

30 YEARS OF VIDEO CODING EVOLUTION - WHAT CAN WE LEARN FROM IT IN TERMS OF QoE?

Tomas MIZDOS¹ , Marcus BARKOWSKY² , Peter POCTA¹ , Miroslav UHRINA¹ 

¹Department of Multimedia and Information-Communication Technologies, Faculty of Electrical Engineering and Information Technology, University of Zilina, Univerzitna 1, 010 26 Zilina, Slovakia

²Department of Interactive Systems and Internet of Things, Faculty of Applied Computer Science, Deggendorf Institute of Technology, Dieter-Gorlitz-Platz 1, 94469 Deggendorf, Germany

tomas.mizdos@uniza.sk, marcus.barkowsky@th-deg.de, peter.pocta@uniza.sk, miroslav.uhrina@uniza.sk

DOI: 10.15598/aece.v20i3.3998

Article history: Received Nov 06, 2020; Revised Nov 04, 2021; Accepted Feb 10, 2022; Published Sep 30, 2022.
This is an open access article under the BY-CC license.

Abstract. From the beginnings of ITU-T H.261 to H.265 (HEVC), each new video coding standard has aimed at halving the bitrate at the same perceptual quality by redundancy and irrelevancy reduction. Each improvement has been explained by comparably small changes in the video coding toolset. This contribution aims at starting the Quality of Experience (QoE) analysis of the accumulated improvements over the last thirty years. Based on an overview of the changes in the coding tools, we analyze the changes in the quantized residual information. Visual comparison and statistical measures are performed and some interpretations are provided towards explaining how irrelevancy reduction may have led to such a huge reduction in bitrate. The interpretation of the results in terms of QoE paves the way towards an understanding of the coding tools in terms of visual quality. It may help in understanding how the irrelevancy reduction has been improved over the decades. Understanding how the differences of the residuals relate to known or yet unknown properties of the human visual system, may enable a closer collaboration between perception research and video compression research.

1. Introduction

The efficient reduction of bitrate in multimedia coding is one of the core applications of research on perception. The encoding of audio signals has been largely improved by audio perception research in both speech and music transmission. For video encoding the gain due to the knowledge transfer seems less significant. No major difference in the coding structure itself can be noticed from the ITU-T H.261 standard in the early nineties to the recent High Efficiency Video Coding standard. The encoding loop is still very similar containing a block by block processing with Discrete Cosine Transformation, quantization, entropy coding step as well as translational motion compensation between successive frames. However, there are many differences that have been accumulated over the years of continuous development. While it may be impossible to judge the perceptual impact of each of the thousands of contributions that have been made to the video encoding efficiency improvement and it may still be hard to interpret the difference between one standard and its successor, the coding performance enhancement over the last thirty years seems worth investigating from a perceptual viewpoint.

While a large part of the aforementioned coding gain may stem from a redundancy reduction without any link to the human visual system, we observe a bitrate reduction of a factor around ten which seems hard to attribute solely to the general progress in a redundancy reduction (i.e. switching to an arithmetic coding) or to an improved representation of the video signal. In this contribution, we aim at identifying and charac-

Keywords

HEVC, H.261, Irrelevancy reduction, Quality of Experience, video coding.

terizing the irrelevancy reduction part. Instead of taking an algorithmic perspective and assessing the perceptual impact of each of the coding steps, we take a signal oriented perspective by comparing the intermediate images in the H.261 and HEVC encoders such as the predicted image. This allows for a more holistic approach with the focus on the perceptual differences in terms of QoE. The results of the complex encoder operations (especially after several frames encoded with consecutive predictions) can thus be perceptually evaluated on well-chosen video content. In this contribution, we present the results of the analysis of the predicted image using both signal based statistics as well as visual inspection of the video frames. To the best of our knowledge, this is the first work which reverses the viewpoint of video coding improvement and human vision research by taking this holistic perspective.

The remainder of the paper is organized as follows. Section 2. provides a brief overview of the evolution of the video coding tools in the last decades. In Sec. 3., the analysis of signal based and QoE interpreted differences between the investigated video coding standards is described. Section 4. presents the results of the statistical and perceptual analysis. Section 5. provides the final conclusions and suggests future work.

2. Evolution of Video Coding

This section sets the context for our research. It firstly introduces the concept of QoE and its measurement, second, it compares both video coding approaches to be later investigated in the paper, i.e. H.261 and H.265. It then briefly provides an evidence on progress, which was made in this context over the decades.

2.1. Measuring Quality of Experience

The concept of QoE contains a variety of influential factors, notably Human, System, and Context Influence Factors [1]. Video coding artifacts that impact on the QoE are usually categorized as System Influence Factor. In most cases, these factors are evaluated in formal, standardized laboratory tests with human observers, thus optimizing the context for the most critical viewing conditions and averaging on human factors. Conducting such tests on a larger scale requires an open international collaboration, which is organized, for example, in the Video Quality Experts Group (VQEG) [2]. Descriptions and results, including video sequences and subjective results, are available for several campaigns that were conducted in the past and that show the behavior of the typical video coding al-

gorithms that were used although the focus of VQEG is on validating objective video quality measurement algorithms. Several measurement algorithms have been standardized by the International Telecommunications Union (ITU) after VQEG's validation procedure, for example J.343 [3] for algorithms that take both the bitstream as well as the decoded video and potentially the reference video as input for their quality prediction.

The number of research papers proposing methods for video quality assessment is very large and continuously growing. Good reviews may be found in [4] and [5]. The research focus in these algorithms is on the prediction of image and video quality for a particular application scope, which includes the content and the usage scenario and, in the case of algorithms that consider the bitstream information, on one specific video coding algorithm. In this paper, however, the focus is on the difference between video codecs and its impact on the perceived quality, which may be considered as an inverse view. When it comes to the quality assessment, databases containing annotated image and video samples are necessary. Some popular and extensively used ones can be found in [6], [7], [8], [9], [10], [11], [12], [13], [14], [15] and [16].

In terms of QoE, it is worth mentioning that the expectation of human observers on video quality in the era of H.261 (early 1990's) was different from today's expectations, thus the analysis should be extended to Human Influence Factors as well.

