

# Comparison of Compression Efficiency between HEVC/H.265, VP9 and AV1 based on Subjective Quality Assessments

Pinar Akyazi and Touradj Ebrahimi

Multimedia Signal Processing Group (MMSPG)

Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

Email: {pinar.akyazi, touradj.ebrahimi}@epfl.ch

**Abstract**—The growing requirements for broadcasting and streaming of high quality video continue to trigger demands for codecs with higher compression efficiency. AV1 is the most recent open and royalty free video coding specification developed by Alliance for Open Media (AOMedia) with a declared ambition of becoming the most popular next generation video coding standard. Primary alternatives to AV1 are the VP9 and the HEVC/H.265 which are currently among the most popular and widespread video codecs used in applications. VP9 is also a royalty free and open specification similar to AV1, while HEVC/H.265 requires specific licensing terms for its use in commercial products and services. In this paper, we compare AV1 to VP9 and HEVC/H.265 from rate distortion point of view in a broadcasting use case scenario. Mutual comparison is performed by means of subjective evaluations carried out in a controlled environment using HD video content with typical bitrates ranging from low to high, corresponding to very low up to completely transparent quality. We then proceed with an in-depth analysis of advantages and drawbacks of each codec for specific types of content and compare the subjective comparisons and conclusions to those obtained by others in the state of the art as well to those measured by means of objective metrics such as PSNR.

**Index Terms**—Subjective quality assessment, video compression, AV1, VP9, HEVC, H.265, WebM, AOMedia.

## I. INTRODUCTION

Video encoding strategies are of major importance within research communities and broadcasting experts, and keep evolving progressively due to dramatically increasing trend in video consumption. Meeting the users' demand of higher resolution and quality is among the main efforts of service providers to improve current video acquisition and display technologies, as well as coding and transmission standards. A critical requirement persists in the video coding domain to maintain higher compression efficiency and lower costs.

The most recent coding specification called AV1, developed by the Alliance for Open Media (AOMedia), aims to be the state of the art format for video coding. AV1 is a royalty free and open codec, similar to its predecessor VP9 developed by Google prior to the emergence of AOMedia. Main competitor of the aforementioned codecs is the High Efficiency Video Coding, often referred to as HEVC/H.265 [1], which has been developed by the Joint Collaborative Team on Video Coding (JCT-VC) established by ITU-T Video Coding Experts Group (ITU-T SG16/Q.6) and ISO/IEC Moving Picture

Experts Group (ISO/IEC JTC1/SC29/WG11). HEVC/H.265 still remains to become as popular and widespread as its predecessor AVC/H.264.

VP9 was released in a similar time frame to HEVC/H.265 with the claim of being a superior alternative, thereby attracting research and professional communities to compare their performance. Most studies relied on objective metrics to measure and compare the compression efficiency of HEVC and VP9, yet conflicting results have been reported. In [2] VP9 was claimed to be inferior to HEVC/H.265 by different statistics, whereas [3] revealed a similar performance between the two. To benchmark the performance of HEVC/H.265 and VP9, both objective and subjective assessments have been carried out in [4] and [5], with bitrates ranging from low to high corresponding to very low up to transparent video quality. In [4] a clear advantage of HEVC/H.265 over VP9 with almost twice as much bitrate savings was reported when using the low delay configuration on high definition (HD) and ultra high definition (UHD) video content. Similar results were obtained in [5] using UHD video content assuming a broadcasting configuration, noting a competitive performance of VP9 for some synthetic video.

Despite its higher bitrate requirements to achieve similar quality when compared to HEVC/H.265, VP9's popularity continued to rise and further studies were conducted to close the compression efficiency gap between VP9 and HEVC/H.265. As a prominent result of collaborative efforts, AV1 has been developed by AOMedia to replace VP9 with a specification projected to be finalized by January 2018. Although the performance of AV1 cannot be tested before the official release of its final specification, it already holds promising advantage thanks to the AOMedia members intention to deploy AV1 encoded contents in browsers, mobile devices and smart TVs. To the authors' best knowledge, no scientific and rigorous comparisons of AV1 with respect to HEVC/H.265 and VP9 have been performed and published in the literature.

