Title:	**Draft Joint Call for Proposals on Video Compression with Capability beyond HEVC**		
Status:	Output document of JVET		
Purpose:	Draft Call for Proposals		
Author(s) or Contact(s):	Andrew Segall	Email:	asegall@sharplabs.com
	Mathias Wien	Email:	wien@ient.rwth-aachen.de
	Vittorio Baroncini	Email:	baroncini@gmx.com
	Jill Boyce	Email:	jill.boyce@intel.com
	Teruhiko Suzuki	Email:	teruhikos@jp.sony.com
Source:	Editors		

---

1 Introduction

This document is a draft Call for Proposals on video coding technology with compression capabilities that significantly exceed those of the HEVC standard the HEVC standard (Rec. ITU-T H.265 | ISO/IEC 23008-2) and its current extensions. This draft Call for Proposals (CfP) has been issued jointly by ITU-T SG16 Q.6 (VCEG) and ISO/IEC JTC 1/SC 29/W G11 (MPEG), and the evaluation of submissions is planned to be conducted in joint collaboration. It is planned to release a final Call for Proposals in October 2017. Please note that the final Call for Proposals may differ from this document.

2 Purpose and procedure

A new generation of video compression technology that has substantially higher compression capability than the existing HEVC standard is targeted. More background information, as well as information about applications and requirements, is given in [1][2].

Companies and organizations that have developed video compression technology that they believe to be better than HEVC are invited to submit proposals in response to this Call.

To evaluate the proposed compression technologies, formal subjective tests will be performed. Results of these tests will be made public (although no direct identification of the proponents will be made in the report of the results unless it is specifically requested or authorized by a proponent to be explicitly identified and a consensus is established to do so). Prior to having evaluated the results of the tests, no commitment to any course of action regarding the proposed technology can be made.

Descriptions of proposals shall be registered as input documents to the proposal evaluation meeting of April 2018 (see the timeline in section 3). Proponents also need to attend this meeting to present their proposals. Further information about logistical steps to attend the meeting can be obtained from the listed contact persons (see section 9).

3 Timeline

The timeline of the Call for Proposals is as follows:

2017/10/31:   Final Call for Proposals

2017/11/01:   Formal registration period opens.

2018/01/15:   Formal registration period ends.

2018/01/25:	Final fee is determined and an invoice for the testing fee (see section 6) will be sent by the test facility.
2018/02/15:	Coded test material shall be available at the test site[1]. By this date, confirmation of the purchase order shall be received.
2018/03/??:	Subjective assessment starts
2018/03/??:	Registration of documents describing the proposals[2]
2018/04/??:	Submission of documents
2018/04/??:	Cross-checking of bitstreams and binary decoders (participation mandatory for proponents)
2018/04/??:	Subjective test results available within standardization body
2018/04/??-20:	Evaluation of proposals at standardization meeting[3]

Anticipated tentative timeline after CfP:

(referring to the first version of an anticipated new standard, which may be extended later in subsequent work):

2018/04	Test model selection process begins
2018/10	Test model selection established
2020/10	Final standard completed

4 Test categories, coding conditions, and anchors

Test categories for standard dynamic range (SDR), high dynamic range (HDR) and 360º content are defined in the three sub-sections below. Proponents are encouraged (but not required) to submit results for all test categories. However, submitters are required to provided results for all test cases in a given test category.

4.1 Standard Dynamic Range

4.1.1 Video test sequence formats and frame rates

All test material is progressively scanned and uses 4:2:0 colour sampling with either 8 or 10 bits per sample.

---

[1] People who formally registered will receive instructions regarding how to submit the coded materials. If material is received later, the proposal may be excluded from testing.

[2] Contact persons will provide information about document submission process. Note that submitted documents will be made publicly available. Exceptions to public availability will be considered on a case by case basis upon request by the contributor.

[3] Proponents are requested to attend this standardization meeting. The starting date has not been finalized; participants will be notified.

The classes of video sequences are:

Class SDR-UHD1:   3840×2160p 60 fps 10 bits: "FoodMarket4","CatRobot1","DaylightRoad2"
                               3840×2160p 50 fps 10 bits: "ParkRunning3"
                               3840×2160p 30 fps 10 bits: "CampfireParty2"

Class SDR-HD1:   1920×1080p 50fps 8 bits: "BasketballDrive", "Cactus"
                               1920×1080p 60fps 8 bits: "BQTerrace"
                               1920×1080p 60fps 10 bits: "RitualDance","MarketPlace"

Class SDR-UHD2:   3840×2160p 30 fps 10 bits: "MountainBay2"

Note – It is anticipated that additional video sequences will be included in the SDR-UHD2 class in the Final Call for Proposals. These sequences may be of lower resolution.

4.1.2 Coding conditions of submissions

Constraint cases are defined as follows:

– Constraint Set 1: not more than 16 frames of structural delay, e.g. 16 pictures "group of pictures (GOP)", and random access intervals of 1.1 seconds or less. A random access interval of 1.1 seconds or less shall be defined as 32 pictures or less for a video sequence with a frame rate of 24, 25 or 30 frames per second, 48 pictures or less for a video sequence with a frame rate of 50 frames per second, 64 pictures or less for a video sequence with a frame rate of 60 frames per second, and 96 pictures or less for a video sequence with a frame rate of 100 frames per second.

– Constraint Set 2: no picture reordering between decoder processing and output, with bit rate fluctuation characteristics and any frame-level multi-pass encoding techniques to be described with the proposal. (A metric to measure bit rate fluctuation is implemented in the Excel file to be submitted for each proposal.)

Submissions shall include encodings for all video sequences in all classes, and each decoding shall produce the full specified number of pictures for the video sequence (no missing pictures).

Submissions shall be made for the test cases (combinations of classes and constraint sets) as listed in **Table 1**.

Table 1 – Combinations of classes and constraint sets for the Standard Dynamic Range category

	Class SDR-UHD1	Class SDR-HD1	Class SDR-UHD2
Constraint set 1	X	X	X
Constraint set 2	–	X	–

– Submissions for Class SDR-UHD1 using Constraint Set 1, and submissions for class SDR-HD1 using Constraint Set 1 and Constraint Set 2 will be evaluated by means of a formal subjective assessment and BD PSNR and rate [3][4] criteria.

– Submissions for Class SDR-UHD2 using Constraint Set 1 will be evaluated by BD PSNR and rate criteria.

For Class SDR-UHD1 using Constraint Set 1 defined above, submissions to this Call shall submit results for the target bit rate as listed in **Table 2**. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in Annex B.

Table 2 – Target bit rate points not to be exceeded for Class SDR-UHD1[4]

	Target bit rates [kbit/s]			
Sequences	**Rate 1**	**Rate 2**	**Rate 3**	**Rate 4**
FoodMarket4	950	1400	2200	3800
CatRobot1	1500	2400	3800	6000
DaylightRoad2	1400	2200	3500	5500
ParkRunning3	2000	3300	6000	10000
CampfireParty2	2000	3300	6000	10000

For Class SDR-HD1 using Constraint Set 1 and Constraint Set 2 defined above, submissions to this Call shall submit results for the target rate points as listed in **Table 3**. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in Annex B.

