SVT Open Content Video Test Suite 2022 – Natural Complexity

Josef Andersson
josef.andersson@svt.se
Team Video Core, Sveriges Television AB
Stockholm, Sweden

Marcus Linder
marcus.linder@svt.se
Team Video Core, Sveriges Television AB
Stockholm, Sweden

Olof Lindman
olof.lindman@svt.se
Team Video Core, Sveriges Television AB
Stockholm, Sweden

Fredrik Lundkvist
fredrik.lundkvist@svt.se
Team Video Core, Sveriges Television AB
Stockholm, Sweden

ABSTRACT

When evaluating the quality and performance of video encoders, high-quality video material is an obvious necessity. This may not be an issue for commercial actors, as they usually have high-quality material in their archives or through production and purchasing agreements. However, the general public does not have this luxury. As a part of our effort to engage more with the video research and development community, we (the Video Core Team at SVT) are releasing our internal test suite Natural Complexity to the video R&D community under a creative commons license.

This document serves as the technical documentation for the suite, with some comments on the capture and post-production processes, and information on how we use the included files to evaluate video encoding quality.

We hope that the content proves to be as useful for the video R&D community at large as it is for us, and appreciate any and all feedback.

KEYWORDS

Datasets, Video Encoding, Open Content, HDR Video, UHD Video

INTRODUCTION

This document contains information about the SVT Open Content Test Suite 2022 – Natural Complexity. All of the sequences in this suite were produced and post processed with the intent to help SVT's internal OTT Video R&D team evaluate and assess the video quality produced by any given encoder-implementation of existing video coding standards. All of the material in this suite may be distributed, modified and used freely (complete or in parts) as per the license described below.

All of the content was shot at a capture resolution of 3840×2160 at 50 frames per second using professional equipment, and the utmost care has been taken during the post-production process to ensure that the maximum amount of visual detail is retained. The data itself is provided in three different formats, of which two are graded and one is ungraded. However, no effort whatsoever has been made to make the graded material aesthetically pleasing; instead the minimum number of modifications has been introduced to each sequence in order to maximize the luminance latitude within a given colour space.

The text in this document has been edited to be concise and to the point, avoiding overstated academic terminology and/or jargon unless it is absolutely necessary for the sake of clarity. In much the same way, the reader can expect the descriptive language found throughout this document to err on the side of being accessibly expressive rather than technically precise. This is important, as we want the document to be as useful as possible regardless of the reader's level of expertise. This data set has time and time again proved to be a valuable asset when benchmarking all sorts of transcoding operations, and we sincerely hope that it will prove beneficial to others as well. If there are any questions at all regarding the sequences, please contact:

The Video Core Team at SVT: **videocore@teams.svt.se**
Open Source at SVT: **opensource@svt.se**

INTENTIONS, LICENSE AND RESTRICTIONS OF USE

TECHNICAL INFORMATION

Production details can be found in table 1. Post-processing information is found in table 2, and information on output formats is found in table 3.

Original on-location Production	June 2021 (Post Production July 2021)
Producers	Sveriges Television AB (SVT), Video Core Team, Olof Lindman, olof.lindman@svt.se
Camera	Blackmagic Design Ursa Mini Pro 4.6K G1
Recording Framerate	50 frames per second
Shutter Speed	1/100 s
Lenses	Zeiss 35 Prime Lens, Xeen 35mm T1.5 Lens
Recording Formats	CDNG Raw Lossless, BMD RAW Q0
Capture Resolution	Width: 3840 pixels, Height: 2160 pixels
Recorded Bit Depth	12 bit **(as in 12 bit log, read as 16 bit linear)**
Known Issues	The camera sensor contained at least one dead pixel

Table 1: Production details

Software	Blackmagic Design Davinci Resolve Studio
Monitor Dynamic Range	SDR (7 Stops)

Table 2: Post processing specifications

Standards for Primaries and Transfer Functions	ACES AP0 (ACEScc, No IDT or ODT), Rec. 709, Rec. 2100
Data Formats	EXR RGB Half Float (ZIP Compression), JPEG2000 Lossless YUV444 12 bit
Bit Depth	16 bit, 12 bit
Resolution	3840×2160
Framerate	50 frames per second
Audio	None
Containers	OpenEXR, QuickTime MOV

Table 3: Output formats

DETERMINING VIDEO QUALITY

Although not strictly a part of this suite, nor the actual data itself, it is important to define what we mean when we refer to video quality within the context of this document. Thus, in order for the reader to be able to properly gauge and understand the descriptions of the sequences and their intended use case, one first has to determine what we are talking about.

