# MSU Codec Comparison 2019

*Part II: FullHD Content, Subjective Evaluation*


---


| | |
|---|---|
| Video group head | Dr. Dmitriy Vatolin |
| Project head | Dr. Dmitriy Kulikov |
| Measurements & analysis | Dr. Mikhail Erofeev, Anastasia Antsiferova, Sergey Zvezdakov, Denis Kondranin, Stanislav Grokholskiy |


Free version

## Codecs:

**H.265**

- Bytedance
- sz265
- Tencent V265 Encoder
- UC265
- x265
- xin265

**Non H.265**

- arowana xvc
- SIF Encoder
- VP9
- WZAurora AV1 Encoder
- x264


CS MSU Graphics & Media Lab, Video Group
November 2, 2019
http://www.compression.ru/video/codec_comparison/index_en.html
videocodec-testing@graphics.cs.msu.ru

# Contents

# 1. ACKNOWLEDGMENTS

The Graphics & Media Lab Video Group would like to thank the following companies for providing the codecs and settings used in this report:

- ByteDance Inc.
- Divideon
- MulticoreWare, Inc.
- Nanjing Yunyan
- Peppa
- SIF Encoder Team
- Tencent
- Ucodec Inc.
- Visionular
- x264 Developer Team

We're also grateful to these companies for their help and technical support during the tests.

# 2. INTRODUCTION

In this report we describe our subjective comparison of video codecs using a method 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 platform (a description of which appears in Section 2.2). We provide a detailed description of the study and score-computation method in Appendix A, as well as a short summary of the study conditions in Section 2.1. To show that our study's crowdsourcing approach is accurate enough to compare video encoders, we replicated a study that Netflix conducted in a controlled laboratory environment, verifying that our results match the laboratory results with a high correlation coefficient (see Appendix C of our prior subjective comparison).

This report complements our prior report, *HEVC/H.265 Video Codecs Comparison 2019*. Study participants had an option to provide us different version of encoder binary and command line arguments for subjective comparison. If study participant choose not to provide updated version of encoder or command line arguments, we used version provided for objective comparison. We limit the scope of our study to the "Ripping" use case (i.e., all codecs should have a mean encoding speed greater than 1 FPS). The complete list of codecs and command-line arguments appears in Appendix C.

## 2.1. Study Conditions

**Encoders under comparison:** Eleven video encoders with preselected command-line arguments that deliver at least a 1 FPS encoding speed. See the complete list in Appendix C.

| Codec | Developer | Version |
|---|---|---|
| arowana xvc | Divideon | 0.2.0.7 |
| Bytedance | ByteDance Inc. | 1.2.3 |
| SIF Encoder | SIF Encoder Team | v1.78.0 |
| sz265 | Nanjing Yunyan | |
| Tencent V265 Encoder | Tencent | 1.3.5.3 |
| UC265 | Ucodec Inc. | v1.0.7 (64 bit) |
| VP9 | The WebM Project | v1.8.0-424-ge50f4e411 |
| WZAurora AV1 Encoder | Visionular | v0.8 |
| x264 | x264 Developer Team | 0.157.2969 d4099dd |
| x265 | MulticoreWare, Inc. | 3.1.1+1-04b37fdfd2dc |
| xin265 | Peppa | 1.0 |

**Table 1:** Short codecs' descriptions

**Test video sequences:** Five Full HD video sequences with frame rates of 24–60 FPS. See the complete list in Appendix B.

| Sequence | Number of frames | Frame rate | Resolution |
|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| 1. | Crowd Run (short) | 500 | 50 | 1920×1080 |
| 2. | Kayak Trip (short) | 900 | 60 | 1920×1080 |
| 3. | Making Alcohol (short) | 360 | 24 | 1920×1080 |
| 4. | Tractor (short) | 375 | 25 | 1920×1080 |
| 5. | Wedding Party (short) | 360 | 24 | 1920×1080 |

**Table 2:** Summary of video sequences

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

**Test hardware:** All codecs ran on an Core i7 8700K (Coffee Lake) @ 3.7Ghz RAM 32 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 25784 valid answers from 732 unique participants and converted pairwise responses to subjective scores using the Bradley-Terry model [1]. A detailed description of this step appears in Section A.2.