2.2. Coding Tools Comparison H.261 vs. H.265

H.261 [17] is an ITU-T video compression standard established in November 1988. It is the first member of the H.26x family of video coding standards in the domain of the ITU-T Video Coding Experts Group (VCEG, later renamed to Specialists Group on Coding for Visual Telephony) and was developed with a number of companies including Hitachi, Picture-Tel, NTT, BT and Toshiba. It was the first video coding standard that was useful in practical terms. It is worth noting that H.261 was originally designed for transmission over ISDN lines on which data rates are multiples of 64 kbps. The coding algorithm was designed to allow operations of video bit rates between 40 kbps and 2 Mbps. The standard supports two video frame sizes, i.e. Common Intermediate Format (CIF) (352×288 luma with 176×144 chroma) and QCIF Quarter Common Intermediate Format (QCIF) (176×144 with 88×72 chroma), using a 4:2:0 sampling scheme. The main elements of the compression are motion compensated prediction, discrete cosine transform (DCT), block transformation and quantization. In contrast to H.261, HEVC also known as

H.265 and MPEG-H Part 2, is the most recent joint video recommendation of the ITU-T video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG) standardization organizations working together in a partnership known as the Joint Collaborative Team on Video Coding (JCT-VC). H.265/HEVC has the same hybrid coding structure as previous standards starting from H.261 and followed by MPEG-2 Video and its successor - H.264 (MPEG-4 Part 10)/Advanced video coding (AVC) compression standard. Compared to H.261, H.265/HEVC contains many incremental improvements such as: more flexible partitioning from large to small partition sizes, greater flexibility in motion prediction blocks sizes and transform block sizes and prediction modes, including intra predicted blocks, more sophisticated interpolation and deblocking filters, an improved signalling of modes and motion vectors and features to support efficient parallel processing. The H.265/HEVC standard was designed to achieve multiple goals, i.e. coding efficiency, ease of transport system integration and data loss resilience, as well as implementability using parallel processing architectures. The video coding layer of H.265/HEVC employs the same hybrid approach (inter-/intrapicture prediction and 2-D transform coding) as deployed in all video compression standards since H.261. The various features involved in hybrid video coding using H.265/HEVC are highlighted as follows: Coding Tree Units and Coding Tree Block (CTB) structure, Coding Units (CUs) and Coding Blocks (CBs), Prediction Units and Prediction Blocks (PBs), TUs and transform blocks, Motion vector signaling, Motion compensation, Intrapicture prediction, Quantization control, various entropy coding modes, in-loop deblocking filtering, and Sample Adaptive Offset (SAO) [18], [19] and [20].

2.3. Evidence on Progress

With every higher video resolution, a new video coding standard is being developed, starting from H.261 up to the newest one - H.265, also called HEVC. Each standardization work started with the ambitious goal that the new video coding standard halves the bitrate at the same perceptual quality compared to its predecessor. This statement is confirmed by performance tests related to the video quality assessment. There are only few publications dealing with a performance of the oldest video codec, i.e. H.261. This may be due the fact that the H.261 codec was primarily designed for transmission over ISDN lines. In [21], the authors evaluated two video codecs, namely H.261 and H.263, using the PSNR metric. The results showed that the coding efficiency of the H.263 codec is very similar to the H.261 compression standard. In [22], a comparison of the H.261, H.263+ and H.264 codecs using the SDSCE (Simultaneous Double Stimulus for Continuous Evalu-

ation) subjective methodology as well as the PSNR and SSIM objective methods was done. The authors used two video sequences in CIF resolution (325×288 pixels) encoded to different bitrates ranging from 512 kbps up to 4096 kbps. The results showed that the H.264 codec outperforms both of the older standard codecs, i.e. the H.261 and H.263. In the case of the performance comparison of the H.264 and HEVC codecs, there are many publications which confirmed that the HEVC coding standard in fact outperforms the H.264 codec by halving the bitrate at the same perceived quality. In several publications [23], [24], [25], [26], [27], [28], [29], [30], [31], [32], [33], [34] and [35] authors examine the performance using objective metrics, in other publications [36] and [37] subjective assessment is used, and finally, there are publications [38], [39], [40] and [41] in which the authors deployed both methods.

It may be difficult to quantitatively name a bitrate improvement factor for all contents. From the goal of the standardization group, their core test experiments and independent tests, at least H.264 and H.265 have likely halved the bitrate, leading to a total factor of four at the same perceptual quality. Conversely, four times the bitrate has been shown in various subjective experiments to change the perceptual quality significantly, and in most cases dramatically. Thus, there seems to be sufficient evidence that the coding performance improvements of the video coding standards has led to a significantly higher perceptual quality at the same bitrate.

3. Approach to Analyzing QoE Differences

The basic idea of hybrid multimedia coding is to combine redundancy and irrelevancy reduction. In redundancy reduction, the multimedia signal is transformed into a representation that minimizes the signal energy. In video coding, this is done by a Differential Pulse Code Modulation (DPCM) loop that includes a predictor that uses previous image contents that have already been transmitted. Those may stem from previously transmitted data of the same image (intra-prediction) or of image patterns that stem from temporally preceding (or following) images and that may be motion compensated (inter-prediction). The resulting residual image contains less energy than the original image in most cases. The energy is further compacted by using a frequency transformation (DCT for H.261, Integer-DCT for H.265). None of the redundancy reduction steps is lossy. The loss occurs in the quantization of the DCT coefficients. This step is supposed to remove irrelevancy, i.e. image information that is not mandatory to perceive important content. The frequency transform helps in identifying the significance of the information

for human, in particular the DC-component. Later on, the signal is entropy coded, thus minimizing the required bitrate, which is another redundancy reduction step.

QoE analysis needs to focus on the irrelevancy reduction as this removes irretrievable information from the signal. Taking a different perspective, the reference video and the decoded video differ to a certain amount that can be interpreted as added noise. The shape of the noise should be optimized such that the QoE impact is minimized.

Two approaches may be used. The first one consists of analyzing the signal difference of the reference video and the decoded video. This is the well-known classical approach of video quality estimation by Full-Reference methods. It is not the goal of this paper. The second one consists of the analysis of the information that is transmitted from the sender to the receiver. In video coding this information is identical to the transmitted bitstream and the closest visual representation is the quantized residual signal. This signal is calculated in the encoder just before entropy coding and is available in the decoder immediately after entropy decoding. From a perceptual point of view, interpreting and understanding the characteristics of this signal gives an approximation of what difference is important for the Human Visual System (HVS) to perceive within a frame's time (e.g. 20 ms at 50 fps). This interpretation and understanding is limited by many factors, notably the assumption that the motion estimation and compensation sufficiently approximates the motion prediction of the HVS and that the transmitted information resembles what the HVS requires in addition to the previous image and the prediction of object movements. The first limitation may be relativized by the fact that block-based object motion estimation algorithms perform reasonably, for the second limitation we propose the hypothesis that a significant improvement towards the requirements of the HVS relates to the efficiency gain in video coding explained earlier, i.e. the difference seen in between the quantized residual information of H.261 and H.265 should hint on the required difference information for the HVS.

4. Analysis

In this section, the results of the statistical and perceptual analysis are presented respectively.

4.1. Statistical Results

We assume that an improvement of quality may be captured by residuals of successive frames. It is worth noting first that the residuals have changed significantly

cantly over the years. Today, the residuals of successive frames carry less data/information than in the past but they still seem capable of covering all important information for the HVS. Therefore, in the following, we compare the residuals of the older H.261 video coding standard to those residuals of the current H.265 video coding standard.