When we use the term **Video Quality**, what we actually mean is: *The subjectively perceived visual quality of a given piece of playing video*

There are three parts to the above definition that are important to observe: first and foremost, we deem subjectively perceived quality as superior to objectively measured quality for end-user viewing purposes. That is to say, while objective measurements are a good first approximation and a useful tool to determine trends at scale, ultimately what our viewers perceive is far more important than any objective metric.

Secondly, since quality is determined to be a subjective matter, what constitutes high and low video quality may vary substantially between different kinds of content. For instance, an encoder might find two pieces of content equally challenging to process and thus, for a given bitrate produce similar QP values (or other relevant parameters) for both of those videos, meanwhile a viewer could very well experience a stark difference in perceived quality between those encodes.

Finally, it is important to stress that a video is a *played* medium, and thus normally viewed as a sequence of moving images. While this might sound obvious in and of itself, it is easily forgotten when actual video quality comparisons are made. Even in professional, as well as amateur settings, there is still a prevalence of frame-by-frame comparison when measuring quality of encodings. Although such methodology is by far the easiest way to compare two videos, frame-by-frame comparisons are hardly representative of what a viewer will notice upon playback. To be clear: We will not deny that analysing individual frames certainly has its uses –especially when pinning down artefacts– but in all fairness the actual quality should primarily be determined using moving images.

SEQUENCES

A list of the sequences and their respective transcoding difficulty, where the complexity is defined by our experience alongside subjective and objective metrics

Name	TRANSCODING DIFFICULTY / COMPLEXITY FROM 1 - 5	DURATION
waterfall - SDR	3 - Challenging	60s
waterfall - HDR	4 - Difficult	60s
smithy - SDR	1 - Easy	60s
smithy - HDR	2 - Moderate	60s
midnight_sun - SDR	5 - Very Difficult	60s
midnight_sun - HDR	5 - Very Difficult	60s
smoke_sauna - SDR	4 - Difficult	60s
smoke_sauna - HDR	5 - Very Difficult	60s
forest_lake - SDR	2 - Moderate	60s
forest_lake - HDR	3 - Challenging	60s
water_flyover - SDR	3 - Challenging	30s
water_flyover - HDR	4 - Difficult	30s
svt_video_quality_test - SDR	5 - Very Difficult	120s

Table 4: List of sequences included in the suite

Sequence descriptions and intended usage at SVT

As mentioned in the introduction, all of the sequences in this suite were selected and prepared with the purpose of testing and benchmarking the qualitative aspects of video encoders aimed at high compression for OTT streaming purposes. To that end, each sequence is specifically created to test a small set of parameters or features. It should once again be stressed that these sequences have NOT been graded/modified to be aesthetically pleasing or even present a specific or uniform look. Instead, as much visual information as possible has been left completely untouched from the recorded state, or in the case of Rec. 709 and Rec. 2100, gone through a minimum of necessary modifications to fit the standard.

This means that these sequences can come across as under- or oversaturated, including seemingly unnatural mixes in hue within colour gradients, and/or either too bright or too dark. While this might sound counter-intuitive when the goal is to test encoders that are more often than not optimized for professionally graded content (i.e., with a look that has creative intent), our experience thus far has been that errors and artefacts found when using this suite are indeed reproducible with actual content.

In the end, the purpose of the suite is to provide a set of sequences that can showcase how an encoder performs when tackling specific challenges. But it is also important to remember that the descriptions below represent how we use these sequences internally, which in turn corresponds to the limitations and factors that govern our particular OTT-video-pipeline from archive to client. Thus, this is by no means the only way to use any of these sequences, and indeed someone else might find completely different use cases for each one of them.

