

Quality Evaluation of HEVC and VP9 Video Compression in Real-Time Applications

Martin Řeřábek, Philippe Hanhart, Pavel Korshunov, and Touradj Ebrahimi
Multimedia Signal Processing Group (MMSPG)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Station 11, Lausanne, Switzerland

Email: martin.rerabek@epfl.ch, philippe.hanhart@epfl.ch, pavel.korshunov@epfl.ch, touradj.ebrahimi@epfl.ch

Abstract—Video consumption over Internet has increased significantly over the recent years and occupies the majority of the overall data traffic. To decrease the load on the Internet infrastructure and reduce bandwidth taken by video, higher efficiency video codecs, such as H.265/HEVC and VP9, have been developed. The availability of these two new competing video coding formats raises the question of which is more efficient in terms of rate-distortion and by how much they outperform the current state-of-the-art coding standard, H.264/AVC. This paper provides an answer to this difficult question for low-delay video applications, e.g., real-time video streaming/conferencing or video surveillance. The benchmarking of HEVC and VP9 video compression was conducted by means of subjective evaluations, assuming web browser playback, an uncontrolled environment, and HD video content. Considering a wide range of bit rates from very low to high bit rates, corresponding to low quality up to transparent quality (when compared to the original video), results show a clear advantage of HEVC with average bit rate savings of 59.5% when compared to AVC and 42.4% when compared to VP9.

Keywords—Video compression, subjective assessment, HEVC, H.265, VP9, WebM, AVC, H.264, Real-time applications

I. INTRODUCTION

Recent dramatic increase in video consumption over IP-networks, with video data taking more than 75% of Internet traffic, prompted for the development of new video compression technologies that would be significantly more efficient than the existing video codecs, including the popular H.264/MPEG 4 Part 10 AVC [1] (referred as AVC in this paper). The development efforts led to the creation of two video codecs. One is the high efficiency video coding (HEVC) [2], also known as H.265/MPEG-H Part 2, jointly developed by ITU-T and ISO/IEC. Another is VP9 video compression, which is an OpenSource alternative to HEVC developed by Google and is positioned as a royalty-free, license-fee-free solution, with the main focus on supporting Internet-based video consumption.

The fact that VP9 was released at a similar time frame to HEVC and that it was announced as a superior alternative raised interest in the research and professional communities. It resulted in several studies comparing these two codecs to each other and to AVC. Most such studies relied on objective metrics to measure coding efficiency and resulted in conflicting conclusions depending on the study performed. In [3], the authors claim that VP9 is inferior to both AVC and HEVC and demonstrate that HEVC provides average

bit rate savings of 43.3% compared to VP9. However, a different study [4] comes to a different conclusion, with VP9 showing similar compression efficiency when compared to HEVC and a significantly higher compression efficiency when compared to AVC. Such conflicting conclusions are mainly caused by different usage scenarios assumed in the papers and by different encoding configurations used. The authors of [3] have further extended their study to a low-delay scenario [5], which is more suitable for real-time video applications, and by using PSNR measurements, and conclude that using HEVC results in average bit rate savings of 32.5% when compared to VP9.

In the above studies, authors relied only on PSNR as objective metric to compare compression efficiency of selected encoding schemes. However, human perception is subjective, and results of subjective assessments performed using standard quality evaluation methodologies is *a priori* a more reliable measure of compression efficiency. Therefore, a subjective evaluation of HEVC, VP9, and AVC codecs was performed in [6] to determine the actual perceived quality of compressed video content. The study assumed a broadcasting scenario using ultra high definition (UHD) video content in a standard test laboratory environment with controlled lighting conditions and a professional UHD reference monitor. According to the subjective evaluation results, HEVC outperformed VP9, showing on average a nearly 50% bit rate reduction for the same subjective quality.

This paper presents the comparison of the compression efficiency between HEVC, VP9, and AVC by means of subjective evaluations assuming a real-time Internet-based streaming scenario. In such a scenario, subjects receive a real-time streamed video content and watch it in a web browser in an uncontrolled environment. High definition (HD) content is typical for current video consumption over the Internet and is compressed using parameters most suitable for Internet-based scenario. In our experiments, a total of 26 subjects took part in a crowdsourcing subjective assessment, evaluating 8 different video contents with resolutions ranging from 720p to 1080p, which were compressed to four different bit rates using HEVC, VP9, and AVC.

