|  |  |
| --- | --- |
| **Joint Collaborative Team on Video Coding (JCT-VC)**  **of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11**  23rd Meeting: San Diego, USA, 19–26 February 2016 | Document: JCTVC-W1018 |

|  |  |  |  |
| --- | --- | --- | --- |
| *Title:* | **Verification test plan for HDR/WCG coding using HEVC Main 10 Profile** | | |
| *Status:* | Output Document from JCT-VC | | |
| *Purpose:* | Test Plan | | |
| *Author(s) or Contact(s):* | R. Sjöberg (Ericsson) V. Baroncini (MPEG Test Chair) A. K. Ramasubramonian (Qualcomm) | Email: | [rickard.sjoberg@ericsson.com](mailto:rickard.sjoberg@ericsson.com) [baroncini@gmx.com](mailto:baroncini@gmx.com) [aramasub@qti.qualcomm.com](mailto:aramasub@qti.qualcomm.com) |
| *Source:* | JCT-VC | | |

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# Introduction

This document contains a plan for an HDR video verification test for single-layer coding to be conducted to test the coding performance of High Dynamic Range (HDR) content using HEVC Main 10 coding as specified in HEVC version 2.

Subjective tests will be conducted in-between the 23rd and 24th JCT-VC meeting. The tests will compare the current state-of-the-art HEVC Main 10 coding of HDR against the MPEG HDR CfE coding configuration from the version 1.0 anchors, but generated with HM 16.7 using new bitrates. The current state-of-the-art coding will use technology described in JCTVC-W1017 [1] and as implemented in JCTVC-W1020 [5].

# Schedule

The following schedule is anticipated for the subjective test:

* Bitstreams encoded and made available by April 1, 2016
* Subjective evaluation starts: April 4, 2016
* Subjective evaluation completed: May 4, 2016
* Subjective evaluation results available: Wednesday May 25, 2016

(For information, JCT-VC is planned to start Thursday May 26, 2016)

# Test sites

The subjective test shall be carried out at multiple test sites. The following test sites have been identified.

* Rome Laboratory, Italy
* Ericsson AB, Sweden

# Test conditions

The following test conditions will be used for the verification test:

1. Number of sequences, resolutions, and frame rates
   1. 6 sequences in 1080p resolution
   2. 3 sequences are 24 or 25 fps and 3 sequences are 50 fps
2. Bitstreams
   1. Anchor 1.0 configuration bitstreams generated as described in Annex B
   2. Anchor 3.2 configuration bitstreams generated as described in Annex B
3. Encoding parameters
   1. Quantization
      1. 4 bit rate points per sequence
      2. Bitrates shall not exceed the target rate point. Rate matching using at most one single QP step change per sequence will be used
   2. Bit depth
      1. 10 bit encoding for all sequences
   3. Coding structure
      1. Random access coding structure using 7 B-pictures in a hierarchical sub-GOP structure
      2. IRAP picture period at approximately 1.0 second
   4. Other settings as in the configuration files as described in Annex B

# Test Sequences

The proposed test sequences to use are listed in Table 1. Bitstreams will be generated using HEVC Main 10 Profile, BT.2020 container, ST 2084 Transfer Function and NCL Y’CbCr colour space conversion.

All the test sequences have the following characteristics:

* Resolution: 1920x1080 progressive
* Original (not coding) colour format: RGB 4:4:4
* Coding format: 10 bit 4:2:0
* Container: BT.2020 (Gamut: BT. 709 or P3D65 depending on the content)

Table 1: Selected test sequences and properties

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Sequence | Frame rate | Gamut | Length (frames) | Rate 1 | Rate 2 | Rate 3 | Rate 4 |
| Market3 | 50 | BT.709 | 400 | 5400 | 2700 | 1700 | 1300 |
| Showgirl | 25 | P3D65 | 339 | 3400 | 1700 | 1000 | 600 |
| EBU\_06 | 50 | BT.709 | 500 | 2700 | 1600 | 800 | 500 |
| EBU\_04 | 50 | BT.709 | 500 | 6500 | 3000 | 1900 | 1100 |
| Garage | 24 | P3D65 | 288 | 2700 | 1400 | 750 | 460 |
| BalloonFestival | 24 | BT.709 | 240 | 4200 | 2600 | 1600 | 1200 |

Note that the EBU\_04 and EBU\_06 are originally shot in 100 fps and will be played out in 50 fps (slow motion).

# Test length

There are 6 sequences, 4 rates, and 2 codecs used in the test. At least two reference vs. reference cases will be added; furthermore a stabilization phase made of at least 3 BTC has to be considered.

