

1. STUDY CONDITIONS

Our subjective comparison of video codecs method is similar to that of our prior objective comparisons. Instead of objective SSIM quality scores, however, we employ subjective scores obtained from a crowdsourced online study conducted using the [Subjectify.us](https://www.subjectify.us) platform (a description of which appears in Section 1.2).

The method comprises three main steps:

1. Video encoders launches (see Section 1.3)
2. Subjective-score estimation (see Section 1.4)
3. Integral-score estimation (see Section 1.5)

1.1. Summary

Encoders under comparison: Ten video encoders with preselected command-line arguments that deliver at least a 1 FPS encoding speed.

Codec	Developer	Version
Kingsoft HEVC Encoder	Kingsoft	2.7.1.1
SIF Encoder	SIF Encoder Team	1.58
sz264	Nanjing Yunyan	0.5
sz265	Nanjing Yunyan	0.5
Tencent Shannon Encoder	Tencent	1.0
UC265	Ucodec Inc.	v1.0.4 (64 bit)
VITEC HEVC GEN2+ Encoder¹	VITEC	
VP9	The WebM Project	v1.7.0-507-g19d31c70a
x264	x264 Developer Team	r2901-7d0ff22
x265	MulticoreWare, Inc.	2.7+19-1fafca24a399

Table 1: Short codecs' descriptions

Test video sequences: Five Full HD video sequences with frame rates of 24–50 FPS. See the complete list in Section 2.

Sequence	Number of frames	Frame rate	Resolution
1. Crowd Run (short)	500	50	1920×1080
2. Ducks Take Off (short)	500	50	1920×1080
3. Mountain Bike (short)	500	24	1920×1080
4. Playground (short)	500	30	1920×1080
5. Red Kayak (short)	500	25	1920×1080

Table 2: Summary of video sequences

¹It is a hardware encoder. We tested it's software implementation. Encoding speed of a hardware implementation was provided by developers.

Encoding bitrates: 1 Mbps, 2 Mbps and 4 Mbps.

Test hardware: All codecs ran on an Core i7 8700K (Coffee Lake) @ 3.7Ghz RAM 16 GB Windows 10.

Computation of subjective quality scores: Using the Subjectify.us platform, we showed study participants pairs of videos encoded at various bitrates by the codecs under evaluation. We asked them to choose the video with the best visual quality from each pair. To filter out responses from participants who made thoughtless decisions, we also asked them hidden quality-control questions. We collected 22542 valid answers from 473 unique participants and converted pairwise responses to subjective scores using the Bradley-Terry model [1]. A detailed description of this step appears in Section 1.4.

Computation of integral quality and speed scores: To summarize an encoder’s performance at multiple bitrates, we computed relative quality and speed scores. The relative quality score is the test encoder’s mean bitrate divided by the reference encoder’s mean bitrate for the same range of quality scores. The relative speed score is the test encoder’s mean encoding speed divided by the reference encoder’s mean speed for the same bitrate range. A detailed description of integral-score computation appears in Section 1.5.

1.2. About Subjectify.us

We obtained the subjective scores for this study using Subjectify.us. This platform enables researchers and developers to conduct crowdsourced subjective comparisons of image- and video-processing methods (e.g., compression, inpainting, denoising, matting, etc.) and carry out studies of human quality perception.



To conduct a study, researchers must apply the methods under comparison to a set of test videos (images), upload the results to Subjectify.us and write a task description for study participants. Subjectify.us handles all the laborious steps of a crowdsourced study: it recruits participants, presents uploaded content in a pairwise fashion, filters out responses from participants who cheat or are careless, analyzes collected results, and generates a study report with interactive plots. Thanks to the pairwise presentation, researchers need not invent a quality scale, as study participants just select the best option of the two. The platform is optimized for comparison of large video files: it prefetches all videos assigned to a study participant and loads them into his or her device before asking the first question. Thus, even participants with a slow Internet connection won’t experience buffering events that might affect quality perception.

