Joint Video Exploration Team (JVET)
of ITU-T SG 16 WP 3 and ISO/IEC JTC 1/SC 29/WG 11
6th Meeting: Hobart, AU, 31 March – 7 April 2017

Document: JVET- F0036

Title:	**AHG8: Rotated Sphere Projection for 360 Video**
Status:	Input Document to JVET
Purpose:	Proposal

Author(s) or Contact(s): Adeel Abbas, David Newman
Email: aabbas@gopro.com
dnewman@gopro.com

Source: GoPro

Abstract

This contribution provides description of Rotated Sphere Projection (RSP) for 360 video. RSP represents 360 video using only two faces that are arranged in two rows. The top face is directly cropped from middle 270°x90° of ERP, while the bottom face is also cropped from middle 270°x90° of ERP after performing a spherical rotation by 180° along Y-axis and 90° along X-axis.

1 Introduction

Omnidirectional or 360/VR video involves capturing all sides around the observer at the same time. Typically, this is accomplished by way of gen-locked multi-lens camera rigs, which are designed to capture video with some overlap. Stitching tools, which are usually aware of camera configuration, try to remove parallax between different views and create a perfectly calibrated 3D model of the scene. Once content is stitched, it must be projected onto a 2D plane, so that codecs can encode this video representation.

The Equirectangular Projection [9] (ERP) has the advantage of being simple and straightforward to visualize. It also has lowest number of discontinuities and is relatively easy to manipulate using existing video editing tools and rendering engines. This is one of the reasons that ERP is currently the most widely deployed projection format. However, ERP exhibits extreme distortion at poles which results in "pixel inefficiencies" – i.e. much higher number of pixels are devoted to poles than equator. Furthermore, traditional codecs were never designed for content with so much geometric distortions at poles, hence they perform poorly and/or need to work harder.

Numerous projection formats and frame packing arrangements have been proposed recently. Some of these include Icosahedral projection [1] (ISP), Segmented Sphere Projection [6] (SSP), Cubemap Projection [8] (CMP), Octahedron Projection [5] (OHP). There has been considerable work underway at JVET to define common objective and subjective test conditions to evaluate different projection formats [1]. While certain projections seem to have advantages over others, there is a clear need to identify important factors that make one projection more desirable over another. Some of these factors are stated below:

1. **Low surface area compared to sphere**. Projections like ISP and SSP get very close to sphere. As a result, they need 25% lower number of pixels compared to ERP to represent sphere. Alternatively, one can pack 25% more resolution under same level constraints, which is equivalent to 5K resolution on sphere. This is important, because most people agree that VR/360 systems need to improve on resolution. In contrast, CMP must lie farther than sphere (in order circumscribe it) and hence it needs 22% more pixels than ERP (38% more pixels than ISP and SSP).

2. **Number of faces and face arrangement**. A projection like SSP (with only 3 faces) performs better with existing coding tools because 2/3rd of the picture looks almost like normal video i.e. has no soft and hard discontinuities. Projections like ISP and OHP have greater number discontinuities where coding tools suffer. As a result, we observe lower compression performance from these

projections. Secondly, having more faces requires one to devote more pixels in storing padded or overlap region, which again sacrifices resolution.

3. **Aspect ratio**. Most hardware encoders and decoders that are deployed today are designed to optimally handle 3840x2160 video with 16:9 aspect ratio. To guarantee compatibility with already deployed codecs and to be aligned with non-360 applications of a future codec, it is desirable to have an aspect ratio of a projection that is closer to 16:9. Projections that have a very wide aspect ratio, for example SSP and NPCM [7] pose several challenges:

   - Most already deployed codecs are designed to handle maximum picture width of 4K, thus their line buffers are designed such that width of the image will never exceed 3840 or 4096 pixels. If we are to signal a projection with wide aspect ratio and be compliant with these codecs, it appears that resolution will have to be sacrificed.
   - The wide aspect ratio looks unfit for most 16:9 monitors, television sets and phone screens. As a result, one is not able to appreciate resolution of the content in native projection format (because width of the display device becomes limiting factor).
   - Wider aspect ratio does not lend well to some parallel processing tools, e.g. slices in H.264/AVC. Most common multi-slice implementations of H.264 encoders will split 6:1 frame into very thin rectangular slices, which are inefficient for compression (slices are most efficient when they are square).

