

An evaluation of the next-generation image coding standard AVIF

Nabajeet Barman and Maria G. Martini

School of Computer Science and Mathematics, Kingston University, London, UK

{n.barman, m.martini}@kingston.ac.uk

Abstract—This paper presents a comparative performance evaluation of the newly proposed AV1 Image File Format (AVIF) vs. other state-of-the-art image codecs, for natural, synthetic and gaming images. The codecs are compared in terms of Rate-quality curves and BD-Rate savings considering different quality metrics. AVIF results in the best overall performance considering both 4:2:0 and 4:4:4 chroma subsampling encoded images for all types of images.

Index Terms—AVIF, JPEG, Image Coding Standard, Synthetic Images, Gaming Images.

I. INTRODUCTION

The human vision system (HVS) is highly responsive to visual aids, such as images and videos. Images today are used across a wide range of applications, from social media, such as Facebook, Instagram and Snapchat, to the field of machine learning and artificial intelligence for various tasks such as pattern recognition and detection of tumours in medical images. The recent advancements and acceptance of such applications are partly made possible due to the increased bandwidth availability and the improvement in image compression and processing capabilities. Image compression is a widely investigated and studied field, with many image compression standards being available for both lossy and lossless encoding with application-specific compression standards and techniques also being proposed [1] [2].

JPEG, which has been in use since 1992, is currently the most widely used lossy image compression standard especially for Internet applications and digital cameras, with almost 70% of websites using it now. Its successor, JPEG 2000 [3], is a discrete wavelet transform based compression standard shown to provide better image quality than JPEG and supporting both lossless and lossy image compression within the same file. WebP, developed in 2010 as a competitor of JPEG for use in web applications, is another image compression standard currently being developed by Google. Comparison studies presented in [4] have shown it to be 25-34% more efficient for the same SSIM value. The High Efficiency Image File Format (HEIF) is a standard that supports the storage of image data encoded using the HEVC standard and is shown to provide up to 25% reduction in bitrate compared to JPEG 2000 for the same objective quality [5]. Similar to HEIF, AVIF, which is the latest image compression standard, allows encapsulating AV1 intra-frame coded content and supports High Dynamic Range (HDR) and Wide Color Gamut (WCG) images as well as Standard Dynamic Range (SDR) [6]. So far, the newly developed AVIF format has been evaluated only on natural images by Netflix [7]. Hence, we present in this paper the first independent comparative evaluation of AVIF image coding format on three different datasets consisting of

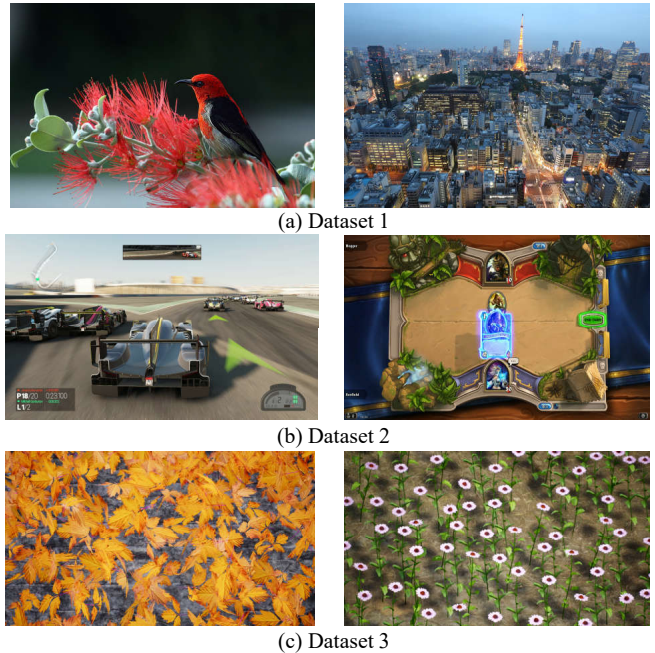


Fig. 1: Sample images from the three datasets.

natural, gaming and synthetic images (see Section II-A). For further investigations and reproducibility of the results, we additionally provide the gaming images as an open-source dataset¹. The rest of the paper is organized as follows. Section II presents the evaluation of dataset and methodology. Section III presents the results and observations of this study and Section IV concludes the paper with a discussion of future work.

II. EVALUATION DATASET AND METHODOLOGY

A. Evaluation Dataset

To evaluate the performance of the newly proposed AVIF codec versus the existing codecs, we selected three different datasets. We report two sample images from each dataset in Figure 1. A description of the considered datasets is reported below.

