Constant quality mode 4k video comparison using AV1 reference tool

Milan Milivojevic, Dragi Dujkovic, and Ana Gavrovska, *Member, IEEE*

***Abstract*— A variety of new coding tools has been developed and are expected to be developed in the near future in order to achieve high-quality video transmission. One of the next-generation video encoding formats is AV1, as the first compression format developed by the Alliance for Open Media, where AV1 is recognized as VP9 successor. In this paper 4k low frame rate video sequences are compared for different constant quality (CQ) values using reference tool libaom. The obtained results are also compared to VP9 and HEVC codecs. Slow AV1 coding is performed using libaom, in order to analyze differences between different CQ settings. The results show compression performance using 2-norm evaluation and time needed for coding.**

***Index Terms*— Video coding, 4k video, constant quality, AV1, libaom, VP9, HEVC.**

I. Introduction

Novel video technologies and standards are inevitable nowadays. The known fact is that most of the telecommunication traffic is related to video. Namely, it is estimated that by 2022, approximately about 82% of global internet will be dedicated to video content [1]. So, the general focus is on video delivery, such as in the case of OTT (Over-the-Top) streaming, wired or wireless communication or TV broadcasting. Particularly important are the coding formats for internet applications and efficient delivery over internet. It is obvious that OTT providers (Netflix, Hulu, etc.) and other immersive media content based industries are interested in developing innovative solutions in video encoding and compression.

In [2] three main video technology trends in 2020 are pointed out. The first one is to increase expectations of viewers by providing higher QoE (Quality of Experience), besides QoS (Quality of Service). Thus, higher quality video should be delivered to consumers, but this should be done in efficient manner. Consequently, the fast delivery of the solutions is the second trend. Moreover, the trend for media content is to find its way quickly to the market, and this should make both services and producing assets more efficient. Finally, the third trend is to improve the return of investments by advanced controlling and managing costs, risks and application complexity.

Even though, H.264 is an older video format from about 2004, it is still in massive usage despite the fact there is more recent one. The recent one is H.265, or HEVC (High Efficiency Video Coding), from about 2013, which gives better performance [3]. Besides HEVC, another UHD (Ultra High Definition) codec has become popular over the years. Open and royalty-free Google's solution like VP9 was widely adopted by different platforms, such as Youtube. Nowadays, AV1 is considered as VP9 successor, and a new open royalty-free format for video coding developed by the AOMedia (Alliance for Open Media) [4-6]. Its first release was in 2018, and, since then, different solutions has been designed for video transmission and video services over internet.

In recent work [7] 4k video traffic has been analyzed using different prediction models, where the sequences were encoded with H.265/HEVC, whereas in [8] variability of the traffic was analyzed. In this paper, analysis of the 4K video traffic is performed using 4k AV1 based on AV1 reference software called libaom. The results are compared to VP9 and HEVC for 4k video from CQ (constant quality or CRF – constant rate factor), coding time and 2-norm traffic standpoint.

The paper is organized as follows. After the introduction, AV1 format is briefly explained in Section II. The simulation and materials used in this paper are explained in Section III. Then, in Section IV the experimental results followed by discussion are given in order to evaluate the compression performance. Finally, in Section V conclusions and future work are mentioned.

II. AV1 video format

From AV1 (AOMedia Video 1), much is expected [3-6]. There is a need for efficient compression standards, and future implementations should balance software and hardware possibilities. Firstly, software experimental analysis is of great importance for realizing complex schemes and architectures. In the case of AV1, there are open implementations. It is royalty-free solution for video coding. The general AV1 architecture is illustrate in Fig.1.

As in every new generation of video coding format there are differences in coding structure, and the gains are obtained using different tools [4]. Motion vector prediction is improved in spatio-temporal domain, where eight main intra prediction directional modes are used. Also, superblocks are introduced.

Milan Milivojevic is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia (e-mail: milansmilivojevic@gmail.com).

Dragi Dujkovic is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia (e-mail: dragi@etf.rs).

Ana Gavrovska is with the School of Electrical Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11020 Belgrade, Serbia (e-mails: anaga777@gmail.com; anaga777@etf.rs).