It should be noted that the comparison of different codecs in a fair and unbiased manner is very hard, especially because of many differences in their actual implementations [4]. Nevertheless, since the development of the first video compression standards, two main coding configurations have been considered. The first one is the low-delay configuration, which

Codec	Version	Parameters
HEVC	HM 16.2	Default main low-delay profile with B frames. IntraPeriod = -1 (only first frame encoded as I frame). List0 reference.
x264	r2491	--profile high --tune psnr --ref 4 --direct auto --weightp 2 --level 5.1 --subme 8 --b-pyramid none --bframes 0 --b-adapt 0 --merange 24 --me tesa --no-fast-pskip --trellis 2 --min-keyint=9999 --keyint=9999 --pass 1 --slow-firstpass --fps < FR > --qp < QP > --psnr -v
VP9	v1.3.0-4786-gbf44117	--good --cpu-used=0 --end-usage=3 --codec=vp9 --kf-max-dist=9999 --kf-min-dist=9999 --lag-in-frames=0 --passes=1 --cq-level=< QP > -w < W > -h < H > --fps=< FR > --psnr -v -t 0

Table I: Selected parameters and settings for the AVC, HEVC, and VP9 codecs.

aims to achieve real time encoding-decoding, exhibiting less than 150ms delay [7]. In this configuration, picture reordering is not allowed and only the first frame is intra-coded. Teleconferencing and video surveillance are typical application scenarios that requires low-delay. The second configuration is referred to as random-access. In this configuration, picture reordering is allowed and intra-coded frames are inserted periodically, e.g., every 0.5-1s. Typical application scenarios include video storage, to be able to easily access the video stream at different positions, and broadcasting, for channel surfing [8]. To reflect the real-time application scenario, we selected low-delay coding configuration. Additionally, we used fixed quantization parameters (QP) and disabled any rate-distortion optimization tools.

The paper is organized as follows. In the next section, the selected coding algorithms compared in our experiments and their main configuration parameters are introduced and briefly described. Section III contains detailed description of the selected test video sequences, the testing environment, the test methodology adopted in our experiment, and brief description of data analysis. Statistical analysis of the collected subjective data and results are presented in Section IV. Finally, Section V concludes the paper.

II. CONFIGURATION OF SELECTED CODECS

For HEVC, the HM reference software [9] was selected, as it is a popular implementation [2]. The latest codec of the WebM Project, i.e., VP9 codec [10], released by Google was selected and used in our experiments. Finally, the x264 library [11] was used to evaluate the performance of AVC based coding scheme as it is fast, publicly available, and one of the most commonly used implementations of AVC. For each codec, the fixed quality parameter was set separately. Such setting allows fair mutual comparison of encoders as it removes all rate control adaptation between video frames. A more detailed description of the selected encoders, including their profiles and parameters configuration, is presented further in this section.

The latest versions of the HM reference software was used for encoding video sequences with HEVC. The Low-Delay (LD) configuration in default main profile with B frames was selected [12]. Low-delay configuration with B frames was selected since it achieves higher coding efficiency (because of bi-prediction), when comparing to low-delay configuration with P frames only. In this configuration, the first frame is encoded as an I frame and subsequent frames are encoded as B frames, while reordering of the B frames is not allowed, i.e., only the reference picture list 0, which references to past frames, is used. Therefore, this configuration introduces

minimal coding delay and can be used for real-time application scenarios.

For this work, the VP9 encoder and decoder were considered as a most recent implementation of the WebM Project. Due to the lack of official documentation and specifications for this encoder, the parameters were set based on recommendations received from the WebM Project lead developers. VP9 encoder allows to set the QP in two different ways. First approach [3], [5] sets the `--min-q` and `--max-q` parameter to the same value. According to the comments of lead developers of VP9, such a setting apparently decreases the compression efficiency [6]. Therefore, the available fixed quality mode `--end-usage=3`, which allows to vary the coding quality factor, was selected for VP9 encoding. Furthermore, the Intra Period parameters (`kf-min-dist` and `kf-max-dist`) were set to very large values to ensure that only the first frame is an I frame, which corresponds to low-delay configuration requirements for real-time scenarios considered in this paper. The selected configuration for VP9 allows comparative testing with AVC and HEVC.

Since the x264 implementation allows low-delay configuration only with P frames, it is only used as an orientation anchor to benchmark the other two next generation codecs.

