JVET-G1030: JVET common test conditions and evaluation procedures for 360° video

Jill M Boyce
Vimmerse

Yan Ye
Assumption University of Thailand

Joint Video Exploration Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
7th Meeting: Torino, IT, 13–21 July 2017

Document: JVET-G1030v2

Title:	**JVET common test conditions and evaluation procedures for 360° video**		
Status:	Output Document of JVET		
Purpose:	Report		
Author(s) or Contact(s):	Name(s) Jill Boyce, Elena Alshina, Adeel Abbas, Yan Ye	Email:	jill.boyce@intel.com, elena_a.alshina@samsung.com, aabbas@gopro.com, Yan.Ye@InterDigital.com
Source:	Editors		

Abstract

This document defines common test conditions and software reference configurations to be used in video codec performance evaluation for 360° video. These 360° video test conditions are also recommended for use in JVET technical contributions.

1 Introduction

360° video common test conditions are desirable to conduct experiments in a well-defined environment and ease the comparison of the outcome of experiments.

2 Testing procedure

2.1 Definitions

Here and after following abbreviations are used to specify formats (projections) representing 360° video, which are further described in JVET-G1003 [2].

ERP – *Equi-rectangular projection*

CISP – *Compact Icosahedral projection*

CMP – *Cube Map projection*

COHP – *Compact Octahedron projection*

SSP –*Segmented sphere projection*

AEP – *Adjusted equal area projection*

ACP – *Adjusted cube map projection*

RSP – *Rotated sphere projection*

EAC – *Equi-Angular Cube Map projection*

ECP – *Equatorial Cylindrical projection*

2.2 Processing chain

High fidelity test materials are provided in YUV 4:2:0 format representing 360° video in ERP. Prior to encoding, these materials are converted to one of the projection formats listed above according to the testing procedure specified by Figure 1.