Figure 2: Stills from the *waterfall* sequence. Original on the left, compressed on the right

Waterfall

Figure 1: Frame from the waterfall sequence

Main features: Quick shifts in mid-to-high frequency information, detail retention, area integrity

The waterfall sequence contains several important features that are very useful when assessing an encoder's ability to perform detail retention, specifically mid-to-high frequency information. Typical use cases involve checking edge detection and edge thresholds (of residuals) for operations such as partitioning and motion estimation.

The rapid movement of water creates a pattern that is observable by the human eye, and thus easier to evaluate, but not deterministic enough for an encoder to easily reproduce using temporal techniques. The perceived sharpness of specific details, such as a droplet of water flying through the air, becomes rather easy to spot against a backdrop of noisier content (see figure in the margin). Likewise, the difference in water volume at various parts of the waterfall stays relatively constant and creates ample opportunity to observe noisy and sharp areas within the video.

All of this enables the viewer to evaluate loss of quality between encoded versions, since even though the water flows and falls rather quickly, there are several areas that as a whole can look blurry or sharp in motion, as seen in the margin figures.

Smithy

Figure 3: Frame from the smithy sequence

Main features: Comparisons between Rec. 709 and Rec. 2100, chrominance retention, colour transforms.

Although it might not be obvious at first, the smithy is a sequence that was selected to effectively showcase how an encoder differs in its handling of Rec. 709 and Rec. 2100 encoding.

To start off, this video contains the widest dynamic range of the whole suite, just above 12 stops (which is technically close to the maximum for the camera sensor), where the brightest parts can be found within the specular reflections outlining some of the foliage and the darkest part is the barely distinguishable tools hanging inside the smithy. Furthermore, this sequence contains several regions of distinct colour that are easily represented by the wide color gamut (WCG) within Rec. 2100 but that in turn cannot be properly replicated in Rec. 709, most notable are the various tones of brown along the outer wall to the right.

In a similar vein, the building itself together with the trees provide a wide range of different but distinct textures offers an excellent opportunity to showcase subsampling artefacts, as seen in the margin figure. Accompanying all of this are a small number of flies zipping across the foreground which can help the acute viewer notice what the encoder determines is detail and not.

Figure 4: Still from the *smithy* sequence. Original on the left, compressed on the right

Figure 6: Stills from the *midnight sun* sequence. Original on the left, compressed on the right

Midnight Sun

Figure 5: Frame from the midnight sun sequence

Main features: Challenging on multiple levels, luminance and chrominance range, darkness thresholds for transforms, upperbound specular highlights

Midnight Sun is primarily used to test an encoder's ability to retain different parts of the luminance and chrominance ranges respectively. It is internally infamous as a hard sequence to transcode properly as an encoder is challenged on several levels. Whilst this might sound more important within the WCG-colour space, we find it to be just as pivotal for regular Rec. 709 content.

Due to the large difference in luminance between the highlights (including specular highlights in the case of HDR) and shadows within the overall video, it looks fairly dark to the human eye. However, the video contains distinct and separate areas of differing luminance such that they present a wide set of thresholds and frequency ranges for an encoder. The mountain range in the background, clad in green forest, is a good example of this where the darkest areas are positioned to the left whereas patches of higher luminance can be found to the right.

Likewise, the sky, punctured by the rotating wind turbines, contains a dynamic range of almost

10 stops, which makes it easier to determine how an encoder quantizes different types of blocks (as in transforming macroblocks or CTUs) given a particular bitrate.

Last but not least, we have the obvious rippling water with small gentle waves. The most striking part of this moving body of water is that sloppy compression always leads to a bad viewing experience where the smooth movement turns into a choppy strafe. Apart from the rich set of sharp details in the water itself, there is a strong contrast, especially in HDR, between the darker areas and the bright reflection of sunlight. Much like the clouds in the sky the dynamic range between the brightest glitter and darkest water is just above 10 stops.

Figure 8: Stills from the *smoke sauna* sequence. Original on the left, compressed on the right

Smoke Sauna

Figure 7: Frame from the smoke sauna sequence

Main features: Moving volumes of noise, blocking artefacts, banding

The smoke sauna sequence provides two defining features that are very difficult to encode properly using standard spatial video compression techniques.

