



Advanced Video Quality Tool (AVQT)

Introduction

The Advanced Video Quality Tool, or AVQT in short, is a macOS command line tool, which estimates perceptual video quality of compressed videos, that might contain video coding and scaling artifacts. The tool expects reference (source) and test (compressed) videos as inputs and outputs frame level and segment level scores of the test video on a scale of 1 to 5, where 1 indicates bad video quality and 5 indicates excellent video quality. You can also use AVQT to compute MSE (Mean Squared Error) and PSNR (Peak Signal to Noise Ratio) scores.

System Requirements

The Advanced Video Quality Tool (AVQT) package requires an Intel based or Apple Silicon based Mac running macOS Big Sur (version 11.0) or later.

Installation

AVQT is made available for download via the developer tools portal (<https://developer.apple.com/download/more/>). The downloaded image consists of AVQT.pkg, README, License and Acknowledgement files. To install AVQT, double click on the AVQT.pkg file and follow the instructions on screen. You can type "man AVQT" in Terminal to read more about the tool.

Usage

AVQT [<Options>] --test <test-video> --reference <reference-video>

Options:

--test (or -t)

Path to test video file. The video file can be containerized (.mov, .mp4, etc.) or raw YUV. (e.g. --test /Users/sam/videos/movie1_hls_tier_1.mov)

--reference (or -r)

Path to reference (source) video file. The video file can be containerized (.mov, .mp4, etc.) or raw YUV. (e.g. --reference /Users/sam/videos/source.mov)

--metrics (or -m)

A list of one or more space separated video quality metrics. In addition to the default metric (AVQT), the tool also supports MSE and PSNR (e.g. --metrics AVQT PSNR MSE).

--segment-duration (or -sd)

Duration of each segment in seconds. Frame scores within a segment will be aggregated using a temporal pooling method to compute segment score. The default value is 6.0 seconds. The allowed range lies between 0.5 and 60 seconds. (e.g. --segment-duration 5.0)

--temporal-pooling (or -tp)

Temporal pooling algorithm to be used for aggregating frame scores within a segment. Options include {ArithmeticMean, HarmonicMean, Worst10%, Worst90%, Best10%, Best90%}. The default algorithm is ArithmeticMean. (e.g. --temporal-pooling HarmonicMean)

--output

Path to the results file containing scores. A time-stamped filename will be used if not specified. (e.g. --output movie1_hls_tier_1.csv)

--output-format (or -of)

Format of the results file containing scores. Options include {CSV, JSON}. The default value is CSV. (e.g. --output-format JSON)

--display-resolution (or -dr)

Resolution of display in <width>x<height> format. The default value is set to 1920x1080. Options include 6016x3384, 5120x2880, 3840x2160, 2560x1440, 1920x1080, 3384x6016, 2880x5120, 2160x3840, 1440x2560, 1080x1920. See *Viewing Distance and Display Resolution* section below to set this parameter correctly. (e.g. --display-resolution 3840x2160)

--viewing-distance (or -vd)

Viewing distance in multiples of display height. Options include {1.5H, 3H, 4.5H, 6H}. The default value is set to 3H. See *Viewing Distance and Display Resolution* section below to set this parameter correctly. (e.g. --viewing-distance 1.5)

--force-analysis-resolution (or --far)

Analysis resolution in <width>x<height> format to be used for all metrics. The display-resolution and viewing-distance values will be ignored.

--force (or -f)

Bypass frame rate, aspect-ratio, video duration and transfer characteristics equality checks between reference and test video. AVQT requires spatial and temporal alignment between the reference and test videos. You can use this flag to disable this requirement.

--test-resolution (or -tr)

Resolution of test video in <width>x<height> format. This is required only if test video is raw YUV. (e.g. --test-resolution 960x540)

--reference-resolution (or -rr)

Resolution of reference video in <width>x<height> format. This is required only if reference video is raw YUV. (e.g. --reference-resolution 1920x1080)

--test-fourcc (or -tf)

Pixel format of test video. Options include {I420 (IYUV), NV12, NV21, YV12, IYUV10}. This is required only if test video is raw YUV. (e.g. --test-fourcc I420)

--reference-fourcc (or -rf)

Pixel format of reference video. Options include {I420 (IYUV), NV12, NV21, YV12, IYUV10}. This is required only if reference video is raw YUV. (e.g. --reference-fourcc IYUV10)

--test-framerate (or -tfr)

Frame rate of test video. This is required only if test video is raw YUV. (e.g. --test-framerate 30.0)

--reference-framerate (or -rfr)

Frame rate of reference video. This is required only if reference video is raw YUV.
(e.g. --reference-framerate 30.0)

--test-transferCharacteristics (or -ttc)

Transfer characteristics of test video. This is required only if test video is raw YUV. Options include {ITU_R_709_2, ITU_R_2100_HLG, SMPTE_ST_2084_PQ, IEC_sRGB, SMPTE_ST_428_1, ITU_R_2020, SMPTE_240M_1995}. (e.g. --test-transferCharacteristics SMPTE_ST_2084_PQ). For more information, see [AVVideoTransferFunctionKey](#)

--reference-transferCharacteristics (or -rtc)

Transfer characteristics of reference video. This is required only if reference video is raw YUV. Options include {ITU_R_709_2, ITU_R_2100_HLG, SMPTE_ST_2084_PQ, IEC_sRGB, SMPTE_ST_428_1, ITU_R_2020, SMPTE_240M_1995}. (e.g. --reference-transferCharacteristics SMPTE_ST_2084_PQ). For more information, see [AVVideoTransferFunctionKey](#)

