Compression Efficiency and Computational Cost Comparison between AV1 and HEVC Encoders

Isis Bender, Daniel Palomino, Luciano Agostini, Guilherme Correa, Marcelo Porto
Video Technology Research Group (ViTech)
Graduate Program in Computing (PPGC) – Federal University Pelotas (UFPel)
Pelotas, Brazil
{idbender, dpalomino, agostini, gcorrea, porto}@inf.ufpel.edu.br

Abstract— High Efficiency Video Coding (HEVC) is the current state-of-the-art standard for video compression, finalized by ISO/IEC and ITU-T in 2013. However, as HEVC is protected by several patents and is subject to royalty policies, more than thirty companies recently joined efforts to develop a royalty-free codec, named AOMedia Video 1 (AV1). In the forthcoming years, AV1 is expected to be adopted by several companies and streaming service providers as the main digital video format. In this paper, the compression efficiency and the computational cost of AV1 is compared to its main concurrent, the HEVC standard. The methodology employed in this work uses equivalent quantization parameters in both encoders to measure compression efficiency in terms of Bjontegaard Delta (BD)-rate more accurately than related works. Experimental results show that the AV1 reference software requires an average computational cost 14.64 times greater than the HEVC Model, with a BD-rate increase of 16.35% in comparison to that encoder.

Keywords—AV1, HEVC, compression efficiency, computational cost

I. Introduction

Several embedded electronic devices, such as smartphones and tablets, currently allow easy access to high definition (HD) and ultra-high definition (UHD) digital videos. It is expected that by 2021 overall video traffic will reach approximately 82% of all IP traffic in the world [1]. As digital videos are represented by a huge amount of data, the success in storing and transmitting this type of media relies on the efficiency of video compression algorithms. This scenario becomes even more challenging when considering UHD videos, such as 4K and 8K resolutions, and the immersive video content, such as 3D. In this context, the continuous development and improvement of video coding standards is essential to drive video-based applications.

The video consumer market is constantly looking for new video coding algorithms, standards and codecs that achieve high compression efficiency. For a long time, standardization groups such as ISO, IEC and ITU-T has dominated the market. Many video coding standards were developed in the last decades, including MPEG-1, MPEG-2/H.262, H.263, MPEG-4 Visual and H.264/AVC [2]. The most recent one, called High Efficiency Video Coding (HEVC), was released in 2013 [3]. It is the current state-of-the-art in video coding and is able to achieve the same visual quality at about half the bitrate when compared to its predecessor, the H.264/AVC standard [3].

Digital videos are often accessed by applications running over the internet. As the main internet technologies are open, such as the Hyper Text Transfer Protocol (HTTP), the Transfer Control Protocol (TCP) and the Internet Protocol (IP), Google embarked on the WebM project [4] to develop open-source and royalty-free video codecs. The first codec released as part of the project was entitled VP8, which was later succeeded by VP9 [5]. In 2015, several high-tech companies, such as Mozilla, Netflix and Cisco, joined the effort and founded the Alliance for Open Media (AOMedia), which is responsible for the creation of a new royalty-free codec to compete with HEVC, named AOMedia Video 1 (AV1). The AV1 encoder development used proprietary VP9 video coding scheme as a starting point [6], but many other compression tools have been included in the project in the last years to increase compression efficiency. The codec and the bitstream specification were released in 2018 [2].

AV1 was designed with the goal of achieving higher compression rates in comparison to HEVC and previous standards. Thus, it is important to compare their encoding efficiency and computational cost under similar scenarios and test conditions. To the best of the authors' knowledge, there are currently three related works by Nguyen [6], Laude [7] and Layek [8] that show that AV1 is far less efficient than HEVC, considering both computational cost and compression efficiency ratio. However, the methodology employed in those works does not guarantee similar encoding scenarios for AV1 and HEVC, which precludes a fair comparison. For instance, the comparisons in [6-8] do not use equivalent quantization parameters for both standards, which may impair the Bjontegaard Delta (BD)-rate results [12] Besides, as the experiments presented in [6-8] considered a version of the reference AV1 encoder prior to the official codec release in June 2018 [13], an updated analysis of the AV1 encoder efficiency is now required.