Table 3 – Target bit rate points not to be exceeded for Class SDR-HD1

	Target bit rates [kbit/s]			
Sequences	**Rate 1**	**Rate 2**	**Rate 3**	**Rate 4**
BQTerrace	400	600	1000	1500
RitualDance	900	1500	2300	3800
MarketPlace	500	800	1200	2000
BasketballDrive	800	1200	2000	3500
Cactus	500	800	1200	2000

For Class SDR-UHD2 using Constraint Set 1, submissions to this Call shall submit results for the target distortion points. The average picture PSNR of the submitted results, calculated over all pictures of the test sequence, shall be within a +/−0.5dB range of the target distortion points. The target distortion points for Class SDR-UHD2 using Constraint Set 1 are listed in **Table 4**.

Table 4 – Target distortion points for Class SDR-UHD2 using Constraint Set 1

	Target PSNR [dB]			
Sequences	**Distortion 1**	**Distortion 2**	**Distortion 3**	**Distortion 4**
MountainBay2	TBD	TBD	TBD	TBD

Note – The values in Table 4 will be updated in the Final Call for Proposals. Additionally, it is anticipated that the target distortion range may also be modified.

---

[4] 1 kbit/s means 10^3 bits per second, and 1 Mbit/s means 10^6 bits per second.

Submissions to this Call shall obey the following additional constraints:

1. No use of pre-processing.

2. Only use post-processing if it is part of the decoding process, i.e. any processing that is applied to a picture prior to its use as a reference for inter prediction of other pictures. Such processing can also be applied to non-reference pictures.

3. Quantization settings should be kept static except that a one-time change of the quantization settings to meet the target bit rate is allowed. When any change of quantization is used, it shall be described. This includes a description of the one-time change.

4. Proponents are discouraged from optimizing encoding parameters using non-automatic means as well as optimizing encoding parameters on a per sequence basis.

5. Proponents shall not use any part of the video coding test set as a training set for training large entropy coding tables, VQ codebooks, etc.

4.1.3 Anchors

Two anchors have been generated by encoding the above video sequences. The first anchor uses the HM 16.16 software package[5]. Static quantization parameter (QP) settings shall be applied, though a one-time change of the quantization parameter from value QP to value QP+1 may be applied in order to meet the defined target bit rates. The quantization parameter settings applied for the anchors will be reported.

The second anchor uses the Joint Exploration Test Model (JEM) software package. The Joint Video Exploration Team (JVET) maintains the JEM software package in order to study coding tools in a coordinated test model [6]. JEM anchor bitstreams will be generated using this software package and will obey the coding conditions in section 4.1.2. It is planned that the JEM 7.0 software package will be used to generate the anchors, though a later version may be used if available.

The purpose of the anchors is to facilitate testing in accordance with BT.500 **Error! Reference source not found.**, providing useful reference points demonstrating the behaviour of well-understood configurations of current technology, obeying the same constraints as imposed on the proposals. The anchors will be among the encodings used in the testing process, however the purpose of the test is to compare the quality of video for proposals to each other rather than to the anchors.

Note – It is anticipated that the configuration files "encoder_randomaccess_main10.cfg" and "encoder_lowdelay_main10.cfg" provided in the HM 16.16 software package will be used to generate the HM anchor bitstreams corresponding to Constraint Set 1 and Constraint Set 2, respectively. Furthermore, it is anticipated that the configuration files "encoder_randomaccess_jvet10.cfg" and "encoder_lowdelay_jvet10.cfg" provided in the JEM7.0 software package will be used to generate the JEM anchor bitstreams corresponding to Constraint Set 1 and Constraint Set 2, respectively. As described in the footnote below, the configuration files will be made available at the same location as the anchor bit-stream for the Final Call for Proposals.

4.2 High Dynamic Range

4.2.1 Video test sequence formats and frame rates

All test material is progressively scanned and uses 4:2:0 colour sampling with 10 bits per sample.

---

[5] Information for accessing the anchors can be obtained from the contact persons. For details, proponents may refer to the config files prepared for the anchors, which are available at the same location as the anchor bitstreams. A summary of the coding conditions used in these config files is provided below.

One class of video sequences is defined:

Class HDR-HD: 1920×1080p 50 fps: "Market3", "Hurdles", "Starting"
                  1920×1080p 25 fps: "ShowGirl2"
                  1920×1080p 24 fps: "Cosmos1"

4.2.2 Coding conditions of submissions

The coding conditions and submissions in 4.1.2 shall apply to the HDR test category with the following exception:

- The quantization settings do not need to be kept static. Instead, a static temporal quantizer structure shall be used. The quantizer may then be adjusted within a frame as a function of the local, average luma value and/or the local, average chroma value. A one-time change of the temporal quantizer structure is also allowed to meet the target bit rate. If either a local adjustment or a one-time change is used, a description of the adjustment scheme shall be provided in the descriptive document submission.

Submissions shall be made for the test cases (combinations of classes and constraint sets) as listed in **Table 5**.

Table 5 – Combinations of classes and constraint sets for the High Dynamic Range category

	Class HDR-HD
Constraint set 1	X
Constraint set 2	–

- Submissions for Classes HDR-HD will be evaluated by means of a formal subjective assessment and BD PSNR and rate criteria.

- Submissions for Class HDR-HD will further be evaluated by the following metrics: weighted PSNR values (at least average of frame wPSNR for each video sequence and encoding point, separate for luma and chroma components), deltaE100 and PSNR-L100, as well as the Bjøntegaard Delta-Rate and Delta-PSNR for each metric. Metric definitions are provided in Annex D.

Submissions to this Call shall, for each of the test cases defined above, submit results for the target bit rate points as listed in **Table 6**. The submitted results shall not exceed the target bit rate points. The calculation of bit rate is further defined in Annex B.

Table 6 – Target bit rate points not to be exceeded for the High Dynamic Range category[6]

Sequences	Target bit rates [kbit/s]			
	Rate 1	**Rate 2**	**Rate 3**	**Rate 4**
Market3	600	900	1500	2400
ShowGirl2	350	500	750	1200
Hurdles	450	800	1500	2000
Starting	450	600	850	1400
Cosmos1	500	900	1500	3000

---

[6] 1 kbit/s means 10^3 bits per second, and 1 Mbit/s means 10^6 bits per second.

Note – It is anticipated that the bit rates for Rate 3 of ShowGirl2 and Starting may be reduced in the Final Call for Proposals. The bit rates for other sequences may be modified as well.

4.2.3 Anchors

The anchor conditions and descriptions in 4.1.3 shall apply to the HDR test category with the following changes:

1. The generation of the anchor does not use a static quantization parameter (QP) setting. Instead, the configuration allows the QP value to vary spatially, where the variation is an explicit function of the average, local luma value. A one-time change of the quantization parameter from value QP to value QP+1 may also be applied in order to meet the defined target bit rates. The quantization parameter settings applied for the anchors will be reported.

4.3 360° Video

4.3.1 Video test sequence formats and frame rates

All test material is progressively scanned and uses 4:2:0 colour sampling with 8 or 10 bits per sample.

The classes of video sequences are:

Class 360:   8192×4096 30 fps 10 bits: "ChairliftRide"
8192×4096 30 fps 8 bits: "KiteFlite", "Harbor", "Trolley"
6144×3072 60 fps 8 bits: "Balboa"
6144×3072 30 fps 8 bits: "Landing"

Note – It is planned that the number of sequences in Class 360 will be reduced to five in the Final Call for Proposals.

4.3.2 Coding conditions of submissions

The coding conditions and submissions in 4.1.2 shall apply to the 360° video test category, with the following exceptions:

Submissions shall be made for the test cases (combinations of classes and constraint sets) as listed in **Table 7**.