--version

Show the version

--help (or -h)

Show help information

Example Invocations

Videos in a container format

```
AVQT --test /Users/sam/videos/test.mov --reference /Users/sam/
videos/source.mov --segment-duration 6.0 --viewing-distance 3H
--display-resolution 1920x1080 --metrics PSNR AVQT --output-
format JSON --output movie1_hls_tier1_avqt_psnr.json
```

Raw YUV videos

```
AVQT --test ./videos/movie1_hls_tier1.yuv --reference /Users/
sam/videos/source.yuv --test-resolution 1280x720 --test-fourcc
I420 --test-framerate 30.0 --test-transferCharacteristics
SMPTE_ST_2084_PQ --reference-resolution 1920x1080 --reference-
fourcc IYUV10 --reference-framerate 30.0 --reference-
transferCharacteristics SMPTE_ST_2084_PQ --output
movie1_hls_tier1_avqt.csv
```

Score Interpretation

- AVQT scores align with the MOS (Mean Opinion Score) scale. The score is a floating point number between 1 and 5, where 1 is BAD, 2 is POOR, 3 is FAIR, 4 is GOOD and 5 is EXCELLENT quality.
- For computing MSE scores, the luma plane in reference and test videos are converted to normalized, full range, float values in the range 0-1. This results in MSE scores in the range 0-1, enabling comparison of videos with different bit-depths.
- Frame level PSNR scores are capped at a maximum of 100 dB.

Viewing Distance and Display Resolution

Viewing setup can affect the video quality perceived by end users. Factors such as viewing distance, screen size and display resolution can mask or exaggerate artifacts in the compressed video. AVQT uses the viewing-distance and display-resolution flags to model the impact of viewing setup while predicting video quality. The viewing distance is measured in multiples of active screen height (H).

AVQT supports viewing distance values of 1.5H, 3H, 4.5H and 6H and display resolutions of 6016x3384, 5120x2880, 3840x2160, 2560x1440 and 1920x1080 in both portrait and landscape orientations. While AVQT is tuned for two viewing setups, 1920x1080 at 3H and 3840x2160 at 1.5H, it captures the trends for other cases correctly. The default values for viewing distance and display resolution are 3H and 1920x1080 respectively. To disable the viewing setup model in AVQT, you can use the force-analysis-resolution flag to set the resolution to be used by all metrics for both reference and test videos.

Performance

Accuracy

Subjective Dataset	Number of test videos	Content category	PCC	SROCC	RMSE
Waterloo IVC 4K [1]	480	Nature, animation, screen content, sports, people, food	0.91	0.90	0.54
VQEG HD3 [2]	72	Sports, nature, advertisement, animation, home video	0.94	0.93	0.36
VQEG HD5 [2]	49	Sports, wildlife, animation, nature	0.81	0.83	0.71

Subjective Dataset	Number of test videos	Content category	PCC	SROCC	RMSE
LIVE Mobile [3]	40	People, faces, nature, advertisement	0.91	0.92	0.50
Netflix Public [4]	70	Animation, people, dance, sports, nature	0.95	0.94	0.40
Apple Internal	64	Premium movie content	0.96	0.95	0.53

*PCC = Pearson Correlation Coefficient

*SROCC = Spearman Rank Ordered Correlation Coefficient

*RMSE = Root Mean Squared Error

[1] Z. Li, Z. Duanmu, W. Liu and Z. Wang, "AVC, HEVC, VP9, AVS2 or AV1? — A Comparative Study of State-of-the-art Video Encoders on 4K Videos", 16th International Conference on Image Analysis and Recognition, Waterloo, Ontario, Canada, August 27-29, 2019.

[2] Video Quality Experts Group (VQEG), VQEG HDTV Phase I, <https://www.its.bldrdoc.gov/vqeg/projects/hdtv/hdtv.aspx>.

[3] A. K. Moorthy, L. K. Choi, A. C. Bovik, and G. De Veciana, "Video quality assessment on mobile devices: Subjective, behavioral and objective studies," IEEE J. Select. Topics Signal Process., vol. 6, no. 6, pp. 652–671, 2012.

[4] Z. Li, A. Aaron, I. Katsavounidis, A. Moorthy, and M. Manohara, "Toward a practical perceptual video quality metric." <http://techblog.netflix.com/2016/06/toward-practical-perceptual-video.html>.

Speed

Video Resolution	Speed (FPS)
3840x2160	63
1920x1080	175
1280x720	261
960x540	335

*Speed results obtained on a 2020 M1 MacBook Pro

Other Notes

- AVQT is designed for videos with compression and scaling artifacts only. As such, network related artifacts such as jitter, frame drop and packet loss are currently not supported.

- AVQT analyzes only the luma plane for all video quality metrics (AVQT, PSNR, MSE). In most cases involving video compression and scaling, there is a high correlation between the artifacts in the luma plane and chroma planes.
- AVQT assumes both reference (source) and test (compressed) videos to be spatially and temporally aligned. For example, pixel shifts and frame drops/duplicates are not handled by AVQT.
- To evaluate Dolby Vision Profile 5 content, currently, both reference (source) and test (compressed) videos have to be Dolby Vision Profile 5.
- AVQT natively supports all video formats and codecs supported by AVFoundation. To extend support for other codecs used in professional video workflows, such as Apple ProRes RAW, Avid DNxHD, AVC-Intra 50, etc., download and install the Pro Video Formats package from <https://support.apple.com/kb/DL2050>

Change History

version 1.0.0 (June 09, 2021)

- First release