AN OVERVIEW OF RECENT VIDEO CODING DEVELOPMENTS IN MPEG AND AOMEDIA

Andre Seixas Dias[1], Saverio Blasi[1],
Fiona Rivera[2], Ebroul Izquierdo[2] and Marta Mrak[1]

[1]BBC, Research & Development Department, UK
[2]Queen Mary University of London, Multimedia and Vision Group, UK

ABSTRACT

Video compression technologies play a key role in the distribution of video content in broadcasting. Many techniques have been proposed to improve the coding efficiency provided by the most recent video compression standard, High Efficiency Video Coding (HEVC). This includes the latest tools being investigated within the Joint Video Exploration Team (JVET) and the tools provided by the royalty-free AV1 codec developed by the Alliance for Open Media. This paper analyses the overall compression capabilities of these two emerging technologies with respect to HEVC, including objective and subjective performance evaluation and the associated encoding and decoding complexities.

INTRODUCTION

Due to the increasing consumption of video content with higher resolutions, the need for more efficient video compression techniques is growing. The first version of the High Efficiency Video Coding (HEVC) standard (1), jointly developed by the ITU-T VCEG and the ISO MPEG, was finalised in 2013. A wide range of products and services support HEVC for video encoding/decoding, especially for Ultra High Definition (UHD) content, where HEVC can provide around 50% bitrate savings for the same subjective quality as its predecessor H.264/AVC (2). However, HEVC is not yet as widely adopted as H.264/AVC.

As an alternative to HEVC, the Alliance for Open Media (AOMedia) initiated the development of the AOMedia Video 1 (AV1) specification (3). AOMedia was founded in 2015 as a consortium of partners from the semiconductor industry, video on demand providers and web browser developers, specifically to create an open, royalty-free video coding specification. AV1 was built using Google's VP9 specification (4) as a base, and similarly to VP9, AV1 follows the typical hybrid block-based approach also used in the standards from the MPEG/ITU family. The AV1 specification was primarily finalised at the end of March 2018, with some further minor details defined shortly after.

At the same time as the above, work on video compression technologies beyond the capabilities of HEVC continued by the MPEG/ITU, with the creation of the Joint Video Exploration Team (JVET) on future video coding in October 2015. Many new coding tools have been proposed in the context of JVET, which eventually led to a Call for Proposals on video coding technologies with video compression capabilities beyond HEVC. The reference software used in the exploration phase of JVET, called Joint Exploration Model (JEM) (5), was leveraged as the base for the majority of responses to the call. Results included responses demonstrating compression efficiency gains of around 40% or more with respect to HEVC (6). This initiated the work by the Joint Video Experts Team (JVET) on the development of a new video coding standard, to be known as Versatile Video Coding (VVC). Even though JEM was created as an exploration software model, it provides clear evidence of video coding technology that can significantly outperform the capabilities of HEVC. Many of its tools are therefore likely to be included in the final version of the VVC specification, expected to be completed by the end of 2020.

Given these new trends, this paper provides a performance comparison between HEVC, AV1 and the JEM software, based on both objective and subjective tests. This analysis provides indicators on how the achieved objective bitrate savings translate into subjective quality advantages. Moreover, the associated encoding and decoding complexity of both technologies is also analysed with respect to the HEVC reference software.

EVALUATION METHODOLOGY

The compression performance of the video coding technologies analysed in this paper is evaluated in terms of rate-distortion trade-off. For each technology, a set of operation points corresponding to different bitrates and respective output qualities are considered. The compressed bitrate and respective decoded video quality are measured for each of the selected operation points. Results of such tests can be visualised in so-called Rate-Distortion (RD) curves, as shown in the example of Figure 1. In the case of Figure 1, it is clear that the performance of technology B is superior to technology A, since the red curve is above the blue curve. More specifically, for the same bitrate, B is able to provide higher decoded video quality, or from a different perspective, B is able to achieve the same quality as A using lower bitrates. In Figure 1, quality was measured using the Mean Opinion Score (MOS), which represents the average quality score given by subjects taking part in subjective tests. The objective distortion between decoded and original video sequences can alternatively be evaluated using PSNR.