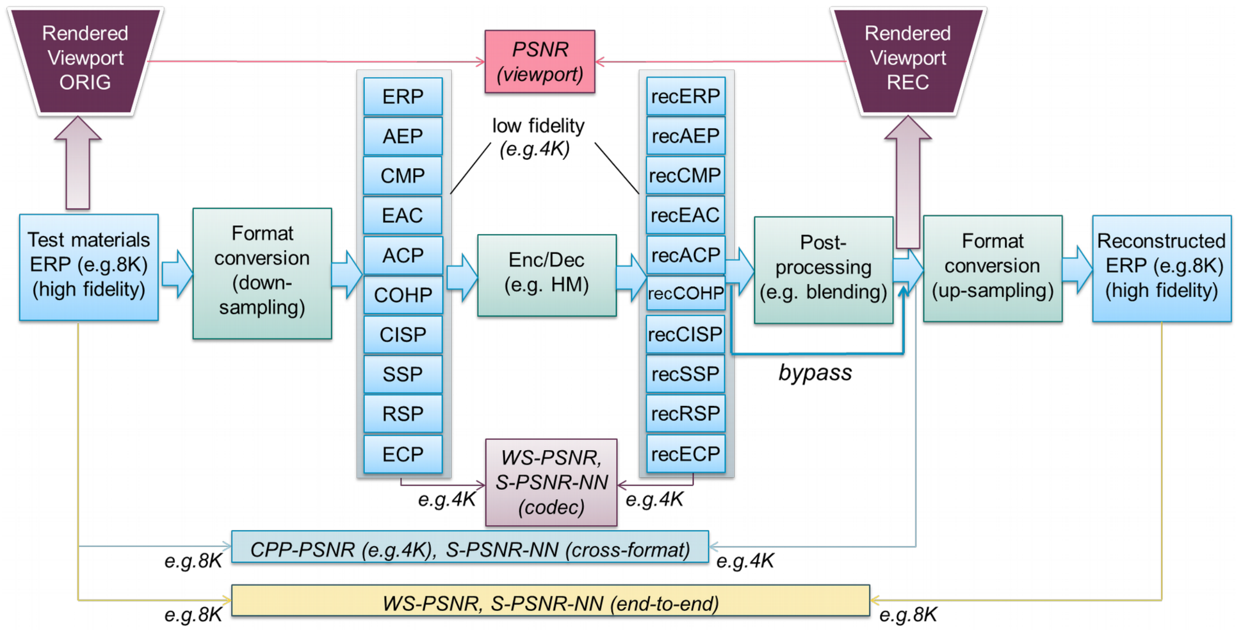

Figure 1 – 360° video testing procedure

3 Software

For conversions and quality measurement, the 360Lib-4.0 software package located at https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-4.0/ should be used.

Version 16.16 of HM HEVC reference software should be used for experiments related to projection formats. For 360°-specific coding tool proposals, JEM7.0 software is recommended.

JVET-G1003 [2] "Algorithm description of projection format conversion and video quality metrics in 360Lib" provides additional information about the projection format conversion process, packing formats, and the video quality metric computations supported in the software package.

4 Test sequences and sizes

4.1 High fidelity input

Table 1 defines the set of test sequences. They are all in YUV 4:2:0 format representing 360° video in ERP. Height (H) of the picture in ERP projection is half of the width (W) reported in Table 1. All frames (as defined by frame count) shall be processed and encoded for all sequences.

Test sequences are available on ftp://jvet@ftp.ient.rwth-aachen.de and ftp://jvet@ftp.hhi.fraunhofer.de in the /testsequences/testset360 directory. Accredited members of VCEG and MPEG may contact the JVET chairs for login information.

Table 1 – High fidelity input test sequences in ERP format

Class	Sequence name	Frame count	Resolution@FPS	Bit-depth	Proposed in
8K	SkateboardInLot	300	8192×4096@30	10	JVET-D0026
8K	ChairLift	300	8192×4096@30	10	JVET-D0026
8K	KiteFlite	300	8192×4096@30	8	JVET-D0039
8K	Harbor	300	8192×4096@30	8	JVET-D0039
8K	Trolley	300	8192×4096@30	8	JVET-D0039
8K	GasLamp	300	8192×4096@30	8	JVET-D0039
6K	Balboa	600	6144x3072@60	8	JVET-D0053
6K	Broadway	600	6144x3072@60	8	JVET-D0053
6K	Landing	300	6144x3072@30	8	JVET-G0147
6K	BranCastle	300	6144x3072@30	8	JVET-G0147
4K	PoleVault_le	300	3840×1920@30	8	JVET-D0143
4K	AerialCity	300	3840×1920@30	8	JVET-D0179

The final versions of the test sequence files names and md5 check sums are given in Table 2:

Table 2 – Test sequences and Md5 checksums

Sequence	Md5
SkateboardInLot_8192x4096_30fps_10bit_420_erp.yuv	e8eae04c43e959060f641fec4892fced
ChairliftRide_8192x4096_30fps_10bit_420_erp.yuv	9126f753bb216a73ec7573ecc4a280c3
KiteFlite_8192x4096_30fps_8bit_420_erp.yuv	18c0ea199b143a2952cf5433e8199248
Harbor_8192x4096_30fps_8bit_420_erp.yuv	aa827fdd01a58d26904d1dbdbd91a105
Trolley_8192x4096_30fps_8bit_420_erp.yuv	25c1082d1e572421da2b16530718156d
GasLamp_8192x4096_30fps_8bit_420_erp.yuv	3417d0b862ffb0fd34f65c3bc810d25c
Balboa_6144x3072_60fps_8bit_420_erp.yuv	1457bb109ae0d47265f5a589cb3464d7
Broadway_6144x3072_60fps_8bit_420_erp.yuv	1baf6667ef84bc3534809463faee8e82
Landing_6144x3072_30fps_8bit_420_erp.yuv	c715e332e78ab30e477445954da66c6b
BranCastle_6144x3072_30fps_8bit_420_erp.yuv	f1b6ade7780c402296a90fd3061cdc36
PoleVault_le_3840x1920_30fps_8bit_420_erp.yuv	d038b3b1497df2ce5894a9f28cb467a8
AerialCity_3840x1920_30fps_8bit_420_erp.yuv	a9eb9e7d44722c76dcbb6a56a8dd7248