Table 7 – Combinations of classes and constraint sets for the 360° video category

	Class 360
Constraint set 1	X
Constraint set 2	–

- Submissions for Class 360 will be evaluated by means of a formal subjective assessment and BD PSNR and rate criteria.

- Submissions for Class 360 will further be evaluated using the following objective metrics: E2E WS-PSNR, E2E S-PSNR-NN, cross-format CPP-PSNR, cross-format S-PSNR-NN, codec WS-PSNR and codec S-PSNR-NN, as described in Annex E.

Submissions to this Call shall, for each of the test cases defined above, submit results for the target bit rate points (which are not to be exceeded) as listed in **Table 8**. The calculation of bit rate is further defined in Annex B.

Table 8 – Target bit rate points not to be exceeded for the 360° video category[7]

	Target bit rates [kbit/s]			
Sequences	**Rate 1**	**Rate 2**	**Rate 3**	**Rate 4**
Balboa	2000	3300	6000	10000
Chairlift	1500	2400	4000	7000
KiteFlite	1200	2400	4000	7000
Harbor	700	1200	2000	3300
Trolley	1500	2400	4000	7000
Landing	1500	2400	4000	7000

Submissions to this Call shall obey the constraints in 4.1.2 with the following change:

1. The quantization settings do not need to be kept static. Instead, the quantization settings may be adjusted within a picture as a function of the geometric position. If local adjustment is used, a description of the adjustment scheme shall be provided in the descriptive document submission.

2. Pre-processing may be used to perform a projection mapping operation, and post-filtering may be used to perform an inverse projection mapping operation. The projection mapping algorithms may allow dynamic changes within a sequence if an automatic selection algorithm is used. The same projection mapping operation and inverse projection mapping operation shall be used for all test sequences in the test case. If projection mapping is used, a description of the projection mapping technique shall be provided in the descriptive document submission. Respondents are asked to provide information regarding at least: (i) the coded resolution of the projection map, (ii) the use of padding and blending, (iii) the use of global rotation, (iv) the use of multi-pass projection mapping, and (v) PSNR values comparing each test sequence to the result of applying the projection mapping algorithm and then converting this result back to the equirectangular projection format without compression.

3. Post-processing after decoding of seam artifacts near discontinuous edges is permitted in submissions, but must be fully described. If post-processing of seam artifacts is performed, the post-processed pictures are used in cross-format and end-to-end objective metrics.

Decoder and/or projection format conversion binaries of submissions for Class 360 must decode and/or projection format convert each bitstream to a 4:2:0 YUV file of the same resolution as the input sequence, either 8192×4096 or 6144x3072. For subjective evaluation, a viewport using a previously undisclosed viewpath per sequence will be extracted from this decoded video using the 360Lib tool.

4.3.3 Anchors

The anchor conditions and descriptions in 4.1.3 shall apply to the 360° video category, with the following exceptions:

1. The equirectangular projection input pictures (at 8192x4096 or 6144x3072 resolution) will be spherically downsampled using the 360Lib software to 4096×2048 equirectangular projection format sequences representing the full 360°×180° sphere, and 8 luma samples of padding on

---

[7] 1 kbit/s means 10^3 bits per second, and 1 Mbit/s means 10^6 bits per second.

each of the left and right sides on the picture will be added by copying samples, yielding 4112×2048 padded equirectangular projection pictures, in which the yaw range exceeds 360°, for encoding by the HM and JEM. After decoding, the 4112×2048 padded equirectangular projection pictures will use linear position-based weighted blending of the overlapping region to convert to 4096×2048 equirectangular projection format without padding, representing 360° of yaw range. 360Lib is used for spherical upsampling to the input resolution, 8K or 6K respectively. E2E and cross-format objective metrics will use the 4096×2048 pictures with blending in their calculations.

Note – The application of the E2E and cross-format objective metrics are further described in JVET-G1003. It is anticipated that this description will be included directly in the Final Call for Proposals.

5   Test sites and delivery of test material

The proposals submission material for the following Classes will be evaluated by means of a formal subjective assessment process: Constraint Set 1 bitstreams of Class SDR-UHD1, Class SDR-HD1, Class HDR-HD and Class 360, and Constraint Set 2 bitstreams of Class SDR-HD1.

The tests will be conducted by the Test Coordinator and one or more sites. The names of the sites will be provided in the Final Call for Proposals.

All proponents need to deliver, by the due date of 2018/02/15, an SSD hard drive to the address of the Test Coordinator (see section 9). The disk shall contain the bitstreams, YUV files, and the decoder executable[8] used by the proponent to generate the YUV files from the bitstreams.

The correct reception of the disk will be confirmed by the Test Coordinator. Any inconvenience caused by unexpected delivery delay or a failure of the disk will be under the complete responsibility of the proponents, but solutions will be negotiated to ensure that the data can still be included in the test if feasible, which means that correct and complete data need to be available before the beginning of the test at the latest.

All the bitstreams, the decoder executable, and the YUV files shall be accompanied by an MD5 check-sum file to verify their correct storing on the disk.

Further technical details on the delivery of the coded material are provided in Annex B.

6   Testing fee

Proponents will be charged a fee per submitted algorithm proposal. An algorithm proposal may consist of a single response to one or more of the Standard Dynamic Range, High Dynamic Range, and/or 360° video categories (but not multiple responses to a single category). Such fee will be a flat charge for each proposal to cover the logistical costs without any profit. The fee is non-refundable after the formal registration is made.

7   Requirements on submissions

More information about file formats can be found in Annex B. Files of decoded sequences and bitstreams shall follow the naming conventions as specified in section C-8 of Annex C.

Proponents shall provide the following; incomplete proposals will not be considered:

   A) Coded test material submission to be received on hard disk drive by February 15, 2018:

      1. Bitstreams for all test cases specified in Table 1, Table 5, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 6, and/or Table 8,

---

[8] The decoder executable must be able to perform the decoding operation on a computer with a Windows 10 operating system.

2. Decoded sequences (YUV files) for all test cases as specified in Table 1, Table 5, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 6, and/or Table 8.

3. Binary decoder executable.

4. MD5 checksum files for A1-A3.

B) Coded test material to be brought for the meeting in April 2018:

1. Bitstreams for all test cases as specified in Table 1, Table 5, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 6, and/or Table 8.

2. Decoded sequences (YUV files) for all test cases as specified in Table 1, Table 5, and/or Table 7 and all bit rates as specified in Table 2 and Table 3, Table 6, and/or Table 8.

3. Binary decoder executable.

4. MD5 checksum files for B1–B3.

C) Document to be submitted before the evaluation meeting in April 2018 shall contain

1. A technical description of the proposal sufficient for full conceptual understanding and generation of equivalent performance results by experts and for conveying the degree of optimization required to replicate the performance. This description should include all data processing paths and individual data processing components used to generate the bitstreams. It does not need to include complete bitstream format or implementation details, although as much detail as possible is desired.

2. An Excel sheet as attached to the final CfP, with all white fields for the respective test cases filled. For objective metric values to be computed per picture, a precision of 2 digits after decimal point shall be used. BD measures [3][4] against the appropriate anchor will be automatically computed from the Excel sheets at the meeting where the evaluation is performed.

3. The technical description shall also contain a statement about the programming language in which the software is written, e.g. C/C++ and platforms on which the binaries were compiled.[8]

4. The technical description shall state how the proposed technology behaves in terms of random access to any picture within the sequence. For example, a description of the GOP structure and the maximum number of pictures that must be decoded to access any picture could be given.