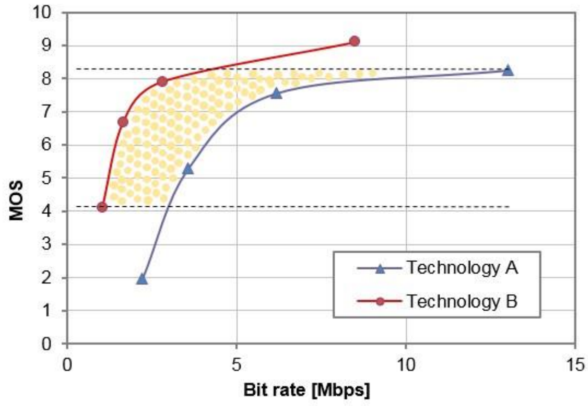
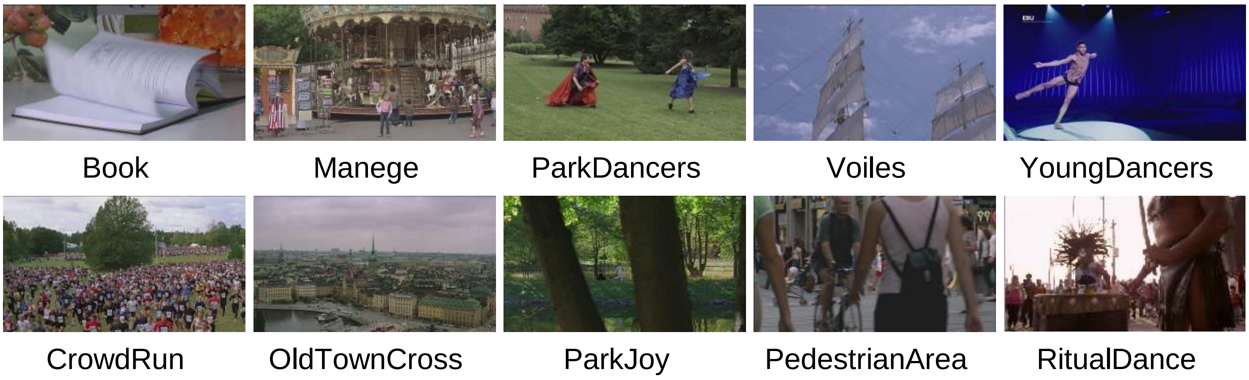


Figure 1 – Example quality interval used in BD-rate computation.

In addition to RD curves, the Bjøntegaard Delta (BD) bitrate (7), denoted as BD-rate, can be used to quantify the average bitrate difference (in percentage) between two techniques for the same video quality, measured objectively with PSNR or subjectively with MOS. BDrates are obtained by computing the difference between the area below the inverse of the interpolated RD curves obtained by the two technologies being compared (i.e. anchor and test), as highlighted in Figure 1. Negative BD-rate values represent compression gains, while positive values correspond to compression losses.

TEST SETTINGS

The video content selected to test the performance of the analysed video compression technologies comprises 5 HD (i.e. a spatial resolution of 1920 × 1080) and 5 UHD (i.e. spatial resolution of 3840 × 2160) sample test sequences with typical broadcasting content, as illustrated in Figure 2. These were selected ensuring that none were used during the development of any of the evaluated video coding technologies. All sequences are at 50 fps, with the exceptions of Manege, Voiles and RitualDance at 60 fps and PedestrianArea at 25 fps. All sequences have a length of 10 seconds, 4:2:0 chroma subsampling and bit depth of 8 bits, with the exception of RitualDance with 10 bits.


Book Manege ParkDancers Voiles YoungDancers

CrowdRun OldTownCross ParkJoy PedestrianArea RitualDance

Figure 2 – UHD (top row) and HD (bottom row) test sequences.