The first and most obvious example is the continuously revolving smoke at the upper-to-mid section of the image. With its varying amount of thickness, it generates sharp details and self-contained volumes of noise moving in a somewhat random pattern. This is particularly useful when testing for blocking artefacts and banding in 8 bit transcodes.

The above image is also a great example as to why the quality of an encode should primarily be determined when the video is actually playing. The severe loss of fine detail within the smoke is much more apparent when it is moving, than in any frame-by-frame comparison.

The second part of the video that is hard to compress properly is the small but very bright fire within the stone oven. With its very sharp outline, mixing saturated colours with high luminance

(especially in Rec. 2100), and flickering movement most block-based encoders have a hard time representing the flames in anything else than I-frames.

While the artefacts produced out of either of the above features is problematic enough in SDR, they become even more apparent in HDR. With a larger gamut there is an increased chance of colour errors within the smoke itself, ranging from slightly disfigured hue to full blown wide spectrum noise. It should be noted that while the HDR variant also contains a dynamic range of 11 stops across the whole video, the general darkness in major parts of the picture tends to introduce quite a lot of noise at high levels of compression. Among all the things that are hard to properly encode in the video we tend to be more forgiving than usual about this particular aspect when testing internally.

Forest Lake

Figure 9: Frame from the forest lake sequence

Main features: I-Frame popping, motion estimation at different scales, framerate conversion

The forest lake sequence has three distinct features that make it ideal to test motion estimation and distortion rate between frame types. The most obvious section is the gently rolling water that moves from left to right across the video. Due to the perspective of the shot the picture presents several areas, of different sizes, in constant motion, ranging from larger waves in the foreground to distant diffuse movements in the background. These movements are typically hard to encode and usually require advanced motion estimation between frames to properly replicate. This property also has another level of complexity since the body of water happens to reflect several colours varying from deep violet to blue and green, which in turn presents an opportunity to showcase how the effect of the above-mentioned motion estimation varies across chrominance.

The rest of the image is divided into two parts, in the middle there is a tree line with minor soft movement from the wind, and at the top a group of clouds move slowly. The tree line offers an easy way to spot the infamous "I-frame popping" that can occur in several block-based encoders. Seeing as the trees are rather still most of the time, an encoder might be tempted to spend less and less bits reproducing them across frames.

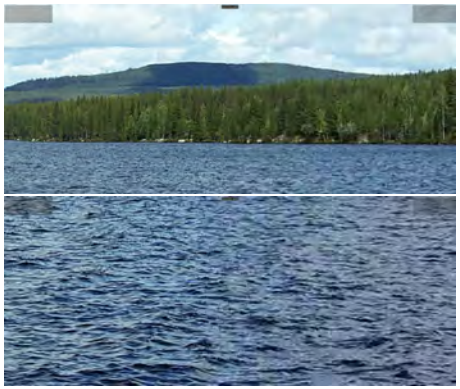

Figure 10: Stills from the *forest lake* sequence. Original on the left, compressed on the right

Water Flyover

Figure 11: Frame from the water flyover sequence

Main features: Tearing artefacts, onion effects, mid frequency thresholds

Just like any proper water sequence the water flyover video focuses on an encoders ability to compress water in motion at different scales/sizes. The video contains a simple downwards/backwards movement over a slowly rolling body of water that shifts from dark to mid-bright. The uniform distribution of waves makes it a lot easier to spot a common tendency among encoders to pay less attention to darker areas of an image, making artefacts more prevalent in the leftmost part of the video.

Much like many other types of content including water, we have found this particular sequence corresponds very well to tearing errors that occur when an encoder keeps producing blocking artefacts over the same area which in turn propagates the error as temporal compression converts artefacts into details. Similarly, if we are able to determine a luminance threshold under which an encoder pays less attention to detail, we can almost always reproduce that error in other content, regardless of complexity.

SVT Video Quality Test (SVQT)

Main features: General torture test, retention of chrominance details (particularly reds), blocking, banding, insufficient chrominance subsampling, text handling

The final sequence in this test suite was not captured by any camera, but rather designed and build in software from the ground up in order create a very difficult “torture test” for video encoders. This sequence is by far the most extreme one in the set, and thus the one least similar to any actual content. However, much like we alluded to in the introduction to this section, the SVQT-sequence has nonetheless proven to pinpoint particular shortcomings in encoders that are later reproducible in other content.