This paper presents a comparison of compression efficiency between HEVC/H.265, VP9 and AV1 by means of subjective assessments assuming a broadcasting scenario. HD video content is presented to different subjects in a controlled environment. A total of 30 subjects took part in the experiments,

TABLE I  
SELECTED PARAMETERS AND SETTINGS FOR THE AV1, HEVC AND VP9 CODECS.

Codec	Version	Parameters
AV1	0.1.0-7203-g0687d3e	-i420 -w <SourceWidth> -h <SourceHeight> -p 1 -aq-mode=0 -fps=<FrameRate> -good -cpu-used=1 -end-usage=3 -codec=av1 -psnr -v -test-decode=warn -cq-level=<QP> -kf-min-dist=<IntraPeriod> -kf-max-dist=<IntraPeriod> -auto-alt-ref=1 -arnr-maxframes=7 -arnr-strength=5 -threads=0 -tile-columns=0 -frame-parallel=0 -o <BitStreamFile> <InputFile>
HEVC	HM 16.17	Default main random access profile with GOP size 8, coding order 8 4 2 1 3 6 5 7, motion search range = 64
VP9	v1.6.1-1481-ge20ca4f	-i420 -w <SourceWidth> -h <SourceHeight> -p 1 -aq-mode=0 -fps=<FrameRate> -good -cpu-used=1 -end-usage=3 -codec=vp9 -psnr -v -test-decode=warn -cq-level=<QP> -kf-min-dist=<IntraPeriod> -kf-max-dist=<IntraPeriod> -auto-alt-ref=1 -arnr-maxframes=7 -arnr-strength=5 -threads=0 -tile-columns=0 -frame-parallel=0 -o <BitStreamFile> <InputFile>

evaluating 5 distinct video contents which were compressed to four different bitrates using HEVC/H.265, VP9 and AV1. To achieve a fair comparison, exclusive encoder enhancement techniques were disabled and fixed quantization parameters (QP) were used.

The paper is organized as follows. In the next section we present the configuration of codecs and discuss the selected parameters. Section III describes the selected video sequences, test environment and experimental methodology in detail. Statistical analysis of the subjective test results and objective measures are presented in Section IV. Finally, the paper is concluded in Section V.

## II. CONFIGURATION OF SELECTED CODECS

The latest version of HM reference software [6] was selected for HEVC/H.265. For VP9, the latest encoder within the WebM Project was employed [7]. Finally, for AV1 the version available as of the 1st of January 2018 was used [8]. Selected versions of each codec used in the experiments are reported in Table I. To ensure fair comparison, the fixed QP is separately selected for each codec. This removes rate control adaptation between video frames, yielding to an unbiased quality evaluation.

The Random Access (RA) configuration was selected for HM assuming a broadcasting scenario. The Group of Pictures (GOP) size was set to 8 pictures. Hierarchical B-frames were used with a QP increase of 1 between each temporal level. The coding order was set to 0 8 4 2 1 3 6 5 7. The configuration parameters were selected according to [5] to avoid assigning priority to HM. More details on the encoder options can be found on Table I.

For VP9 the configurations were selected based on [5] due to the lack of official specification of the encoder and decoder parameters. The preferred options have been selected based on recommendations of the WebM Project developers. Similarly, in choosing the parameters for AV1, leading developers from AOMedia were consulted. The adaptive quality mode was disabled for both VP9 and AV1 since this feature is not present in HM. To fix the QP, the available fixed quality mode *-end-usage=3* was preferred according to expert recommendations as described in [9]. More details on VP9 and AV1 encoding parameters used can be found in Table I.

## III. SUBJECTIVE EVALUATION

This section describes the dataset used for subjective assessment, as well as the testing environment, equipment and methodology. The processing of collected scores is also explained here.