The evaluated coding technologies were each used to encode all test sequences with 4 Quantisation Parameters (QPs), defining the operation points for the rate-distortion assessment. Both subjective and objective rate-distortion evaluations were performed using MOS and PSNR, respectively. For the subjective quality assessment, the ITU-R recommendation BT500-13 (8) was used to define the structure and set-up of the tests. The Double Stimulus Impairment Scale (DSIS) (8) was adopted, where for each test session a series of Basic Test Cells (BTC) were shown to the subjects. In each BTC, the original video clip is first displayed, followed by the encoded version of the same clip, as shown in Figure 3. A voting period of 6 seconds is given to the subjects, to vote according to the distortions perceived in the compressed video with respect to the original. The subjects vote using a scale ranging from 0 (lowest quality) to 10 (highest quality).

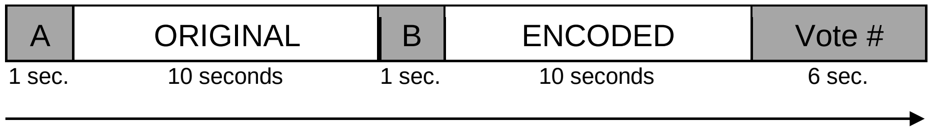


Figure 3 – Structure of the adopted BTC.

The BTCs generated with the 3 technologies were shown to 16 subjects in arbitrary order at a viewing distance of 3 and 1.5 times the height of the active part of the display for HD and UHD content, respectively. Excluding an initial training session comprising 6 BTCs, each subject participated in 4 different test sessions (2 HD and 2 UHD). Each session included an adaptation phase of 3 BTCs (8) and 32 main BTCs. All test subjects were tested for colour blindness and normal or corrected to normal visual acuity.

Encoding configurations

The software codebases used for the technologies under evaluation are shown in Table 1, along with the respective QPs. In the case of JEM and the HEVC reference software, HM (9), the same QPs used during the development of the HEVC standard were selected (10) (11). In the case of AV1, for which the QP range spans from 0 to 63, 4 QPs were selected within the range of the 5 normally used during the development of the AV1 specification.

	Software	QPs
HEVC	HM 16.10	22, 27, 32, 37
AV1	Commit 5f4f73 (9th Jan.18)	25, 36, 47, 58
JEM	JEM 7.1	22, 27, 32, 37

Table 1 – Software versions and QPs used in the tests.

The remaining encoding configurations were similar to those used during the development of each coding technology, in order to test their best performance. In the case of HM and JEM, the Random Access (RA) configuration (10) was adopted. For AV1, the default options of the version used in these tests were used, apart from the following parameters which were modified to allow a fair comparison with HM and JEM:

“--end-usage=q --threads=1 --passes=1 --lag-in-frames=0 --bit-depth=12”.

The parameter “--end-usage=q” was set to force fixed QP encoding according to the QPs in Table 1 and “--threads=1” was used to run the encoder in single-thread mode. The parameters “--passes=1” and “--lag-in-frames=0” were set to run AV1 in single pass mode without the possibility of looking ahead in the video sequence before encoding. Finally, the internal bit depth of the codec was set to 12 as typically used during the AV1 development. Finally, for all encoding technologies, each sequence was split into chunks of one Intra period (approximately 1 second, as defined in the RA configuration for HM and JEM), which allowed each chunk to be independently encoded in parallel. This coding configuration was adopted to reduce the overall time needed to encode with AV1, instead of encoding each 10 second sequence sequentially.

PERFORMANCE RESULTS

The obtained bitrate-MOS plots and respective confidence intervals are shown in Figure 4 and Figure 5. As can be seen, in most cases AV1 and HM provide similar performance for the test conditions defined in this paper. Some exceptions include the sequences OldTownCross (for which HM shows higher performance at higher bitrates), and the sequence ParkDancers (for which AV1 outperforms HM at medium to high bitrates). JEM, achieves higher performance than the other technologies, especially for lower bitrates. This is more evident in sequences like CrowdRun, Book, RitualDance or