This paper presents a comparison between AV1 and HEVC in terms of compression efficiency and computational cost based on a methodology that guarantees the use of equivalent quantization parameters for both standards. Thus, by using the closest possible bitrates for both encoders, a fairer compression efficiency comparison is possible using the Bjontegaard Delta (BD) measures. The experiments were performed using five video sequences recommended in [10]. In addition to the four QPs evaluated for the HEVC, an average of eighteen CQ values were tested for video sequence in the AV1. The paper also presents a computational cost analysis in terms of encoding time for AV1, providing an updated comparison between the latest version of the new codec with the HEVC reference software.

This paper is organized as follows. Section II presents related works that present compression efficiency and computational cost comparisons between AV1 and HEVC. Section III present the methodology used for the analyses presented in this work. Section IV presents the experimental results, comparisons and discussions. Finally, section V concludes this work.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

II. Related Works

Some works that report studies regarding the new AV1 codec have been published recently. Some of them are focused on coding tools optimization, hardware design, parallelism exploration and image quality assessment [15-18]. There are also published works that compare AV1 with other codecs available in the market. Such type of comparisons are usually important contributions to the research field, especially when they focus on objectively evaluating new codecs in terms of compression efficiency and computational cost in relation to their competitors, as shown in recent works as [6], [7] and [8]. However, each of these works employs a different methodology, yielding different results.

Layek [8] compares encoding efficiency and computational cost between the AV1 reference software (AOM) [9] and the x265 codec, which follows the HEVC standard. The work presents a comparison under two different setups. The first one uses the default parameters in both encoders, only varying the target bitrate. In the second setup, the x265 encoder is configured in the *placebo* mode, providing the best visual quality per bitrate. Experimental results were reported for objective quality in terms of PSNR (Peak Signal Noise Ration), SSIM (Structural Similarity Index Metric) and VIFp (Visual Information Fidelity) metrics. BD measures were not reported in the experiments.

More recently, Nguyen [6] and Laude [7] also compared encoding efficiency and computational cost between AV1 and HEVC. Nguyen [6] presents results for the HEVC reference software (HEVC Model – HM) 16.18 and the AOM version with Git hash code described in [6] (Jan. 31, 2018), while Laude [7] presents results for HM 16.16 and AOM 0.1.0-5913. Both works employ the Random Access (RA) configuration for the HEVC encoder with four quantization parameters. Also, in both of them the Variable Bit Rate (VBR) mode was used for AV1, which employs a rate control mechanism where the codec dynamically chooses CQ values within the range 0-63. Nguyen [6] also presents results for AV1 when configured in the quantization mode using the base CQs 31, 39, 47 and 55.

Both [6] and [7] employed the VBR mode when encoding with AV1. However, Nguyen [6] states that the use of rate control algorithms lead to encoding decisions that hinder a direct comparison between standards, since these algorithms are not standardized. In fact, the use of a fixed quantization parameter (QP) structure is more recommended for this type of comparison, since the effect of rate control algorithms can be disregarded. The authors in [6] present a brief comparison considering the quantization mode, using four base CQs. However, the used CQ values are not justified, so that they can favor either AV1 or HEVC in the comparison.

In this context, this paper presents a comparison between AV1 and HEVC that uses equivalent quantization parameters between AV1 and HEVC, targeting a fairer comparison between them. This paper also presents a computational cost analysis (in terms of encoding time) for the AV1 reference software (AOM) in its most recent version. Notice that the three related works [6-8] present encoding time results for older versions of the AOM codec, launched before the official codec release in June 2018.

III. Comparison Methodology

The eight video sequences recommended in [9] were initially considered for the comparisons between AV1 and HEVC. However, due to the great encoding time demanded by the AOM reference software, which can spend more than 13 minutes to encode a single 1280x720 frame, a smaller set of five sequences were selected. To guarantee that the five sequences differ from each other in both spatial and temporal characteristics, an analysis of the spatial activity index (SI) and the temporal activity index (TI) [14] was performed. Fig. 1 shows the resulting SI (x axis) and TI (y axis) of the eight video sequences (1280×720 resolution). The five sequences selected for the comparisons were chosen according to their SI and TI. Notice that video sequences *Boat_hdr_amazon*, *Dark*, *KristenAndSara*, *Netflix_DrivingPOV* and *Netflix_RollerCoaster* are those that differ the most from one another in both indexes. Thus, the first 30 frames of each sequence were encoded in the HM-16.20 reference software and in AOM v1.0.0-943-ga7f959f06 (Nov. 19, 2018).