5. The technical description shall specify the expected encoding and decoding delay characteristics of the technology, including structural delay e.g. due to the amount of picture reordering and buffering, the degree of picture-level multi-pass decisions and the degree by which the delay can be minimized by parallel processing.

6. The technical description shall contain information suitable to assess the complexity of the implementation of the technology, including the following:

   - Encoding time [9] (for each submitted bitstream) of the software implementation. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software and algorithm optimizations undertaken, if any, is welcome.

   - Decoding time [10] for each bitstream running the software implementation of the proposal, and for the corresponding constraint case anchor bitstream(s)[11] run on the

---

[9] For example, using ntimer for Windows systems.

[10] For example, using ntimer for Windows systems.

[11] The decoder source code to be used to process the anchor bitstreams will be provided to proponents and must be compiled as-is, without modification of source code, compiler flags, or settings.

same platform. Proponents shall provide a description of the platform and methodology used to determine the time. To help interpretation, a description of software optimisations undertaken, if any, is encouraged.

- Expected memory usage of encoder and decoder.

- Complexity characteristics of Motion Estimation (ME) / Motion Compensation (MC): E.g. number of reference pictures, sizes of picture (and associated decoder data) memories, sample value wordlength, block size, and motion compensation interpolation filter(s).

- Description of transform(s): use of integer/floating point precision, transform characteristics (such as length of the filter/block size).

- Degree of capability for parallel processing.

    Furthermore, the technical description should point out any specific properties of the proposal (e.g. additional functionality such as benefit for 4:4:4 coding, error resilience, scalability).

D) Optional information

    Proponents are encouraged (but not required) to allow other committee participants to have access, on a temporary or permanent basis, to their encoded bitstreams and binary executables or source code.

8 Subsequent provision of source code and IPR considerations

Proponents are advised that, upon acceptance for further evaluation, it will be required that certain parts of any technology proposed be made available in source code format to participants in the core experiments process and for potential inclusion in the prospective standard as reference software. When a particular technology is a candidate for further evaluation, commitment to provide such software is a condition of participation. The software shall produce identical results to those submitted to the test. Additionally, submission of improvements (bug fixes, etc.) is certainly encouraged.

Furthermore, proponents are advised that this Call is being made subject to the common patent policy of ITU-T/ITU-R/ISO/IEC (see http://www.itu.int/ITU-T/dbase/patent/patent-policy.html or ISO/IEC Directives Part 1, Appendix I) and the other established policies of the standardization organizations.

9 Contacts

Prospective contributors of responses to the Call for Proposals should contact the following people:

Prof. Dr. Jens-Rainer Ohm
RWTH Aachen University, Institute of Communications Engineering
Melatener Str. 23, 52074 Aachen, Germany
Tel. +49-241-8027671, Fax. +49-241-8022196, email ohm@ient.rwth-aachen.de

Dr. Gary Sullivan
Microsoft Corporation
One Microsoft Way, Redmond, WA 98052
Tel. +1-425-703-5308, Fax. +1-425-936-7329, email garysull@microsoft.com

Dr. Vittorio Baroncini (Test Coordinator)
Technical Director
GBTech
Viale Castello della Magliana, 38, 00148 – Rome, Italy
Tel. +39-3335474643, email baroncini@gmx.com

10 References

[1] "Requirements for Future Video Coding (FVC)", ITU-T SG16/Q6 VCEG 56th meeting, Torino, Italy, July 2017, Doc. VCEG-BD03 (available at http://wftp3.itu.int/av-arch/video-site/1707_Tor/).

[2] "Requirements for a Future Video Coding Standard v5", ISO/IEC JTC1/SC29/WG11 MPEG, 119th meeting, Torino, Italy, July 2017, Doc. N17074 (available at http://mpeg.chiariglione.org/meetings/119).

[3] Gisle Bjontegaard, "Calculation of Average PSNR Differences between RD curves", ITU-T SG16/Q6 VCEG 13th meeting, Austin, Texas, USA, April 2001, Doc. VCEG-M33 (available at http://wftp3.itu.int/av-arch/video-site/0104_Aus/).

[4] Gisle Bjontegaard, "Improvements of the BD-PSNR model", ITU-T SG16/Q6 VCEG 35th meeting, Berlin, Germany, 16–18 July, 2008, Doc. VCEG-AI11 (available at http://wftp3.itu.int/av-arch/video-site/0807_Ber/).

[5] International Telecommunication Union – Radio Communication Sector, *Methodology for the subjective assessment of the quality of television pictures*, Recommendation ITU-R BT.500-13 (available at https://www.itu.int/rec/R-REC-BT.500).

[6] "Algorithm Description of Joint Exploration Test Model 7 (JEM7)", Joint Video Exploration Team (JVET) of ITU-T VCEG (Q6/16) and ISO/IEC MPEG (JTC 1/SC 29/WG 11), 7th Meeting, Torino, Italy, July 2017, Doc. JVET-G1001 (available at http://phenix.it-sudparis.eu/jvet/).

Annex A: Detailed description of test sequences

Table A-1 – SDR UHD test sequence example pictures

FoodMarket4	CatRobot1	DaylightRoad2	ParkRunning3
CampfireParty2	MountainBay2		

Table A-2 – SDR HD test sequence example pictures

BQTerrace	RitualDance	MarketPlace	BasketBallDrive
Cactus			

Table A-3 – SDR test sequences

Sequence name	Resolution	Frame count	Frame rate	Chroma format	Bit depth
FoodMarket4	3840×2160	720	60	4:2:0	10
CatRobot1	3840×2160	600	60	4:2:0	10
DaylightRoad2	3840×2160	600	60	4:2:0	10
ParkRunning3	3840×2160	500	50	4:2:0	10
CampfireParty2	3840×2160	300	30	4:2:0	10
MountainBay2	3840×2160	300	30	4:2:0	10
BQTerrace	1920×1080	600	60	4:2:0	8
RitualDance	1920×1080	600	60	4:2:0	10
MarketPlace	1920×1080	600	60	4:2:0	10
BasketballDrive	1920×1080	500	50	4:2:0	8
Cactus	1920×1080	500	50	4:2:0	8

Table A-4 – SDR test sequence md5sums

Sequence name	MD5Sum	Source (or Owner)
FoodMarket4	a378b34190f54f688d048a9a8b46a8ac	Netflix
CatRobot1	03a89792693fd9ecfd72ef2590025e97	B<>COM
DaylightRoad2	bf1d22643afb41b1521749d2749fb5f0	Huawei
ParkRunning3	e7a1d1ebff269767ec4bffd2998d1154	Huawei
CampfireParty2	63d3d9f9e4e8b5c344e89840e84e6428	SJTU
MountainBay2	f27b6b70244fb083baac546958fcf696	DJI
BQTerrace	efde9ce4197dd0b3e777ad32b24959cc	NTT DOCOMO Inc.
RitualDance	a3cb399a7b92eb9c5ee0db340abc43e4	Netflix
MarketPlace	dc668e7f28541e4370bdbdd078e61bba	Netflix
BasketballDrive	d38951ad478b34cf988d55f9f1bf60ee	NTT DOCOMO Inc.
Cactus	3fddb71486f209f1eb8020a0880ddf82	EBU/RAI

Note – Copyright information is included in the sequence ZIP container.