More detailed information about the configuration of all investigated encoders can be found in Table I.

III. SUBJECTIVE EVALUATION

This section describes the dataset and methodology used in the crowdsourcing-based subjective quality assessment, as well as the processing of the collected raw subjective scores.

A. Dataset preparation and description

Ten video sequences were used in the experiments, with different spatial and temporal characteristics, resolutions, and frame rates. Eight sequences were used for the subjective tests and two sequences were used for training. Figure 1 shows a representative frame sample of each video sequence. Each video sequence was ten seconds long and stored as raw, progressively scanned video file, with YCbCr 4:2:0 color sampling and 8 bits per sample. Furthermore, each video file was encoded with all three evaluated codecs at four bit rates. Since fixed QP configuration was used to control the quality of AVC, HEVC, and VP9 compressed bitstreams, the sequences were first encoded at various QP values. Then, an expert screening session was conducted to select the lower and upper QP bounds for each content separately (including training), by targeting bit rates defined in [13] and trying to cover the full quality scale for each content. Table II reports the final sets

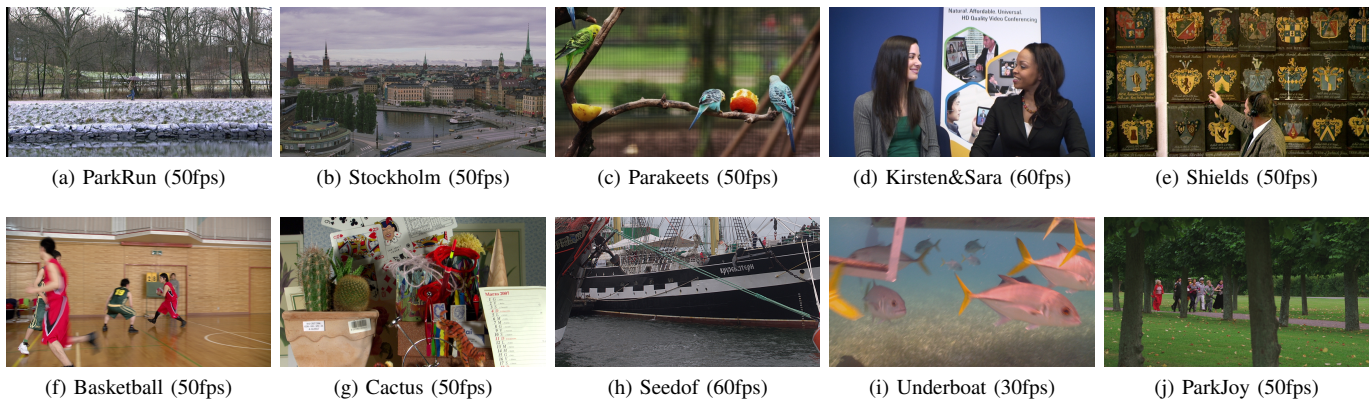


Figure 1: Sample frames of individual contents used in the subjective evaluation campaign. Contents (a) - (e) and (f) - (j) have a resolution of 1280×720 and 1920×1080 pixels, respectively. Contents Shields (e) and ParkJoy (j) were used for training. Contents (c) - (d) and (f) - (j) were used in HEVC verification test, whereas the remaining contents were downloaded from [14].

Sequence	R1'		R1				R2'		R2				R3'		R3				R4'		R4							
	[kbps]	QP	AVC [kbps]	HEVC [QP]	VP9 [QP]	VP9 [kbps]	[kbps]	QP	AVC [kbps]	HEVC [QP]	VP9 [QP]	VP9 [kbps]	[kbps]	QP	AVC [kbps]	HEVC [QP]	VP9 [QP]	VP9 [kbps]	[kbps]	QP	AVC [kbps]	HEVC [QP]	VP9 [QP]	VP9 [kbps]				
1280x720p																												
ParkRun	384	50	374	45	376	62	773	512	48	530	43	603	61	990	850	46	818	41	941	60	1193	1500	43	1551	39	1420	58	1655
Stockholm	256	45	269	33	228	53	255	384	40	395	31	328	45	369	512	38	480	29	473	40	529	850	34	823	27	682	35	839
Parakeets	256	48	250	34	240	53	261	384	42	394	31	356	45	418	512	38	530	29	471	42	506	850	33	896	25	817	35	823
Kirsten&Sara	256	43	251	32	252	49	254	384	38	379	29	402	42	395	512	35	508	27	558	39	492	850	31	811	25	801	33	828
1920x1080p																												
Basketball	2000	43	2014	33	2148	52	2117	3000	39	3020	31	2889	47	2961	4500	35	4722	30	3373	41	4540	7000	32	6800	26	6446	35	7320
Cactus	2000	41	1919	32	2059	49	2077	3000	37	2954	30	2818	43	3117	4500	33	4793	27	4585	39	4193	7000	30	7405	26	5519	33	6951
Seedof	2000	43	2119	37	2062	56	2336	3000	41	2865	35	2958	54	2897	4500	38	4817	33	4187	49	4508	7000	36	6858	30	7353	45	6489
Underboat	1000	40	1061	30	1000	43	1029	1600	36	1522	27	1692	39	1445	2500	32	2429	25	2510	33	2609	4000	28	4243	23	3997	29	3899