Try Subjectify.us in your research project at www.subjectify.us. This [demo video](#) shows an overview of the Subjectify.us workflow.

1.3. Running Codecs

In this study we compare the same encoders as in our previous report, [HEVC/H.265 Video Codecs Comparison 2018](#), using the same command-line arguments from the “Ripping” use case (i.e., a minimum encoding speed of 1 FPS).

We applied software encoders with preselected command-line arguments (see the full list of codecs in Appendix ??) to five Full HD test video sequences (see the full list in Appendix 2) at three bitrates: 1 Mbps, 2 Mbps and 4 Mbps.

All encoders ran on a computer with an Core i7 8700K (Coffee Lake) @ 3.7Ghz RAM 16 GB Windows 10. The source and encoded files, the encoder executable, and the operation system all resided on an SSD. We ran every encoder three times and recorded the duration of the fastest run for each.

1.4. Subjective-Score Estimation

To conduct an online crowdsourced comparison, we uploaded encoded streams from the previous step to Subjectify.us. The platform hired study participants and showed the upload streams to them in pairs. Each pair consisted of two variants of the same test video sequence encoded by various codecs at various bitrates. Videos from each pair were presented to study participant sequentially (i.e., one after another) in full-screen mode. After viewing each pair, participants were asked to choose the video with the best visual quality. They also had the option to play the videos again or to indicate that the videos have equal visual quality. We assigned each study participant 10 pairs, including 2 hidden quality-control pairs, and each received money reward after successfully completing the task. The quality-control pairs consisted of test videos compressed by the x264 encoder at 1 Mbps and 4 Mbps. Responses from participants who failed to choose the 4 Mbps sequence for one or more quality-control questions were excluded from further consideration. Study participants could take part in the study again after 12 hours break. In total we collected 22542 valid answers from 473 unique participants.

To convert the collected pairwise results to subjective scores, we used the Bradley-Terry model [1]. Thus, each codec run received a quality score. We then linearly interpolated these scores to get continuous rate-distortion (RD) curves, which show the relationship between the real bitrate (i.e., the actual bitrate of the encoded stream) and the quality score.

1.5. Integral-Score Estimation

To compare not just individual encoder runs but all runs for a single sequence, and to obtain an overall score for each encoder, we computed integral relative scores: the relative quality score and the relative speed score.

Relative quality score is the test encoder's mean bitrate divided by a reference encoder's mean bitrate for the same range of quality scores. The relative quality score $x\%$ for test encoder A means the following: A must deliver $x\%$ of the reference encoder's bitrate to achieve the same visual quality. To compute this score, we employ the following procedure:

1. Transpose the RD curve for both the test codec and reference codec (see Figures 1a and 1b).
2. Find the RD curves' projection onto the quality axis—that is, the largest quality range for which both curves are defined (see Figure 1b).
3. Compute the area under the curves for the quality range from the previous step (see Figure 1c).
4. Define the relative quality score as the area under the test codec's RD curve divided by the area under the reference codec's RD curve.
5. Additionally, to score the estimate from the previous step, define the confidence as the length of the quality range used to compute the area divided by the length of quality range for which the reference RD curve is defined.