Table A-5 – HDR test sequence example pictures

Market3	ShowGirl2	Hurdles	Starting
Cosmos1			

Table A-6 – HDR test sequences

Sequence name	Resolution	Frame count	Frame rate	Chroma format	Bit depth
Market3	1920×1080	400	50	4:2:0	10
ShowGirl2	1920×1080	339	25	4:2:0	10
Hurdles	1920×1080	500	50	4:2:0	10
Starting	1920×1080	500	50	4:2:0	10
Cosmos1	1920×856	240	24	4:2:0	10

Note: The capture frame rate of the HDR3 (Hurdles) and HDR4 (Starting) sequences was 100fps. However, these sequences are treated as 50fps sequences for the evaluation processes defined in this document.

Table A-7 – HDR test sequence md5sums

Sequence name	MD5Sum	Source (or Owner)
Market3	c97abe47455fd12f6d6436cecfad7c7d	Technicolor
ShowGirl2	44f1974d68f7799c71eea29fb72b245b	Stuttgart Media University
Hurdles	bc3cba849d6f4ee74d39056600722aa5	EBU
Starting	1cbc416696cb0dfcf4da9886eeb6a4a2	EBU
Cosmos1	da4a2488c249720da0535f01c3693efa	Netflix

Note – Copyright information is included in the test sequence ZIP container.

Table A-8 – 360º test sequences

Balboa	ChairliftRide	KiteFlite
Harbor	Trolley	Landing

Table A-9 – 360º video test sequences

Sequence name	Input resolution	Anchor resolution	Coded luma sample count of anchors	Frame count	Frame rate	Chroma format	Bit depth
Balboa	6144x3072	4096×2048	8388608	300	60	4:2:0	8
ChairliftRide	8192×4096	4096×2048	8388608	300	30	4:2:0	10
KiteFlite	8192×4096	4096×2048	8388608	300	30	4:2:0	8
Harbor	8192×4096	4096×2048	8388608	300	30	4:2:0	8
Trolley	8192×4096	4096×2048	8388608	300	30	4:2:0	8
Landing	6144x3072	4096×2048	8388608	300	30	4:2:0	8

Note – The sequences are omnidirectional 360º × 180º degree video and are stored in an equirectangular projection (ERP) format. The number of coded luma samples in the anchor is lower than the resolution of the input sequence.

Table A-10 – 360° video test sequence md5sums

Sequence	MD5Sum	Source (or Owner)
Balboa	1457bb109ae0d47265f5a589cb3464d7	InterDigital
ChairliftRide	9126f753bb216a73ec7573ecc4a280c3	GoPro
KiteFlite	18c0ea199b143a2952cf5433e8199248	InterDigital
Harbor	aa827fdd01a58d26904d1dbdbd91a105	InterDigital
Trolley	25c1082d1e572421da2b16530718156d	InterDigital
Landing	c715e332e78ab30e477445954da66c6b	GoPro

Note – Copyright information is included in the sequence ZIP container.

Annex B: Distribution formats for test sequences and decoded results, Delivery of Bitstreams and Binary Decoders, Utilities and Cross-check Meeting

Distribution of original video material files containing test sequences is done in YUV files with extension ".yuv". Colour depth is 10 bit per component for all sequences. A description of the YUV file format is available at http://www.fourcc.org/ web site, designated as "planar iyuv".

HEVC Anchor bitstreams are provided with extension ".hevc". JEM Anchor bitstreams are provided with extension ".jem". Bitstream formats of proposals can be proprietary, but must contain all information necessary to decode the sequences at a given data rate (e.g. no additional parameter files). The file size of the bitstream will be used as a proof that the bit rate limitation from Table 2, Table 3, Table 6, or Table 8 has been observed. The file extension of a proposal bitstream shall be ".bit".

Decoded sequences shall be provided in the same ".yuv" format as originals, with the exception that the colour depth shall be 10 bits per component for all sequences.

All files delivered (bitstreams, decoded sequences and binary decoders) must be accompanied by an MD5 checksum file to enable identification of corrupted files. An MD5 checksum tool that may be used for that purpose is typically available as part of UNIX/LINUX operating systems; if this tool is used, it should be run with option "-b" (binary). For the Windows operating systems, a compatible tool can be obtained from http://www.pc-tools.net/win32/md5sums/. This tool should be run with additional option "-u" to generate the same output as under UNIX.

For the 360° video category, a binary projection format convertor may be provided, to be applied to the decoded YUV files. The output of the binary decoder, or the projection format convertor, if provided, must decode and/or convert each bitstream to a 4:2:0 YUV file of the same resolution as the input sequence.

The hard disk drive should be shipped (for handling in customs) with a declaration "used hard disk drive for scientific purposes, to be returned to owner" and low value specification (e.g. 20€). The use of a hard disk drive with substantially larger size than needed is discouraged. The hard disk drive should be a 3½-inch SATA solid state drive without any additional enclosure (no case, no power supply, no USB interface etc.), NTFS file format shall be used.

Before the evaluation meeting, a one-day cross-check meeting will be held. Proponents shall bring another hard disk drive, which can be connected via USB 3.0 to a Windows PC, containing original and decoded sequences in YUV format, bitstreams, binary decoder executables and all related checksum files. An adequate computer system shall also be brought to this meeting. Proponents shall specify the computing platform (hardware, OS version) on which the binary can be run. Should such a computing platform not be readily available, the proponent shall provide a computer adequate for decoder verification at this meeting. Further information will be exchanged with the proponents after the registration deadline.

Annex C: Description of testing environment and methodology

The test procedure foreseen for the formal subjective evaluation will consider two main requirements:

- to be as much as possible reliable and effective in ranking the proposals in terms of subjective quality (and therefore adhering the existing recommendations);

- to take into account the evolution of technology and laboratory set-up oriented to the adoption of FPD (Flat Panel Display) and video server as video recording and playing equipment.

Therefore, two of the test methods described in Recommendation ITU-R BT.500 [5] are planned to be used, applying some modification to them, relating to the kind of display, the video recording and playing equipment.

C.1   Selection of the test method

The anticipated test methods are:

1. DSIS (Double Stimulus Impairment Scale)
2. DSCQS (Double Stimulus Continuous Quality Scale)

C.1.1 DSIS

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 three 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.

All the video material used for these tests will consist of video clips of 10 seconds duration.

The structure of the Basic Test Cell (BTC) of DSIS 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 C-1)
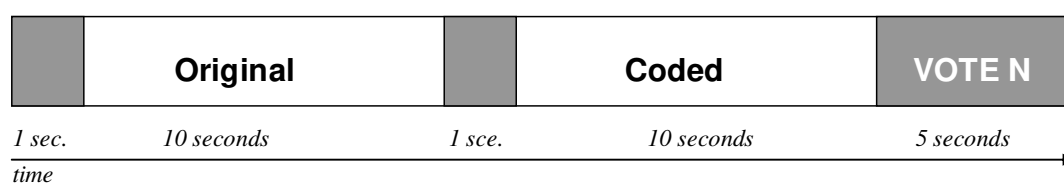
	Original		Coded	VOTE N
1 sec.	*10 seconds*	*1 sce.*	*10 seconds*	*5 seconds*

time

Figure C-1 – DSIS BTC

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

C.1.2 DSCQS

This test method is typically selected when the range of visual quality presented to the viewing subjects is in the upper part of the quality scale.

In this method, the original and the coded samples of a video clip are presented two times.