Table II: Targeted R' and actual R bit rates including the corresponding QP values for each codec.

of targeted ($R1'$ - $R4'$) and actual ($R1$ - $R4$) bit rates, with corresponding QPs, for each codec.

B. Test methodology

The single-stimulus (SS) methodology [15] was chosen for evaluations. A five-grade quality scale (1: *Bad*; 2: *Poor*; 3: *Fair*; 4: *Good*; 5: *Excellent*) was used. The subjects were asked to judge the overall quality of the evaluated video sequence. To reduce contextual effects, the stimuli orders of display were randomized applying different permutation for each subject, while the same content was never shown consecutively.

To display the video sequences and collect individual scores, a slightly modified version of the QualityCrowd 2 framework [16] was used. QualityCrowd 2 uses a simple scripting language that allows for the creation of test campaigns with high flexibility, e.g., specific pages for instructions, mixing different methodologies, etc. However, QualityCrowd 2 uses a fixed task order for each batch. To overcome this drawback, a plugin was developed to redirect each worker to a different batch, with a different display order for stimuli. Additionally, the VLC web plugin was used instead of the default Flash player, as it offers fullscreen playback. Fullscreen was automatically enforced for full HD video sequences.

All decoded video sequences were re-encoded with AVC,

since transmitting uncompressed video data to remote workers is impractical and there is no browser video plugin capable of reliable real-time decoding and displaying for all evaluated codecs and resolutions, especially for HEVC and VP9 full HD content. The 720p contents at 50 fps were compressed at 20 Mbit/s, which is commonly considered as perceptually transparent quality for video broadcasting. For other frame rates and resolutions, the bit rate was set proportionally to their frame rate and resolution corresponding to the above mentioned bit rate. For example 1080p contents at 50fps were encoded at 45 Mbit/s. A two-pass encoding was used and the deblocking filter was disabled to preserve the original blockiness artifacts when encoded at low bit rates. Expert viewing session was conducted prior to the main subjective assessment and the expert viewers evaluated the quality of this second encoding as visually lossless.

To mimic the realistic real-time application scenario, the subjective tests should ideally be conducted in form of crowd-sourcing. Nevertheless, as it is relatively difficult to find online subjects equipped with a full HD monitor and because of the relatively large amount of transmitted video data, the experiments were conducted at EPFL in an uncontrolled lab room with desktop computers. Therefore, the workers demographic was limited to university students participating on voluntary basis, and thus they were not remunerated for their effort.

However, this approach helps to focus the subjective tests to compression part of the transmission chain only, as it limits the artifacts due to network transmission, transport protocol and playback settings.

A total of 26 subjects participated in the study. Each subject evaluated all test stimuli. Half of the subjects evaluated the 720p contents first, while the other half evaluated the 1080p contents first. To minimize visual fatigue effects, subjects took 10 minutes break between the two tasks.

Before the experiments, short written training instructions were provided to the subjects to explain their tasks. Additionally, three training samples, representative of *Excellent*, *Fair*, and *Bad* quality, were displayed to familiarize subjects with the assessment procedure. The training instructions and samples were presented using QualityCrowd 2.

C. Data processing

For the objective comparison of encoders, the rate distortions curves of luma component were used. The resulting PSNR_Y value for each video sequence was computed as the average of all measurements across frames.