### A. Dataset preparation and description

Six distinct video sequences with different spatial and temporal characteristics were used for tests. Five of the contents (one training) were selected from video available in SJTU dataset [10] and downsampled to HD resolution using the Lanczos scaling filter of FFmpeg software version 3.4.1 [11]. One synthetic content from Sintel, a computer generated movie [12], was also included. Figure 1 depicts the spatial (SI) versus temporal (TI) index on the luminance component of each content, where the indexes SI and TI are determined from the analysis of spatial and temporal perceptual information of each content [13]. For a comprehensive evaluation, the selected contents have varying SI and TI measures from relatively small to relatively large values, as indicated on Figure 1. 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. The representative sample frame of each content is shown in Figure 2.

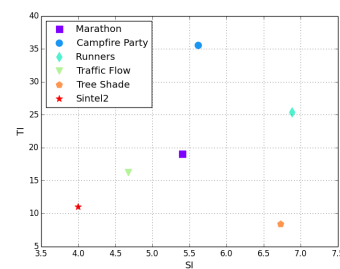


Fig. 1. Spatial information (SI) versus temporal information (TI) indexes of selected contents

Each video file was encoded with all three evaluated codecs using the parameters previously described, at four bitrates ranging from low to high corresponding to very low up to transparent quality. Targeted bitrates were selected separately for different contents due to their varying spatial and temporal

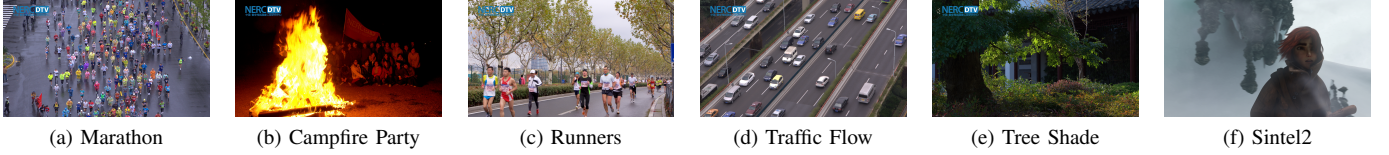


Fig. 2. Sample frames of contents used in subjective assessment test. Contents (a)-(e) have a resolution of  $1920 \times 1080$  and frame rate of 30fps. Synthetic content (f) has a resolution of  $1920 \times 872$  and frame rate of 24fps.

characteristics, as well as the presence of natural and synthetic contents at the same time. An expert screening session was conducted to determine the lower and upper QP bounds covering the full quality scale for each content. The complete final sets of targeted ( $R1' - R4'$ ) and actual ( $R1 - R4$ ) bitrates as well as the corresponding QPs are listed in Table II for each codec.

### B. Test methodology

The Double Stimulus Impairment Scale (DSIS) Variant I [14] was chosen to compare the quality of different coding schemes. In this test, the stimulus under assessment and the reference are presented simultaneously to the subject. The subject is asked to rate the degree of annoyance of the visual distortions in the stimulus under assessment with respect to the reference. The degree of annoyance is divided into five different levels labeled as *Very annoying*, *Annoying*, *Slightly annoying*, *Perceptible but not annoying* and *Imperceptible*, corresponding to a quality scale ranging from 1 to 5, respectively.

Each stimulus under assessment and the respective reference were cropped to half their original width prior to being displayed side by side simultaneously. The mid-section of each stimuli was cropped using FFmpeg version 3.4.1 [11]. The stimuli were randomized such that the same content is never presented consecutively. Two dummy sequences were included in each test, about which the subjects were not informed. To display the video sequences, a 30 inch Eizo ColorEdge CG301W monitor with a resolution of  $2560 \times 1600$  was used. The monitor was calibrated using an EyeOne Display2 color calibration device according to the guidelines described in [15]. Same guidelines were followed to set up the controlled environment for viewing with a mid gray level background behind both the displayed contents and the monitor.

