig. 6(b) and Fig. 6(c). In the Tab. 6, the values represent the quality difference expressed in percentage between the sequences with and without GoP setting measured by the subjective methods are shown. The same values are shown in the Fig. 7.

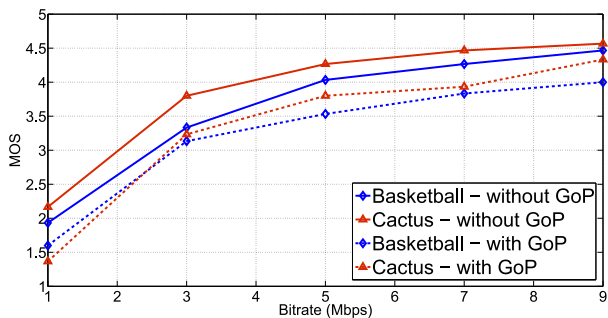
According to the graphs the same as in case of the objective assessment can be said - the quality increases logarithmically with increasing bitrate - in low bitrates the quality grows swifter than in high bitrates, which means the degradation influenced by compression is



(a) DSIS.



(b) DSCQS.



(c) ACR.

Fig. 6: The relationship between the video quality measured by the subjective methods and the bitrate. The curves represent the test sequences with and without GoP.

more visible in low bitrates than in the high ones. This fact also recognized the observers. Regarding the quality between the sequences with and without GoP setting, it can be said that the people saw the difference between these two types of sequences. The observers rated the quality of sequences without GoP setting with higher marks than the sequences with GoP setting. The table, as well as the figure, shows that the difference in quality between the two mentioned sequences is by all metrics biggest in low bitrate - by 1 Mbps. By the others bitrates by the DSIS and DSCQS methods the difference values are quite similar - they move between 4.63 % and 14.71 %. Only by the DSCQS method the difference values are quite different. It can be caused by the type of this method, where the observers do not know which one was the reference and

which one the test sequence, so they could rate test sequences with higher marks than the reference ones.

Tab. 6: The quality difference (in percentage) between the sequences with and without GoP setting by the subjective methods.

Mbps	Basketball difference (%)			Cactus difference (%)		
	DSIS	DSCQS	ACR	DSIS	DSCQS	ACR
1	39.62	35.42	20.83	56.00	70.73	58.54
3	4.63	7.84	6.38	0.00	24.73	17.53
5	10.71	27.55	14.15	9.65	14.02	12.28
7	14.05	13.56	11.30	4.65	12.93	13.56
9	4.65	32.69	11.67	6.15	17.86	5.38

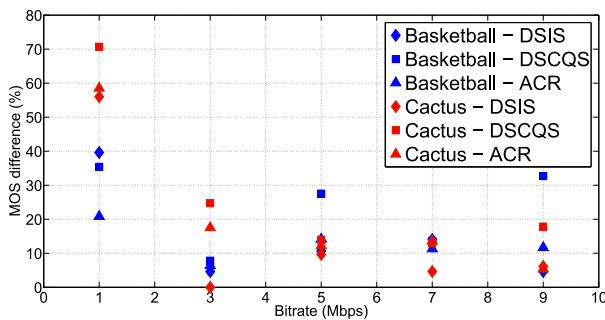


Fig. 7: The quality difference (in percentage) between the sequences with and without GoP setting measured by the subjective methods.

Afterwards the Pearson correlation coefficients of the differences between all objective and subjective methods for both test sequences were calculated. It was done using the formula:

$$r_{xy} = \frac{k_{xy}}{d_x d_y}, \quad (4)$$

where k_{xy} is the covariance and d_x and d_y are the standard deviations of the two variables. The correlation coefficients for both test sequences are reported below in the Tab. 7.

Tab. 7: The correlation coefficients of the differences between all objective and subjective methods for both test sequences.