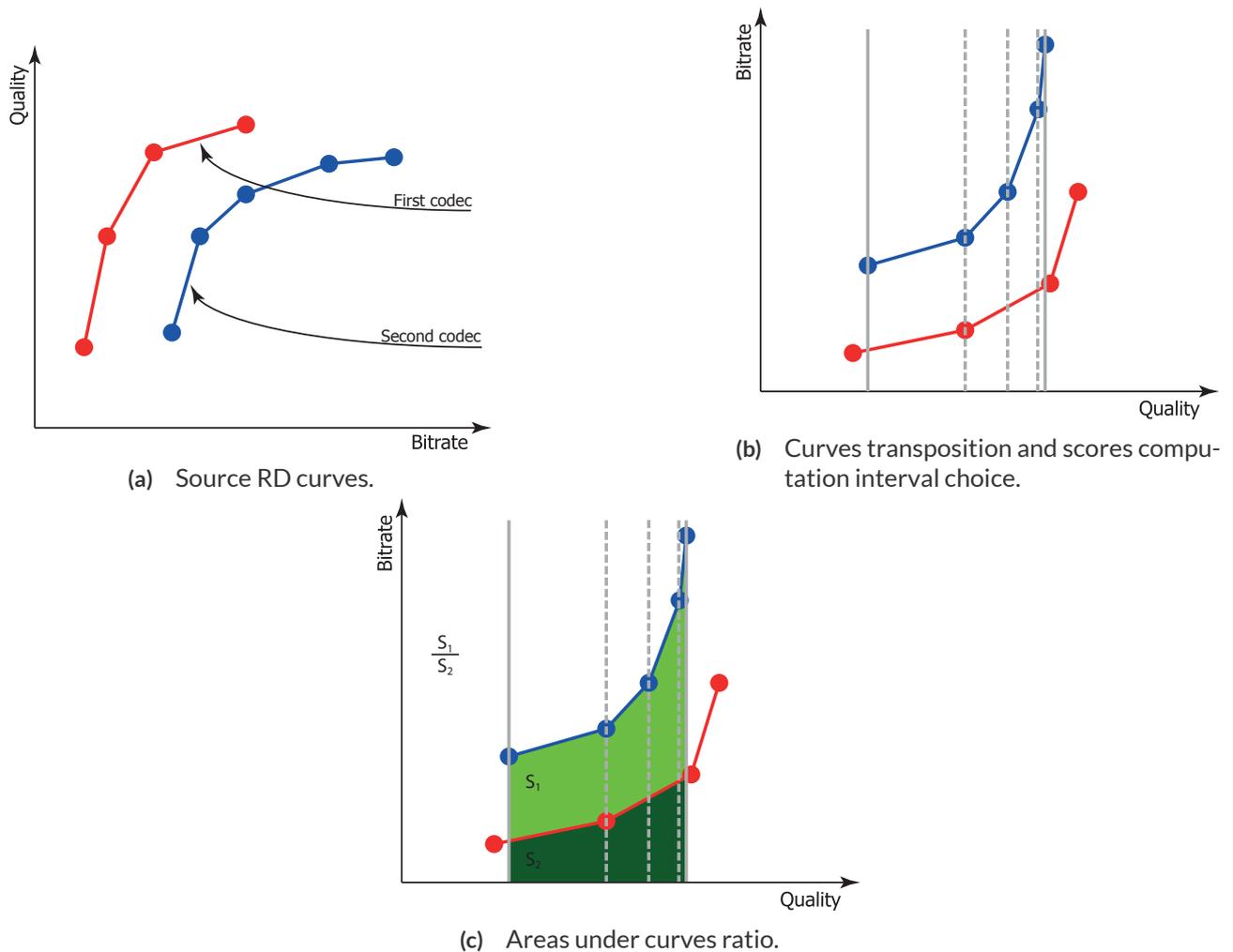


Figure 1: Relative quality score computation.