   There has been a suggestion to rotate SSP by 90° and encode the image in 1:6 aspect ratio. This solves the issues of line buffer and slice parallel processing, but creates new challenges:

   - Most already deployed codecs are designed to handle content in wider aspect ratio (with more horizontal than vertical motion e.g. camera pans). This tall and rotated orientation may result in sub-optimal codec performance (in terms of compression efficiency)
   - With 1:6 aspect ratio, the codec pipeline (which was designed to handle 4K line buffers) may now be sub-optimally used, as the picture is taller. This may also result in worst throughput.

From above discussion, it seems that a good projection would have a more "compliant" aspect ratio, lower surface area compared to sphere and lesser face discontinuities.

2 Rotated Sphere Projection (RSP) Explained

The Rotated Sphere Projection is composed of only two faces arranged in two rows such that aspect ratio of the final picture is 3:2. The first face (or row) can be obtained by directly cropping middle 270°x90° from Equirectangular (ERP) image, as shown by white dotted line in Figure 1.

Figure 1 – Cropping top face of RSP

Since ERP can be viewed as a cylinder surrounding a sphere, the above operation extracts a partial cylinder. The 90°x270° from top, bottom and back side must now be extracted and packed into RSP. To do so, ERP image can be rotated such that pixels at the poles are brought to the equator and pixels corresponding to back face are brought to the front. The new image is shown in Figure 2.

Figure 2 – Cropping bottom face of RSP

The image along dotted line is cropped to get top, back and bottom field of view information. After stacking both the cropped images, final RSP frame is shown in Figure 3.

Figure 3 –RSP frame

The RSP projected video has same height as original ERP video, but width is reduced by 25%. For instance, an ERP frame of size 4096x2048 would convert to size 3072x2048 in RSP format. In Figure 3, since corners of each strip overlap with other strip's middle section, they can be greyed out (to mark them as inactive). Since JVET rules require each projection to have same number of active pixels on sphere, the RSP resolution can be set to 3648x2432 to have same number of active pixels as ERP.

Much like a tennis ball as shown in Figure 4, RSP divides the sphere into two segments (or faces). The face that captures front side is called top face, while the face that captures back side, along with top and bottom is called bottom face. The size and geometry of both faces is identical and as a result, the shape is symmetric.

Figure 4 –Showing RSP faces arranged around sphere like a tennis ball

3 Mathematical Formulation

A typical representation of a sphere in 3D XYZ coordinate system is shown in Figure 5. X axis points toward the front view of the sphere from the center of sphere. Y axis points toward the top view of the sphere from the center of sphere. Z axis points toward the right side of the sphere from the center of sphere.

The sphere can be sampled with longitude (ϕ) and latitude (θ). The longitude ϕ in the range $[-\pi, \pi]$ while latitude θ in the range $[-\pi/2, \pi/2]$. Using convention shown, longitude is defined by the angle starting from X axis in counter-clockwise. The latitude is defined by the angle from equator plane toward Y axis.

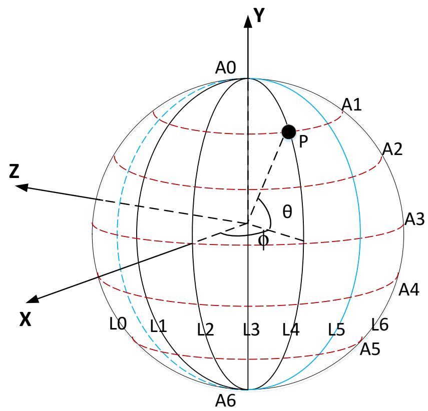

Figure 5 –Sphere representation in 3D XYZ Coordinate System

The top face derivation is trivial, since it is directly cropped from ERP. For the bottom face, we perform following rotation on the sphere:

- 180° rotation along Y-axis (to bring back side to the front)
- 90° along X-axis (to bring polar data to equator).