We decided to analyze video sequences in CIF (352×288) resolution for an investigation of the residual's behavior of the H.261 and H.265 video codecs. This choice has been made because such low resolution can be easily depicted in a publication for visual inspection and because smaller images are faster to process. The processing time not only concerns the encoding process but also the extraction of residuals and the calculation of signal based statistics presented in the following.

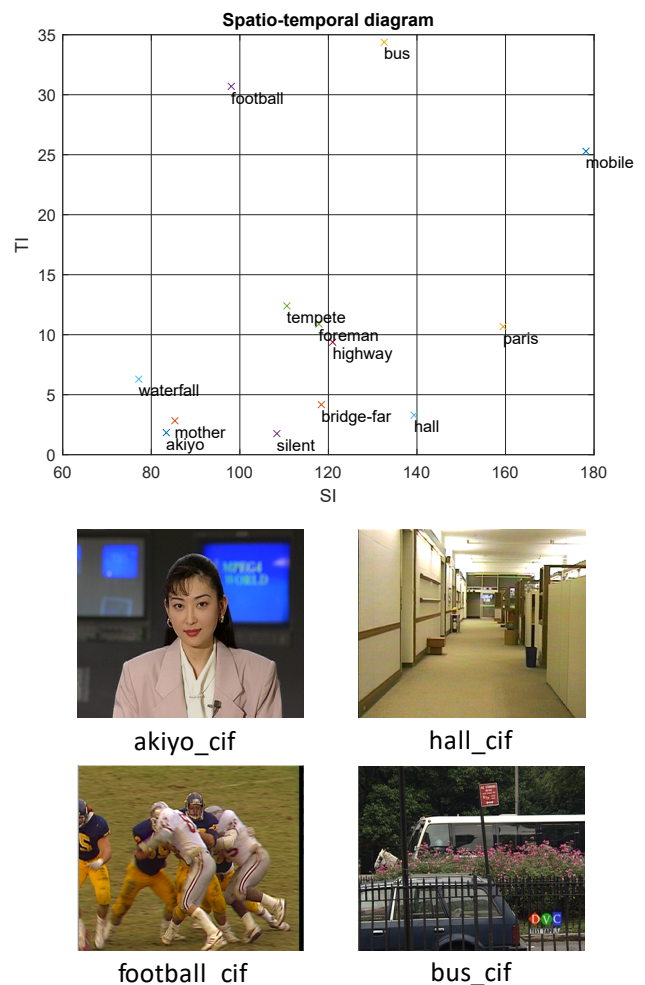


Fig. 1: Spatial-temporal diagram of the deployed sequences and their thumbnails.

Sequences were downloaded from the Video trace library [42] in YUV_420 format. We selected four different contents namely Akiyo, Bus, Football and Hall on the basis of the

spatial-temporal indicator (SI/TI) diagram, see Fig. 1 for more detail, in order to cover a broad range of content in terms of spatial and temporal information. Only the first five frames were used for the creation of the SI/TI diagram and for the analysis. It is worth noting here that the first five frames represent the behavior of the whole sequence in terms of the SI/TI activity. Each sequence was encoded by a simple MATLAB implementation of the H.261 encoder and also by the HM reference software (version 16.20) of the H.265 standard. The encoding process was restricted to involve only the Intra (I) and Predictive (P) frames to minimise the difference between the standards. The remaining parameters, except for the Quantization Parameter (QP), were kept as default. As the deployed standards utilize different representation of Quantization parameter. In order to specify the severeness of compression, we adjusted the parameters of the quantization setting such that a similar quantization step size for both standards H.261 and H.265 was used. To cover a wide range of quality degradation, five different quantization steps were selected. More specifically, we chose quantization steps of 2, 16, 32, 64 and 128 which we related to the Quantization Parameters 1, 8, 16, 32, 64 of the H.261 codec and in the H.265 case we used the Quantization Parameters 10, 28, 34, 40 and 46. During the encoding process the residuals from both the H.261 and H.265 sequences were extracted.

The first analysis consists of the visual comparison of the quantized residuals of both standards. We show the residuals of the first two frames of each sequence, which appeared very similar to the residuals of the other frames. As the residuals represent differences, they may contain positive and negative values. In order to visualize the data as an image, we added the mid-gray value 128 to each residual pixel value. When we compare the residuals visually, see Fig. 4, it can be seen that more blocks are distinguishable with increasing quantization step size. This is more pronounced in the case of the H.261 codec than in the case of the H.265 codec. Thus, the higher bitrate of H.261 seems to be partly due to a higher energy and thus requires more bitrate for the residual information. Furthermore, the coded (non-zero) blocks are often positioned in different locations when it comes to the H.261 and H.265 codecs. This is caused by the fact that the DC component of the H.261 codec is always quantized by a division by eight independently of the selected quantization step. That results in more blocks being visibly different from the surrounding as the DC value is less often quantized to zero. For the observer, this may finally lead to a distortion recognized as a blockiness effect and to a quality degradation. These coded blocks cause also increasing residuals energy as can be seen in Tab. 1. It is evident from the table that H.265 residuals have a lower energy in comparison to the H.261 resid-

uals and the difference increases with higher quantization steps. The energy was computed by the following Eq. (1):

$$E = \frac{\sum(A_{x,y}^2)}{(m \cdot n)}, \quad (1)$$

where $A_{x,y}$ is the value of residual at the coordinates x, y and m, n are width and height of the image respectively.

Tab. 1: Energy of residues.

	-	Qs 2	Qs 16	Qs 32	Qs 64	Qs 128
Akiyo	H.261	0.81	1.02	2.38	3.82	8.79
	H.265	0.64	0	0	0	0
Bus	H.261	157.77	144.74	139.38	114.42	83.63
	H.265	80.81	44.13	27.36	18.48	1.62
Football	H.261	186.5	168.25	158.08	125.23	76.93
	H.265	108.42	83.42	70.33	51.53	44.02
Hall	H.261	9.05	4.76	5.96	9.19	19.87
	H.265	25.01	0.41	0	0	0

There is also a difference in sizes of the blocks. The flexibility of block size selection in H.265 compared to H.261 enables the H.265 to use smaller block sizes, in particular at higher quantization steps. When we take a look at frequencies stored in the blocks of the residuals, we can see that the H.265 residuals encoded with the higher quantization steps contain lower frequencies in comparison to the H.261 residuals. It is also evident that the residuals contain more information (higher energy) when the scene contains faster motion.