This allows the viewing subjects to evaluate the sequences they are seeing more accurately.

An important aspect of DSCQS is that the viewing subject is asked to evaluate the two original and the coded clips separately. Furthermore, the viewer does not know which of the two clips is the original or the coded one.

The position of the original and the coded one is randomly changed for each BTC.

The structure of the Basic Test Cell of the DSCQS method therefore contains two consecutive pairs of presentations.

The first presentation is announced by a mid-grey screen with the letter "A" in the middle displaying for one second; the second presentation is announced by another mid grey screen with the letter "B"; these are repeated during the second pair of presentations changing A and B into A* and B*; the message "VOTE" displays on the screen for 5 seconds after the second pair of presentations is done (see Figure C-2)

A	original (coded)	B	coded (original)	A*	original (coded)	B*	coded (original)	VOTE N
1 sec.	10 seconds	1 sec.	10 seconds	1 sec.	10 seconds	1 sec.	10 seconds	5 seconds

time
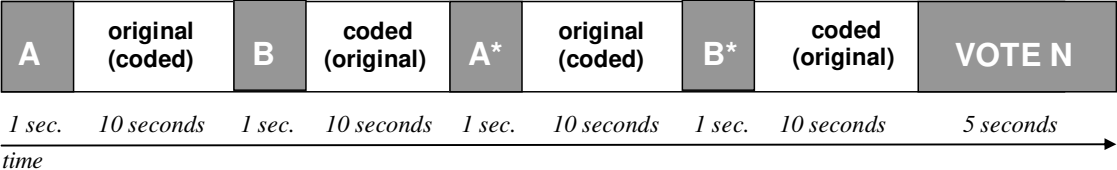
Figure C-2 – DSCQS BTC

C.2 How to express the visual quality opinion with DSIS

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

The scoring sheet for a DSIS test is made of a section for each BTC; each section is made of a column of 11 vertically arranged boxes, associated to a number from 0 to 10 (see Figure C-3).

The viewers have to put a check mark on one of the 11 boxes; checking the box "10" the subject will express an opinion of "best" quality, while checking the box "0" the subject will express an opinion of the "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 column of the scoring sheet.

VOTE 01	VOTE 02	VOTE 03	VOTE 04	VOTE 05	VOTE 06	VOTE 07	VOTE 08
☐ 10	☐ 10	☐ 10	☐ 10	☐ 10	☐ 10	☐ 10	☐ 10
☐ 9	☐ 9	☐ 9	☐ 9	☐ 9	☐ 9	☐ 9	☐ 9
☐ 8	☐ 8	☐ 8	☐ 8	☐ 8	☐ 8	☐ 8	☐ 8
☐ 7	☐ 7	☐ 7	☐ 7	☐ 7	☐ 7	☐ 7	☐ 7
☐ 6	☐ 6	☐ 6	☐ 6	☐ 6	☐ 6	☐ 6	☐ 6
☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5
☐ 4	☐ 4	☐ 4	☐ 4	☐ 4	☐ 4	☐ 4	☐ 4
☐ 3	☐ 3	☐ 3	☐ 3	☐ 3	☐ 3	☐ 3	☐ 3
☐ 2	☐ 2	☐ 2	☐ 2	☐ 2	☐ 2	☐ 2	☐ 2
☐ 1	☐ 1	☐ 1	☐ 1	☐ 1	☐ 1	☐ 1	☐ 1
☐ 0	☐ 0	☐ 0	☐ 0	☐ 0	☐ 0	☐ 0	☐ 0

Figure C-3 – Example of DSIS test method scoring sheet

C.3 How to express the visual quality opinion with DSCQS

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

The scoring sheet for a DSCQS test is made of a section for each BTC; each section is made of two continuous columns with 100 horizontal marks, (see Figure C-4).

The viewers have to put a check mark for each of the two vertical lines; checking the upper side of the bar the subject will express an opinion of "best" quality, while checking the lower side of the bar the subject will express an opinion of the "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 column of the scoring sheet.
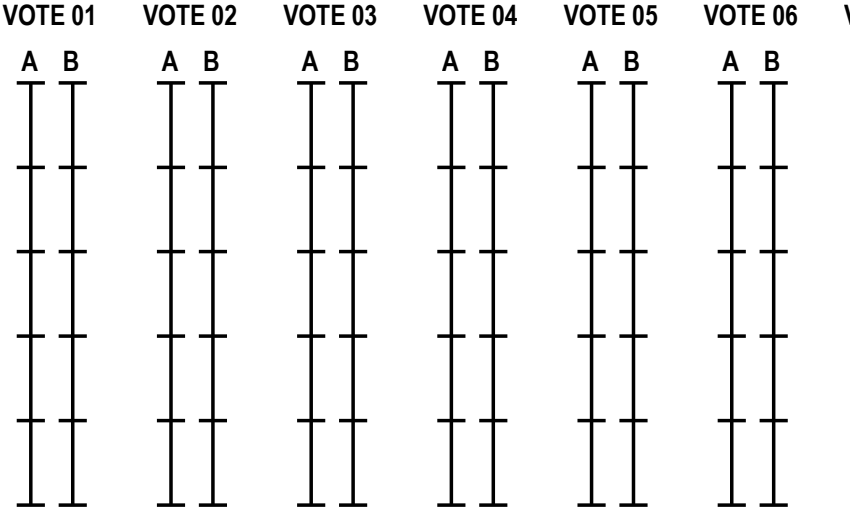
Figure C-4 – Example of DSCQS test method scoring sheet

C.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.

C.5 The laboratory setup

The laboratory for a subjective assessment is planned to be set up following Recommendation ITU-R BT.500 [5], except for the selection of the display and the video play server.

High quality displays with controllable colour rendition will be used.

When a video clip is shown at a resolution lower than the native resolution of the display itself, the video has to be presented in the centre of the display; the active part of the display (i.e. that is actually showing the video signal) must have a dimension equal in rows and columns to the raster of the video; the remaining part of the screen has to be set to a mid-grey level (128 in 0-255 range). This constraint guarantees that no interpolation or distortion artefacts of the video images will be introduced.

The video play server, or the PC used to play video has to be able to support the display of both and UHD video formats, at 24, 30, 50 and 60 frames per second, without any limitation, or without introducing any additional temporal or visual artefacts.

C.5.1 Viewing distance, seats and monitor size

The viewing distance varies according to the physical dimensions of the active part of the video; this will lead to a viewing distance varying from 2H to 4H, where H is equal to the height of the active part of the screen.

The number of subjects seating in front of the monitor is a function of the monitor size and type; for example, a monitor equal or superior to 30" is expected to allow the seating of two or three subjects at the same time; monitors with 24" would allow two subjects; monitors of 21" would allow just one subject. Monitors with diagonal lower than 21" should not be used. In any case monitors must support a wide screen aspect ratio without any picture adaptation (e.g. letter box etc.); i.e. they must give native support for the 16:9 aspect ratio.

C.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; this light must have an intensity as specified in Recommendation ITU-R BT.500. No light source has to be directed to the screen or create 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 D65.

C.6 Example of a test schedule for a day

This is an example of the schedule of the planned test activity for one day.

A time slot of one hour is dedicated every morning and afternoon to welcome, to screen and to train the viewers.

After the subjects' screening for visual acuity and colour blindness, the subjects will be grouped in testing groups.

In the following example, four groups of subjects are created, according to the laboratory set-up[12] and to the time constraints.