The video contains several layers of semi-transparent slowly revolving noise, which in turn contains random sub-revolutions of varying sizes. These layers are further overlaid by four colour gradients residing in the four corners of the video, each reaching their peak value in their respective quadrant. This layout makes it immediately clear whether the chrominance components are sufficiently quantized at a given bitrate, since the red and blue quadrant will loss much more detail than their yellow and green counterparts.

Furthermore, the sequence contains a rectangular hexagon mesh with just a few pixels between each shape. A black gradient of varying strength moves across this mesh, creating higher contrast background towards the underlying noise pattern. This is useful to see how an encoder handles edge thresholds for colour in dark and bright areas. Finally, the sequence contains three small fields of moving details. Two of them contain filling gradients, and oscillating from black and white and the other from black to transparent. This is particularly useful when checking for banding artefacts. The third and final field has moving text towards a transparent background.

POST PRODUCTION DETAILS

The post-production was conducted in July of 2021, using the software Blackmagic Design Davinci Resolve Studio. As of the current version of this document, the up-to-date suite was exported in February 2022 using version 17.4.3 Build 10. Although the process itself was rather straight forward, it is nonetheless important for the sake of transparency and clarity to specify how that process was conducted and exactly which options were used.

One such important notion is the fact that the chosen storage container used for the high-quality mezzanines in this suite ended up being **QuickTime File Format (.mov)**. Usually, the Video Core team at SVT prefers to use the **Material Exchange Format (.mxf)** whenever it is possible. In this instance however, we were unable to successfully export our intermediate files as .mxf without incorrect representation of colour space data. We cannot say for certain what caused this issue, but it only occurred when using a YUV444 12 bit sampling schema.

Another important factor is that each sequence was shot with the camera mounted on stable tripod. This means that each scene is still, in the sense that there is no camera motion within the sequence. This is a deliberate choice in order to focus on exposure of specific motifs over a short time period. The exports of the suite are divided into three groups, corresponding to different but equally important use cases.

ARCHIVAL – ACES

The first and in many ways most future proof group of exports in this suite are the archival versions utilizing the Academy Color Encoding System (ACES). These exports verifiably embody the closest thing to a mathematically lossless conversion of the captured RAW-data into stored media files. Our intention and hope is that while these files might not be interesting in the short term, seeing as neither software nor hardware tends to be optimized for proper ACES workflows just yet, the fact that they contain the complete RGB information will make them more useful in the future. The colour settings used within Davinci Resolve to produce the ACES exports are shown in figure 12.

Apart from specifying **ACEScc**, using the latest version of ACES, opting to neither apply IDT nor ODT, the settings were left to their default values. Our adjustments were made to maximize compatibility regardless of application. Since these files are meant for archival purposes, no further adjustments were made to the material.