In the next step, an analysis of the correlation of the residual signal is presented. Concerning the correlation between the horizontally adjacent pixels, see Fig. 2, it seems that the H.265 residuals show a higher correlation compared to those of the H.261 codec. In Fig. 2, a cross-like structure appears in the middle of the scatter plot for the H.265 residues with the higher quantization steps. This indicates that pixels requiring a large residual value are co-located to zero pixels. This behavior is not observed in the case of the H.261 residuals. Values of the correlation to adjacent pixels for each quantization step are presented in Tab. 2. It was calculated according to the equation for 2-D auto-correlation:

$$r = \frac{\sum_m \sum_n (A_{x,y} - \bar{A})(A_{x+1,y} - \bar{A})}{\sqrt{(\sum_m \sum_n (A_{x,y} - \bar{A})^2)(\sum_m \sum_n (A_{x+1,y} - \bar{A})^2)}} \quad (2)$$

where $A_{x,y}$ is the value of residual in position x, y , and $\bar{A}$ is the average values of the residual.

The next analysis concerns the inter-standard pixel-by-pixel correlation. We compare the residual values of H.261 to those at the same identical pixel position in H.265. In Fig. 3, visualization is depicted in the form

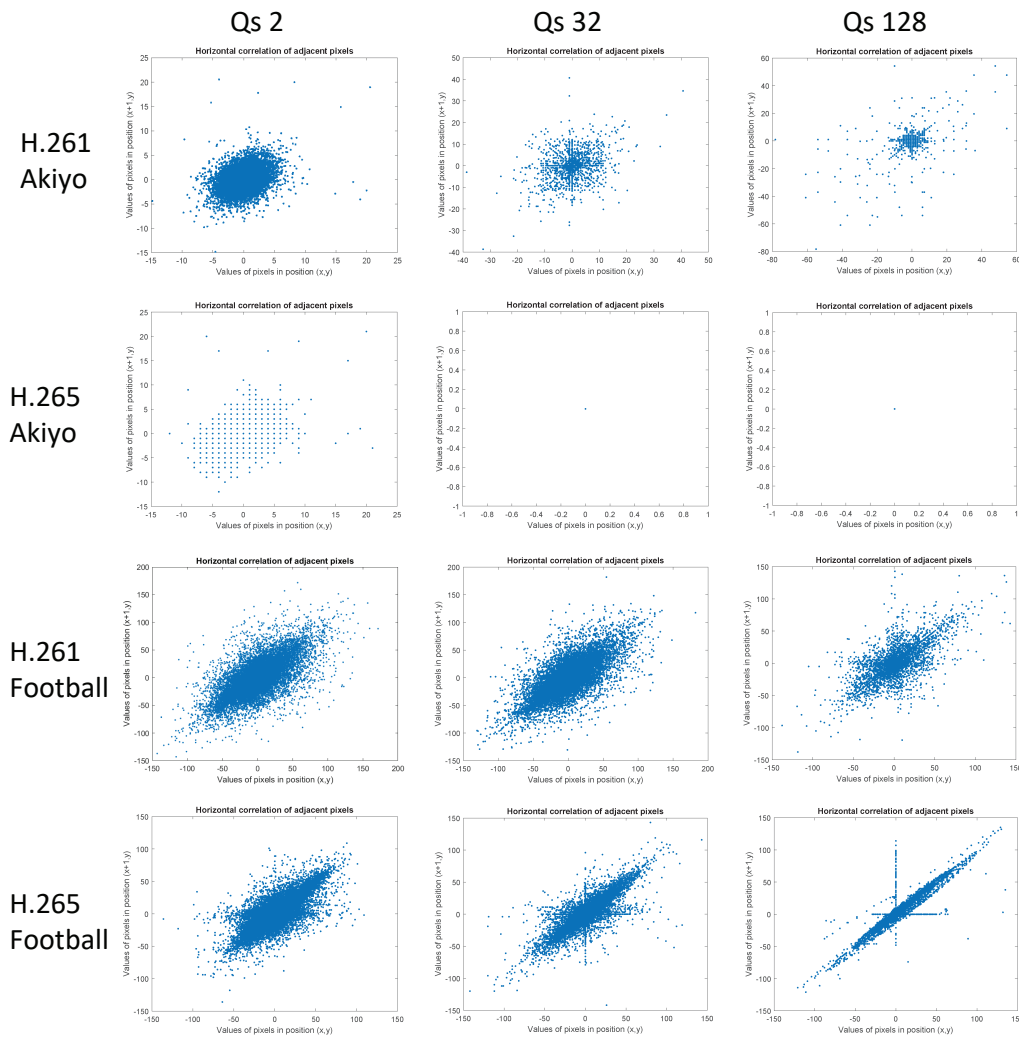


Fig. 2: Horizontal correlation between adjacent pixels in residuals of the first two frames of the sequences "Akiyo" and "Football" encoded with Qs 2, 32 and 128.

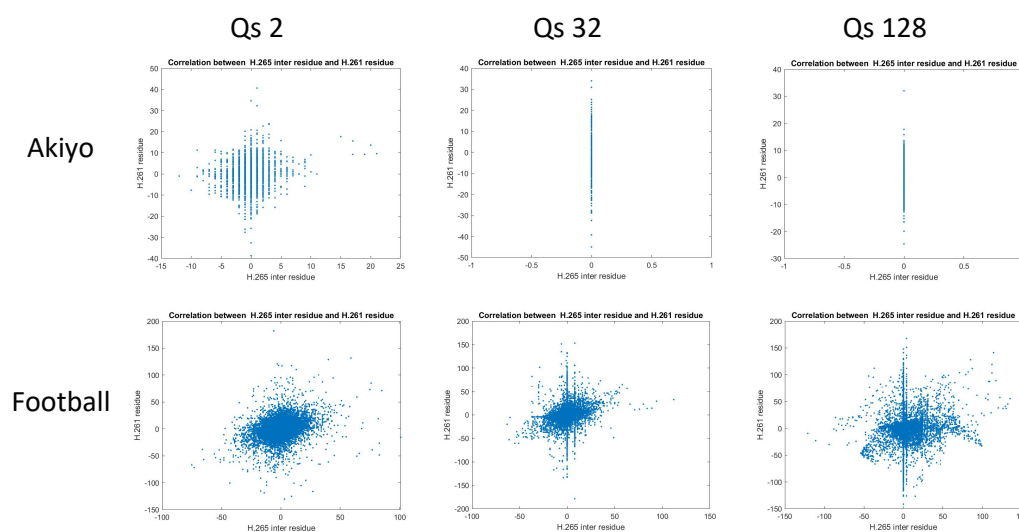


Fig. 3: Correlation between H.265 inter residual and H.261 residual of the sequences "Akiyo" and "Football" encoded with Qs 2, 32 and 128.