A total of 30 consenting subjects participated in the study. Each subject evaluated all test stimuli. Half of the subjects were presented with the reference content on the right of the display, whereas for the remaining half the reference content was displayed on the left. Before the experiments, written instructions were provided and tasks were explained verbally to each subject. Three training samples from the Marathon sequence [10] were displayed in the beginning of each session, representative of *Imperceptible*, *Slightly annoying* and *Very annoying* levels of impairment scale in order to familiarize the subjects with the assessment procedure.

### C. Data Processing

The rate distortion curves of luma and chroma components were used for objective comparison of encoders. The com-

bined  $PSNR_{YUV}$  value for each video sequence was calculated as an average over measurement across all frames. For the evaluation of perceived quality, a standard outlier detection was performed on all sets of raw scores to remove subjects whose ratings deviated strongly from others [14]. None of the subjects were identified as outliers in our experiments. The mean opinion score (MOS) and 95% confidence intervals (CIs) assuming a Student's  $t$ -distribution of the scores were computed for each test condition [16]. To determine and compare the differences among MOS obtained for different codecs and bitrates, a one-sided Welch test at 5% significance level was performed on the scores.

## IV. RESULTS AND DISCUSSION

Figure 3 shows the rate-distortion (R-D) curves based on PSNR measurements for all tested contents on the left column. According to objective metrics, both AV1 and HEVC/H.265 outperform VP9 between 0.5 to 2.0 dB depending on the content and bitrate. For sequence Campfire Party, AV1 outperforms HEVC/H.265 and VP9, whereas for the synthetic content Sintel2 the gain of HEVC/H.265 is the highest.

The middle column of Figure 3 shows the R-D curves based on subjective ratings, i.e. MOS and CI for all tested contents. Relatively small confidence intervals indicate reliable subjective results with small variations across subjects. Moreover, the objective and subjective measurements show a similar trend for the tested contents. At very high bitrates, there is no sufficient statistical evidence showing differences in performance for any of the codecs. While CIs are overlapping only for the highest and lowest bitrates tested, at intermediate bitrates VP9 is largely outperformed both by AV1 and HEVC/H.265. Qualitative results show that AV1 has achieved a good level of improvement over its predecessor VP9. On the other hand, CIs of AV1 and HEVC/H.265 are overlapping for all bitrates, for all contents. This indicates a competitive perceived quality between the two codecs. Present-day applications require fully or close to transparent quality coded video, therefore a performance analysis of the three codecs for operating points that are close to transparent quality in terms of perceived quality, i.e. a MOS level of 4 or higher, provides better insight in terms of pairwise comparison. All codecs achieve a perceived quality close to transparent, i.e.  $MOS \geq 4$  for high bitrates.

The last column of Figure 3 presents a multiple comparison test for all possible combinations of codecs and bitrates ( $R1$  through  $R4$ ) for each content separately. In each plot, the color of the squares indicates the result of the significance test between the mean opinion scores related to the two

TABLE II  
TARGETED BITRATES  $\mathbf{R}'[\text{KBPS}]$  AND ACTUAL BITRATES  $\mathbf{R}[\text{KBPS}]$  WITH CORRESPONDING QP VALUES FOR EACH CODEC, FOR EACH CONTENT.

Sequence	R1'	R1						R2'	R2						R3'	R3						R4'	R4					
		AV1		VP9		HEVC			AV1		VP9		HEVC			AV1		VP9		HEVC			AV1		VP9		HEVC	
		[kbps]	QP	[kbps]	QP	[kbps]	QP		[kbps]	QP	[kbps]	QP	[kbps]	QP		[kbps]	QP	[kbps]	QP	[kbps]	QP		[kbps]	QP	[kbps]	QP	[kbps]	QP
Campfire Party	1200	63	1005	62	1325	40	1323	2500	55	2547	55	2678	34	2625	5000	47	4705	47	5064	29	5187	10000	35	11527	36	11596	24	10705
Runners	1200	63	1296	63	1187	40	1131	2500	55	2857	56	2525	34	2617	5000	47	5233	47	5204	29	5270	10000	35	11982	36	11688	24	10637
Sintel2	200	59	174	58	202	33	189	300	51	293	51	312	30	276	500	43	502	43	535	26	467	2000	23	2154	24	2134	16	2079
Traffic Flow	350	63	338	63	373	38	348	700	55	663	56	666	32	716	2000	39	2281	40	2128	24	2229	8000	23	8195	24	8244	17	7751
Tree Shade	1200	59	1163	58	1173	35	1248	2500	51	2319	50	2436	30	2438	5000	43	4319	42	4708	26	4975	1000	31	10068	31	10859	21	10720