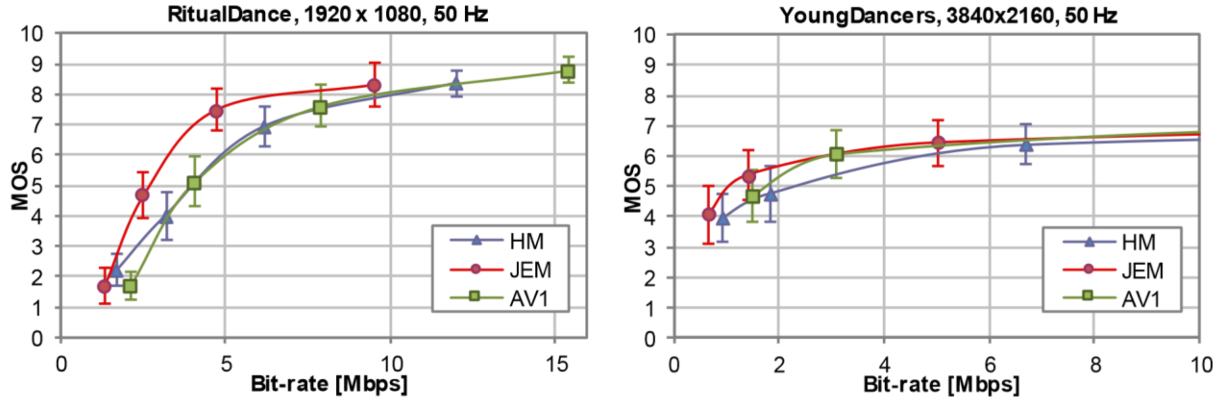
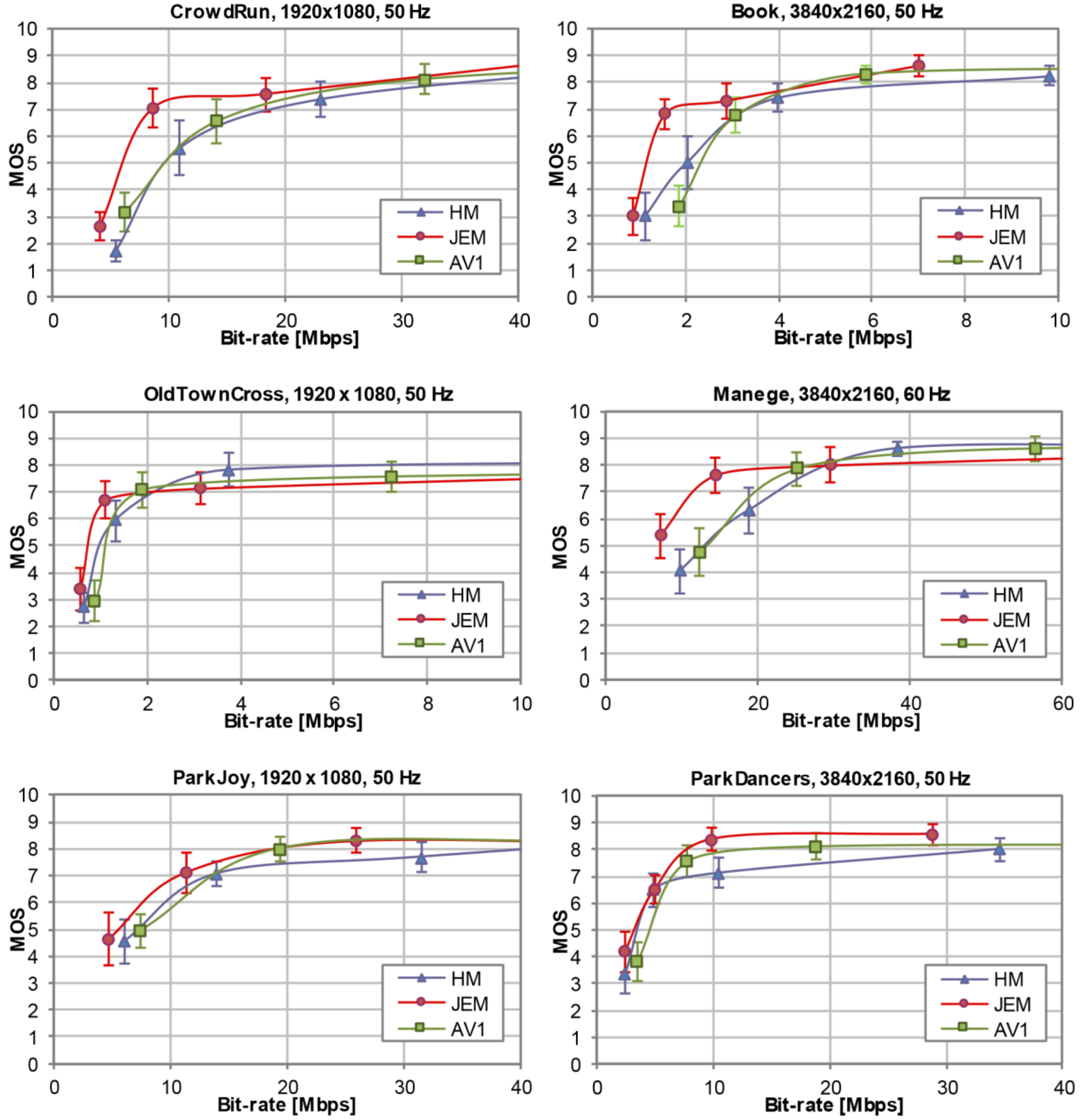