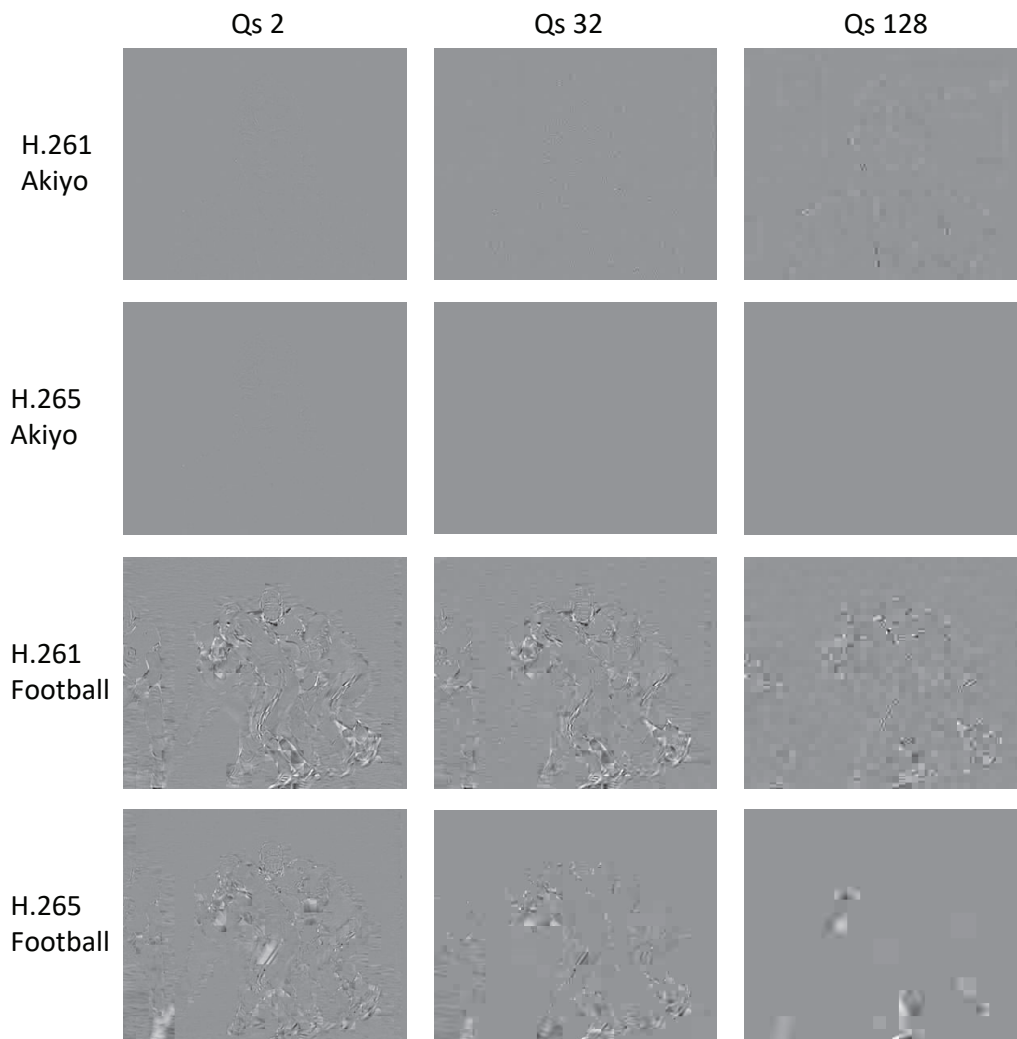


Fig. 4: H.261 and H.265 residuals of the first two frames of the sequences "Akiyo" and "Football" encoded with Quantization Steps (Qs) 2, 32 and 128.

Tab. 2: Horizontal correlation between adjacent pixels of residues.

	-	Qs 2	Qs 16	Qs 32	Qs 64	Qs 128
Akiyo	H.261	0.33	0.30	0.52	0.67	0.78
	H.265	0.32	-	-	-	-
Bus	H.261	0.62	0.62	0.63	0.68	0.74
	H.265	0.71	0.72	0.74	0.86	0.98
Football	H.261	0.69	0.71	0.73	0.74	0.80
	H.265	0.75	0.76	0.82	0.88	0.93
Hall	H.261	0.33	0.35	0.41	0.53	0.70
	H.265	0.71	0.24	-	-	-

of scatter plot. Please note that H.265 discerns two types of residual values. The first type is the residual of intra-frame predicted blocks, which therefore contains differences within frames. The second type is the residual information of inter-predicted blocks is also present in H.261, which therefore covers differences between the frames. The H.265 encoder decides whether a particular block is represented by the intra and inter

prediction. Fig. 3 depicts the inter-standard pixel-by-pixel correlation only for the inter-predicted blocks of the H.265. The points created by the straight vertical lines are noteworthy: They denote pixels that have zero values in the H.265 residual but non-zero values in the H.261 residual.

4.2. Perceptual Interpretation

The above mentioned differences in residuals may also lead to a different impact on observers when asked for QoE evaluation. The residuals of the H.261 codec often contain a non-zero average value in each block, caused by the DC coefficient, which is quantized by a comparably small, constant value of 8. For a human, general image understanding is often based on lower frequencies and this processing of the image was probably introduced in the H.261 standard to guarantee that the image content would be recognizable although the

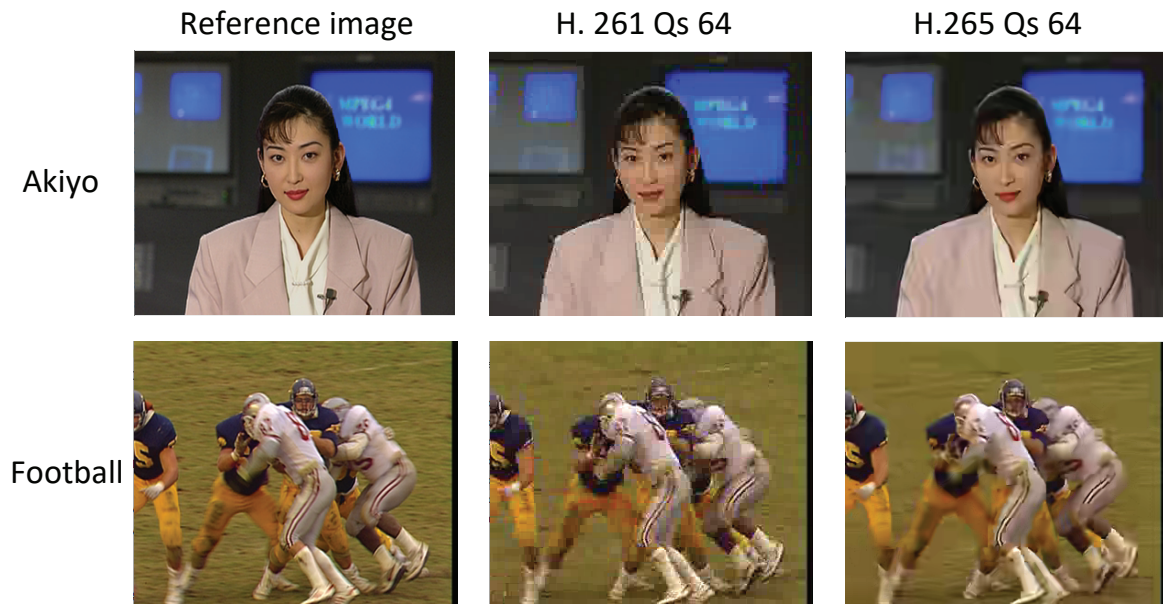


Fig. 5: Preview of distortions appearance caused by the investigated standards.