	DAY 1		DAY 2	
9:00 – 10:00	Screening / training		Screening / training	
10:00 -10:40	G1-S1		G1-S1	
10:40 -11:20		G2-S1		G2-S1
11:20 -12:00	G1-S2		G1-S2	
12:00 -12:40		G2-S2		G2-S2
13:00 – 14:00	Screening / training		Screening / training	
14:00 -14:40	G3-S3		G3-S3	
14:40 -15:20		G4-S3		G4-S3

---

[12] The viewing rooms of the laboratories could be different according to the test material and/or to the design of the laboratory. Large displays (e.g. monitor equal or wider than 50″) will allow to seat three (or more) subjects at the same time; a laboratory set-up, in which three wide monitors are available, will allow the creation of wider groups of viewers (three or more).

| 15:20 -16:00 | G3-S4 | | G3-S4 | |
| 16:00 -16:40 | | G4-S4 | | G4-S4 |

C.7 Statistical analysis and presentation of the results

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

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

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.

C.8 Proponents identification and file names

Each Proponent submitting to the CfP will be identified with a two digit code preceded by the letter "P" (e.g. P01 P02 … Pnn).

Each coded video file provided for a submission will be identified by a name formed by the below listed combination of letters and numbers:

PnnSxxRyCz.<filetype>

where:

- Pnn identifies the Proponent,

- Sxx identifies the original video clip used to produce the coded video, as identified in the tables of Annex A;

- Ry identifies the bit rate y, as identified in Table 2, Table 3, Table 6, or Table 8;

- Cz identifies the constraint set z (z=1 or z=2), as identified in section 4.1.2 (Table 1 and above)

- <filetype> identifies the kind of file:

     o .bit = bitstream

     o .yuv = decoded video clip in YUV format

Note – It is anticipated that additional information describing the testing methodology for the 360° video category will be provided in the Final Call for Proposals.

Annex D: Description of HDR video category test metrics

In this Annex, further information about the wPSNR and deltaE100, and PSNR-L100 metrics are provided. The metrics deltaE100 and PSNR-L100 shall be calculated using the HDRMetrics v0.15 utility available at https://gitlab.com/standards/HDRTools.

D.1 wPSNR

The wPSNR, or weighted PSNR, metric is calculated from the weighted mean squared error of the pixel values. An implementation of the metric is provided in the JEM software, and that implementation shall be used for calculation of the metric using the weighting functions provided below.

As an informative description of the metric follows:

The wPSNR metric is calculated as:

$$wPSNR = 10 * log\frac{X^2}{wMSE},$$

where X is the maximum pixel value for the specific bit depth and *wMSE* is given as

$$wMSE = \sum\nolimits_{all\ pixels\ i\ in\ block} w_i\left(luma\left(x_{orig,i}\right)\right)*(x_{orig,i}\text{-}x_{dec,i})^2,$$

where $w_i\left(luma\left(x_{orig,i}\right)\right)$ is a weight that is a function of the luma value corresponding to the pixel $x_{orig,i}$, $x_{orig,i}$ is the original value at location *i*, and $x_{dec,i}$ is the reconstructed value at location *i*.

The calculation of the weight is computed as:

$y_i = 0.015*luma(x_{orig,i}) − 1.5 − 6;$

$y_i = y_i < −3\ ?\ −3 : (y_i>6\ ?\ 6 : y_i);$

$w_i(luma(x_{orig,i})) = pow(2.0, y_i÷3.0);$

In all cases, the metric is calculated individually for a single luma or chroma channel and then used to compute a Bjøntegaard Delta-Rate and Delta-PSNR rate.

D.2 deltaE100-based metric

An implementation of the deltaE100 distortion metric is provided in the HDRTools software, and that implementation shall be used for calculation of the metric using the HDRMetric utility and configuration provided with the HDRTools software.

An informative description of the metric follows.

The original and test material must first be converted to linear-light 4:4:4 RGB EXR. For example, if the material is in the YCbCr BT.2100 4:2:0 PQ 10 bit format, it must be converted to a 4:4:4 RGB BT.2100 OpenEXR file using configuration provided with the HDRTools software.

Subsequently, the following steps should be applied for each (R,G,B) sample, within the content to be compared, i.e. original source (content 1) and test material (content 2) :

- Convert the content to the XYZ colour space

- Convert the content from the XYZ to Lab space using the following equations. Computations are performed using double floating point precision.

- o invYn = 1.0 ÷ Yn, with Yn = 100

- o invXn = invYn ÷ 0.95047

- o invZn = invYn ÷ 1.08883

- o yLab = convToLab(y * invYn)

- o L = 116.0 * yLab − 16.0

- o a = 500.0 * (convToLab(x * invXn) − yLab)

- o b = 200.0 * (yLab − convToLab (z * invZn))

    with convToLab(x) is defined as

    if x >= 0.008856

    $$convToLab(x) = x^{(1 \div 3)}$$

    else

    convToLab(x) = 7.78704 * x + 0.137931

- The deltaE100 distance DE between two samples (L1,a1,b1) and (L2,a2,b2) is then computed as follows:

    cRef = sqrt(a1 * a1 + b1 * b1)

    cIn = sqrt(a2 * a2 + b2 * b2)

    cm = (cRef + cIn) ÷ 2.0

    $$g = 0.5 * (1.0 − sqrt(cm^{7.0} \div (cm^{7.0} + 25^{7.0})))$$

    aPRef = (1.0 + g) * a1

    aPIn = (1.0 + g) * a2

    cPRef = sqrt(aPRef * aPRef + b1 * b1)

    cPIn = sqrt(aPIn * aPIn + b2 * b2)

    hPRef = arctan(b1, aPRef)

    hPIn = arctan(b2, aPIn)

    deltaLp = L1 − L2

    deltaCp = cPRef − cPIn

    deltaHp = 2.0 * sqrt(cPRef * cPIn) * sin((hPRef − hPIn) ÷ 2.0)

    lpm = (L1 + L2) ÷ 2.0

    cpm = (cPRef + cPIn) ÷ 2.0

    hpm = (hPRef + hPIn) ÷ 2.0

$$rC = 2.0 * sqrt(cpm^{7.0} \div (cpm^{7.0} + 25^{7.0}))$$

$$dTheta = DEG30 * exp(-((hpm - DEG275) \div DEG25) * ((hpm - DEG275) \div DEG25))$$

$$rT = -sin(2.0 * dTheta) * rC$$

$$t = 1.0 - 0.17 * cos(hpm - DEG30) + 0.24 * cos(2.0 * hpm) +$$
$$0.32 * cos(3.0 * hpm + DEG6) - 0.20 * cos(4.0 * hpm - DEG63)$$

$$sH = 1.0 + (0.015 * cpm * t)$$

$$sC = 1.0 + (0.045 * cpm)$$

$$sL = 1.0 + (0.015 * (lpm - 50) * (lpm-50) \div sqrt(20 + (lpm - 50) * (lpm - 50)))$$

$$deltaLpSL = deltaLp \div sL$$

$$deltaCpSC = deltaCp \div sC$$

$$deltaHpSH = deltaHp \div sH$$

$$deltaE100 = sqrt(deltaLpSL * deltaLpSL + deltaCpSC * deltaCpSC +$$
$$deltaHpSH * deltaHpSH + rT * deltaCpSC * deltaHpSH)$$

with

DEG275 = 4.7996554429844

DEG30 = 0.523598775598299

DEG6 = 0.1047197551196598

DEG63 = 1.099557428756428

DEG25 = 0.436332

After this process, the deltaE100 values for each frame are averaged within the "Distortion" specified window. Finally, a PSNR based value is derived as:

$$PSNR\_DE = 10 * log10(PeakValue \div deltaE100)$$

where PeakValue is set to 10,000.