So, one may use blocks of 128x128 pixels and 64x64 pixels. Higher precision is expected (ten or twelve bits), besides eight bit depth. For transform domain rectangular DCT (Discrete Cosine Transform) and asymmetric DST (Discrete Sine Transform) are used in AV1, while new quantization parameters and filtering techniques are adopted [4-6].

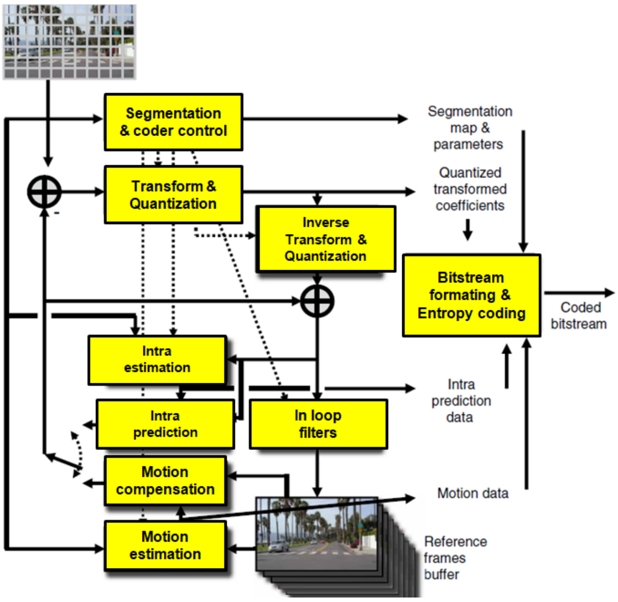

Fig.1. Illustration of general AV1 architecture [4].

Recursive partitioning of superblocks of 128x128 pixels introduced in AV1 is illustrated in Fig.2. Handling hierarchical and recursive techniques and much more in AV1 format led to developing different versions optimized for various purposes. Libaom was introduced as reference AV1 codec (coder-encoder), and enabled the main insight into the AV1 possibilities. It made progress in fast delivering new solution to the public.

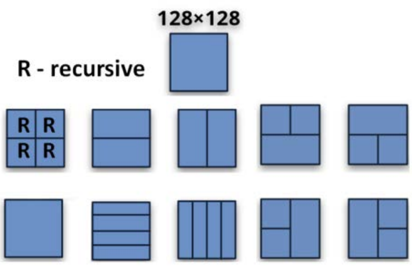

Fig.2. Recursive partitioning of superblock of 128x128 pixels introduced in AV1.

There are already different solutions of AV1 available for specific aims besides the reference software. Typical example is another AV1 codec dav1d for low CPU (Central Processing Unit) processing. Others variations are also available (rav1e, svt-av1). The idea is to develop solutions for video coding for optimized high-performance tasks [5].

In [5] it is claimed that AV1-libaom performs better than HEVC. Even about 43% improvement is reported from PSNR (Peak Signal-to-Noise ratio). AV1 is compared to HEVC since its general architecture looks alike.

III. Simulation

Sequences in mp4 format of 4k resolution are used for the experiment. The illustration of the tested content is shown in Fig. 3. Details about the 4k data are given in Table I. The sequences are of lower frame rate (LFR), so even higher length than common ITU 10 seconds can be used to observe the packet variability (here thirty seconds). Having in mind the spatial resolution like 4k (DCI – Digital Cinema Initiatives or UHD – Ultra High Definition), as well as time resolution, i.e. frame rate, the time needed for testing should be reasonable, and the length is often decreased by video trimming (e.g. to five seconds in [3]). Moreover, one should have in mind higher bit depth based content expected to be a part of future common traffic. Initial attempts when choosing relatively reasonable sequence length showed very slow procedure of applying the reference tool (libaom) in the LFR 8 bit cases, which are considered in this paper.

Materials are prepared according to Constant Quality (CQ) or crf factor. For each sequence four CQ/crf values are used: CQ-20, CQ-24, CQ-30, CQ-34. It was not possible to apply CQ-10 or CQ-40 in the experiment.