4.2 Coding projections

Prior to the encoding, high fidelity test materials shall be converted to one of the coding projections listed above. Frame sizes for each projection, as listed in Table 3 will be used for generation of HM anchors.

For new projection format proposals, the number of coded samples should not exceed 101% of the ERP coded format.

For all conversions between projections prior to the encoding and after decoding, the conversion software shall use 6-taps Lanczos filter for Luma and 4-taps Lanczos filter for Chroma.

Table 3 – Frame sizes for different projections for 8K, 6K and 4K input resolution

Class	8K and 6K Input ERP Resolution				4K Input ERP Resolution			
Format	faceW	faceH	W	H	faceW	faceH	W	H
ERP (coded)	4096	2048	4096	2048	3328	1664	3328	1664
Padded ERP	4096	2048	4112	2048	3328	1664	3344	1664
CISP	952	824	2496	3320	776	672	2056	2712
CMP compact	1184	1184	3552	2368	960	960	2880	1920
COHP	1552	1344	2688	3144	1256	1088	2176	2552
SSP	1184	1184	1184	7136	960	960	960	5792
AEP	4096	2048	4096	2048	3328	1664	3328	1664
ACP	1184	1184	3552	2368	960	960	2880	1920
RSP	1184	1184	3552	2368	960	960	2880	1920
EAC	1184	1184	3552	2368	960	960	2880	1920
ECP	1184	1184	3552	2368	960	960	2880	1920

Referring to Figure 1, the input projection is firstly converted to the coding projection at resolutions specified in Table 3. The output of conversion process is YUV 4:2:0 10 bits video. Therefore, input bit-depth for encoding is 10 bits. Encoding and decoding shall be performed using 10 bits internal bit-depth.

In order to measure objective performance, PSNR based quality metrics for spherical video listed below shall be used at various places in the chain, as defined in JVET-G1003 [2].

Codec Level	WS-PSNR, S-PSNR-NN
Cross-Format	CPP-PSNR, S-PSNR-NN
End-To-End	S-PSNR-NN, WS-PSNR
Viewport	PSNR

After encoding and decoding (i.e. at codec level) WS-PSNR and SPSNR-NN are calculated between original and reconstructed video, using 360Lib.

Between reconstructed coding projection after post-processing (e.g. blending if applicable) and high-fidelity ERP (i.e. at cross-format level), CPP-PSNR and S-PSNR-NN shall be calculated. For CPP-PSNR, the metric is calculated at the coded resolution (i.e. resolution of the reconstructed coding projection) according to Figure 1.

Reconstructed coding projection is converted back to ERP format with picture size equal to the high-fidelity ERP original video. WS-PSNR and SPSNR-NN are measured between original and reconstructed

ERP in high fidelity resolution (this method is also referred to as end-to-end). When padding is used with ERP format, blending post-processing using a linear weighting is applied prior to calculation of the cross-format and end-to-end metrics.

Objective metrics are calculated for dynamic viewports within a Field Of View (FoV), calculated based on 2D rectilinear viewport pictures generated from 360°×180° video. The viewport rendered from the reconstructed coding projection (reconstructed viewport) will be compared with an anchor generated from the high fidelity ERP (original viewport). For each test sequence, 2 dynamic viewports shall be generated, with a field of view size of 75°×75°. Viewport parameters for each test sequence are listed in Table 4. A range of (yaw, pitch) offset positions are provided for the first and last pictures of the sequence, with linear interpolation of the offset position for the other pictures in the sequence. 360Lib software is used to generate the dynamic viewports. The output of viewport generation is 10 bits 4:2:0 YUV video of the size listed in Table 4. Between original viewport and reconstructed viewport PSNR will be measured.