The X', Y' and Z' coordinates on sphere after performing this rotation are defined below:

$$X' = -X$$

$$Y' = -Z$$

$$Z' = -Y$$

In 2D (u,v) coordinate system, as shown in Figure 6, a point contributed by the top face to the sphere, indicated by coordinates (x^t, y^t) is given as:

$$x^t = \frac{W}{2}\left(\frac{\phi}{\pi}+1\right), \phi \in \left(-\frac{3\Pi}{4}, +\frac{3\Pi}{4}\right]$$
$$y^t = \frac{H}{2}\left(1-\frac{4\theta}{\pi}\right), \theta \in \left(-\frac{\Pi}{4}, +\frac{\Pi}{4}\right]$$

For the bottom face, as shown in Figure 6, a point specified by coordinates (x^b, y^b) is given by equation:

$$x^b = \frac{W}{2}\left(\frac{\phi'}{\pi}+1\right), \phi' \in \left(-\frac{3\Pi}{4}, +\frac{3\Pi}{4}\right]$$
$$y^b = \frac{H}{2}\left(1-\frac{4\theta'}{\pi}\right), \theta' \in \left(-\frac{\Pi}{4}, +\frac{\Pi}{4}\right]$$

The above derivations are same as ERP. The longitude (ϕ') and latitude (θ') after rotation can be expressed in terms of longitude (ϕ) and latitude (θ) before rotation as follows:

$$\theta' = \arcsin(\cos\theta \sin\phi)$$
$$\phi' = \arctan(-\frac{\tan\theta}{\cos\phi})$$

The derivation of above formulae is simple and is not reported for the sake of brevity.

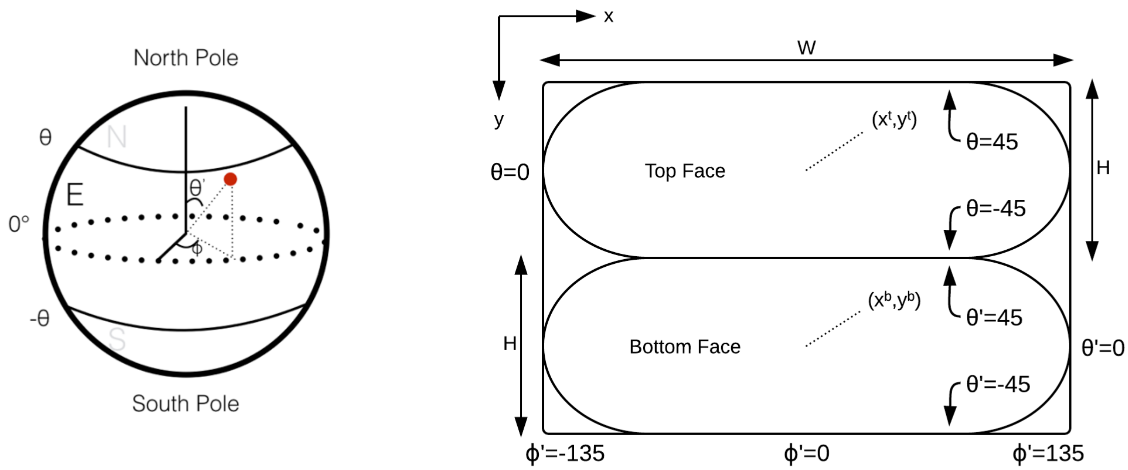

Figure 6 –RSP projection to sphere

4 RSP Conversion Software and Testing Methodology

Conversion tools have been developed that convert ERP to RSP and RSP to ERP. The tools do not have any custom algorithms, rather they call 360Lib [2] for rotation and rescaling. The ERP to RSP tool crops out top facet from ERP and crops bottom facet after performing spherical rotation on ERP. For the bottom facet, rotation and resampling is done in one step (as in 360Lib), to minimize interpolation loss. The RSP to ERP tool performs opposite rotation to put back bottom facet, while top facet is placed as-is.

All settings are same as common test conditions, described in [1]. This includes face sizes, metric computation and encoder software/settings. To fairly compare with other projections, all settings of 360Lib were kept as default. For example, 6-taps Lanczos filter is used for Luma and 4-taps Lanczos filter is used for Chroma. Four end to end (E2E) metrics are reported, namely SPSNR-NN-E2E, SPSNR-I-E2E, CPP-PSNR-E2E and WS-PSNR-E2E. Viewport extraction from RSP is not yet implemented, therefore viewport PSNR numbers have not been reported. Although one can construct viewport from high fidelity reconstructed ERP, it will be unfair comparison because RSP viewport would have had two interpolations.