Given that each Basic Test Cell takes 27 seconds (see figure 1 below), the test time for both codecs and including a training phase of 5 BTC becomes approximately {[(6 \* 4 \* 2) + 5] \* 27 seconds} = 23’ 51”.

The above time suggests using two test sessions.

Note that training period is not included.

It may be considered to use the DCR Variant II method (A B A B Vote).

The final decision whether to use method I or method II will be taken by the Test Chair in agreement with the Participant to this experiment.

### Annex A

### *Description of testing environment and methodology*

One of the test methods described in [2] are planned to be used, applying some modification to them, in relation to the kind of display, the video recording and play-back equipment.

A.1 Test method

The test method adopted for this evaluation is DCR (Degradation Category Rating) [2].

A.1.1 Degradation Category Rating (DCR)

This test method is commonly adopted when the material to be evaluated shows a range of visual quality that well distributes across all quality scales.

This method will be used under the schema of evaluation of the quality (and not of the impairment); for this reason a quality rating scale made of 11 levels will be adopted, ranging from "0" (lowest quality) to "10" (highest quality). The test will be held in different laboratories located in countries speaking different languages: This implies that it is better not to use categorical adjectives (e.g. excellent good fair etc.) to avoid any bias due to a possible different interpretation by naive subjects speaking different languages.

The structure of the Basic Test Cell (BTC) of DCR method is made by two consecutive presentations of the video clip under test; at first the original version of the video clip is displayed, immediately afterwards the coded version of the video clip is presented; then a message displays for 5 seconds asking the viewers to vote (see *Figure* ***1***)



*Figure 1 - DCR BTC*

The presentation of the video clips will be preceded by a mid-grey screen displaying for one second.

A.2 How to express the visual quality opinion with DCR

The viewers will be asked to express their vote putting a mark on a scoring sheet.

The scoring sheet for a DCR test is made of a section for each BTC; each section has a box wherein which the viewer shall write the score ranging from 0 to 10. By writing a score of “10”, the subject will express an opinion of “best” quality, while by writing a score of “0” the subject will express an opinion of “worst” quality.

The vote has to be written when the message "Vote N" appears on the screen. The number "N" is a numerical progressive indication on the screen aiming to help the viewing subjects to use the appropriate box of the scoring sheet.

A.4 Training and stabilization phase

The outcome of a test is highly dependent on a proper training of the test subjects.

For this purpose, each subject has to be trained by means of a short practice (training) session.

The video material used for the training session must be different from those of the test, but the impairments introduced by the coding have to be as much as possible similar to those in the test.

The stabilization phase uses the test material of a test session; three BTCs, containing one sample of best quality, one of the worst quality and one of medium quality, are duplicated at the beginning of the test session. By this way, the test subjects have an immediate impression of the quality range they are expected to evaluate during that session.

The scores of the stabilization phase are discarded. Consistency of the behaviour of the subjects will be checked inserting in the session a BTC in which original is compared to original.

A.5 The laboratory set-up

The laboratory for a subjective assessment will be set up according to [2], except for the selection of the display and the video play-out server.

SIM2 monitors will be used in this test. The monitors should be appropriately calibrated. The video play server, or the PC, used to play video has to be able to support the display of 1080p at all frame rates in the test without any limitation, and without introducing any additional temporal or visual degradation.

A.5.1 Viewing distance

The viewing distance should be 2.5 H, where H is equal to the height of the screen. Maximum two subjects will be seated in front of the SIM2 monitor at the same time.

A.5.2 Viewing environment.

The test laboratory has to be carefully protected from any external visual or audio pollution.

Internal general light has to be low (just enough to allow the viewing subjects to fill out the scoring sheets) and a uniform light has to be placed behind the monitor, in a way no direct light hits the viewing subjects seated in front of the screen; the light behind he monitor must be dimmed to an intensity of around 25 lumen. No other light source is admitted, and in particular any light source directed to the screen or creating reflections; ceiling, floor and walls of the laboratory have to be made of non-reflecting material (e.g. carpet or velvet) and should have a colour tuned as close as possible to dark or mid grey.

A.6 Statistical analysis, viewers screening and presentation of the results

The data collected from the score sheets, filled out by the viewing subjects, will be stored in an Excel spread sheet.

For each coding condition the Mean Opinion Score (MOS) and associated Confidence Interval (CI) values will be given in the spread-sheets.

The MOS and CI values will be used to draw graphs. The Graphs will be drawn grouping the results for each video test sequence. No graph grouping results from different video sequences will be considered.

A screening of the viewing subjects will be done prior and after the test. Pre-screening will check for visual acuity (Snellen Chart) and colour blindness (Ishihara tables). Post screening will be done computing the correlation index of each individual viewer against the MOS values. A threshold of 0.75 will be set as lower correlation index.