test conditions in the corresponding row and column, i.e. condition A and condition B. A white (black) square indicates the MOS corresponding to test condition A is statistically significantly better (worse) than the MOS corresponding to the test condition B, whereas a gray square indicates no statistically significant distinction between the MOS values of the two test conditions. For all contents except Tree Shade, the performance of AV1 is significantly better than VP9 except for the highest bitrate R4 for which no significant distinction is detected. In the case of content Tree Shade, which has the smallest TI value, no statistically significant difference between AV1 and VP9 is observed at any bitrate value tested. For the same content, the performance of HEVC/H.265 and VP9 are also not significantly different, except for the second lowest bitrate R2 where HEVC/H.265 performs better. For contents Runners and Tree Shade, which have relatively high SI values, there is no statistical difference between the performance of AV1 and HEVC/H.265 at similar bitrates. Meanwhile at the lowest bitrate R1, HEVC/H.265 is significantly outperforming AV1 for Traffic Flow and Sintel2, which are the two contents with relatively low SI and TI values. A corresponding trend in favor of HEVC/H.265 is evident for higher bitrates for the same contents, where HEVC/H.265 performs either statistically similar or better than AV1, but the contrary cannot be observed. The only instance where AV1 performs significantly better than HEVC/H.265 is at bitrate level R2 for the content Campfire Party with the highest TI value.

Figure 4 shows pairwise comparison for the actual bitrates (R1-R4) corresponding to targeted bitrates (R1'-R4') indicated in Table II. For the lowest bitrate, HEVC/H.265 outperforms AV1 by two contents whereas AV1 fails to outperform HEVC/H.265. Nevertheless, AV1 and HEVC/H.265 are both superior to VP9 for 3 contents at the lowest bitrates tested. As the bitrate is increased to intermediate levels AV1 and HEVC/H.265 outperform VP9 for 4 contents out of 5. For intermediate bitrates, HEVC/H.265 outperforms AV1 for two contents in total while AV1 manages to outperform HEVC/H.265 for only one content. None of the codecs outperforms the other at the highest bitrate.

The bitrate reduction of one codec over another for a similar quality is estimated using the Bjøntegaard Delta PSNR (BD-PSNR) [17]. Using the combined  $\text{PSNR}_{\text{YUV}}$  measures and the corresponding bitrates as inputs, BD-PSNR model determines the average bitrate difference between a pair of codecs. In Table III the percentage BD-PSNR results are presented. AV1

provides an average bitrate reduction of 28.1% relative to VP9 over all contents, however on average 1.9% more bitrate is required for AV1 to achieve a similar objective quality with HEVC.

TABLE III  
COMPARISON OF INVESTIGATED CODECS IN TERMS OF BITRATE REDUCTION FOR SIMILAR  $\text{PSNR}_{\text{YUV}}$  AND MOS. NEGATIVE VALUES INDICATE ACTUAL BITRATE REDUCTION.

Content	AV1 vs HEVC/H.265		AV1 vs VP9		HEVC/H.265 vs VP9	
	BD-PSNR	BD-MOS	BD-PSNR	BD-MOS	BD-PSNR	BD-MOS
Campfire Party	-23.2%	-19.0%	-30.9%	-35.4%	-9.3%	-19.6%
Runners	-2.6%	1.5%	-24.1%	-40.8%	-22.2%	-41.5%
Traffic Flow	5.7%	4.9%	-35.2%	-47.7%	-38.5%	-55.6%
Tree Shade	-5.8%	-9.3%	-26.6%	-21.8%	-22.8%	-14.7%
Sintel2	35.6%	38.0%	-23.6%	-26.7%	-43.7%	-46.7%
Average	<b>1.9%</b>	<b>3.2%</b>	<b>-28.1%</b>	<b>-34.5%</b>	<b>-27.3%</b>	<b>-35.6%</b>