The parameters that control quantization and bitrate in HEVC and in AV1 are called quantization parameter (QP) and constant quality (CQ), respectively. The QP in HEVC varies between 0 and 51, whereas the CQ in AV1 varies between 0 and 63. The HEVC Common Test Conditions (CTC) recommends the use of QPs 22, 27, 32 and 37 for [11] comparison purposes. Daede [9] recommends the use of CQs 20, 32, 43 and 55 in AV1. Thus, the first tests performed in this work considered the use of these recommended QPs/CQs to compare HEVC and AV1 in terms of compression efficiency. The obtained bitrate and PSNR results for all encoded video sequences are presented in Fig. 2. The recommended QPs/CQs are shown along the curves in the figure.

Notice that the use of these QP/CQ values is not the best way of comparing compression efficiency among the encoders, since each QP/CQ pair can result in very different bitrates or image quality. This difference is especially grater for smaller quantization values, as in the pair 22/20 (QP/CQ). For each QP/CQ pair, AV1 achieves better PSNR results than HEVC, but it always requires a much higher bitrate. On average, AV1 requires a bitrate 49.15% higher than HEVC for *Boat_hdr_amazon*, 100.68% higher for the *Dark* sequence, 95.22% higher for *KristenAndSara*, 70.92% higher for *Netflix_DrivingPOV* and 32.27% higher for the *Netflix_RollerCoaster* sequence. Notice that, except for the *Dark* sequence, HEVC is able to encode at small bitrates

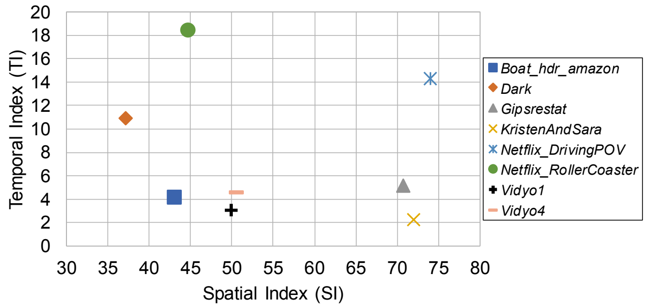

Fig. 1. Spatial and temporal activity indexes for video sequences recommended in [10] (1280x720 resolution).

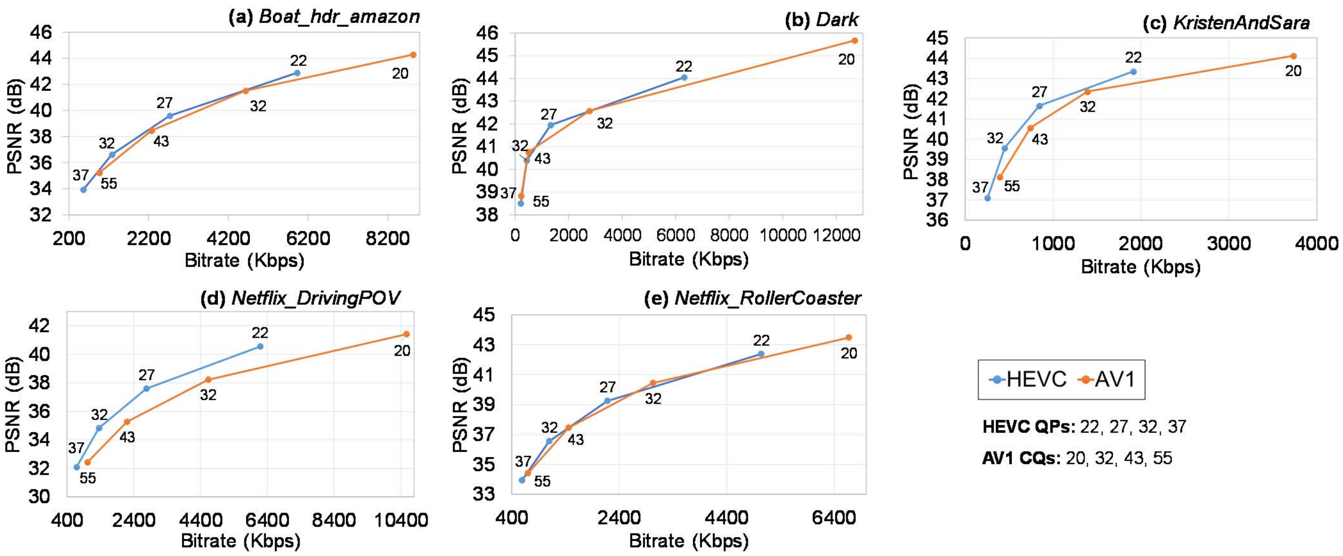

Fig. 2. Compression efficiency results for AV1 and HEVC when encoding with quantization parameters (QP) and constant quality (CQ) parameters recommended in [10].