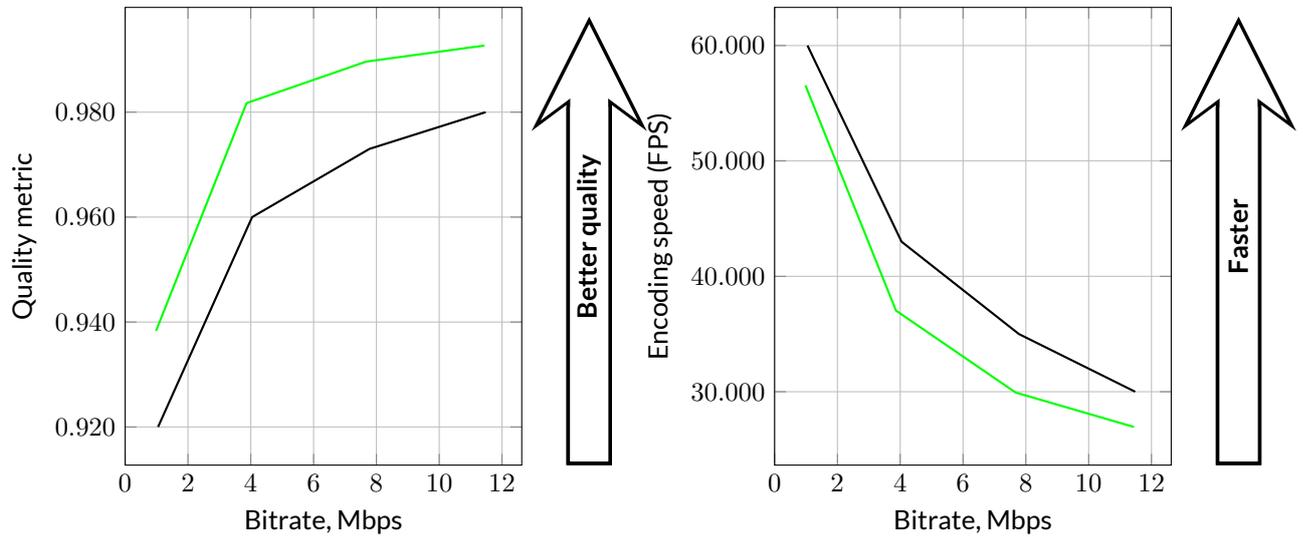
To compute an overall quality score for the test encoder, we average its relative quality scores for the individual test sequences.

Relative speed score is the mean encoding speed of the test encoder divided by the mean encoding speed of the reference encoder for the same bitrate range. To compute this score, we employ the same procedure as above but compute the area under the encoding-speed curves, rather than the RD curves, along the bitrate axis.

1.5.1. Speed/Quality Trade-Off Plots

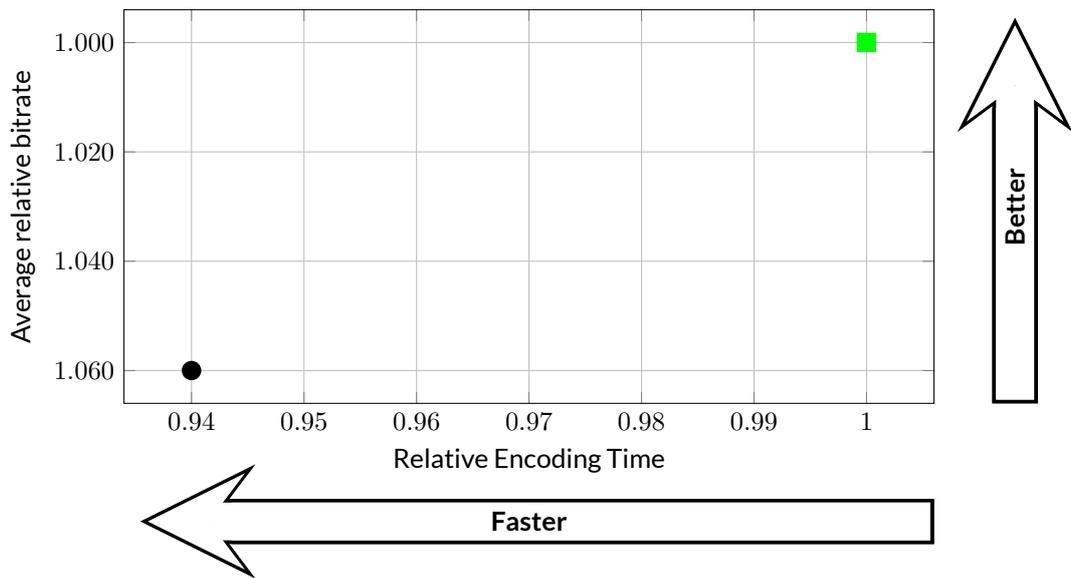
To compare both relative quality and relative speed scores, we show speed/quality trade-off plots (the x-axis corresponds to the speed score and the y-axis to quality score). These plots enable us to see whether a codec won first place in both categories (speed and quality). If no absolute winner emerges, the plot helps in finding Pareto-optimal encoders (i.e., encoders for which no competitor has a higher score in both speed and quality).