resolution would be reduced by a factor of 8. However, this can also cause the occurrence of visible block borders or the visibility of distinct blocks, often called blockiness. This occurs especially in image areas with a smooth color gradient when the gradient is substituted by clearly distinguishable block with different average values (DC coefficients). It can be also captured by a higher energy of the H.261 residuals (Tab. 1). Contrary to the older H.261, the current H.265 standard uses a more sophisticated quantization, when taking into consideration the size and type of the transform block and the transform coefficients representing different frequencies. This may lead to a lower difference between the adjacent blocks and ultimately to a less pronounced blockiness effect. A preview of distortions caused by the investigated video compression standards is depicted in Fig. 5. In our analysis this known fact was evidenced by the intra-standard correlation analysis. It was seen that the effect gets more pronounced when the quantization step size increases. The numerical characterization (correlation analysis) of the visual artifact (blockiness) executed by comparing the same video content in between an older standard (H.261) and a new one (H.265) may be used to further improve the coding tools for the residual information in upcoming standards.

From the visual inspection of the above quantized residuals, it also seems that the H.265 codec produces distortions that resemble more to a blurring effect than to a blockiness effect. This may be caused by the above mentioned quantization difference as well as the variation in block sizes. Blurriness may be perceived as a more commonplace effect in daily life and may thus

be more acceptable when viewed by humans [43]. However, the same authors found in [44] that blockiness and blurriness is detected at a similar level and perceived with similar annoyance when the resulting signal distortion is similar. As this change appears in the residual information of the inter-predicted blocks, a temporal effect of these distortions needs to be considered. In H.261, the stronger amount of blockiness may lead to a flickering of blocks or a structured noise in consecutive frames. The blurriness in the residual may also lead to degradation but this noise may appear more as a random noise, such as camera noise, rather than a structured, localized signal distortion. Further evaluating this effect on the residual may help in a better understanding of the QoE gain that H.265 has achieved over H.261. In the long term, as mentioned before, the intentional shaping of the residual towards a desired signal may further increase the perceptual quality.

5. Conclusions

In this paper, we focused on introducing a possible path towards the QoE analysis of differences in between video coding standards. The quantized residual in the video encoders (or decoders) was analyzed as the visual representative of the introduced degradation. The signal based analysis helped in quantifying the perceived difference in the residuals of H.261 and H.265 codecs. Interpretation of the visualization of the residual and the signal based characteristics towards an understanding of QoE differences was provided. Although only very basic signal processing and visual analysis has been performed, some interesting

results were documented that hint to the potential of the approach, both for understanding the gain in irrelevancy reduction in video coding as well as the possibility to design video coding tools towards an improved QoE.

Future work should focus on a subjective assessment of the human perception of quantized residuals. This would not only lead to an improved understanding of the same perceptual quality can be reached at a lower bitrate today but would also pave the way towards further improvements of video coding by shaping the information contained in the quantized residual by optimized video coding tools.

Acknowledgment

This publication is the result of the project implementation: Centre of excellence for systems and services of intelligent transport II, ITMS 26220120050 supported by the Research & Development Operational Programme funded by the ERDF.

References

- [1] BRUNNSTROM, K., S. BEKER, K. DE MOOR, A. DOOMS, S. EGGER, M.-N. GARCIA, T. HOSSFELD, S. JUMISKO-PYYKKO, C. KEIMEL, C. LARABI, B. LAWLOR, P. LE CALLET, S. MOLLER, F. PEREIRA, M. PEREIRA, A. PERIS, J. PIBERNIK, A. PINHEIRO, J. PIBERNIK, A. RAAKE, P. REICHL, U. REITER, R. SCHATZ, P. SCHELKENS, L. SKORIN-KAPOV, D. STROHMEIER, C. TIMMERER, M. VARELA, I. WECHSUNG, J. YOU and A. ZGANK. Qualinet White Paper on Definitions of Quality of Experience. In: *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)* [online]. 2012, Available at: <http://www.qualinet.eu/resources/qualinet-white-paper/>.
- [2] Video Quality Experts Group. In: *VQEG* [online]. 2022. Available at: <https://www.its.bldrdoc.gov/vqeg/vqeg-home.aspx>.
- [3] J.343 *Hybrid perceptual bitstream models for objective video quality measurements*. Geneva: ITU-T. 2014.
- [4] CHIKKERUR, S., V. SUNDARAM, M. REISSLEIN and L. J. KARAM. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison. *IEEE Transactions on Broadcasting*. 2011, vol. 57, iss. 2, pp. 165–182. ISSN 1557-9611. DOI: 10.1109/TBC.2011.2104671.
- [5] SHAHID, M., A. ROSSHOLM, B. LOVSTROM and H.-J. ZEPERNIK. No-reference image and video quality assessment: a classification and review of recent approaches. In: *EURASIP Journal on Image and Video Processing*. 2014, vol. 2014, iss. 1, pp. 1–32. ISSN 1687-5281. DOI: 10.1186/1687-5281-2014-40.
- [6] GHADIYARAM, D. and A. C. BOVIK. Crowd-sourced study of subjective image quality. In: *48th Asilomar Conference on Signals, Systems and Computers*. Pacific Grove: IEEE, 2015, pp. 84–88. ISBN 978-1-4799-8297-4. DOI: 10.1109/ACSSC.2014.7094402.
- [7] SINNO, Z. and A. C. BOVIK. Large Scale Subjective Video Quality Study. In: *25th IEEE International Conference on Image Processing (ICIP)*. Athens: IEEE, 2018, pp. 276–280. ISBN 978-1-4799-7061-2. DOI: 10.1109/ICIP.2018.8451467.
- [8] SINNO, Z. and A. C. BOVIK. Large-Scale Study of Perceptual Video Quality. In: *IEEE Transactions on Image Processing*. 2019, vol. 28, iss. 2, pp. 612–627. ISSN 1941-0042. DOI: 10.1109/TIP.2018.2869673.
- [9] BAMBAPIS, C. G., Z. LI, A. K. MOORTHY, I. KATSAVOUNIDIS, A. AARON and A. C. BOVIK. Study of Temporal Effects on Subjective Video Quality of Experience. *IEEE Transactions on Image Processing*. 2017, vol. 26, iss. 11, pp. 5217–5231. ISSN 1941-0042. DOI: 10.1109/TIP.2017.2729891.
- [10] SESHADRINATHAN, K., SOUNDARARAJAN, R., BOVIK, A. C. and L. K. CORMACK. Study of Subjective and Objective Quality Assessment of Video. In: *IEEE Transactions on Image Processing*. 2010, vol. 19, iss. 6, pp. 1427–1441. ISSN: 1941-0042. DOI: 10.1109/TIP.2010.2042111.
- [11] CHEN, C., L. K. CHOI, DE G. VECIANDA, C. CARAMANIS, R. W. HEATH and A. C. BOVIK. Modeling the Time-Varying Subjective Quality of HTTP Video Streams With Rate Adaptations. *IEEE Transactions on Image Processing*. 2014, vol. 23, iss. 5, pp. 2206–2221. ISSN 1941-0042. DOI: 10.1109/TIP.2014.2312613.
- [12] MOORTHY, A. K., L. K. CHOI, A. C. BOVIK and G. DE VECIANA. Video Quality Assessment on Mobile Devices: Subjective, Behavioral and Objective Studies. In: *IEEE Journal of Selected Topics in Signal Processing*. 2012, vol. 6, iss. 6, pp. 652–671. ISSN 1941-0484. DOI: 10.1109/JSTSP.2012.2212417.