Figure 4 – RD curves obtained from the subjective quality tests.



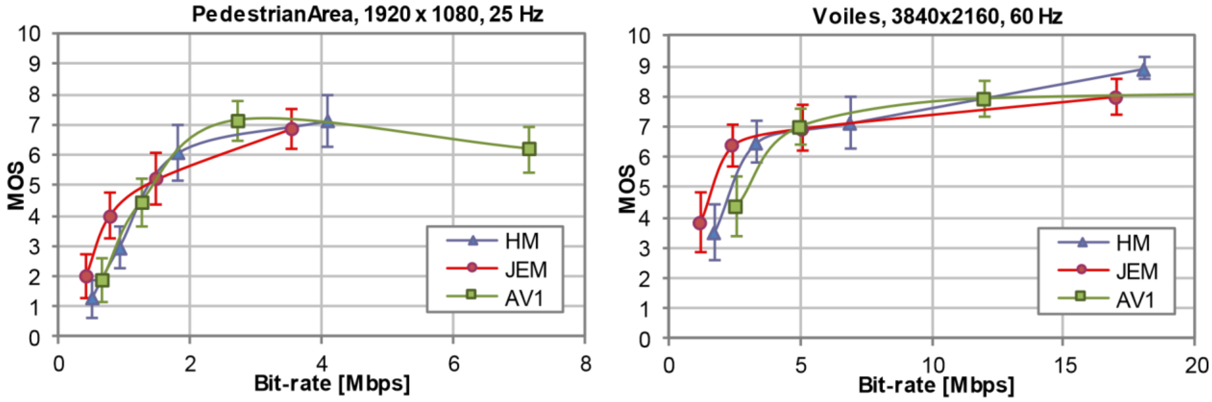


Figure 5 – RD curves obtained from the subjective quality tests.

Manege, even though the confidence intervals of the MOS scores are considerable in some points. For higher bitrates, the differences between the technologies are generally more difficult to evaluate. This is expected since at higher qualities it is more difficult to subjectively distinguish differences in visual quality.

One important aspect to clarify is the limitation on the visualisation of the bitrate axis for some sequences in Figure 4 and Figure 5. In some cases, the bitrate/quality operation points obtained with the lowest QP used, in particular for AV1, generated higher bitrates when compared with HM and JEM. A higher QP should have ideally been used in these cases. Given the very high encoding times associated with AV1 encoding (as detailed further in this section), it was not possible to re-run the encoder to tune these operation points. Nevertheless, these correspond to quality saturation points where subjective differences are difficult to perceive. Moreover, the associated bitrates are not relevant in the context of distribution scenarios, which are the target of the encoder configurations and profiles considered in this paper. Therefore, some of these AV1 test points at very high rates were intentionally left outside the plots in order to improve clarity, focusing on areas more relevant to the performance analysis. All points were however considered in the computation of the BD-rate values shown further in Table 2, both for PSNR and MOS.