In order to determine the average bitrate difference between different codecs, a more realistic approach is presented in [18] taking the MOS values into account rather than the PSNR measures. Using the model in [18] the average bitrate savings for a similar perceived quality are presented also in Table III. The average bitrate reduction of AV1 over VP9 is 34.5% based on subjective ratings, while AV1 is slightly inferior to HEVC/H.265 in terms of average bitrate savings as 3.2% more bitrate is required for AV1 to reach similar perceived quality to that of HEVC/H.265. However, this small difference could be more a function of the contents used than an actual difference, generally speaking. Compression efficiency of AV1 and HEVC/H.265 over VP9 based on objective measures is similar to the gain observed from subjective ratings. Bit rate savings of HEVC/H.265 over AV1, however, is emphasized at a greater extent in average, in terms of subjective measures compared to objective metrics. For most contents, the perceived quality appears to be better than objective measures for similar bitrates when comparing AV1 and HEVC/H.265 to VP9. The opposite effect is observed for all contents except Tree Shade within the comparison AV1 vs HEVC/H.265, predicting higher objective gain than perceived quality.

## V. CONCLUSION

In this paper, a detailed description of the subjective quality evaluation tests performed to benchmark the compression efficiency of AV1, HEVC/H.265 and VP9 codecs is presented, assuming a broadcasting scenario. Pairwise and mutual comparisons have been carried out between the codecs, using five different HD video contents with varying spatio-temporal