Table 4 – Dynamic viewport parameters

Class	Sequence name	Viewport Size	V1 start (yaw, pitch)	V1 end (yaw, pitch)	V2 start (yaw, pitch)	V2 end (yaw, pitch)
8K	SkateboardInLot	1816×1816	−180, 35	−270, 5	−65, −5	25, 25
8K	ChairliftRide	1816×1816	−45, −15	45, 15	−201, −73	−111, −43
8K	KiteFlite	1816×1816	−114, −53	−24, −23	13, −41	103, −11
8K	Harbor	1816×1816	−83, −56	7, −26	−140, 12	−50, 42
8K	Trolley	1816×1816	210, −18	300, 12	30, −44	120, −14
8K	GasLamp	1816×1816	−51, −11	39, 19	61, −15	151, 15
6K	Balboa	1816×1816	-4, 3	86, 33	-116, -24	-26, 6
6K	Broadway	1816×1816	50, -20	140, 10	-111, -41	-21, -11
6K	Landing	1816×1816	-45, 0	45, 30	-225, 0	-135, 30
6K	BranCastle	1816×1816	116, -15	206, 15	-145, 0	-55, 30
4K	PoleVault	856×856	−45, −15	45, 15	32, −53	122, −23
4K	AerialCity	856×856	−45, −15	45, 15	77, −54	167, −24

In addition, for each projection format, an end-to-end WS-PSNR calculation of the projection mapping process without compression should also be provided, calculated between the original and reconstructed video, but skipping the Enc/Dec step in Figure 1. These results will be referred to as conversion-only.

For coding tool proposals, test results for ERP are mandatory to be reported. Test results for other projection formats are optional.

4.3 Subjective quality check using "evil viewports"

For each projection format, a subjective quality check test should be performed to evaluate the impact of discontinuous edges, as described in [3]. Bitstreams or decoded video should be provided for viewing by the AHG, according to the following procedure.

Two vertex locations will be identified for each projection format, as indicated in Table 5. The first vertex will be along the picture border at a discontinuous edge. The second vertex will be within the picture, not along the picture border, at a position with a maximum number of discontinuous edges. For ERP, the second "vertex" is at the (0, 0) center of the picture, and does not align with a discontinuous edge, as the only discontinuous edge in the format is already represented by the first vertex [4].

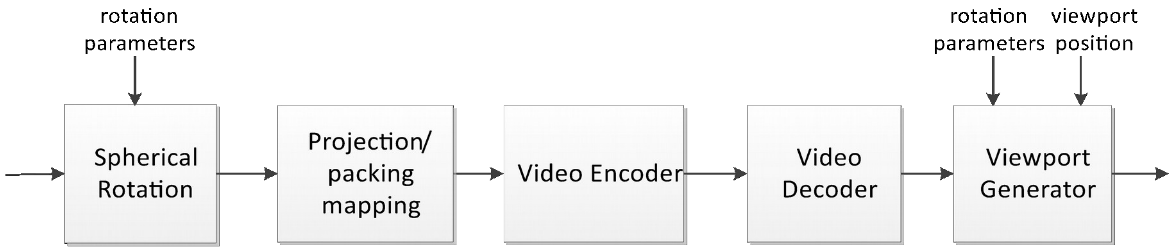

Figure 2 – 360° video subjective testing procedure with “evil viewports”

Table 5 – Vertex positions for “evil viewports”

Format	Vertex 1 (yaw, pitch) (in degrees)	Vertex 2 (yaw, pitch) (in degrees)
ERP	−180, 0	0, 0
CMP	−135, 35	45, −35
CISP	0, 27	−141, −90
COHP	−180, 45	0, 45
SSP	−180, 45	90, −45
AEP	−180, −90	0, −90
ACP	−135, 35	45, −35
RSP	−135, 0	0, −45
EAC	−135, 35	45, −35
ECP	−135, 35	45, −35