Finally, it is worth noting the unexpected behaviour of the AV1 curve for the sequence PedestrianArea and ParkJoy. For these sequences, higher bitrate test points were given a lower average MOS than points at a lower bitrate. Such unexpected behaviour can happen in subjective tests, especially for higher qualities. In the case of PedestrianArea, some of the content in the background is out of focus which may mislead the subjects to vote with a

lower score, even if the compressed sequence is identical to the original.

In order to compare the results of the subjective tests with the objective performance of each compression technology, Table 2 shows the BD-rate values obtained for AV1 and JEM, respectively, using HM as anchor. The BD-rate values are shown in both cases for the PSNR of the Y, Cb and Cr components, along with the MOS BD-rates computed using the curves in Figure 4 and Figure 5.

Table 2 – Objective and subjective BD-rate results for AV1 and JEM with respect to HM.

	AV1 vs HM	JEM vs HM

Sequence	PSNR BD-rate Y	PSNR BD-rate Cb	PSNR BD-rate Cr	MOS BD-rate	PSNR BD-rate Y	PSNR BD-rate Cb	PSNR BD-rate Cr	MOS BD-rate
Book	2%	-15%	-22%	1%	-44%	-61%	-66%	-41%
Manege	0%	-10%	-8%	-9%	-36%	-44%	-41%	-37%
ParkDancers	-1%	-24%	-52%	-5%	-28%	-53%	-52%	-23%
Voiles	29%	-17%	-14%	14%	-39%	-63%	-56%	-19%
YoungDancers2	-6%	-11%	-23%	-34%	-38%	-65%	-63%	-45%
CrowdRun	2%	7%	11%	-9%	-25%	-30%	-26%	-42%
OldTownCross	15%	-37%	-30%	22%	-30%	-60%	-46%	-11%
ParkJoy	8%	22%	-17%	-19%	-19%	-20%	-17%	-32%
PedestrianArea	-1%	-4%	0%	-8%	-29%	-38%	-36%	-18%
RitualDance	-1%	-7%	-9%	3%	-29%	-34%	-38%	-32%
Average	**5%**	**-10%**	**-16%**	**-4%**	**-32%**	**-47%**	**-44%**	**-30%**
Encoding time	**~106x**				**~5x**			
Decoding time	**~4x**				**~5x**			

It can be observed from Table 2 that the objective BD-rate results are in general in accordance with the subjective test results. However, for sequences like Voiles, for example, both Y PSNR and MOS show lower performance for AV1 compared to HM, even though this lower performance seems to be less significant subjectively. Also, in the case of the sequence ParkJoy, AV1 seems to perform better than HM subjectively, even though objectively HM outperforms AV1 in the Y component. A possible explanation for these cases is the fact that, objectively, AV1 shows a consistently significant higher performance for the chroma components, suggesting that more bits are invested in the quality of the chroma components in AV1. When weighted altogether subjectively, this may compensate for the lower performance in terms of Y BD-rate. From the average of both objective and subjective BD-rate scores, it can be concluded that AV1 and HM provide similar overall performance. On the other hand, JEM shows a significantly better performance in terms of BD-PSNR. This is confirmed by the subjective MOS-BD-rate, showing an overall 30% bitrate reduction with respect to HM for the selected test material and coding conditions.

The average encoding and decoding times associated with AV1 and JEM relative to HM were also analysed and shown in Table 2. Compared with HM, encoding with AV1 and JEM takes, on average, 106 and 5 times longer, respectively. As for the decoding time, AV1 and JEM took approximately 4 and 5 times longer than HM, on average. Given the much higher encoding times associated with AV1, an additional test was performed using 3 additional faster encoding configurations of the AV1 encoder (setting the parameter "--cpu-used" to 1, 2 and 8), as reported in Figure 6, along with the respective performance in terms of Y PSNR BD-rate with respect to HM.

Significant encoding time reductions (from 124.3 to 64.7 times the encoding time of HM) can be achieved for speed preset 1, without major performance degradations (from 4.5% to 6.2% BDrate losses). For the fastest speed preset 8, however, even though encoding times are comparable to HM, the performance losses are significant. It is important to note though that the AV1 specification is very recent and more efforts into designing speed-up algorithms can be expected. Finally, it is also important to consider that neither HM and JEM are practical video encoders. This means that significantly faster encoders that retain most of the compression performance are also possible for these technologies. More relevant comparisons with respect to the encoding time/performance trade-off will certainly be of interest when the technology for AV1 encoding becomes more mature.