To evaluate perceived quality, standard statistical indicators describing the score distribution across subjects for each test condition (combination of content, codec, and bit rate) were computed. First, outlier detection was applied to remove subjects whose scores deviated strongly from others. Assuming the reliability of subjects participating on voluntary basis, no crowdsourcing measures, such as honeypots, were used to detect the outliers. However, the outlier detection was performed according to the guidelines described in Section 2.3.1 of Annex 2 of [15]. In our experiments, none of the subjects was detected as an outlier for any of the test sessions. Then, the mean opinion score (MOS) and 95% confidence intervals (CI), assuming a Student's t -distribution of the scores, were computed for each test condition. Finally, a multiple comparison analysis [17] was performed to identify statistically significant differences among MOS obtained for different codecs and bit rates. For this purpose, a one-way ANOVA and multiple comparison tests were performed. No correction was applied to correct for the multiple comparisons.

IV. RESULTS AND DISCUSSIONS

Figures 2 (a) and (c) depict the rate-distortion (R-D) curves for the 720p and 1080p contents, respectively. The R-D curves based on PSNR measurements are plotted with dashed lines, whereas the subjective ratings, i.e., MOS and CI, are plotted with solid lines. Based on PSNR measurements, HEVC outperforms VP9 by 0.5 to 2 dB, while VP9 provides a gain ranging from 0.5 to 6 dB when compared to AVC. For all contents and bit rates, objective measurements show that HEVC outperforms both coding algorithms.

The subjective results show relatively small confidence intervals, indicating a high reliability of the results and rather small variation across subjects. The ratings show similar trend to objective measurements: HEVC provides the best visual quality for a similar bit rate and largely outperforms AVC in most cases. Also, VP9 achieves better visual quality than AVC, except for contents ParkRun and Seedof, where CIs

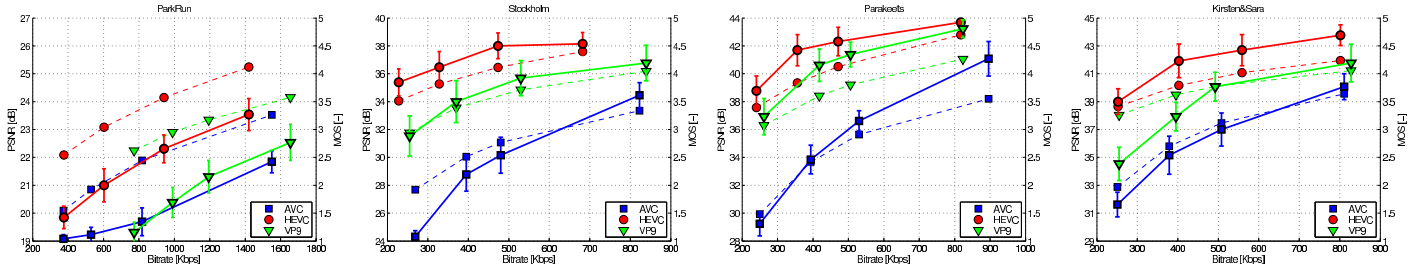
overlap significantly. However, in some cases (in particular, at high bit rates), HEVC and VP9 have similar ratings and there is no sufficient statistical evidence indicating differences in performance between these codecs at these bit rates. Finally, both HEVC and VP9 codecs can achieve Good to Excellent quality, i.e., $\text{MOS} \geq 4$, at the highest bit rates used in our study, with the only exception of content ParkRun. Lower ratings for ParkRun content can be explained by the large values of spatial and temporal indices, computed according to [18], implying that this content contains areas with high level of details and a lot of motion, and thus it is very demanding in bit rate.

Figures 2 (b) and (d) show the results comparing all possible conditions for the 720p and 1080p contents, respectively. Comparing HEVC and AVC at similar bit rates, HEVC always provides statistically better visual quality when compared to AVC for contents Stockholm, Parakeets, Kirsten&Kara, and Seedof. For the other contents, there is not sufficient statistical evidence to show that HEVC outperforms AVC, especially at high bit rates. Looking at HEVC vs. VP9, HEVC is significantly better at the three lowest bit rates for contents Seedof, Kirsten&Sara, and ParkRun, whereas there are no statistical differences on contents Parakeets and UnderBoat. For the other contents, HEVC only outperforms VP9 at the lowest bit rates. Regarding the comparison between VP9 and AVC, VP9 outperforms AVC at the three lowest bit rates on the 720p contents, except for content ParkRun where there is no statistical difference. Similarly, VP9 shows better performance to AVC on the two lowest bit rates for the 1080p contents, with the exception of content Seedof, where there is not sufficient statistical evidence to show that VP9 outperforms AVC. Note that there is no case where AVC nor VP9 outperform HEVC, or when AVC outperforms VP9.