Table 6 – Static viewport center before rotation

Class	Sequence name	Viewport Size	Static viewport 1 center (yaw, pitch)	Static viewport 2 center (yaw, pitch)
8K	SkateboardInLot	1816×1816	−225, 20	−20, 10
8K	ChairLift	1816×1816	0, 0	−156, −58
8K	KiteFlite	1816×1816	−69, −38	58, −26
8K	Harbor	1816×1816	−38, −41	−95, 27
8K	Trolley	1816×1816	255, −3	75, −29
8K	GasLamp	1816×1816	−6, 4	106, 0
6K	Balboa	1816×1816	41, 18	-71, -9
6K	Broadway	1816×1816	95, -5	-66, -26
6K	Landing	1816×1816	0, 15	-180, 15
6K	BranCastle	1816×1816	161, 0	-110, 15
4K	PoleVault	856×856	0, 0	77, −38
4K	AerialCity	856×856	0, 0	122, −39

As the sequence is rotated prior to encoding, as illustrated in Figure 2, it requires a separate encoding at the lowest rate point for each identified vertex and each identified static viewport. Note that the vertex indicated in Table 5 is the destination viewport center in the rotated projection format, and the static viewport center defined in Table 6 is the source viewport center in the original picture before rotation. After decoding, static viewports centered at the vertex defined in Table 5 should be extracted for subjective quality check. For a particular projection format, 4 decoded evil viewports shall be provided for each sequence, including (a) Static viewport 1 centered at Vertex 1, (b) Static viewport 1 centered at Vertex 2, (c) Static viewport 2 centered at Vertex 1, and (d) Static viewport 2 centered at Vertex 2. To force the identified vertex to be located in the center of the identified viewport, two steps of rotations are needed by using the 360lib software. Specifically, for a given vertex in (dst_yaw, dst_pitch) Table 5 and a given source viewport center in (src_yaw, src_pitch) Table 6, firstly, the input ERP is rotated to let the source viewport center to be located at the (0, dst_pitch) center of the picture, secondly, it is further rotated to let the (0, dst_pitch) center of the picture to the identified vertex (dst_yaw, dst_pitch). The first step is a pure ERP-to-ERP rotation and the output frame size is the same with the input. The second step is done together with the encoding and format conversion (down-sampling).

5 Encoder configurations and quantization parameter values

Encoder configuration files and their usage, as well as software configuration for both HM and JEM are described in [1]. For each video sequence four quantization parameter values are to be used: 22, 27, 32 and 37.

Aligned with JVET common test condition document [1], four tests, reflecting intra-only, random-access, and low-delay settings are defined:

- Intra, 10 bit
- Random access, 10 bit
- Low delay, 10 bit
- Low delay, P slices only, 10 bit

Among them, Random access is mandatory while remaining are optional. To be aligned with JVET and JCT-VC CTC, in random access tests motion search range is set to 256, GOP size is set to 16 for both HM and JEM.

6 Test results

Templates for reporting test results for HM and JEM coding can be found in https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/.

7 References

[1] K. Suehring, X. Li, "JVET common test conditions and software reference configurations", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-B1010, 2nd Meeting, Feb. 2016.

[2] Y. Ye, E. Alshina, J. Boyce, "Algorithm descriptions of projection format conversion and video quality metrics in 360Lib", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-G1003, 7th Meeting, July 2017.

[3] J. Boyce, Z. Deng, "AHG8: Subjective testing of 360º video projection/packing formats", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-F0021, 6th Meeting, Apr. 2017.

[4] Z. Deng, L. Xu, J. Boyce, "AHG8: Subjective test pilot study of 360º video projection/packing formats", Joint Video Exploration Team of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11, JVET-F0083, 6th Meeting, Apr. 2017.