- 1) *Dataset 1 (D1)*: We used a total of 52 Images with resolution 2040×1346 from the DIV2K dataset, which consists of a wide range of natural images in .jpg format depicting real-world scenes such as monuments and landscape [8] and is similar to the ones used in [7].
- 2) *Dataset 2 (D2)*: Since the performance of quality assessment metrics is different for gaming content [9] as well as because gaming content is perceived differently from natural content [10], we use in this work a gaming

¹<https://kingston.box.com/s/q6rsdzjg53ur61kqfve9vye1r3ovqtwg>

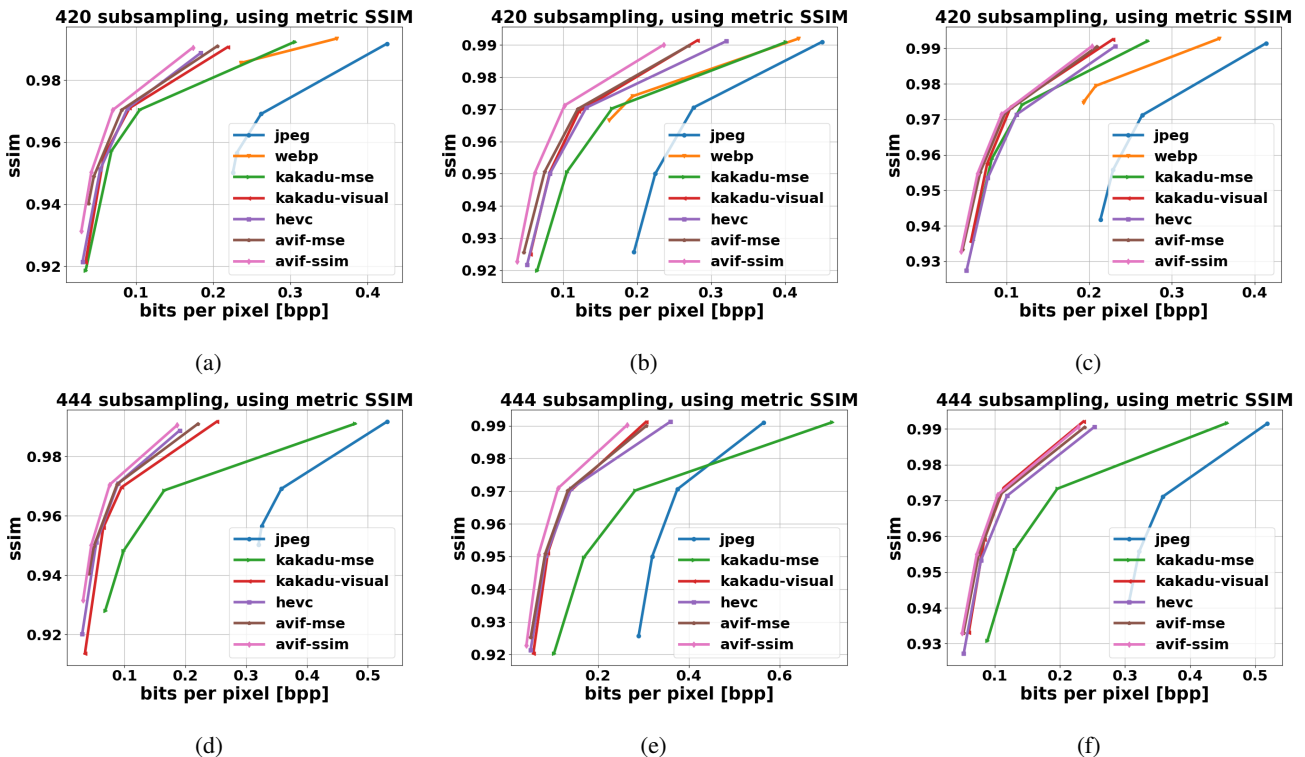


Fig. 2: Rate-Quality curves (SSIM vs. bpp). (a) D1, 4:2:0; (b) D2, 4:2:0; (c) D3, 4:2:0; (d) D1, 4:4:4 ; (e) D2, 4:4:4; (f) D3, 4:4:4.

images dataset. The dataset was created by extracting frames from the reference videos from the open-source gaming videos dataset GamingVideoSET [11]. GamingVideoSET consists of 24 reference gaming video sequences from 12 different games of resolution 1920×1080 in YUV format from which we first extracted all frames in *.yuv* format. From these, we selected 50 distinct images which were then converted to *.png* format using FFmpeg.