Consider the simplified example in Figure 2. As Figure 2a shows, the “green” codec outperforms the “black” codec in quality scores. But the black codec is faster, according to Figure 2b. We can make the same observation at a glance by considering the speed/quality trade-off plot in Figure 2c: green earned a higher quality score by (possi-



(a) RD curve.

(b) Encoding speed (frames per second).



(c) Trade-off plot.

Figure 2: Speed/quality trade-off example.

2. SEQUENCES

2.1. Crowd Run (short)

Sequence title	Crowd Run (short)
Resolution	1920×1080
Number of frames	500
Color space	YV12
Frames per second	50
Source	https://media.xiph.org/video/derf/
Source resolution	FullHD
Bitrate	1244.16

A crowd of sportsmen runs while the camera slowly moves left and right.



Figure 3: Crowd Run (short) sequence, frame 50

2.2. Ducks Take Off (short)

Sequence title	Ducks Take Off (short)
Resolution	1920×1080
Number of frames	500
Color space	YV12
Frames per second	50
Source	https://media.xiph.org/video/derf/
Source resolution	FullHD
Bitrate	994.334

The flock of ducks takes off the pond.

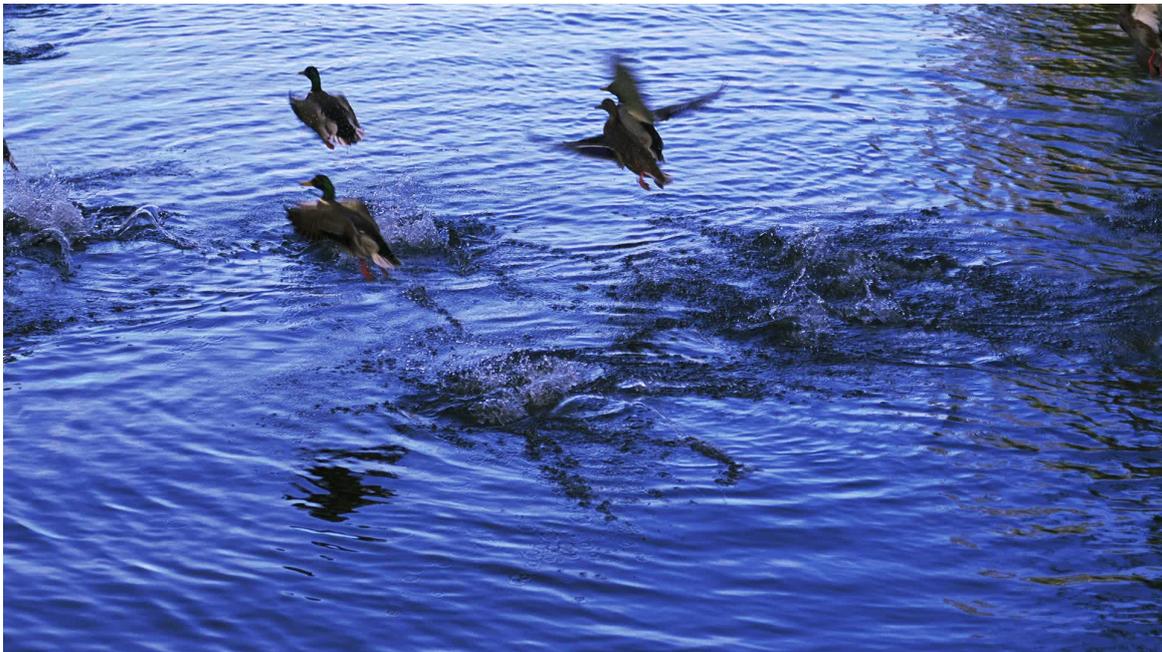


Figure 4: Ducks Take Off (short) sequence, frame 110

2.3. Mountain Bike (short)

Sequence title	Mountain Bike (short)
Resolution	1920×1080
Number of frames	500
Color space	YV12
Frames per second	24
Source	https://vimeo.com/188799676#t=38
Source resolution	FullHD
Bitrate	71.226

The sequence films bikers riding in the forest. Consists of quadcopter shooting, slowmotion and close-up shots.



Figure 5: Mountain Bike (short) sequence, frame 25

2.4. Playground (short)

Sequence title	Playground (short)
Resolution	1920×1080
Number of frames	500
Color space	YV12
Frames per second	30
Source	https://vimeo.com/219044636#t=63
Source resolution	FullHD
Bitrate	143.927

Children are walking on a playground with grass and playing in a sandbox. Captured with hand-held camera.



Figure 6: Playground (short) sequence, frame 344

2.5. Red Kayak (short)

Sequence title	Red Kayak (short)
Resolution	1920×1080
Number of frames	500
Color space	YV12
Frames per second	25
Source	https://media.xiph.org/video/derf/
Source resolution	FullHD
Bitrate	994.334

Scenes with splashing water and a man who is kayaking on a river.



Figure 7: Red Kayak (short) sequence, frame 390

3. ABOUT THE GRAPHICS & MEDIA LAB VIDEO GROUP