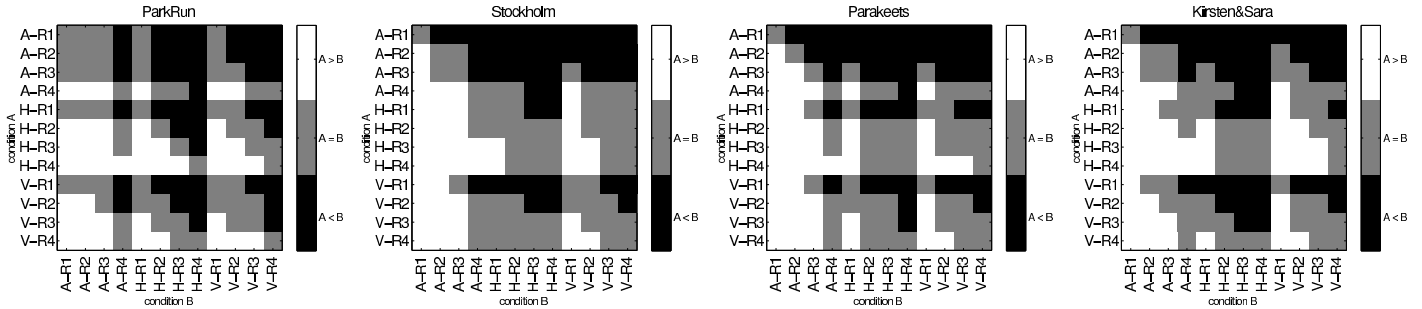
The bit rate reduction of one codec over another for a similar quality is estimated using the Bjøntegaard Delta Rate (BD-Rate) [19]. The Bjøntegaard model relies on PSNR measurements to determine the average bit rate difference. In this paper, only the luma component was considered for the PSNR computations. However, a more realistic estimate of the coding efficiency can be obtained by considering subjective ratings instead of PSNR values. For this purpose, the model proposed in [20] was used. This model reports the average bit rate difference, ΔR , for a similar perceived visual quality. Table III reports the BD-Rate and ΔR results.

Results based on objective measurements show that the average bit rate reduction of HEVC relative to AVC and VP9 is 57.3% and 33.6%, respectively. Although we used different encoders, different parameters (i.e., quality control parameters for VP9), and different metric PSNR_Y , the results comparing HEVC to VP9 correspond to findings of [3], where authors claim 32.5% bit rate savings in favor of HEVC. In other studies [3], [4], [6], authors used random access encoders configuration, and therefore mutual comparison of our results to those works is irrelevant.

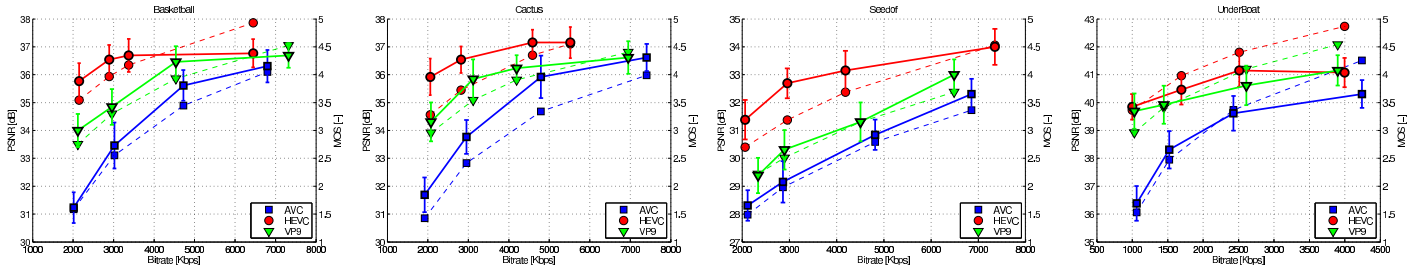
On the other hand, results based on the subjective ratings indicate an average bit rate saving of 59.5% and 42.4% for HEVC when compared to AVC and VP9, respectively. Furthermore, the bit rate reduction achieved by VP9 relative to AVC is 40.4% and 33.3% based on estimated and perceived quality, respectively. These results show that the compression efficiency of HEVC over AVC predicted based on