Face sizes for RSP were set to 1216x1216 and 1008x1008 for 8K and 4K input respectively. This equates to a frame resolution of 3648x2432 and 3024x2016 for 8K and 4K input respectively. With these sizes and the fact that RSP has inactive pixels on the corners, the number of active pixels is nearly same as ERP.

4.1 Conversion Only Results

The conversion-only testing pipeline is shown in Figure 7. This pipeline tests the efficiency of going into and going out of a projection, without effects of compression. This data is important, as it gives an overview of non-codec losses. In [10], 360Lib-1.0 software was used to generate similar data for CTC resolutions and original high fidelity resolutions. For this test, 360Lib-2.1 software was used, along with updated projection arrangements e.g. SSP vertical configuration, compact ISP [4] and OHP1 [5]. Only results with CTC resolutions are being reported here.

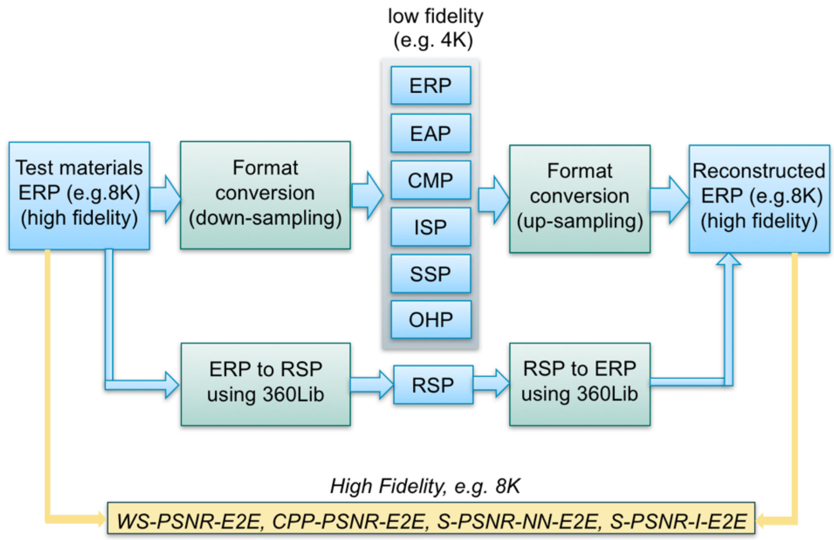

Figure 7 –Conversion-only test flow

In Table 1, average conversion-only results for all projections relative to ERP are shown.

	S-PSNR-NN-E2E			S-PSNR-I-E2E			CPP-PSNR-E2E			WS-PSNR-E2E		
	Y	U	V	Y	U	V	Y	U	V	Y	U	V
CMP	0.67	0.32	0.30	0.48	0.18	0.16	0.52	0.17	0.15	0.68	0.33	0.31
EAP	-1.51	-0.35	-0.09	-1.90	-0.56	-0.32	-1.87	-0.55	-0.31	-1.49	-0.34	-0.09
ISP	2.26	0.46	0.50	2.38	0.37	0.41	2.38	0.37	0.41	2.26	0.46	0.50
OHP	1.25	0.51	0.51	1.12	0.35	0.33	1.14	0.36	0.34	1.28	0.53	0.52
SSP	2.93	1.09	1.07	2.81	0.81	0.78	2.85	0.83	0.80	2.95	1.11	1.09
RSP	2.96	1.14	1.11	2.83	0.84	0.81	2.87	0.85	0.81	2.97	1.15	1.12

Table 1 – Average conversion-only results for all projections

In Table 2, conversion-only results for RSP for different sequences are shown.

