AVM Common Test Condition Latest Test Result Update

Ryan Lei, AOM Testing Sub-Group

AOMedia Workshop @ ICIP 2024

Abu Dhabi, United Arab Emirates, October 30, 2024

Outline

- AOM Testing Subgroup Activities

- AVM v8.0 Latest CTC Test Result

- AVM Subjective Video Quality Testing Result

AOM Testing Subgroup Activities

- Define Common Test Conditions for evaluating coding gain of AVM tools
  - Test sequences, encoding configurations, etc.
  - Quality and performance metrics.
  - Test script and infrastructure.
  - CTC Test Sequences can be found at: https://media.xiph.org/video/aomctc/test_set/
  - Latest AOM CTC v7.0 can be found at:
    https://groups.aomedia.org/g/wg-codec/files/InputDocuments/E2024/CWG-E083/CWG-E083_AOM_CTC_v7.pdf
- Conduct regular test and Tools on/off test after each anchor release.
- Define and conduct subjective test.

AVM CTC Encoding Configurations

- All Intra and Still Image:
  - Mainly used for encoding Key frames and still images.
  - Encoding first 30 frames from all classes as Intra frames

- Low Delay:
  - Mainly used for low latency applications, such as video conferencing, real-time communications, etc.
  - Encoding 130 frames (with only 1 Key frame) without future reference frames

- Random Access:
  - Mainly used for high latency applications, such as VOD, etc.
  - Encoding 130 frames (2 Closed GOPs) with future reference frames

- Adaptive Streaming:
  - Mainly used for convex-hull based high latency applications, such as VOD, adaptive streaming, etc.
  - 4K sequences in A1 class are downscaled to 5 different resolutions.
  - Similar encoding configuration as Random Access is used for encoding.

AVM v8.0 Test Result: All Intra

- AVM v8.0 anchor was released in September 2024. Can be accessed at https://gitlab.com/AOMediaCodec/avm/-/tree/research-v8.0.0?ref_type=tags

	Averages Over research-alt-v1-anchor_r3.0					
	PSNR		SSIM	VMAF	Run-time (%)	
Summary	Y	YUV	Y	Y	Enc-time	Dec-time
Class A1_4K	-15.54%	-17.60%	-12.61%	-17.04%	2538%	334%
Class A2_2K	-14.37%	-16.48%	-12.58%	-15.56%	4730%	311%
Class A3_720p	-15.97%	-18.13%	-15.04%	-16.70%	6552%	296%
Class A4_360p	-13.56%	-16.69%	-15.35%	-13.76%	5990%	329%
Class A5_270p	-11.72%	-13.66%	-12.96%	-14.26%	7117%	303%
Class B1_SYN	-19.23%	-21.40%	-17.75%	-19.26%	4432%	291%
Overall Mandatory w/o B2	-15.37%	-17.60%	-14.21%	-16.32%	4732%	309%
Class B2_SCC	-34.15%	-36.05%	-37.02%	-34.36%	2407%	287%

Intra

AVM v8.0 Test Result: Random Access

Summary	Averages Over research-alt-v1-anchor_r3.0					
	PSNR		SSIM	VMAF	Run-time (%)	
	Y	YUV	Y	Y	Enc-time	Dec-time
Class A1_4K	-25.74%	-27.38%	-23.00%	-30.81%	2168%	589%
Class A2_2K	-23.32%	-24.96%	-21.86%	-28.21%	2395%	487%
Class A3_720p	-25.55%	-27.46%	-24.54%	-29.42%	2321%	584%
Class A4_360p	-23.81%	-25.99%	-24.06%	-27.64%	2633%	673%
Class A5_270p	-23.27%	-24.52%	-23.18%	-28.69%	3960%	521%
Class B1_SYN	-22.72%	-24.66%	-23.80%	-27.38%	2231%	444%
Overall Mandatory w/o B2	-23.93%	-25.70%	-23.10%	-28.58%	2431%	526%
Class B2_SCC	-36.28%	-37.67%	-37.83%	-38.27%	1271%	277%

RA

AVM v8.0 Test Result: Low Delay

	PSNR		SSIM	VMAF	Run-time (%)	
Averages Over research-alt-v1-anchor_r3.0						
Summary	Y	YUV	Y	Y	Enc-time	Dec-time
Class A2_2K	-16.67%	-18.58%	-15.45%	-18.72%	2926%	238%
Class A3_720p	-17.94%	-19.73%	-16.72%	-18.50%	2372%	234%
Class A4_360p	-19.07%	-21.54%	-19.79%	-18.04%	2939%	271%
Class A5_270p	-15.77%	-17.28%	-15.50%	-19.77%	4408%	267%
Class B1_SYN	-17.37%	-19.73%	-18.10%	-18.91%	2786%	239%
Overall Mandatory w/o B2	-17.27%	-19.29%	-16.79%	-18.73%	2895%	244%
Class B2_SCC	-37.92%	-39.17%	-39.19%	-41.00%	1087%	207%