From the “raw” data subject reliability should be calculated and the method used to assess subject reliability should be reported. Some criteria for subjective reliability are given in [3] and [4].

### Annex B

### *Instructions for producing test bitstreams*

The following steps describe how to produce the anchor 3.2 configuration bitstreams:

1. Download HDRTools using SVN from <http://wg11.sc29.org/svn/repos/Explorations/XYZ/HDRTools/tags/0.10/>
2. Download the updated CTC package. Go to <http://wg11.sc29.org/> using a browser. Click on “MPEG-content” and navigate to [http://wg11.sc29.org/content/Explorations/HDR/CEs\_Software\_113/CE1/](http://wg11.sc29.org/content/Explorations/HDR/CEs_Software_113/CE1/updatedCTCsPackage_20160127.zip). Download updatedCTCsPackage\_20160127.zip.
3. Update HDRTools and its config files with HDRTools-0.10-src.patch and HDRTools-0.10-cfg.patch from the updated CTCs package
4. Convert source sequences to Y’CbCr using the patched HDRTools
5. Compile HM\_anchor3.2 from the updated CTC package and encode the sequences using config files encoder\_randomaccess\_main10\_classAGH\_vui\_lumaQP\_r1.cfg for the BT.709 sequences and encoder\_randomaccess\_main10\_classBCD\_vui\_lumaQP\_r1.cfg for the P3 sequences.

The following steps describe how to produce the anchor 1.0 configuration bitstreams:

1. Download HDRTools using SVN from <http://wg11.sc29.org/svn/repos/Explorations/XYZ/HDRTools/tags/0.10/>.
2. Use the downloaded (unpatched) HDRTools with the same (patched) HDRTools config files as for the anchor 3.2 configuration.
3. Convert the source sequences to Y’CbCr using HDRTools using ClosedLoopConversion=0
4. Download HM-16.7 from <https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/HM-16.7> using SVN.
5. Apply the following patch to HM-16.7:  
     
   Index: source/App/TAppEncoder/TAppEncCfg.cpp  
    const Int defaultWhitePointCode[2] = { 16667, 16667 };  
   - SMultiValueInput<Int> cfg\_DisplayPrimariesCode (0, 50000, 3, 3, defaultPrimaryCodes, sizeof(defaultPrimaryCodes )/sizeof(Int));  
   + SMultiValueInput<Int> cfg\_DisplayPrimariesCode (0, 50000, 6, 6, defaultPrimaryCodes, sizeof(defaultPrimaryCodes )/sizeof(Int));  
    SMultiValueInput<Int> cfg\_DisplayWhitePointCode (0, 50000, 2, 2, defaultWhitePointCode, sizeof(defaultWhitePointCode)/sizeof(Int));
6. Compile HM on linux using the command line “make all\_highbitdepth”. This will make HM support 10-bit encoding.
7. Encode the sequences using encoder\_random\_access\_main10.cfg with the following additional parameters:  
     
   Level : 4.1   
   #============ VUI ================  
   VuiParametersPresent : 1  
   VideoSignalTypePresent : 1  
   VideoFullRange : 0  
   ColourDescriptionPresent : 1  
   ColourPrimaries : 9  
   TransferCharacteristics : 16  
   MatrixCoefficients : 9  
   SEIMasteringDisplayColourVolume : 1  
   SEIMasteringDisplayMaxLuminance : 40000000  
   SEIMasteringDisplayMinLuminance : 47  
   SEIMasteringDisplayPrimaries : 13250,34500,7500,3000,34000,16000  
   SEIMasteringDisplayWhitePoint : 15635,16450  
   ChromaLocInfoPresent : 1   
   ChromaSampleLocTypeTopField : 2  
   ChromaSampleLocTypeBottomField : 2

# References

1. Jonatan Samuelsson, “Conversion and Coding Practices for HDR/WCG Video, Draft 1”, JCTVC-W1017, San Diego, USA, Feb. 2016
2. International Telecommunication Union Standardization Sector; Recommendation ITU-T P.910 “Subjective video quality assessment methods for multimedia applications”
3. Pseudo Isochromatic Plates, engraved and printed by The Beck Engraving Co., Inc., Philadelphia and New York, United States.
4. KIRK (R.E.): Experimental Design – Procedures for the Behavioural Sciences, 2nd Editions, Brooks/Cole Publishing Co., California, 1982.
5. E. Francois, J. Sole, P. Yin, J. Ström, “Updated Common Test Conditions for HDR/WCG Video Coding Experiments”, JCTVC-W1020, San Diego, USA, Feb. 2016