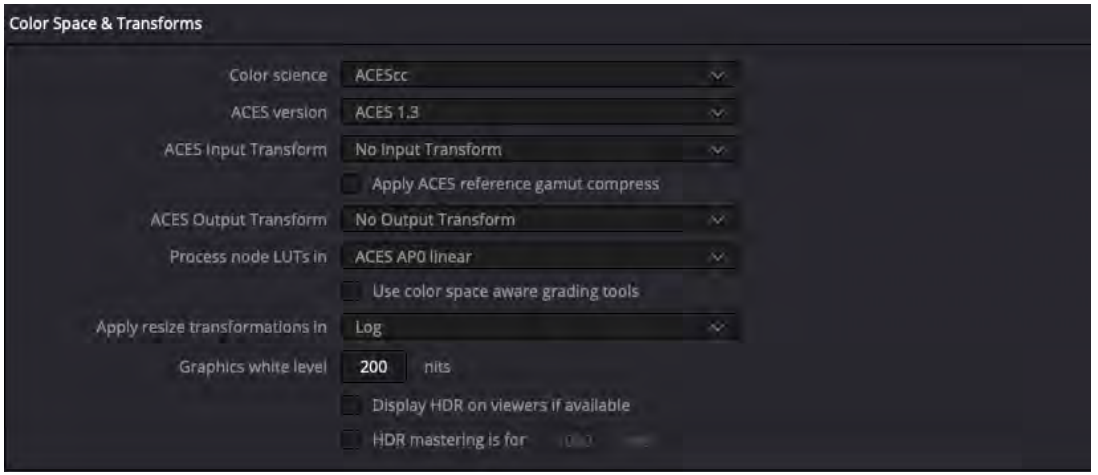

Figure 12: ACES export settings

SDR – Rec. 709

The second group of exports are the Rec. 709 compliant mezzanine files. These files are supposed to be used as reference and benchmarking material for regular SDR transcodes. It is important to note that while the compression itself, provided by the Kakadu JPEG2000 encoder in Davinci Resolve, is lossless and converted to YUV without subsampling, the transform into Rec. 709 certainly results in a loss of information. This means that we have made minor adjustments to fit as much as possible of the interesting information within the limited colour gamut and narrow dynamic range.

The colour management settings used for these exports are shown in figure 13. Although the color processing mode indicates HDR, the Davinci Wide Gamut Intermediate simply allows for more fidelity in the grading procedure and is equally appropriate for SDR exports. However, as mentioned above small adjustments had to be made in order to achieve adequate luminance latitude and focus on important visual details. Using a single node (fig. 14) to alter the information with the curves tool the visual range of RGB was altered to fit within 7 stops peaking at around 100 nits (fig. 15).

Figure 13: Export setting for Rec. 709 mezzanines

Figure 14: Node graph for Rec. 709 export

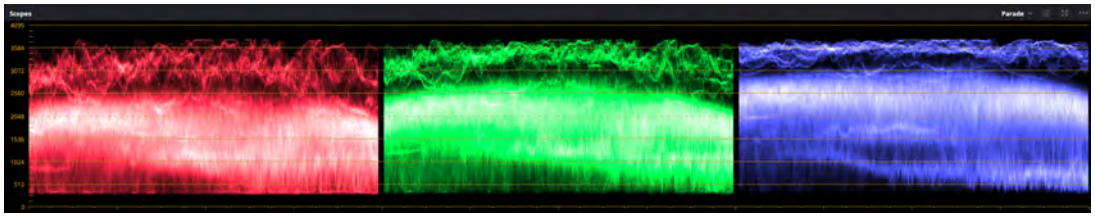

Figure 15: RGB curves for Rec. 709 export

HDR – Rec. 2100

The third and final group of exports are mezzanines exported according to the Rec. 2100 standard, aiming for HDR and WCG tests. Our goal is that these files will be useful when testing an encoders ability to properly handle high fidelity colour information and hefty shifts in luminance, particularly high frequency parts. Much like the Rec. 709 equivalent, the Rec. 2100 does impose a limit on the amount of information that carries over from the source material. Due to the wide colour gamut however, these limitations are far less severe than with Rec. 709.

The colour management settings used for these exports were almost identical to the ones used for Rec. 709, see figure 16: The two main differences are the output color space, which is set to Rec. 2100 with the PQ curve defined by ST2084 and the associated nits value which was set to 1000. Similarly

Figure 16: Export setting for Rec. 2100 mezzanines

to the Rec. 709 exports a single node, containing a curves adjustment, was utilized to fit the source content within the standard. This adjustment mostly affects the RGB channels since the dynamic range provided by the camera never exceeded 12 stops and thus the latitude that was pulled from the capture was easy to fit, with minor tweaks, within the provided nits range.

ACKNOWLEDGEMENTS

We would like to thank the organization Finnbygder i samverkan (FINNSAM) for their help in providing filming locations. Likewise we would like to thank our colleagues at NRK for discovering and pointing out an error with the initial Rec.2100 exports, these have been corrected as of 13/4/2022.

DOCUMENT VERSIONS

Version 1.0 – 15th of February 2022 - Initial Release
Version 1.1 – 17th of February 2022 - Fixed typos and table titles
Version 1.2 – 13th of April 2022 - Uploaded new Rec.2100 exports
Version 1.3 - 15th of January 2023 - Added embedded HDR10-metadata to Rec.2100 sequences

REFERENCES

[1] [n.d.]. Creative Commons License Deed. https://creativecommons.org/licenses/by/4.0/