	Basketball - correlation		
	PSNR	SSIM	VQM
DSIS	0.41730	0.48802	-0.52190
DSCQS	-0.56884	-0.50753	0.44658
ACR	-0.66659	-0.60317	0.54014
	Cactus - correlation		
	PSNR	SSIM	VQM
DSIS	0.38975	0.63191	-0.61787
DSCQS	0.65813	0.83973	-0.79080
ACR	0.85083	0.88905	-0.90623

According to the Tab. 7, very high correlation is between the ACR method and PSNR as well as the SSIM metrics by the Basketball sequence and between the ACR method and all objective metrics by the Cactus

sequence. It follows that the results obtained from the ACR subjective assessment mapped well the results obtained by objective evaluation and that the ACR subjective method should be used in the future research.

8. Conclusion

This paper dealt with the assessment of the Group of Pictures (GoP) impact on the video quality of the VP9 compression standard. The aim of this paper was to research the difference in the video quality between sequences with and without GoP setting and to find out the correlation of the differences between all used methods. The assessment was done using selected objective and subjective methods for two types of Full HD sequences depending on the content. The results showed that the sequences without GoP setting reach better quality than the sequences with typically GoP setting, especially in low bitrates. Afterwards, the correlation of the differences of all objective and subjective methods for both test sequences was calculated. According to the results, it can be said that very high correlation is between the ACR method and PSNR as well as the SSIM metrics by the Basketball sequence and between the ACR method and all objective metrics by the Cactus sequence. All results are part of a new model that is still being created and will be used for predicting the video quality in networks based on IP.

References

- [1] VOZNAK, M., J. SLACHTA and J. ROZHON. Performance analysis of virtualized real-time applications. *International Journal of Mathematical Models and Methods in Applied Sciences*. 2012, vol. 2, iss. 6, pp. 305–313. ISSN 1998-0140.
- [2] VOZNAK, M. E-model modification for case of cascade codecs arrangement. *International Journal of Mathematical Models and Methods in Applied Sciences*. 2011, vol. 5, iss. 8, pp. 1439–1447. ISSN 1998-0140.
- [3] FRNDA, J., M. VOZNAK, J. ROZHON and M. MEHIC. Prediction Model of QoS for Triple Play Services. In: *21st Telecommunications Forum TELFOR*. Belgrade: IEEE, 2013, pp. 733–736. ISBN 978-1-4799-1419-7. DOI: 10.1109/TELFOR.2013.6716334.
- [4] GROIS, D., D. MARPE, A. MULAYOFF and O. HADAR. Performance Comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC Encoders. In: *Picture Coding Symposium (PCS)*. San Jose: IEEE,