(QP/CQ pair: 37/55) that AV1 cannot reach with any CQ value tested.

In summary, the comparison between HEVC and AV1 is not fair using this methodology, since the QP and CQ values are not aligned. To overcome this, Bjontegaard Delta (BD) measures [12] are usually employed to calculate a rate-distortion (R-D) curve approximation when the points belonging to the two compared curves are not aligned [12]. Then, for BD-PSNR the average PSNR difference between two R-D curves can be approximated by the difference between the integrals of the fitted curves divided by the integration interval. The same applies for BD-rate, considering the average bitrate difference between two R-D curves.

In order to obtain a fairer comparison between both codecs and to avoid the BD measure calculation based on approximated curves of not-aligned QP/CQ pairs, we performed a wide exploration on the CQ range to find more suitable values for the comparison. Thus, we propose the use of CQ values in AV1 that lead to bitrates as close as possible to the bitrates obtained with the recommended QPs in the HEVC CTCs [11]. When considering a higher number of CQs (and a wider range) than those ones recommended in [9], it is possible to find better aligned QP/CQ pairs.

Fig. 3 (a) shows the QP/CQ pair exploration for the *KristenAndSara* and Fig. 3 (b) for *Netflix_RollerCoaster* sequences. Along the orange line, the CQ values recommended in [9] are presented. Besides, CQ values that lead to the closest bitrate achieved in HEVC with a recommended QP is presented (underlined CQs). For the *KristenAndSara* sequence, eighteen different CQ values were tested to find the best QP/CQ pairs, while for *Netflix_RollerCoaster* sequence twenty-one values of CQ were tested. One can be notice in Fig. 3 that, differently from Fig. 2(c) and 2(e), Fig. 3 shows that AV1 can operate at lower bitrates then HEVC.

It is important to mention that the differences between the recommended CQs in [10] and the best aligned CQs (considering the closest bitrate obtained with a recommended QPs) are significant, especially for low CQs. For example,

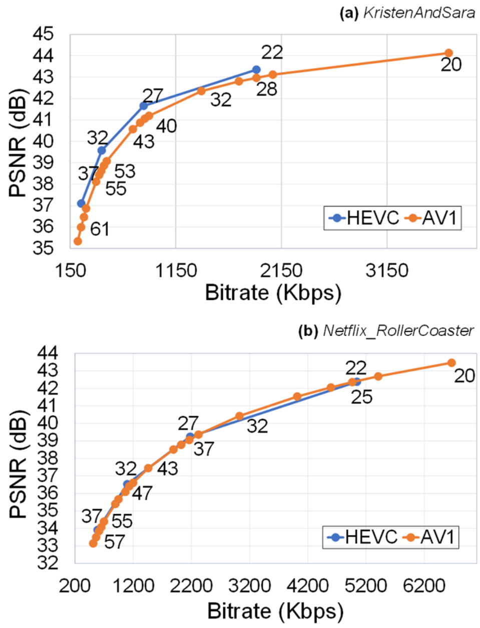

Fig. 3. Compression efficiency results for AV1 and HEVC with all CQ values tested for *KristenAndSara* (a) and *Netflix_RollerCoaster* (b) video sequences.

Fig. 3(a) shows that the recommended QP 22 in HEVC leads to a bitrate close to the one obtained by AV1 using CQ 28. In this case, the bitrate difference in the lower QP/CQ pair (22/28) decreased from 48.78% to 0,03%. Considering the four CQs values presenting closer bitrate values to the HEVC QPs, the AV1 BD-rate increase is 25.47%, while the BD-rate calculated with the recommended QP/CQ values in [10] is 23.64%. This difference of 1.83% is caused by the BD-rate curve approximations. Already the Fig. 3(b) shows that the recommended QP 22 in HEVC leads to a bitrate close to the one obtained by AV1 using CQ 25. In this case, the bitrate difference in the lower QP/CQ pair (22/25) decreased from 24.39% to 1.56%. Considering the four CQs values presenting closer bitrate values to the HEVC QPs, the AV1 BD-rate increase is 4.27%, while the BD-rate calculated with the recommended QP/CQ values in [10] is 3.48%. This difference of 0.79% is too caused by the BD-rate curve approximations