D.3 PSNR-L100

An implementation of the PSNR-L100 distortion metric is provided in the HDRTools software, and that implementation shall be used for calculation of the metric using the HDRMetric utility and configuration provided with the HDRTools software.

An informative description of the metric follows:

PSNR-L100 represents the distortion in the lightness domain of the CIELab colour space. The derivation of Lab values from the linear representation of the signal is similar as given in the description of deltaE100. The mean absolute error (MAE) in the L domain is used to compute the PSNR-L100 as follows:

$$PSNR\text{-}L100 = 10 * log10(PeakValue \div MAE)$$

where PeakValue is set to 10,000.

Annex E: Description of 360° video objective test metrics

The following objective metrics will be provided for sequences in the 360° video category: E2E WS-PSNR, E2E S-PSNR-NN, cross-format CPP-PSNR, cross-format S-PSNR-NN, codec WS-PSNR and codec S-PSNR-NN. Figure E-1 illustrates the processing chain for computation of objective metrics.
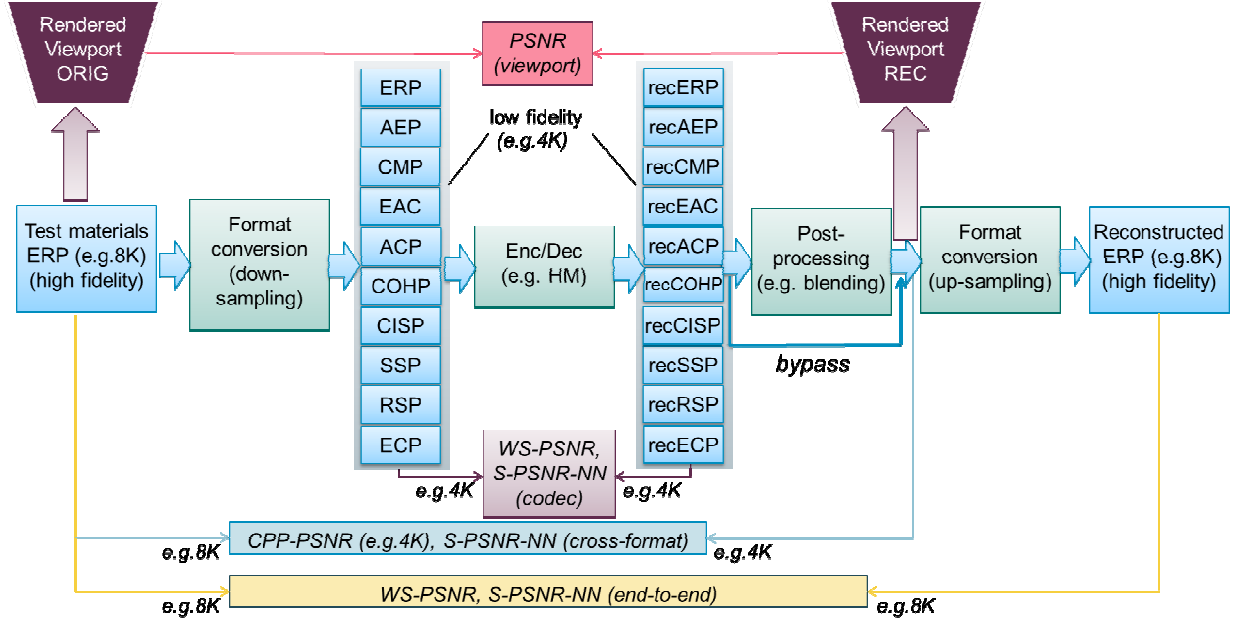

Figure E-1 – Processing chain for 360° video objective metrics

WS-PSNR calculates PSNR using all image samples on the 2D projection plane. The distortion at each position (i, j) is weighted by the spherical area covered by that sample position. For each position (i, j) of an $M \times N$ image on the 2D projection plane, denote the sample values on the reference and test images as $y(i, j)$ and $y'(i, j)$, respectively, and denote the spherical area covered by the sample as $w(i, j)$. The weighted mean squared error (WMSE) is first calculated as:

$$WMSE = \frac{1}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} w(i,j)} \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} (y(i,j) - y'(i,j))^2 * w(i,j) \qquad \text{(E-1)}$$

The WS-PSNR is then calculated as:

$$WS\_PSNR = 10\, log\left(\frac{MAX_I^2}{WMSE}\right) \qquad \text{(E-2)}$$

where MAX_I is the maximum intensity level of the images.

Since WS-PSNR is calculated based on the projection plane, different weights are derived for different projection formats. A description of the WS-PSNR weight derivation is to be provided for the projection formats in the submissions.

E2E WS-PSNR metric will use the ERP format at the same resolution as the input sequence.

For an $M \times N$ image in the ERP format, the weight $w(i, j)$ at position (i, j) is calculated as:

$$w(i,j)_{ERP} = cos\frac{(j + 0.5 - N/2)\pi}{N} \qquad \text{(E-3)}$$

The cross-format quality metrics S-PSNR-NN, S-PSNR-I, and CPP-PSNR measure the quality difference between the original 360° video and the reconstructed video output from the decoder, which may be in a projection format and/or resolution that is different from the original video.

The cross-format S-PSNR-NN in 360Lib-3.0 is calculated following the steps below.
1. Map point S (S is from the set of 655,362 point) to point g in the original video projection plane;
2. Round point g to nearest neighbor point g' at integer sampling position in the original video projection plane;
3. Perform inverse coordinate mapping of the point g' back onto the sphere at S(g');
4. Perform coordinate mapping from the spherical coordinate S(g') to the reconstructed video projection plane at the position r;
5. Apply interpolation to derive the sample value at point r using its neighboring sample values at integer sampling positions (if r is not located at integer sampling position);
6. Calculate error between the sample value at point g' in the original video and the interpolated sample value at point r in the reconstructed video.

Craster parabolic projection PNSR is a generalization of omnidirectional image with a uniform samples distribution close to the distribution on the sample on a surface of unit sphere.

1. No additional weighting is applied to transformed samples.

2. Spherical coordinates are mapped on the 2D plane using following equations:

$$m = R\theta\left[\left(2\cos\frac{2\phi}{3}\right) - 1\right]; \quad n = \pi R \sin\frac{\phi}{3}; \qquad \text{(E-4)}$$

3. where m and n are horizontal and vertical integer luma sample position in CPP projection respectively; θ and ϕ are longitude and latitude in spherical coordinate system; sphere has $R = 1$. Corresponding sampling position in origin projection may be derived using map3Dto2D() method of corresponding projection class.

4. CPP dimensions are derived from input projection parameters in a manner that no additional samples are introduced. Thus CPP width (w_{CPP}) and CPP height (h_{CPP}):

$$w_{CPP} = w_{ERP}^{*}, \qquad \text{(E-5)}$$

$$h_{CPP} = \frac{w_{CPP}}{2}; \qquad \text{(E-6)}$$

where w_{ERP}^{*} is the coding ERP width for cross-projection metrics and hi-fidelity ERP width for end-to-end metrics.