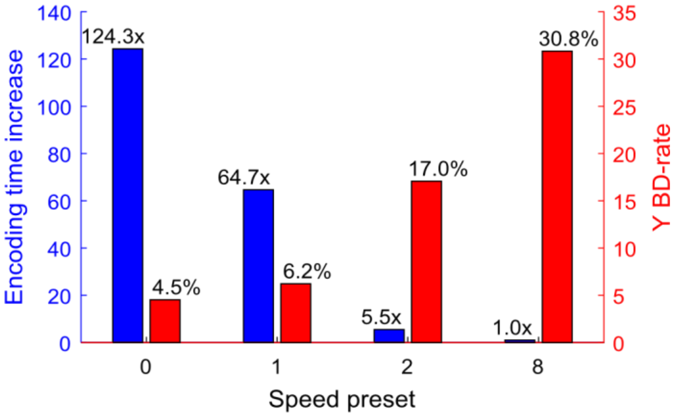


Figure 6 – Encoding time increase and BD-rate losses of AV1 with respect to HM for 4 different speed presets.

CONCLUSION

This paper analysed some of the most promising emerging video compression technologies which are likely to be used on a wide scale in the near future. The tested technologies include the well-established HEVC standard, the recently finalised royaltyfree AV1 specification, and the JEM software used in the JVET exploration activities on video compression capabilities beyond HEVC. The results obtained show, in general, a similar performance between AV1 and the reference HEVC software, HM, both objectively and subjectively, for the test conditions considered in this paper. JEM, on the other hand, seems to outperform both HM and AV1 by approximately 30%, proving that higher compression efficiency can be achieved and will likely be available as part of a future video coding standard in a longer term. As for complexity, both AV1 and JEM show higher decoding times than HM (approximately 4 and 5 times higher, respectively) and the encoding complexity of AV1 is much higher than the remaining reference implementations tested. As practical AV1 implementations get more and more mature, it is expected that faster implementations can reach realistic encoding times, ideally without significantly compromising the compression performance provided by this specification.

REFERENCES

1. Sullivan G. J., Ohm J.-R., Han W.-J. and Wiegand T., 2012. Overview of the High Efficiency Video Coding (HEVC) standard. IEEE Trans. Circuits Syst. Video Technol.. vol. 22, no. 12, pp. 1649 to 1668.
2. Tan T., Weerakkody R., Mrak M., Ramzan N., Baroncini V., Ohm J.-R. and Sullivan, G., 2016. Video quality evaluation methodology and verification testing of HEVC compression performance. IEEE Trans. Circuits Syst. Video Technol. vol. 26, no. 1, pp. 76 to 90.

3. AV1 specification. Available: https://aomedia.org/av1-bitstream-and-decoding-processspecification/.
4. VP9 specification. Available: https://www.webmproject.org/vp9/.
5. JEM software. Available: https://jvet.hhi.fraunhofer.de/svn/svn_HMJEMSoftware/.
6. Baroncini, V., 2018. Results of subjective testing of responses to the joint CfP on video compression technology with capability beyond HEVC. JVET-J0080, 10th JVET meeting, San Diego, April 2018.
7. Bjøntegaard G., 2001. Calculation of average PSNR differences between RD-curves. VCEG-M33, 13th VCEG meeting, Austin, TX, April 2001.
8. Recommendation ITU-R BT.500-13, 2012. Methodology for the subjective assessment of the quality of television pictures.
9. HM software. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.
10. Bossen F., 2013. Common HM test conditions and software reference configurations. JCTVC-L1100, 12th JCTVC meeting, Geneva, January 2013.
11. Sühring, K., Li, X., 2017. JVET common test conditions and software reference configurations. JVET-H1010, 8th JVET meeting, Macau, October 2017.

ACKNOWLEDGEMENTS

The work described in this paper has been conducted within the project COGNITUS. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 687605.