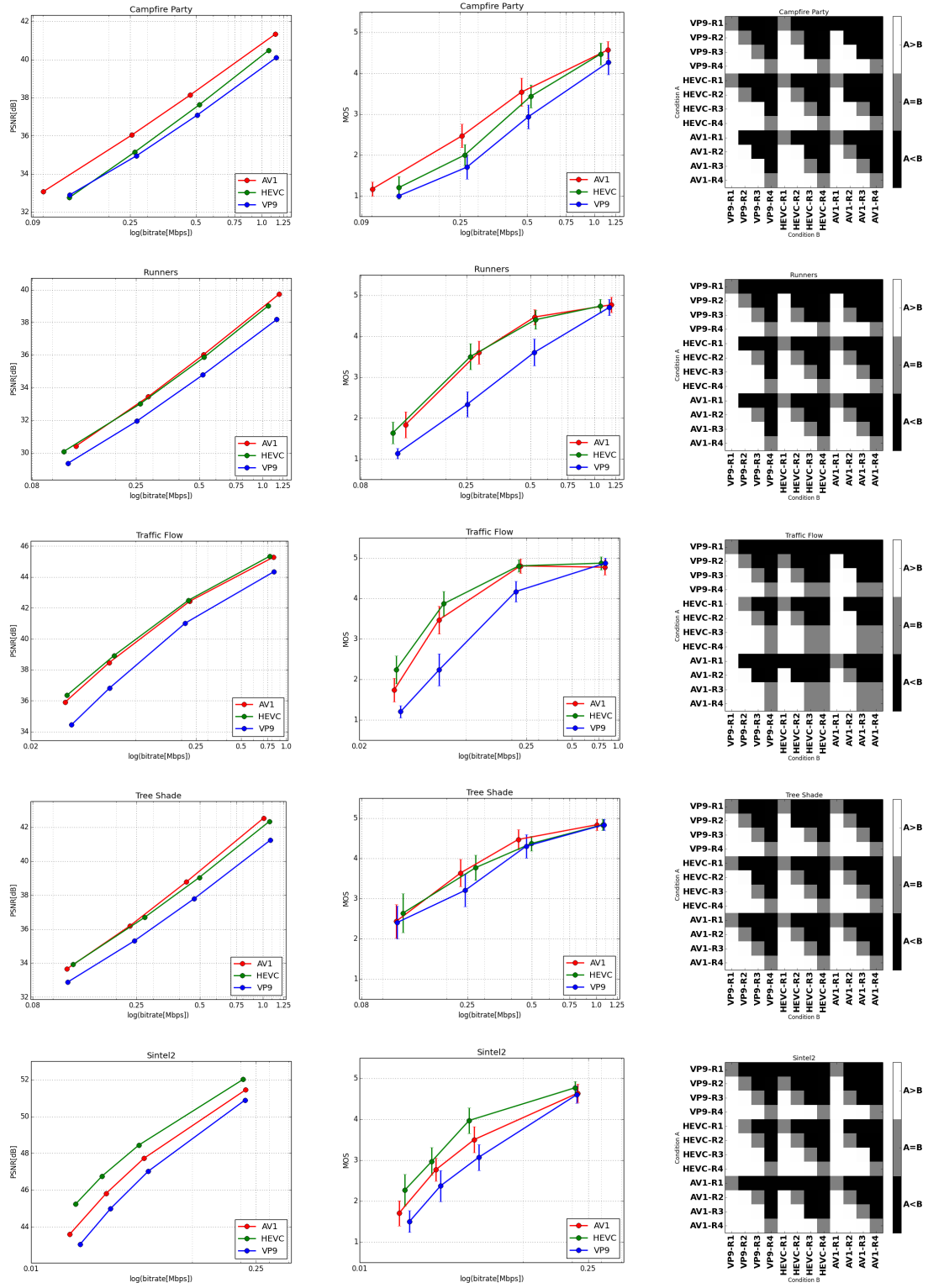


Fig. 3. PSNR R-D curves (left column) and subjective ratings showing MOS and CIs (middle column) for each bitrate and for each content (along rows). Multiple comparison test results for all possible combinations of codecs and bitrates (R1-R4) for each content (along rows). For the multiple comparison result, the color of each square shows the result of the significance test between MOS related to the two test conditions depicted in the corresponding row and column, A and B, respectively.

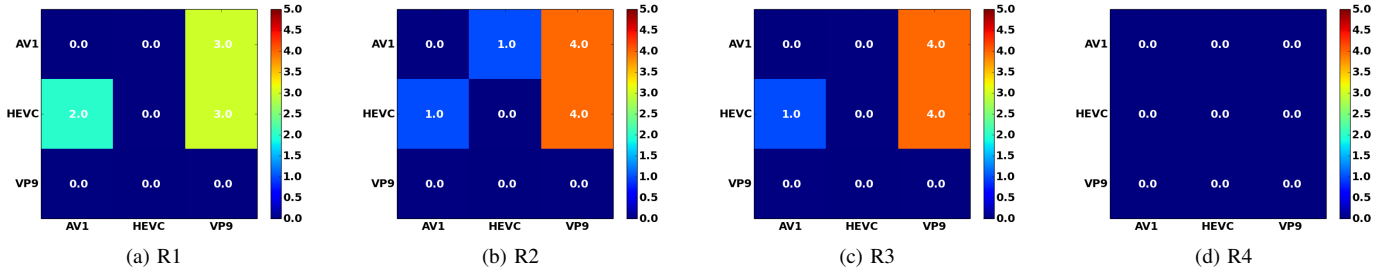


Fig. 4. Pairwise comparison results for bitrates (R1-R4). Each cell contains the number of contents for which the codec indicated in the row performs better than the codec indicated in the column, according to subjective ratings.

characteristics. Each content was compressed at four bitrates ranging from very low to high, and presented to the subjects in a controlled environment, in random order. Subjective quality scores of 30 subjects related to each content, codec and quality parameter have been collected and analyzed. The obtained results maintain high consistency, allowing an accurate comparison of the performance of investigated codecs.