AVM v8.0 Test Result: Still Images

Overall (Still in		Averages Over research-alt-v1-anchor_r3.0						
		PSNR		SSIM	VMAF	Run-time (%)	
	Summary	Y	YUV	Y	Y	Enc-time	Dec-time
	F1 High Resolutions	-8.46%	-10.76%	-7.80%	-10.56%	3319.2%	265.4%
	F2 Medium Resolution	-10.92%	-12.77%	-10.61%	-11.13%	6152.2%	284.2%

AVM Subjective Testing Result

- Two independent subjective tests were conducted by Google and AOM Testing Subgroup
- AOM Testing Subgroup Test (CWG-E004)
  - 8 HD and UHD sequences are used
  - AVM v5.0 vs AV1, matching quality.
  - Using Degradation Category Rating (DCR) methodology with 11 grade impairment scale.
  - Expert viewing on PCs with almost lossless compressed mp4 files.

Video / BD-rates	PSNR Y (%)	PSNR U (%)	PSNR V (%)	PSNR YUV (%)	VMAF (%)	MOS (%)
BarScene_1920x1080	-21.73	-38.35	-38.53	-23.24	-27.09	-24.54
GregoryCactus_1080x1920	-14.73	-36.34	-37.81	-17.25	-18.57	-19.63
Metro_1920x1080	-18.55	-28.68	-37.09	-20.08	-20.98	-30.08
Marathon2_3840x2160	-16.76	-25	-23.6	-17.9	-20.07	-32.45
Meridian_3840x2160	-29.28	-37.11	-45.89	-30.32	-40.13	-47.45
MountainBay2_3840x2160	-15.81	-31.59	36.01	-18.21	-27.76	-25.68
TallBuildings2_3840x2160	-19.51	-47.12	-59.23	-24.13	-20.4	-29.97
YonseiS01_3840x2160	-19.23	-21.16	-26.59	-19.58	-21.85	-33.21
Average	**-19.45**	**-33.17**	**-29.09**	**-21.34**	**-24.61**	**-30.38**

AVM Subjective Testing Result

- Google Test ([CWG-E196](CWG-E196))
  - 5 UHD and 3HD test sequences are used, 6 seconds to 10 seconds duration.
  - AVM v8 anchor vs AV1, bitrate matching by adjusting encoding QPs.
  - Degradation Category Rating (DCR) methodology was used.
  - Onsite viewing on UHD 10bit TVs with decoded YUV.

Sequences	PSNR BD-rate				VMAF BD-rate	VMAF neg BD-rate	On-site MOS	
	Y	U	V	YUV (14:1:1)			BD-rate	CI
Marathon2	-18.78%	-26.88%	-24.22%	-19.79%	-21.94%	-20.76%	**-20.04%**	0.55
MountainBay2	-19.86%	-38.22%	-40.98%	-22.49%	-29.78%	-27.32%	**-19.01%**	0.51
TallBuildings2	-26.42%	-56.13%	-68.70%	-31.38%	-26.17%	-25.70%	**-31.22%**	0.57
YonseiS01	-20.17%	-25.67%	-29.87%	-20.76%	-23.71%	-22.51%	**-42.79%**	0.58
meridian	-29.66%	-42.66%	-63.27%	-31.50%	-40.83%	-37.05%	**-56.15%**	0.55
Average	**-22.98%**	**-37.91%**	**-45.41%**	**-25.19%**	**-28.49%**	**-26.67%**	**-33.84%**	**0.55**

Sequences	PSNR BD-rate				VMAF BD-rate	VMAF neg BD-rate	On-site MOS	
	Y	U	V	YUV (14:1:1)			BD-rate	CI
BarScene	-26.07%	-41.04%	-43.36%	-27.57%	-29.95%	-28.47%	**-23.35%**	0.46
GregoryCactus	-21.81%	-48.44%	-48.01%	-25.30%	-25.60%	-25.02%	**-29.22%**	0.48
Metro	-18.83%	-28.16%	-40.66%	-20.53%	-19.77%	-19.49%	**-24.73%**	0.55
Average	**-22.24%**	**-39.21%**	**-44.01%**	**-24.47%**	**-25.10%**	**-24.32%**	**-25.77%**	**0.50**

Reference

- AOM CTC v7.0:
  https://groups.aomedia.org/g/wg-codec/files/InputDocuments/E2024/CWG-E083/CWG-E083_AOM_CTC_v7.pdf
- Encoder and Decoder code base:
  - AOM Video Model: https://gitlab.com/AOMediaCodec/avm
  - Latest release is v.8.0.0.
- HDRTools is used for downscaling and upscaling in adaptive streaming configuration.
  - https://gitlab.com/standards/HDRTools/-/tree/v0.22
- VMAF v2.3.1 is used for quality metrics calculation.
- High level python script for CTC Test is available under
  - https://gitlab.com/AOMediaCodec/avm/-/tree/main/tools/convexhull_framework
- Excel spreadsheet template for presenting and reporting the CTC result is available under
  - https://gitlab.com/AOMediaCodec/avm/-/tree/main/tools/convexhull_framework/bin
- SVT-AV1: https://gitlab.com/AOMediaCodec/SVT-AV1
- AWCY: https://beta.arewecompressedyet.com/

Thank You