**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 A.3.

## 2.2. 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.

# A. STUDY METHOD

The goal of our study is to rank modern video codecs according to the subjective visual quality of the compressed videos they produce. We therefore employed the method that prior MSU objective-codec comparisons employed, but we replaced the objective SSIM quality scores with subjective scores estimated using Subjectify.us. For the sake of completeness, however, we provide a full description of the method below (including internal details of Subjectify.us).

The method comprises three main steps:

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

## A.1. Running Codecs

In this study we compare mostly the same encoders as in our previous report, MSU Codec Comparison 2019 Part I: FullHD Content, Objective Evaluation, using the same command-line arguments with encoding speed greater than 1 FPS. Study participants had an option to provide us different version of encoder binary and command line arguments for subjective comparison. If study participant choose not to provide updated version of encoder or command line arguments, we used version provided for objective comparison.

We applied software encoders with preselected command-line arguments (see the full list of codecs in Appendix C) to five Full HD test video sequences (see the full list in Appendix B) 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 32 GB Windows 10. The source and encoded files, the encoder executable, and the operation system all resided on an SSD.

## A.2. 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 25784 valid answers from 732 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

| | A | B |
|---|---|---|
| A | 100% ☺ | 75% ☹ |
| B | 134% ☹ | 100% ☺ |


**Table 3:** Example of relative quality analysis table.

(RD) curves, which show the relationship between the real bitrate (i.e., the actual bitrate of the encoded stream) and the quality score. Section ”RD Curves” shows these curves.

## A.3. Integral-Score Estimation

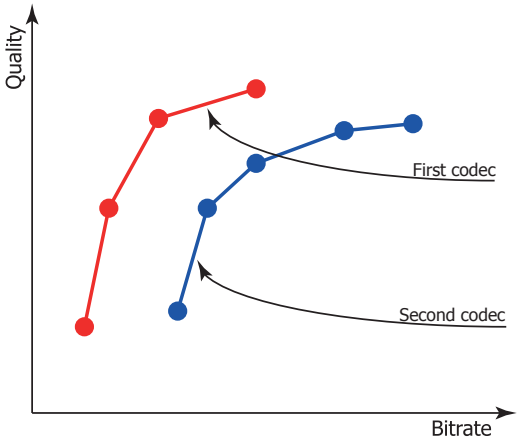
To compare not just individual encoder runs but all runs for a single sequence, and to obtain an overall score for each encoder, we computed integral relative scores: the relative quality score and the relative speed score.

**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:

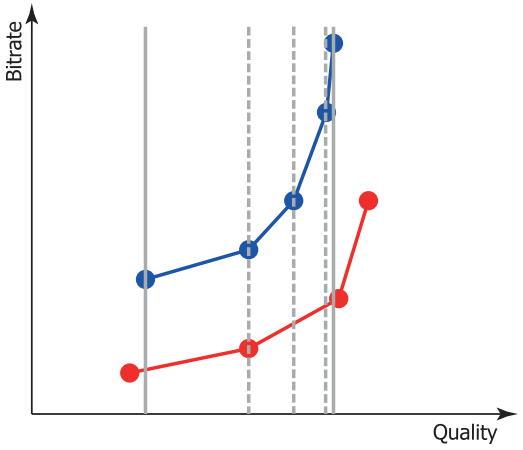
1. Transpose the RD curve for both the test codec and reference codec (see Figures 1a and 1b).
2. Find the RD curves’ projection onto the quality axis—that is, the largest quality range for which both curves are defined (see Figure 1b).
3. Compute the area under the curves for the quality range from 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. Section ”Relative Quality Analysis” depicts these scores with emoticons.