The test results clearly highlighted a substantial improvement of compression efficiency for AV1 compared to VP9. Except for the highest bitrate tested where no significant difference was observed, the performance of AV1 was found statistically better than VP9. The same trend was detected also for HEVC/H.265 as it provided a significant gain in bitrate savings compared to VP9, except for the highest bitrate where a significant difference was not present. On average AV1 and HEVC/H.265 are shown to have 28.1% and 27.3% more bitrate savings than VP9 based on objective measures. These ratios increase to 34.5% and 35.6%, respectively, based on subjective quality.

The compression efficiency of AV1 was slightly below HEVC/H.265 on average over all contents tested, with 1.9% and 3.2% more bitrate requirements based on objective and subjective metrics, respectively. These small values, however, remain quite insignificant and highly dependent on the contents used in tests.

According to leading developers, AV1 is expected to deliver a 30% improvement over VP9 and HEVC/H.265. Although the claim has been verified for AV1's compression efficiency against VP9, a much less average difference was observed between the performances of AV1 and HEVC, with HEVC/H.265 very slightly leading the performance gain, although the latter conclusion is influenced by the contents used in tests.

The presented study focused on HD broadcasting scenarios with a limited set of selected video contents. Further quality assessment tests of the codecs should include Internet streaming scenarios and span a much wider set of contents with resolutions up to 4K. This more complex approach will be investigated in the future, following the expected code freeze in AV1.

#### ACKNOWLEDGMENT

This paper reports a research performed under the framework of project Digital Eye: Deep Learning Video Quality

Assessment Technology, funded by The Swiss Commission for Technology and Innovation (CTI) under the grant 27403.1 PFES-ES.

#### REFERENCES

- [1] G. J. Sullivan, J. 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*, vol. 22, no. 12, pp. 1649–1668, 2012.
- [2] D. Grois, D. Marpe, A. Mulayoff, B. Itzhaky, and O. Hadar, "Performance comparison of h. 265/mpeg-hevc, vp9, and h. 264/mpeg-avc encoders," in *Picture Coding Symposium (PCS), 2013*. IEEE, 2013, pp. 394–397.
- [3] 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 *Picture Coding Symposium (PCS), 2013*. IEEE, 2013, pp. 390–393.
- [4] M. Rerabek, P. Hanhart, P. Korshunov, and T. Ebrahimi, "Quality evaluation of hevc/h. 265 and vp9 video compression in real-time applications," in *Quality of Multimedia Experience (QoMEX), 2015 Seventh International Workshop on*. IEEE, 2015, pp. 1–6.
- [5] M. Rerabek and T. Ebrahimi, "Comparison of compression efficiency between hevc/h. 265 and vp9 based on subjective assessments," in *Applications Of Digital Image Processing Xcvi*, vol. 9217, no. EPFL-CONF-200925. Spie-Int Soc Optical Engineering, 2014.
- [6] HEVC reference software. [Online]. Available: <https://hevc.hhi.fraunhofer.de/trac/hevc/browser/trunk>
- [7] The WebM Project. Vp9 video codec. [Online]. Available: <https://www.webmproject.org/vp9/>
- [8] Alliance for Open Media. [Online]. Available: <https://aomedia.googlesource.com/aom/>
- [9] P. Hanhart, M. Rerabek, F. De Simone, and T. Ebrahimi, "Subjective quality evaluation of the upcoming hevc video compression standard," in *Proceedings of SPIE*, vol. 8499, no. EPFL-CONF-180494, 2012.
- [10] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, "The sjtu 4k video sequence dataset," in *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*. IEEE, 2013, pp. 34–35.
- [11] FFmpeg. [Online]. Available: <http://ffmpeg.org>
- [12] Sintel. [Online]. Available: <https://durian.blender.org/>
- [13] ITU-R P.910, "Subjective video quality assessment methods for multimedia applications," 1992.
- [14] ITU-R BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," January 2012.
- [15] ITU-R BT.2022, "General viewing conditions for subjective assessment of quality of sdtv and hdtv television pictures on flat panel displays," August 2012.
- [16] F. De 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.
- [17] G. Bjontegaard, "Improvements of the bd-psnr model," in *ITU-T 16/Q6, 35th VCEG Meeting, July 2008*, 2008.
- [18] 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.