- 2013, pp. 394–397. ISBN 978-1-4799-0292-7. DOI: 10.1109/PCS.2013.6737766.
- [5] RAO, K. Video coding standards: AVS China, H.264/MPEG-4 PART 10, HEVC, VP9, DIRAC and VC-1. In: *Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. Poznan: IEEE, 2013, pp. 1–11. ISBN 978-94-007-6741-6.
- [6] KIM, K. I., S. LEE, Y. PIAO and J. CHEN. Coding Efficiency Comparison of New Video Coding Standards: HEVC vs VP9 vs AVS2 Video. In: *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. Chengdu: IEEE, 2014, pp. 1–6. ISBN 978-1-4799-4716-4. DOI: 10.1109/ICMEW.2014.6890700.
- [7] RAMZAN, N., Z. PERVEZ and A. AMIRA. Quality of Experience Evaluation of H.265/MPEG-HEVC and VP9 Comparison Efficiency. In: *26th International Conference on Microelectronics (ICM)*. Doha: IEEE, 2014, pp. 220–223. ISBN 978-1-4799-8153-3. DOI: 10.1109/ICM.2014.7071846.
- [8] RERABEK, M., P. HANHART, P. KORSHUNOV and T. EBRAHIMI. Quality Evaluation of HEVC and VP9 Video Compression in Real-Time Applications. In: *7th International Workshop on Quality of Multimedia Experience (QoMEX)*. Pylos-Nestoras: IEEE, 2015, pp. 1–6. ISBN 978-1-4799-8958-4. DOI: 10.1109/QoMEX.2015.7148088.
- [9] KUFA, J. and T. KRATOCHVIL. Comparison of H.265 and VP9 Coding Efficiency for Full HDTV and Ultra HDTV Applications. In: *25th International Conference Radioelektronika*. Pardubice: IEEE, 2015, pp. 168–171. ISBN 978-1-4799-8117-5. DOI: 10.1109/RADIOELEK.2015.7128999.
- [10] RERABEK, M. and T. EBRAHIMI. Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments. In: *Applications of Digital Image Processing XXXVII*. San Diego: SPIE, 2014, pp. 1–13. ISBN 978-162841244-4. DOI: 10.1117/12.2065561.
- [11] UHRINA, M., J. FRNDA, L. SEVCIK and M. VACULIK. Impact of H.265 and VP9 Compression Standards on the Video Quality for 4K Resolution. In: *22nd Telecommunications Forum TELFOR*. Belgrade: IEEE, 2014, pp. 905–908. ISBN 978-1-4799-6190-0. DOI: 10.1109/TELFOR.2014.7034551.
- [12] VP9 Compression Standard and Vpxenc tool. *WebM: an open web media project* [online]. 2015. Available at: <http://www.webmproject.org/>.
- [13] WINKLER, S. *Digital Video Quality: Vision Models and Metrics*. Montreux: John Wiley and Sons Ltd., 2005. ISBN 0-470-02404-6.
- [14] WANG, Z., A. C. BOVIK, H. R. SHEIKK and E. P. SIMONCELLI. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*. 2004, vol. 13, iss. 4, pp. 600–612. ISSN 1057-7149. DOI: 10.1109/TIP.2003.819861.
- [15] WATSON, A. B. Toward a Perceptual Video Quality Metric. In: *Proceedings of SPIE 3299: Human Vision and Electronic Imaging III*. San Jose: SPIE, 1998, pp. 139–147. ISBN 978-0-8194-2031-2. DOI: 10.1117/12.320105.
- [16] LOKE, H. M., P. E. ONG, W. LIN, Z. LU and S. YAO. Comparison of video quality metrics on multimedia videos. In: *Image Processing IEEE 2006*. Georgia: IEEE, 2013, pp. 457–460. ISBN 1-4244-0480-0. DOI: 10.1109/ICIP.2006.312492.
- [17] RECOMMENDATION ITU-R BT.500-13. *Methodology for the subjective assessment of the quality of television pictures*. Geneva: ITU-T, 2013.
- [18] RECOMMENDATION ITU-T P.910. *Subjective video quality assessment methods for multimedia applications*. Geneva: ITU-T, 2008.
- [19] Test Sequence. *Ultra Video Group* [online]. 2014. Available at: <http://ultravideo.cs.tut.fi/#testsequences>.
- [20] ROMANIAK, P., L. JANOWSKI, M. LESZUK and Z. PAPIR. Perceptual quality assessment for H. 264/AVC compression. In: *IEEE Consumer Communications and Networking Conference (CCNC)*. Las Vegas: IEEE, 2012, pp. 597–602. ISBN 978-1-4577-2070-3 . DOI: 10.1109/CCNC.2012.6181021.
- [21] FFmpeg. *FFmpeg tool* [online]. 2015. Available at: <https://www.ffmpeg.org/>.
- [22] Everything about the data compression. *MSU Measurement Tool Pro version* [online]. Available at: http://compression.ru/video/quality_measure/vqmt_pro_en.html#start.

About Authors

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.

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.

Martin VACULIK was born in 1951. He received his M.Sc. and Ph.D. in Telecommunications at the University of Zilina, Slovakia in 1976 and 1987 respectively. In 2001 he was habilitated as associate professor of the Faculty of Electrical Engineering at the University of Zilina in the field of Telecommunications. Currently he works as a head of Department of Telecommunications and Multimedia. His interests cover switching and access networks, communication network architecture, audio and video applications.