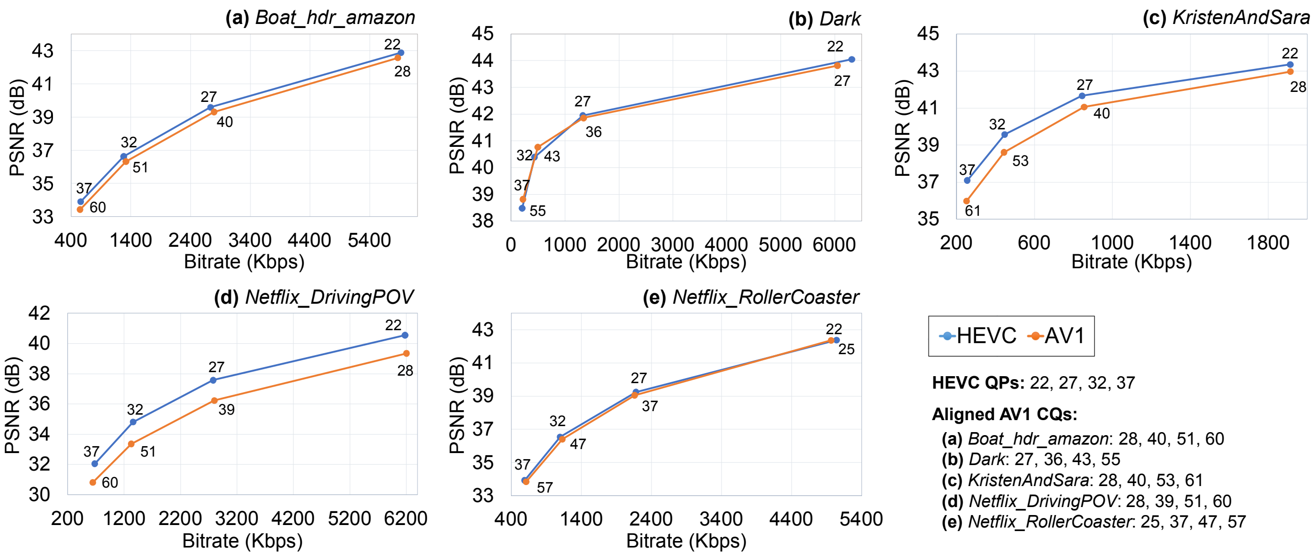

Fig. 4. Compression efficiency results for AV1 and HEVC when encoding with quantization parameters (QP) recommended in [10] and corresponding aligned constant quality (CQ) parameters.

Considering all the videos sequences, 33 different CQ values, within a range from 20 to 61, were tested to find the four CQs that better fit the bitrate obtained when using the recommended HEVC QPs. As the best aligned QP/CQ pair differ from one video sequence to another, the best aligned four CQs were tested for each one of the five sequences, totalizing 92 encodings with the AV1 reference software.

IV. Experimental Results

After obtaining the aligned QP/CQ pairs according to the methodology presented in section III, each video sequence was encoded with both AOM and HM reference encoders four times (once with each QP/CQ value). The Random Access configuration was used in the HM encoder, whereas the Low Latency CQP [10] configuration was used in the AOM reference encoder. All experiments were conducted on an Intel Xeon E5-2640v3 2.4 GHz processor server, with 32 GB of RAM.

Fig. 4 shows experimental results obtained with the four best-aligned QP/CQ pairs for each video sequence. Note that the four CQ values are different for each sequence. For the *Boat_hdr_amazon* sequence, the best CQs are 28, 40, 51 and 60, while for *Dark* the values are 27, 36, 43 and 55. *KristenAndSar*a uses CQs 28, 40, 53 and 61, whereas the *Netflix_DrivingPOV* uses 28, 39, 51 and 60 as best-aligned CQs. Finally, for *Netflix_RollerCoaster* the chosen CQs are 25, 37, 47 and 57. Note that just in the *Dark* sequence one of the best-aligned CQs corresponds to a value recommended in [10] (CQ 43).