- 3) *Dataset 3 (D3)*: BVI-SynTex is a synthetic video texture dataset, that was generated using a computer graphics imagery environment, consisting of 196 video sequences of resolution 1920×1080 of various texture types [12]. Following a similar procedure as for Dataset 2, we extracted *.yuv* frames from the videos and selected 50 distinct image sequences which were then converted to *.png* format.

B. Evaluation Methodology

For evaluation, we use the image compression comparison framework presented by Netflix in [13] using the default settings parameters. The images were compressed using the image compression standards discussed in Section 1. Different profiles and codecs are considered for each compression method, as discussed in [7]:

- JPEG: jpeg-mse (minimise MSE), jpeg-ms-ssim (maximize MS-SSIM), jpeg-im (ImageMagic implementation), jpeg-hvs-psnr (maximize HVS-PSNR);
- WebP (Webp codec);
- JPEG 2000 Kakadu implementations kakadu-mse (minimise MSE) and kakadu-visual (maximize visual quality); JPEG 2000 libopenjpeg implementation, openjpeg;

- HEVC (HM reference software - hevc intra-frame coding);
- AVIF: avif-mse (min MSE), avif-ssim (max SSIM).

The images were encoded at four SSIM (0.92, 0.95, 0.97, 0.99) and VMAF values (75, 80, 85, 90, 95) at both 4:2:0 and 4:4:4 chroma subsampling. The chroma subsampling were chosen based on the observation that 4:4:4 is usually the original, capture subsampling format and 4:2:0 is one of the most widely used subsampling based on the fact that the HVS is more sensitive to the luma component compared to the chroma component.

III. RESULTS

Figure 2 presents the SSIM vs. bits per pixel (bpp) curves for all the three datasets considering both 4:2:0 and 4:4:4 chroma subsampling. Similar curves are also obtained for the image sequences encoded at four target VMAF values. These are not reported here for brevity, but are included as part of the dataset.

Based on the figure, it can be observed that for all three datasets, AVIF results in the best performance for both subsampling modes, which is similar to the results and observations reported in [7]. For D3, the performance of both Mean Square Error (MSE) and SSIM optimized AVIF encoding results in almost similar encoding while for D1 and D2, SSIM optimized AVIF version seems to provide better quality encodes.

It is to be noted that the evaluation presented in [7] does not present the performance results considering VMAF as the objective quality metric. Therefore, we present in Table I the mean percentage of BD-Rate savings results for the images compressed using the different image codecs at different VMAF values for both 4:2:0 and 4:4:4 chroma subsampling. for all three datasets. The codec represents

TABLE I: Mean Bjontegaard-Delta (BD) rate with respective metric (percentage) for three datasets for both 4:2:0 and 4:4:4 chroma subsampling. The best performing codec for each metric is highlighted in bold.

Dataset1													
Mean BDRate-420							Mean BDRate-444						
Codec	VMAF	SSIM	MS_SSIM	VIF	PSNR_Y	PSNR_AVG	Codec	VMAF	SSIM	MS_SSIM	VIF	PSNR_Y	PSNR_AVG
jpeg-mse	18.68	32.43	24.92	4.21	-8.38	-8.08	jpeg-mse	14.87	25.46	19.67	3.61	-7.63	-5.73
jpeg-ms-ssim	-2.26	-6.35	2.08	4.4	-0.7	-2.83	jpeg-ms-ssim	5.05	0.5	8.29	12.58	7.01	-0.8
jpeg-im	-2.13	-4.8	0.17	2.23	0.34	-1.65	jpeg-im	1.97	-1.11	3.41	6.49	4.45	-0.87
jpeg-hvs-psnr	-5.75	-14.42	-3.42	3.46	4.61	2.63	jpeg-hvs-psnr	-1	-9.25	0.3	7.6	8.34	1.8
webp	-8.26	-18.12	-14.93	-25.09	-34.99	-34.25	webp	NA	NA	NA	NA	NA	NA
kakadu-mse	-33.38	-25.03	-27.78	-34.21	-43.88	-42.42	kakadu-mse	-11.48	-4.36	-5.76	-10.17	-24.39	-20.1
kakadu-visual	-41.99	-44.18	-43.9	-31.64	-28.69	-28.62	kakadu-visual	-52.79	-55.74	-55.52	-43.98	-41.54	-42.95
openjpeg	-24.84	-8.34	-20.1	-27.41	-36.89	-34.76	openjpeg	-9.51	5.91	-5.83	-10.94	-23.38	-15.96
hevc	-33.45	-37.66	-38.6	-48.57	-56.25	-55.15	hevc	-47.2	-51.05	-51.7	-58.36	-64.68	-64.58
avif-mse	-38.65	-40.77	-41.89	-52.28	-59.27	-58.59	avif-mse	-50.41	-52.59	-53.41	-60.71	-66.61	-67.33
avif-ssim	-38.76	-45.98	-46.1	-52.67	-58.39	-57.57	avif-ssim	-50.51	-56.43	-56.61	-61	-65.83	-66.39