- [13] ZHANG, F., F. M. MOSS, R. BADDELEY and D. R. BULL. BVI-HD: A Video Quality Database for HEVC Compressed and Texture Synthesized Content. In: *IEEE Transactions on Multimedia*, 2018, vol. 20, iss. 10, pp. 2620–2630. ISSN 1941-0077. DOI: 10.1109/TMM.2018.2817070.
- [14] PAPADOPOULOS, M. A., F. ZHANG, D. AGRAFIOTIS and D. BULL. A video texture database for perceptual compression and quality assessment. In: *2015 IEEE International Conference on Image Processing (ICIP)*. Quebec City: IEEE, 2015, pp. 3407–3411. ISBN 978-1-4799-8339-1. DOI: 10.1109/ICIP.2015.7351309.
- [15] MACKIN, A., F. ZHANG and D. BULL. A study of subjective video quality at various frame rates. In: *2015 IEEE International Conference on Image Processing (ICIP)*. Quebec City: IEEE, 2015, pp. 2781–2785. ISBN 978-1-4799-8339-1. DOI: 10.1109/ICIP.2015.7351436.
- [16] RAO R. R. R., S. GOERING, W. ROBITZA, F. BERNHARD and A. RAAKE. AVT-VQDB-UHD-1: A Large Scale Video Quality Database for UHD-1. In: *2019 IEEE International Symposium on Multimedia (ISM)*. San Diego: IEEE, 2019, pp. 17–177. ISBN 978-1-7281-5606-4. DOI: 10.1109/ISM46123.2019.00012.
- [17] ITU-T Recommendation H.261. *Video codec for audiovisual services at px64 kbits*. Geneva: ITU-T, 1993.
- [18] SULLIVAN, G. J., J.-R. OHM, W.-J. HAN and T. WIEGAND. Overview of the High Efficiency Video Coding (HEVC) Standard. *IEEE Transactions on Circuits and Systems for Video technology*. 2012, vol. 22, iss. 12, pp. 1649–1668. ISSN 1051-8215. DOI: 10.1109/TCSVT.2012.2221191.
- [19] WINKEN, M., P. HELLE, D. MARPE, H. SCHWARZ and T. WIEGAND. Transform Coding in the HEVC Test Model. In: *2011 18th IEEE International Conference on Image Processing*. Brussels: IEEE, 2011, pp. 3693–3696. ISBN 978-1-4577-1303-3. DOI: 10.1109/ICIP.2011.6116521.
- [20] RICHARDSON, I. *HEVC: An Introduction to High Efficiency Video Coding*. In: *Vcodex* [online]. 2013. Available at: <https://www.vcodex.com/hevc-an-introduction-to-high-efficiency-coding/>.
- [21] GIROD, B., E. STEINBACH and N. FARBER. Performance of the H.263 Video Compression Standard. *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*. 1997, vol. 17, iss. 2, pp. 101–111. ISSN 1939-8115. DOI: 10.1023/A:1007942603800.
- [22] HEBAR, M. and P. PLANINSIC. Comparison of video codecs and coded video sequences quality using the latest objective and subjective assessment methods. *Electrotechnical Review*. 2007, vol. 74, iss. 4, pp. 171–176. ISSN 0013-5852.
- [23] RAM, C. and S. PANWAR. Performance comparison of high efficiency video coding (HEVC) with H.264 AVC. In: *13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. Jaipur: IEEE, 2018, pp. 303–310. ISBN 978-153864283-2. DOI: 10.1109/SITIS.2017.58.
- [24] LAYEK, M. A., N. Q. THAI, M. A. HOSSAIN, N. T. THU, L. P. TUYEN, A. TALUKDER, T. CHUNG and E.-N. HUH. Performance analysis of H.264, H.265, VP9 and AV1 video encoders. In: *19th Asia-Pacific Network Operations and Management Symposium (APNOMS)*. Seoul: IEEE, 2017, pp. 322–325. ISBN 978-153861101-2. DOI: 10.1109/APNOMS.2017.8094162.
- [25] GROIS, D., T. NGUYEN and D. MARPE. Coding efficiency comparison of AV1/VP9, H.265/MPEG-HEVC, and H.264/MPEG-AVC encoder. In: *Picture Coding Symposium (PCS)*. Nuremberg: IEEE, 2016, pp. 1–5. ISBN 978-1-5090-5966-9. DOI: 10.1109/PCS.2016.7906321.
- [26] BIENIK, J., M. UHRINA, M. KUBA and M. VACULIK. Performance of H.264, H.265, VP8 and VP9 Compression Standards for High Resolutions. In: *19th International Conference on Network-Based Information Systems (NBIS)*. Ostrava: IEEE, 2016, pp. 246–252. ISBN 978-1-5090-0979-4. DOI: 10.1109/NBIS.2016.70.
- [27] MENGZHE, L., J. XIUHUA and L. XIAOHUA. Analysis of H.265/HEVC, H.264 and VP9 coding efficiency based on video content complexity. In: *IEEE International Conference on Computer and Communications (ICCC)*. Chengdu, IEEE, pp. 420–424. ISBN 978-1-4673-8126-0. DOI: 10.1109/CompComm.2015.7387608.
- [28] BASRI, A. A. and N. ZAINAL. Comparison of high efficiency video coding (HEVC) performance with H.264 advanced video coding (AVC). *Journal of Engineering Science and Technology*. 2015, vol. 10, iss. Special Issue on 4th International Technical Conference (ITC), pp. 102–111. ISSN 1823-4690.
- [29] KOUMARAS, H., M.-A. KOURTIS and C. SKIANIS. Compression performance and video quality comparison of HEVC and AVC. In: *IEEE 19th*