Notice in the charts of Fig. 4 that when the QP/CQ values are aligned in the *x* axis, AV1 presents worse PSNR results than HEVC for all videos sequences analyzed. This means that on average AV1 still presents a worse encoding efficiency in comparison to HEVC. Table 1 summarizes average comparison results between AV1 and HEVC in terms of encoding efficiency (BD-rate) and computational cost (AOM/HM time ratio) for the five video sequences, taking the HM encoder as the baseline in all cases. The numbers show that HM achieves better average results in both encoding efficiency and computational cost than AOM.

Table 1. AV1 compression efficiency considering the four recommended AV1 CQs [10] and the four AV1 CQs aligned to the recommended QPs.

Video Sequence	BD-rate (%) (fixed CQs)	BD-rate (%) (aligned CQs)	BD-rate difference	AOM/HM time ratio
Boat_hdr_amazon	9.82	10.45	+0.63	13.44×
Dark	11.7	1.41	-10.29	17.03×
KristenAndSara	23.64	25.47	+1.83	10.88×
Netflix_DrivingPOV	40.42	40.13	-0.29	17.30×
Netflix_RollerCoaster	3.48	4.27	+0.79	14.54×
Average	17.82	16.35	-1.47	14.64×

Table 2. Comparison results in this work and related works.

Work / Mode	BD-rate (%)	AOM/HM time ratio
This work	**16.35**	**14.64×**
Nguyen et al. [6] / Quantization mode	30.5	32.45×
Nguyen et al. [6] / VBR mode	2.3	20.95×
Laude et al. [7] / VBR mode	38.3	32.6×

On average, AOM requires an encoding time that is 14.64 times higher than the required by HM to encode the same video sequence. This cost is high for all sequences, varying from 10.88 times (*KristenAndSara*) to 17.3 times (*Netflix_DrivingPOV*). Considering compression efficiency calculated over the four best-aligned CQ/QP pairs, the use of AV1 incurs in a BD-rate increase of 16.35% over HEVC. Considering the four points recommended in [10], the average BD-rate increase is 17.82%. The average difference between both BD-rate values calculated for each video sequence is 1.47. For the *Netflix_DrivingPOV* sequence, a much smaller BD-rate is achieved with the best-aligned QP/CQ pair than with the recommended CQs. From Fig. 4(b), it is clear that a small BD-rate value reflects more precisely the small difference between the two R-D curves for this sequence.

Table 2 shows a comparison in terms of encoding efficiency (BD-rate) and encoding time between this work and two other related works found in the literature. Even though they are not exactly comparable because related works use a different version of the AOM codec, comparing them is important to show the encoder recent evolution. In this work, the obtained results show an average BD-rate increase smaller than the observed by Nguyen [6] (quantization mode) and Laude [7] (VBR mode), which resulted in 30.5% and 38.3%, respectively. When compared to the values observed by Nguyen [6] (VBR mode), the obtained BD-rate are significantly higher. However, as previously mentioned, the use of a rate control algorithm (VBR mode) introduces aspects that hinder a direct comparison between codecs, so that the use of a fixed QP/CQ-based approach such as the used in this work and in Nguyen [6] (quantization mode) leads to more reliable results. Thus, this work presents a more accurate methodology for comparing AV1 and HEVC than [6] and [7], since no rate control algorithm is used and the CQ/QP pairs are adjusted to the best-aligned bitrates.

Finally, this work also showed that the encoding time of AOM has decreased significantly in comparison to previous analyses, coming to be 2.2 times faster than the time reported by Nguyen [6] under the quantization mode. Still, the tests confirm the significantly high complexity of AV1 in comparison to HEVC. The several tools introduced in the new codec still require a very high processing power, despite all the recent optimizations performed over the AOM reference encoder. Thus, according to the results obtained in this work, there is still no advantage in using AV1 over HEVC in what concerns compression efficiency and computational cost.

V. Conclusions

This work presented a comparison between the AV1 and HEVC reference encoders in terms of compression efficiency and computational time. As the use of recommended constant quality (CQ) and quantization parameter (QP) values can result in not aligned bitrates in the AV1 and HEVC bitstreams, respectively, this work proposes a comparison approach that searches for the best-aligned QP/CQ pair in terms of bitrate to calculate Bjontegaard Delta measures more accurately. Average results show that the use of AOM incurs in a BD-rate increase of 16.35% in comparison to HM, with a computational cost that is 14.64 times larger. Still, when comparing the obtained results to previous works, it is possible to notice that the AOM complexity has decreased by 55% in recent versions. Finally, despite the advantages of using AV1 due to royalty costs, the obtained results allow us concluding that there is still no advantage in using AV1 over HEVC in what concerns compression efficiency and computational cost.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and also by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

References

[1] Cisco, "Cisco Visual Networking Index: Forecast and Methodology,2016–2021 - Cisco." [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.html#_Toc484813989.