The Graphics & Media Lab Video Group is part of the Computer Science Department of Lomonosov Moscow State University. The Graphics Group began at the end of 1980's, and the Graphics & Media Lab was officially founded in 1998. The main research avenues of the lab include areas of computer graphics, computer vision and media processing (audio, image and video). A number of patents have been acquired based on the lab's research, and other results have been presented in various publications.

The main research avenues of the Graphics & Media Lab Video Group are video processing (pre- and post-, as well as video analysis filters) and video compression (codec testing and tuning, quality metric research and codec development).

The main achievements of the Video Group in the area of video processing include:

- High-quality industrial filters for format conversion, including high-quality deinterlacing, high-quality frame rate conversion, new, fast practical super resolution and other processing tools.
- Methods for modern television sets, such as a large family of up-sampling methods, smart brightness and contrast control, smart sharpening and more.
- Artifact removal methods, including a family of denoising methods, flicking removal, video stabilization with frame edge restoration, and scratch, spot and drop-out removal.
- Application-specific methods such as subtitle removal, construction of panorama images from video, video to high-quality photo conversion, video watermarking, video segmentation and practical fast video deblur.

The main achievements of the Video Group in the area of video compression include:

- Well-known public comparisons of JPEG, JPEG-2000 and MPEG-2 decoders, as well as MPEG-4 and annual H.264 codec testing; codec testing for weak and strong points, along with bug reports and codec tuning recommendations.
- Video quality metric research; the MSU Video Quality Measurement Tool and MSU Perceptual Video Quality Tool are publicly available.
- Internal research and contracts for modern video compression and publication of MSU Lossless Video Codec and MSU Screen Capture Video Codec; these codecs have one of the highest available compression ratios.

The Video Group has also worked for many years with companies like Intel, Samsung and RealNetworks.

In addition, the Video Group is continually seeking collaboration with other companies in the areas of video processing and video compression.

E-mail: video@graphics.cs.msu.ru



4. References

- [1] Ralph Allan Bradley and Milton E Terry. "Rank analysis of incomplete block designs: I. The method of paired comparisons". In: *Biometrika* 39.3/4 (1952), pp. 324–345.