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

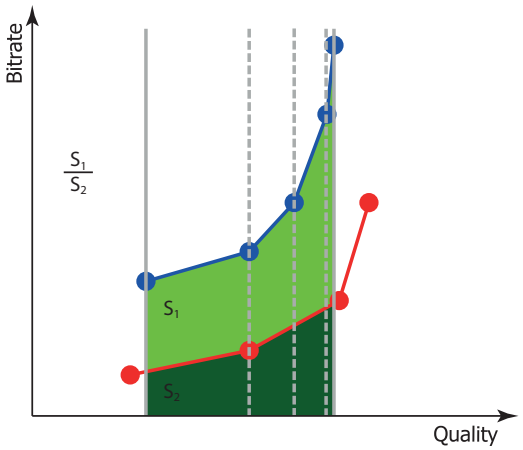
**Relative speed score** is the mean encoding speed of the test encoder divided by the mean encoding speed of the reference encoder for the same bitrate range. To compute this score, we employ the same procedure as above but compute the area under the encoding-speed curves, rather than the RD curves,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




(a) Source RD curves.




(b) Curves transposition and scores computation interval choice.



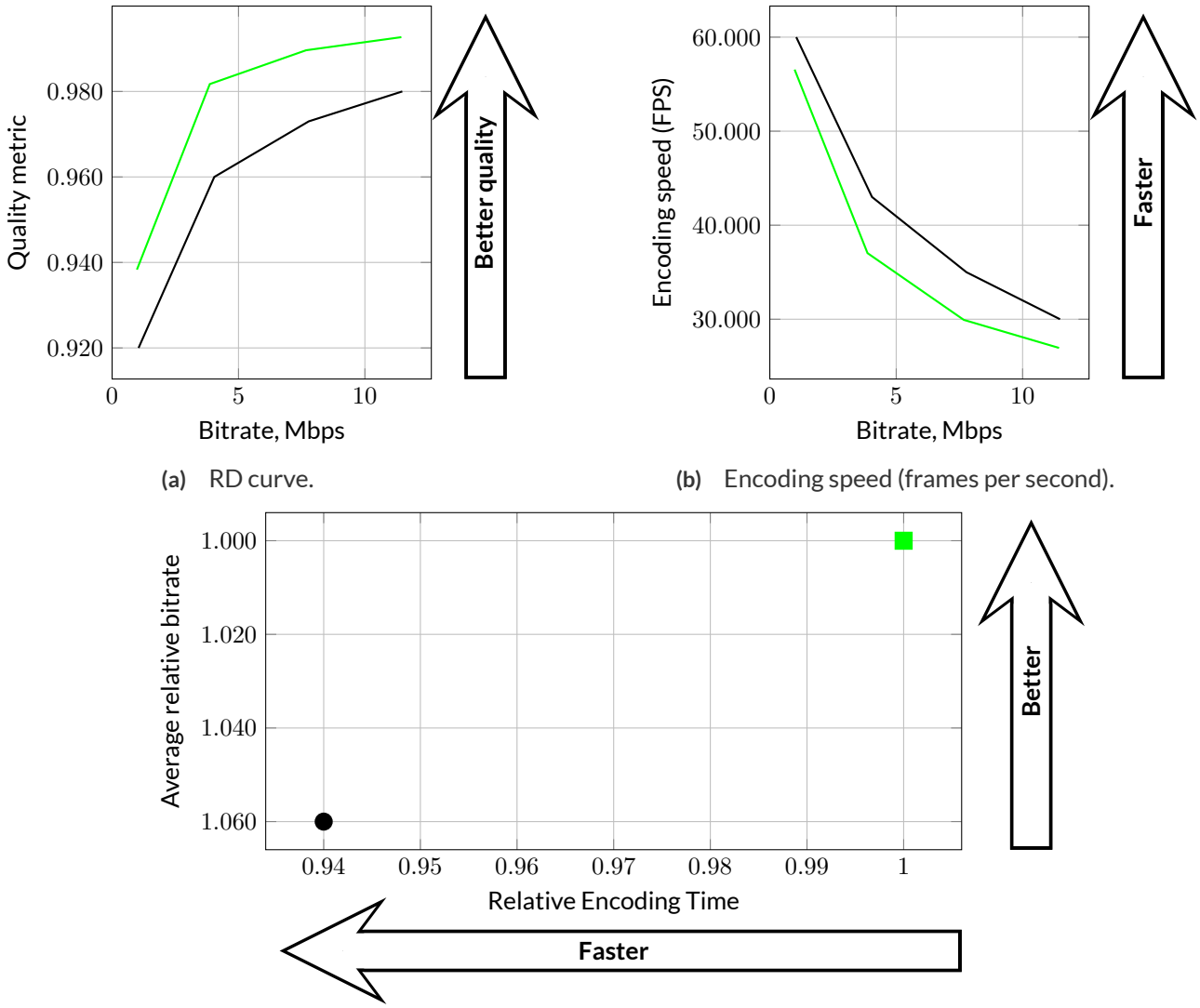

(c) Areas under curves ratio.

**Figure 1:** Relative quality score computation.

### A.3.1. Speed/Quality Trade-Off Plots

To compare both relative quality and relative speed scores, in Section "Speed/Quality Trade-Off" 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 (possibly) sacrificing speed, thus falling short of black in that category. In this example, neither competitor is the absolute winner. Both, however, are Pareto-optimal candidates.




(a) RD curve.

(b) Encoding speed (frames per second).

(c) Trade-off plot.

**Figure 2:** Speed/quality trade-off example.

### A.3.2. Relative Quality Analysis

In Section ”Relative Quality Analysis” we compare encoders pairwise (i.e., one versus another). Using each codec as a reference, we then compute quality scores for all the others relative to that codec. These results are useful when comparing two particular encoders.

In Table 3 we show a simplified example comparison of hypothetical codecs *A* and *B*. Consider row *B*, column *A*, which contains the value 134%. This value means that to produce an encoded stream of the same quality as *A*, *B* must deliver 134% of *A*’s bitrate. The emoticon in the cell depicts the confidence for this estimate.

The plot in Section ”Relative Quality Analysis” depicts data from the table: each line on the plot corresponds to a table row.

# B. SEQUENCES

Direct download links to video sequences used in this comparison can be found in "MSU HEVC Codec Comparison Report 2019" (Enterprise version)