The analysis is performed on 64bit Windows 10 Pro Intel(R) Core(TM) i5-8500 CPU, 3GHz, 8GB using ffmpeg 4.3.1 [9].

TABLE I
4K DATA DETAILS

Source (abbrevation)	Resolution	Frame rate
Big Buck Bunny (bbb)	3840 x 2160	30 fps
Tears of Steel (tos)	3840 x 1714	24 fps

(a)

(b)

Fig. 3. Video sequences: (a) Big Buck Bunny (bbb file), and (b) Tears of Steel (tos file).

By using two files eight xml sequences are obtained (four xml sequences are generated for each crf value). In the first part of the analysis, time is measured in minutes while coding due to slow libaom performance. In the second phase, each sequence is represented by the vector magnitude, regardless of the frame type. The magnitude or 2-norm (Euclidean norm) is calculated as:

$$norm(x) = \|x\| = \left(\sum_k |x_k|^2 \right)^{1/2}, \tag{1}$$

where x is an array corresponding to obtained sequence of frames, and k is frame index. The relative ratio for time, $t(x)$, and 2-norm values, $norm(x)$, are calculated for the comparison reasons. The p ratios are defined as:

$$p_{crf(N)}^{time} = t(x)_{crf(N)} / t(x)_{crf(N_{ref})}, \tag{2}$$

$$p_{crf(N)}^{norm} = norm(x)_{crf(N)} / norm(x)_{crf(N_{ref})}, \tag{3}$$

Where N is current and N_{ref} is reference crf value (here N_{ref} = 20). The above procedure is repeated for VP9 and HEVC video format.

IV. Experimental Results

In Table II the results of using AV1-libaom are presented. Namely, time needed for slow coding is presented for each CQ/crf value. Time spent for coding using standard VP9 and HEVC codecs, are presented in Table III and Table IV, respectively.

TABLE II
TIME FOR CODING TO AV1-LIBAOM FORMAT

No.	1	2	3	4
Coding time [min]	CQ -20	CQ -24	CQ -30	CQ -34
bbb	276	260	249	214
tos	510	417	266	196

TABLE III
TIME FOR CODING TO VP9 FORMAT

No.	1	2	3	4
Coding time [min]	CQ -20	CQ -24	CQ -30	CQ -34
bbb	6	6	5	5
tos	12	10	7	6

TABLE IV
TIME FOR CODING TO HEVC FORMAT

No.	1	2	3	4
Coding time [min]	CQ -20	CQ -24	CQ -30	CQ -34
bbb	10	9	8	7
tos	13	8	7	5

For the comparison, relative coding time is calculated and presented in Fig.4 and Fig.5 for bbb and tos sequences, respectively. The obtained sequences for AV1-libaom coding are presented in Fig.6 and Fig.7, for bbb and tos, respectively.

The new codec requires long time to perform coding. Testing shows that relative time difference for higher quality (between crf20 and crf24) resembles VP9 standard for both bbb and tos sequences. On the other hand, relative time difference needed for lower coding quality between crf30 and crf34 seems similar. Around 10% is the difference between the two highest crf values.

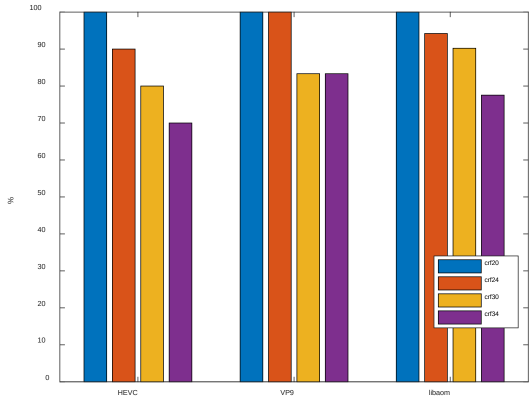

Fig. 4. Relative coding time for bbb sequences.