[2] Y. Chen, D. Murherjee, J. Han, A. Grange, Y. Xu, Z. Liu, S. Parker, C. Chen, H. Su, U. Joshi, C. Chiang, Y. Wang, P. Wilkins, J. Bankoski, L. Trudeau, N. Egge, J. Valin, T. Davies, S. Midtskogen, A. Norkin, P. Rivaz, "An Overview of Core Coding Tools in the AV1 Video Codec", IEEE Picture Coding Symposium, pp. 41-45, California, 2018.

[3] L. Guo, J. D. Cock, A. Aaron, "Compression Performance Comparison of x264, x265, libvpx and aomenc for On-Demand Adaptive Streaming Applications", IEEE Picture Coding Symposium, pp. 26-30, California, 2018.

[4] D. Mukherjee, J. Han, J. Bankoski, R. Bultje, A. Grange, J. Koleszar, P. Wilkins, Y. Xu, "A Technical Overview of VP9 - The Latest Open-Source Video Codec", In: SMPTE Annual Technical Conference & Exhibition, pp. 1-17, Hollywood, 2013.

[5] S. Parker, Y. Chen, J. Han, Z. Liu, D. Mukherjee, H. Su, Y. Wang, J. Bankoski, S. Li, "On Transform Coding Tools Under Development for VP10", Applications of Digital Image Processing XXXIX, Proc. of SPIE, vol. 9971, pp. 1-10, 2016.

[6] T. Nguyen, D. Marpe, "Future Video Coding Technologies: A Performance Evaluation of AV1, JEM, VP9 and HM", IEEE Picture Coding Symposium, pp. 31-35, California, 2018.

[7] T. Laude, Y. G. Adhisantoso, J. Voges, M. Munderloh, J. Ostermann, "A Comparison of JEM and AV1 with HEVC: Coding Tools, Coding Efficiency and Complexity", IEEE Picture Coding Symposium, pp. 36-40, California, 2018.

[8] M. A. Layek, M. Q. Thai, M. A. Hossain, N. T Thu, L. P. Tuyen, A. Talukder, T. Chung, E. Huh, "Performance Analysis of H.264, H.265, VP9 and AV1 Video Encoders" IEEE Asia-Pacific Network Operations and Management Symposium, pp. 322-325, Seoul, 2017.

[9] "AV1 Reference Software Implementation," Online: http://aomedia.org/press-release/, 2017.

[10] T. Daede, A. Norkin, I. Brailovskiy, "Video Codec Testing and Quality Measurement draft-ietf-netvc-testing-07", jul. 2018.

[11] F. Bossen, "Common test conditions and software reference configurations", Tech.Rep. JCTVC-J1100, JCTVC, 2012.

[12] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," Tech. Rep. VCEG-M33, Video Coding Experts Group from ITU-T, 2001.

[13] Mozilla Research. Available: https://research.mozilla.org/av1-media-codecs

[14] A. Webster, S. Wolf, "Spatial and Temporal Information Measures for Video Quality", ANSI T1Q1 Contribution T1Q1.5/92-113. January, 1992. [Online]. Available: https://its.bldrdoc.gov/publications/2617.aspx

[15] L. Trudeau, N. Egge, D. Barr, "Predicting Chroma from Luma in AV1", IEEE Data Compression Conference, pp. 374-382, Snowbird, 2018.

[16] C. Chiang, J. Han, S. Vitvitskyy, D. Mukherjee, Y. Xu, "Adaptative Interpolation Filter Scheme in AV1", IEEE International Conference on Image Processing, pp. 934-937, Athens, 2018.

[17] Y. Zhao, J. Wen, "Wavefront Parallel Processing for AV1 Encoder", IEEE Picture Coding Symposium, pp. 101-105, San Francisco, 2018.

[18] Z. Deng, I. Moccagatta, "Hardware-Friendly Inter Prediction Techniques for AV1 Video Coding", IEEE International Conference on Image Processing, pp. 948-952, Beijing, 2017.