Dataset2													
Mean BDRate-420							Mean BDRate-444						
Codec	VMAF	SSIM	MS_SSIM	VIF	PSNR_Y	PSNR_AVG	Codec	VMAF	SSIM	MS_SSIM	VIF	PSNR_Y	PSNR_AVG
jpeg-mse	16.56	35.07	30.2	5.14	-9.56	-9.24	jpeg-mse	14.57	29.62	25.82	4.85	-8.74	-6.35
jpeg-ms-ssim	-2.93	0.18	3.17	5.58	-0.91	-3.71	jpeg-ms-ssim	3.67	6.15	9.16	12.78	6.21	-2.26
jpeg-im	-1.31	0.04	1.81	3.98	1.74	-1.31	jpeg-im	2.12	2.67	4.36	7.18	5.08	-1.94
jpeg-hvs-psnr	-5.5	-8.59	-4.33	3.77	4.73	2.01	jpeg-hvs-psnr	-1.34	-4.54	-0.63	7.47	8.5	1.17
webp	-19.88	-20.01	-22.57	-31.73	-45.08	-43.57	webp	NA	NA	NA	NA	NA	NA
kakadu-mse	-39.43	-22.03	-27.18	-36.3	-50.3	-47.96	kakadu-mse	-13.69	2.66	-1.69	-10.21	-30.03	-28.11
kakadu-visual	-49.53	-44.42	-44.82	-33.23	-36.05	-34.91	kakadu-visual	-56.66	-53.48	-53.66	-42.91	-45.31	-45.26
openjpeg	-33.95	-9.21	-18.48	-32.04	-44.79	-41.22	openjpeg	-18.21	5.38	-2.6	-12.42	-28.66	-21.98
hevc	-45.41	-38.64	-43.64	-55.89	-65.48	-63.94	hevc	-53.25	-48.48	-52.19	-61.31	-69.71	-69.57
avif-mse	-51.29	-43.35	-48.1	-59.76	-68.32	-67.63	avif-mse	-57.36	-51.66	-55.13	-64.07	-71.75	-72.45
avif-ssim	-50.77	-50.27	-53.2	-60.07	-66.87	-65.92	avif-ssim	-57.28	-57.5	-59.57	-64.62	-70.63	-71.03

Dataset3													
Mean BDRate-420							Mean BDRate-444						
Codec	VMAF	SSIM	MS_SSIM	VIF	PSNR_Y	PSNR_AVG	Codec	VMAF	SSIM	MS_SSIM	VIF	PSNR_Y	PSNR_AVG
jpeg-mse	13.58	22.76	17.26	2.5	2.42	0.38	jpeg-mse	12.56	21.09	15.36	1.48	1.43	-0.45
jpeg-ms-ssim	-1.7	1.93	3.87	5.09	3.32	-1.12	jpeg-ms-ssim	4.36	7.66	9.8	11.7	10.28	-1.12
jpeg-im	-0.94	0.79	1.91	3.37	2.73	-1.28	jpeg-im	1.66	2.85	4.07	6.14	5.74	-2.03
jpeg-hvs-psnr	-5.36	-4.67	-1.11	3.39	2.3	-0.87	jpeg-hvs-psnr	-1.9	-1.36	2.19	7.21	6.22	-1.94
webp	-10.63	-16.72	-21.78	-27.04	-30.63	-30.34	webp	NA	NA	NA	NA	NA	NA
kakadu-mse	-40.4	-37.05	-37.95	-39.34	-42.28	-41.21	kakadu-mse	-10.19	-10.67	-9.89	-8.01	-12.83	-12.84
kakadu-visual	-43.22	-47.2	-45.66	-33.16	-34.24	-33.34	kakadu-visual	-50.15	-53.95	-52.43	-40.76	-41.23	-41.36
openjpeg	-28.34	-18.64	-25.55	-29.6	-28.03	-24.76	openjpeg	-3.3	4.8	-4.08	-3.05	-1.83	4.5
hevc	-37.65	-40.54	-44.85	-49.74	-52.39	-51.38	hevc	-43.77	-47.04	-50.18	-53.82	-56.1	-57.71
avif-mse	-45.61	-45.63	-49.94	-55.13	-57.89	-57.69	avif-mse	-50.69	-50.98	-54.16	-58.23	-60.68	-62.91
avif-ssim	-44.8	-47.28	-51.1	-54.95	-57.31	-57.07	avif-ssim	-50.01	-52.61	-55.36	-58.21	-60.24	-62.4

the various image compression standard along with the objective metric it is optimized for. For a more detailed description, we refer the reader to the table in [7]. The best performing codec for each objective metric is highlighted, from which the following observations can be drawn:

- 1) For D1, in terms of VMAF, Kakadu-visual results in the highest saving, while for other metrics AVIF results in the best performance.
- 2) For D2, AVIF results in the highest bitrate savings for all metrics.
- 3) For D3, except for SSIM metric for 4:4:4 chroma subsampling images, AVIF results in the best performance.
- 4) Overall, AVIF MSE optimized appears to result in more bitrate savings for D3, while for D1 and D2 it seems to vary equally between MSE and SSIM optimized.
- 5) The best performing codecs have higher bitrate savings for 4:4:4 subsampling encodes as compared to the 4:2:0 subsampling encodes.
- 6) Bitrate saving is the highest for gaming images (D2) while it is of similar magnitude for natural and synthetic images (D1 and D3).

IV. CONCLUSION

In this paper, we presented an evaluation of the newly developed AVIF image compression standard for natural, gaming and synthetic images. Our evaluation showed that overall AVIF results in the highest bitrate savings across all objective metrics for both 4:2:0 and 4:4:4 chroma subsampling encoded images for all types of images. While we limited our study here to objective quality evaluation, subjective quality evaluation methodologies are the most reliable way to measure QoE of users. Our future work will include a comparison of compression speed as well as subjective quality evaluation of the encoded images.

REFERENCES

- [1] F. Y. Shih and Y.-T. Wu, "Robust watermarking and compression for medical images based on genetic algorithms," *Information Sciences*, vol. 175, no. 3, pp. 200–216, 2005.
- [2] R. Kumar, K. Singh, and R. Khanna, "Satellite image compression using fractional fourier transform," *International Journal of Computer Applications*, vol. 50, no. 3, pp. 0975–8887, 2012.
- [3] JPEG.org, "Overview of JPEG 2000," <https://jpeg.org/jpeg2000/>, [Online; accessed 22-March-2020].
- [4] Google, "WebP Compression Study," https://developers.google.com/speed/webp/docs/webp_study, [Online; accessed 22-March-2020].

- [5] J. Lainema, M. M. Hannuksela, V. K. M. Vadakital, and E. B. Aksu, "HEVC still image coding and high efficiency image file format," in *2016 IEEE International Conference on Image Processing (ICIP)*, Sep. 2016, pp. 71–75.
- [6] AOM, "AV1 Image File Format (AVIF)," <https://aomediacodec.github.io/av1-avif/>, [Online: accessed 22-March-2020].
- [7] Netflix, "AVIF for Next-Generation Image Coding," <https://netflixtechblog.com/avif-for-next-generation-image-coding-b1d75675fe4>, [Online: accessed 22-March-2020].
- [8] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [9] N. Barman, S. Schmidt, S. Zadtootaghaj, M. G. Martini, and S. Möller, "An evaluation of video quality assessment metrics for passive gaming video streaming," in *Proceedings of the 23rd Packet Video Workshop*, ser. PV '18, Amsterdam, Netherlands, 2018, pp. 7–12.
- [10] N. Barman, S. Zadtootaghaj, M. G. Martini, S. Möller, and S. Lee, "A Comparative Quality Assessment Study for Gaming and Non-Gaming Videos," in *Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, Sardinia, Italy, May 2018.
- [11] N. Barman, S. Zadtootaghaj, S. Schmidt, M. G. Martini, and S. Möller, "GamingVideoSET: A Dataset for Gaming Video Streaming Applications," in *2018 16th Annual Workshop on Network and Systems Support for Games (NetGames)*, Amsterdam, Netherlands, 2018, pp. 1–6.
- [12] A. V. Katsenou, G. Dimitrov, D. Ma, and D. Bull, "BVI-SynTex: A Synthetic Video Texture Dataset for Video Compression and Quality Assessment," *IEEE Transactions on Multimedia*, pp. 1–1, 2020, early Access.
- [13] Netflix, "Image Compression Comparison Framework," https://github.com/Netflix/image_compression_comparison, [Online: accessed 22-March-2020].