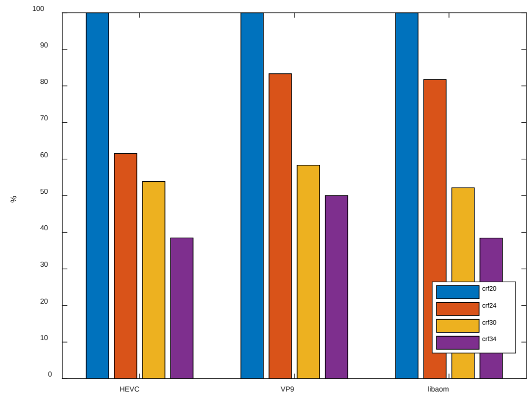

Fig. 5. Relative coding time for tos sequences.

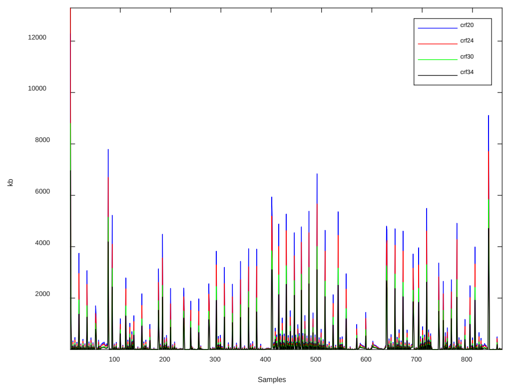

Fig. 6. Obtained sequences using reference AV1-libaom for bbb video.

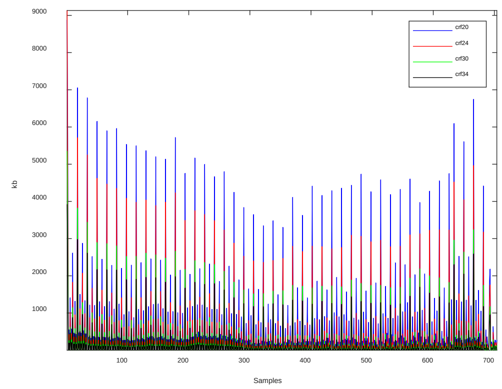

Fig. 7. Obtained sequences using reference AV1-libaom tool for tos video.

It is well known that AV1 generally suppresses VP9 by about 30% [3-6]. In Fig.8 and Fig.9 relative 2-norm values are shown for bbb and tos sequences, respectively. Lower values for HEVC are expected since for HEVC bidirectional frames exist. In Fig.10 obtained predicted frames are shown for libaom and VP9 in the case where crf equals 24.

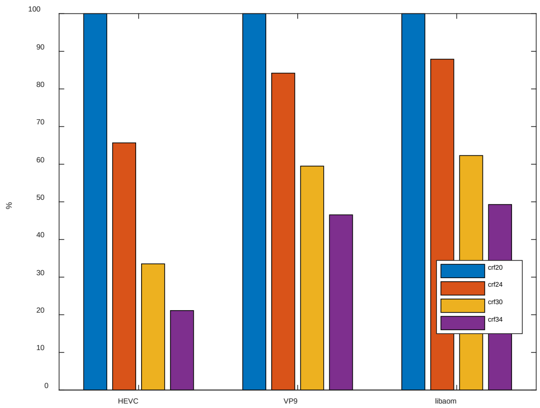

Fig. 8. Relative 2-norm values for bbb sequences.

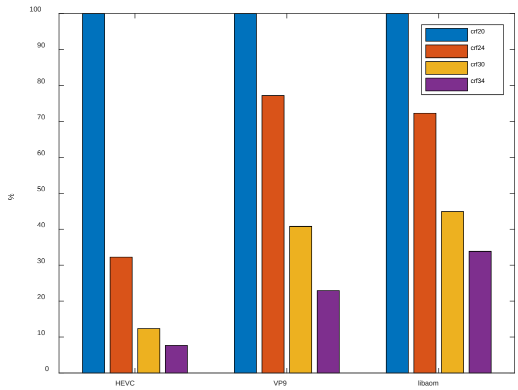

Fig. 9. Relative 2-norm values for tos sequences.