- International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. Athens: IEEE, 2014, pp. 75–79. ISBN 978-1-4799-5725-5. DOI: 10.1109/CAMAD.2014.7033209.
- [30] UHRINA, M., J. FRNDA, L. SEVCIK and M. VACULIK. Impact of H.264/AVC and H.265/HEVC Compression Standards on the Video Quality for 4K Resolution. *Advances in Electrical and Electronic Engineering*. 2014, vol. 12, iss. 4, pp. 368–376. ISSN 1804-3119. DOI: 10.15598/aece.v12i4.1216.
- [31] 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, 2014, pp. 394–397. ISBN 978-1-4799-0294-1. DOI: 10.1109/PCS.2013.6737766.
- [32] MILANOVIC, D., Z. MILICEVIC and Z. BOJKOVIC. MPEG video deployment in digital television: HEVC vs. AVC codec performance study. In: *11th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services (TELSIKS)*. Nis: IEEE, 2013, pp. 101–104. ISBN 978-1-4799-0902-5. DOI: 10.1109/TELSIKS.2013.6704900.
- [33] RERABEK, M., P. HANHART, P. KORSHUNOV and T. EBRAHIMI. Quality evaluation of HEVC and VP9 video compression in real-time applications. In: *Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*. Pilos: IEEE, 2015, pp. 1–6. ISBN 978-1-4799-8958-4. DOI: 10.1109/QoMEX.2015.7148088.
- [34] KOUMARAS, H., KOURTIS M.-A. and D. MARTAKOS. Benchmarking the encoding efficiency of H.265/HEVC and H.264/AVC. In: *Future Network & Mobile Summit (FutureNetw)*. Berlin: IEEE, 2012, pp. 1–7. ISBN: 978-1-905824-30-4.
- [35] AKYAZI, P. and T. EBRAHIMI. Comparison of Compression Efficiency between HEVC/H.265, VP9 and AV1 based on Subjective Quality Assessments. In: *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. Cagliari: IEEE, 2018, pp. 1–6. ISBN 978-1-5386-2605-4. DOI: 10.1109/QoMEX.2018.8463294.
- [36] BIENIK, J., M. UHRINA, M. VACULIK and T. MIZDOS. Perceived Quality of Full HD video - Subjective Quality Assessment. *Advances in Electrical and Electronic Engineering*. 2016, vol. 14, iss. 4, pp. 437–444. ISSN 1336-1376. DOI: 10.15598/aece.v14i4.1831.
- [37] DUMIC, E., S. GRGIC, D. FRANK and K. SAKIC. Subjective quality assessment of H.265 versus H.264 Video Coding for High-Definition Video Systems. In: *13th International Conference on Telecommunications (ConTEL)*. Graz: IEEE, 2015, pp. 1–7. ISBN 978-1-4799-8972-0. DOI: 10.1109/ConTEL.2015.7231203.
- [38] BAJCINOVIĆ, V., M. VRANJES, D. BABIC and B. KOVACEVIC. Subjective and objective quality assessment of MPEG-2, H.264 and H.265 videos. In: *International Symposium ELMAR*. Zadar: IEEE, 2017, pp. 73–77. ISBN 978-953-184-225-9. DOI: 10.23919/ELMAR.2017.8124438.
- [39] KOVACEVIC, S. 36.4: Subjective and Objective Visual Quality Evaluation of 4K Video Using AVC and HEVC Compression. *SID Symposium Digest of Technical Papers*. 2012, vol. 43, iss. 1, pp. 481–484. ISSN 0097-966X. DOI: 10.1002/j.2168-0159.2012.tb05822.x.
- [40] LI, Z., Z. DUANMU, W. LIU and Z. WANG. AVC, HEVC, VP9, AVS2 or AV1? — A Comparative Study of State-of-the-Art Video Encoders on 4K Videos. *16th International Conference on Image Analysis and Recognition (ICIAR)*. Waterloo: Springer, 2019, pp. 162–173. ISBN 978-3-030-27202-9. DOI: 10.1007/978-3-030-27202-9_14.
- [41] 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.
- [42] Video Trace Library. In: *Arizona State University* [online]. 2022. Available at: <http://trace.eas.asu.edu/>.
- [43] FARIAS, M. C. Q., J. M. FOLEY and S. K. MITRA. Some Properties of Synthetic Blocky and Blurry Artifacts. In: *SPIE Human Vision and Electronic Imaging*. Santa Clara: SPIE, 2003, pp. 128–136. ISBN 978-0-819-44807-1. DOI: 10.1117/12.477381.
- [44] FARIAS, M. C. Q., J. M. FOLEY and S. K. MITRA. Detectability and Annoyance of Synthetic Blockiness, Blurriness, Noisiness, and Ringing in Video Sequences. In: *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. Philadelphia: IEEE, 2005, pp. 553–556. ISBN 7803-8874-7. DOI: 10.1109/ICASSP.2005.1415464.

About Authors

Tomas MIZDOS was born in 1993 in Poprad, Slovak Republic. He received his B.Sc. degree in Multimedia technologies at the of Multimedia and Information-Communication Technologies, Faculty of Electrical Engineering and Information Technology, University of Zilina in 2015. Currently he is a Ph.D. student at the same department. His main area of interest is Functionality and Quality of Multimedia Services, Quality of Experience, Video Coding and Digital Signal Processing.

Marcus BARKOWSKY received his Dr.-Ing. degree from the University of Erlangen-Nuremberg in 2009. He joined the University of Nantes and was promoted to an associate professor in 2010. In 2018, he obtained the professorship on interactive systems and internet of things at the Deggendorf Institute of Technology, University of Applied Sciences. His activities range from designing 3-D interaction and measuring visual discomfort using psychometric measurements to computationally modeling spatial and temporal effects of the human perception.

Peter POCTA was born in 1981. He received his M.Sc. and Ph.D. degrees from University of Zilina, Faculty of Electrical Engineering, Slovak Republic in 2004 and 2007, respectively. He is currently an associate professor at the Department of Multimedia and Information-Communication Technologies of the University of Zilina and is involved with International Standardization through the ETSI TC STQ as well as ITU-T SG12. His research interests include speech, audio, video and audiovisual quality assessment, speech intelligibility, multimedia communication and QoE management.

Miroslav UHRINA was born in 1984 in Zilina, Slovak Republic. He received his M.Sc. and Ph.D. degrees in Telecommunications at the Department of Multimedia and Information-Communication Technologies, Faculty of Electrical Engineering and Information Technology, at the University of Zilina in 2008 and 2012, respectively. Currently, he is an assistant professor and a head of Multimedia section in the same department. His research interests include Quality of Experience (QoE), audio and video compression, TV broadcasting and IP networks.