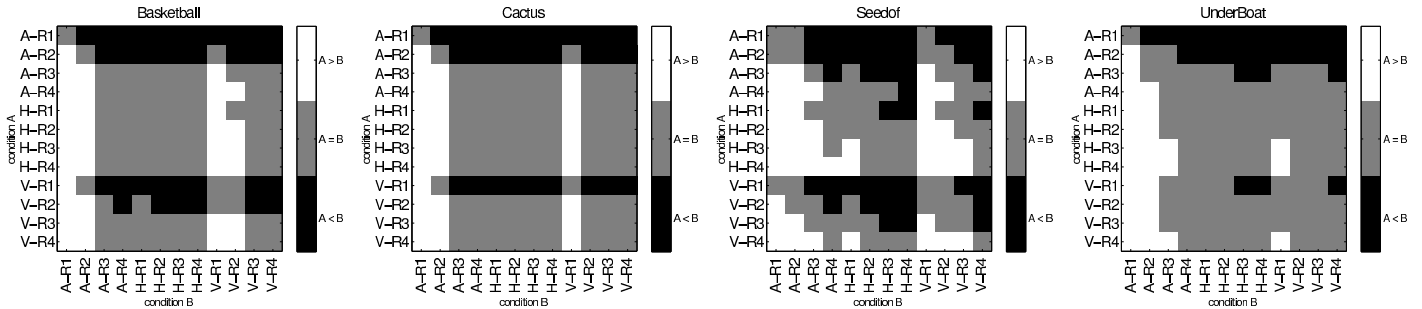
(a) PSNR (dashed line) and MOS and CI (solid line) for 720p contents.



(b) Multiple comparison test results for 720p contents.



(c) PSNR (dashed line) and MOS and CI (solid line) for 1080p contents.



(d) Multiple comparison test results for 1080p contents.

Figure 2: (a) and (c): PSNR R-D curves and subjective ratings showing MOS and CI values for each bit rate and content separately. (b) and (d): Multiple comparison test results for all possible combinations of codecs (A stands for AVC, H stands for HEVC, and V stands for VP9 coding algorithm) and bit rates (R1 to R5), for each test content separately. In each plot, the color of each square shows the result of the significance test between the mean opinion scores related to the two test conditions reported in the corresponding column and row. A white (black) square indicates that the MOS corresponding to condition A is statistically significantly better (worse) than the MOS corresponding to condition B, while a grey square indicates that there is no sufficient statistical evidence indicating differences between the two MOS values.

Content	HEVC vs AVC		HEVC vs VP9		VP9 vs AVC	
	BD-Rate	ΔR	BD-Rate	ΔR	BD-Rate	ΔR
ParkRun	-54.8 %	-53.0 %	-44.0 %	-46.2 %	-18.3 %	-7.4 %
Stockholm	-	-	-46.1 %	-54.7 %	-55.9 %	-49.7 %
Parakeets	-69.1 %	-62.4 %	-32.1 %	-28.0 %	-55.5 %	-48.5 %
Kirsten&Sara	-60.1 %	-62.6 %	-20.8 %	-43.8 %	-52.8 %	-30.8 %
Basketball	-55.8 %	-59.3 %	-38.3 %	-45.6 %	-31.5 %	-28.3 %
Cactus	-54.3 %	-57.5 %	-23.6 %	-43.0 %	-42.9 %	-31.9 %
Seedof	-52.8 %	-61.8 %	-36.0 %	-51.1 %	-26.9 %	-21.9 %
Underboat	-54.2 %	-60.1 %	-27.6 %	-26.6 %	-39.2 %	-48.1 %
Average	-57.3%	-59.5%	-33.6%	-42.4%	-40.4%	-33.3%

Table III: Comparison of investigated coding algorithms in terms of bit rate reduction for similar PSNR_Y and MOS. Negative values indicate actual bit rate reduction. Note that the bit rate difference between HEVC and AVC on content Stockholm could not be computed as the R-D curves have no horizontal overlapping.

PSNR values is similar to the gain observed from subjective ratings. However, the performance of VP9 computed based on objective measurements seems to be overestimated, as the compression efficiency estimated from subjective ratings shows lower values. These results indicate that previous studies relying only on objective evaluations might have overestimated the performance of VP9.