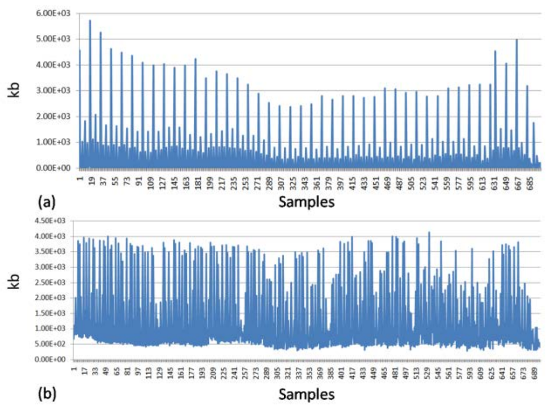

Fig. 10. (a) AV1-libaom and (b) VP9 predicted frames for tos video and crf24.

In this relative estimation it can be seen that the magnitude is similar between VP9 and its successor. As for av1-libaom, smaller difference is obtained for lower quality, i.e. between crf30 and crf34, compared to VP9. This can particularly be observed for tos video in Fig. 9.

In Fig.10(a) it is shown for libaom that frames can be easily differentiated from the size standpoint. In ther words, there are relatively small and relatively high sample values presented. The naturaleness is more visible for VP9 in Fig.10(b), where such differentiation is not obvious. This shows higher control of the traffic in the case of AV1-libaom. The traffic behaviour for different video quality is changed in this way.

V. Conclusion

In this paper low frame rate experiment with constant quality AV1-libaom codec is performed. The obtained results show slow coding performance of the reference software, as well as similar relative magnitude to VP9. Nevertheless, the traffic using new codec shows different behavior in comparison to VP9.

In future work further experiments should be performed in order to evaluate the performance of AV1 standard and its specific implementations, as well as to analyze the behavior in high frame rate and high dynamic range cases. Moreover, other formats are expected to have a great impact on the market, and are expected to be a part of future experiments.

Acknowledgment

This research is supported and funded by Serbian Ministry of Education, Science and Technological Development.

References

[1] CISCO, “CISCO visual networking index: forecast and methodology, 2017–2022,” November 2018.

[2] J. Shulman, “Bitmovin’s video top Video Technologz Trends 2020”, Bitmovin, March 2020. https://bitmovin.com/video-technology-trends/ (last accessed 15.07.2020.)

[3] F. Zhang, A.V. Katsenou, M. Afonso, G. Dimitrov, and D.R. Bull, “Comparing VVC, HEVC and AV1 using Objective and Subjective Assessments,” *arXiv preprint arXiv:2003.10282*, 2020. https://arxiv.org/pdf/2003.10282.pdf (last accessed 15.07.2020.)

[4] I. Trow, “AV1: Implementation, Performance, and Application,” *SMPTE Motion Imaging Journal*, *129*(1), pp.51-56. 2020.

[5] Y. Chen, D. Mukherjee, J. Han, A. Grange, Y. Xu, S. Parker, C. Chen, H. Su, U. Joshi, C.H. Chiang, and Y. Wang, “An Overview of Coding Tools in AV1: the First Video Codec from the Alliance for Open Media,” *APSIPA Transactions on Signal and Information Processing*, Vol. 9, 2020.

[6] Alliance for Open Media – An Alliance of Global Media Innovators, http://aomedia.org/ (last accessed 15.07.2020.)

[7] D. R. Marković, A. M. Gavrovska, I. S. Reljin, “4K Video Traffic Prediction using Seasonal Autoregressive Modeling“, *Telfor Journal*, Telecommunications Society, Belgrade, Vol. 9, No. 1, pp. 8-13, 2017.

[8] A. M. Gavrovska, M. S. Milivojevic, G. Zajic and I. S. Reljin, “Video traffic variability in H.265/HEVC video encoded sequences,“ *13th Symposium on Neural Network Applications in Electrical Engineering* (NEUREL), 22-24. November, Belgrade, Serbia, pp. 109-112, 2016.

[9] Ffmpeg, https://ffmpeg.org/ (last accessed 12.08.2020.)