## B.1. Crowd Run (short)

| Sequence title | Crowd Run (short) |
|---|---|
| Resolution | 1920×1080 |
| Number of frames | 500 |
| Color space | YV12 |
| Frames per second | 50 |
| Source resolution | FullHD |
| Bitrate | 1186.52 |

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

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

## B.2. Kayak Trip (short)

| Sequence title | Kayak Trip (short) |
|---|---|
| Resolution | 1920×1080 |
| Number of frames | 900 |
| Color space | YV12 |
| Frames per second | 60 |
| Source resolution | 4K |
| Bitrate | 117.18 |

Different people kayak in the sea and lakes. Filmed from a drone, a kayak, and from under water.


**Figure 4:** Kayak Trip (short) sequence, frame 180

## B.3. Making Alcohol (short)

| Sequence title | Making Alcohol (short) |
|---|---|
| Resolution | 1920×1080 |
| Number of frames | 360 |
| Color space | YV12 |
| Frames per second | 24 |
| Source resolution | FullHD |
| Bitrate | 95.26 |

People watching after the boiler at the stake in different weather in tropics.

**Figure 5:** Making Alcohol (short) sequence, frame 120

## B.4. Tractor (short)

| Sequence title | Tractor (short) |
|---|---|
| Resolution | 1920×1080 |
| Number of frames | 375 |
| Color space | YV12 |
| Frames per second | 25 |
| Source resolution | FullHD |
| Bitrate | 593.26 |

A tractor rides across a field. The camera zooms in, then zooms out.

**Figure 6:** Tractor (short) sequence, frame 233

## B.5. Wedding Party (short)

| Sequence title | Wedding Party (short) |
|---|---|
| Resolution | 1920×1080 |
| Number of frames | 360 |
| Color space | YV12 |
| Frames per second | 24 |
| Source resolution | 4K |
| Bitrate | 161.08 |

A party after a wedding, many guests dance in a dark club.

**Figure 7:** Wedding Party (short) sequence, frame 72

# C. CODECS

All tested encoders presets can be found in “MSU HEVC Codec Comparison Report 2019” (Enterprise version)

# D. 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

## E. 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.