V. CONCLUSION

In this paper, a detailed description of the subjective quality evaluation tests conducted to benchmark the performance of HEVC and VP9 video codecs for real-time video applications has been presented, using low-delay coding configurations, as defined by MPEG. These codecs have been compared to an AVC anchor using eight different HD contents encoded at four bit rates. Subjective quality scores of 26 subjects related to each content, codec and quality parameter have been collected via a crowd-based platform. High consistency of achieved results allows an accurate comparison of the performance of the investigated codecs.

The test results show that HEVC offers improvements in compression performance when compared to VP9 and AVC, if one considers a wide range of bit rates from low to high, corresponding to video with low to transparent quality. More specifically, objective based measurements show that HEVC achieves average bit rate savings of 57.3% versus AVC and 33.6% versus VP9, whereas an average bit rate reduction of HEVC based on perceived quality reaches 59.5% versus AVC and 42.4% versus VP9.

ACKNOWLEDGMENT

This work has been performed in the framework of the FP7 EC EUROSTAR funded Project - Transcoders Of the Future TeleVision (TOFuTV) and COST IC1003 European Network

REFERENCES

- [1] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, July 2003.
- [2] G. J. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [3] D. Grois, D. Marpe, A. Mulyoff, B. Itzhaky, and O. Hadar, "Performance comparison of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders," in *PCS*, 2013, pp. 394–397.
- [4] D. Mukherjee, J. Bankoski, A. Grange, J. Han, J. Koleszar, P. Wilkins, Y. Xu, and R. Bultje, "The latest open-source video codec VP9-an overview and preliminary results." in *PCS*, 2013, pp. 390–393.
- [5] D. Grois, D. Marpe, T. Nguyen, and O. Hadar, "Comparative assessment of H.265/MPEG-HEVC, VP9, and H.264/MPEG-AVC encoders for low-delay video applications," in *Proc. SPIE*, vol. 9217, 2014, pp. 92 170Q–92 170Q–10.
- [6] M. Řeřábek and T. Ebrahimi, "Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments," in *Proc. SPIE*, vol. 9217, 2014, pp. 92 170U–92 170U–13.
- [7] B. Haskell, A. Puri, and A. Netravali, *Digital Video: An Introduction to MPEG-2: An Introduction to MPEG-2*, ser. Digital multimedia standards series. Springer, 1997. [Online]. Available: <http://books.google.ch/books?id=4qee0Fv8LkIC>
- [8] D. Le Gall, "Mpeg: A video compression standard for multimedia applications," *Commun. ACM*, vol. 34, no. 4, pp. 46–58, Apr. 1991. [Online]. Available: <http://doi.acm.org/10.1145/103085.103090>
- [9] "HEVC reference software," Online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.
- [10] "Webm™: an open web media project," Online: <http://www.webmproject.org/vp9/>.
- [11] "Projects from VideoLAN™, x264 software library and application," Online: <http://www.videolan.org/developers/x264.html>.
- [12] "HEVC reference software default low delay configuration file." Online: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/trunk/cfg/encoder_lowdelay_main.cfg.
- [13] F. D. Simone, L. Goldmann, J.-S. Lee, and T. Ebrahimi, "Towards high efficiency video coding: Subjective evaluation of potential coding technologies," *Journal of Visual Communication and Image Representation*, vol. 22, no. 8, pp. 734 – 748, 2011, emerging Techniques for High Performance Video Coding.
- [14] "Derf's dataset," Online: <https://media.xiph.org/video/derf/>.
- [15] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, January 2012.
- [16] C. Keimel, J. Habigt, C. Horsch, and K. Diepold, "Qualitycrowd - a framework for crowd-based quality evaluation," in *Picture Coding Symposium (PCS)*, 2012. IEEE, 2012, pp. 245–248.
- [17] G. W. Snedecor and W. G. Cochran, *Statistical Methods*. Iowa State University Press, 1989.
- [18] ITU-R, "P.910: Subjective video quality assessment methods for multimedia applications," ITU-R, Tech. Rep. P.910, 1992.
- [19] G. Bjontegaard, "Improvements of the BD-PSNR model," ITU-T SG16/Q6, Berlin, Germany, Tech. Rep. VCEG-A111, July 2008.
- [20] P. Hanhart and T. Ebrahimi, "Calculation of average coding efficiency based on subjective quality scores," *Journal of Visual Communication and Image Representation*, vol. 25, no. 3, pp. 555–564, 2014.