	S-PSNR-NN-E2E			S-PSNR-I-E2E			CPP-PSNR-E2E			WS-PSNR-E2E		
RSP	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Trolley	2.78	0.61	0.67	2.95	0.51	0.56	2.97	0.51	0.57	2.81	0.64	0.69
Gaslamp	2.68	0.66	0.70	2.73	0.54	0.59	2.77	0.55	0.59	2.71	0.68	0.72
SkateboardInLot	3.00	1.29	1.07	3.22	1.39	1.17	3.25	1.40	1.18	3.01	1.32	1.11
ChairliftRide	2.64	0.95	0.93	2.83	1.00	0.99	2.88	1.01	1.00	2.68	0.96	0.94
KiteFlite	2.68	0.70	0.70	2.81	0.58	0.59	2.80	0.58	0.60	2.67	0.73	0.73
Harbor	2.77	0.69	0.73	2.84	0.57	0.61	2.92	0.58	0.62	2.81	0.71	0.75
PoleVault	3.91	2.11	1.94	3.40	1.70	1.45	3.40	1.74	1.50	3.91	2.11	1.95
AerialCity	1.69	1.54	1.48	1.16	0.89	0.87	1.20	0.88	0.86	1.68	1.54	1.48
DrivingInCity	3.61	1.17	1.18	3.01	0.49	0.52	3.06	0.49	0.51	3.60	1.17	1.17
DrivingInCountry	3.81	1.66	1.64	3.41	0.72	0.71	3.45	0.72	0.72	3.82	1.65	1.64
Overall	2.96	1.14	1.11	2.83	0.84	0.81	2.87	0.85	0.81	2.97	1.15	1.12

Table 2 – RSP conversion-only results for different sequences

4.2 Encoding Results

Encoder testing pipeline is shown in Figure 8. All encoder settings were set to be same as CTC. Only random access configuration results are reported.

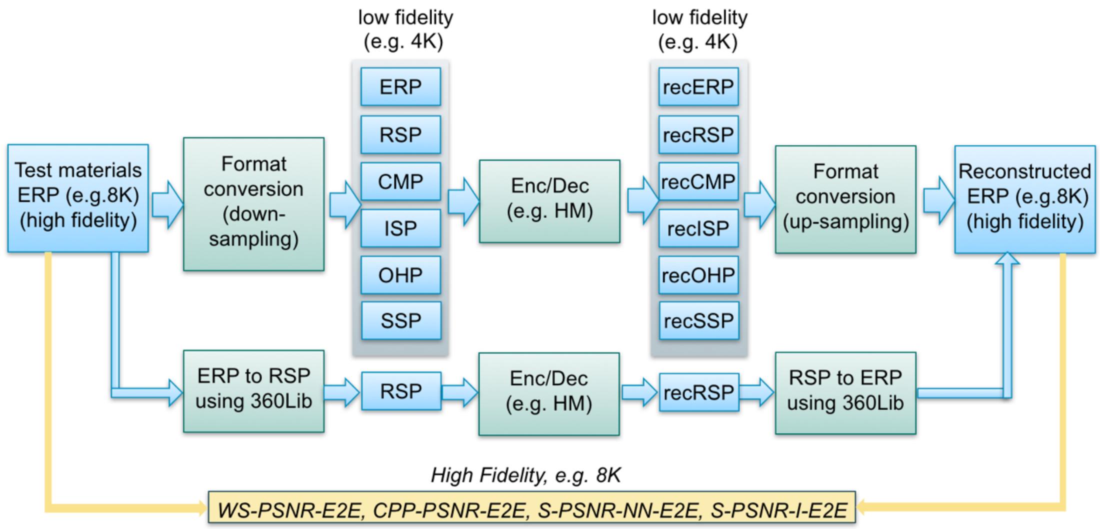

Figure 8 – Encoding test flow

In the first experiment, HM-16.15 was used to run encoding. Results of all projections compared against ERP anchors are reported in Table 3.

	SPSNR-NN-E2E			SPSNR-I-E2E			CPP-PSNR-E2E			WS-PSNR-E2E		
PROJECTION	Y	U	V	Y	U	V	Y	U	V	Y	U	V
CMP	-3.9%	-2.6%	-2.5%	-3.8%	-2.7%	-2.7%	-3.8%	-2.8%	-2.8%	-3.8%	-2.7%	-2.6%
EAP	11.7%	-2.3%	-3.1%	11.9%	-2.4%	-3.2%	11.6%	-2.4%	-3.3%	11.5%	-2.3%	-3.2%
OHP1	-1.3%	3.3%	2.3%	-1.3%	3.0%	2.0%	-1.4%	3.0%	2.0%	-1.4%	3.3%	2.3%
ISP1	-5.0%	-0.4%	-1.2%	-4.9%	-0.6%	-1.3%	-5.0%	-0.6%	-1.4%	-5.2%	-0.5%	-1.3%
SSP(vert)	-10.2%	-4.1%	-4.6%	-10.1%	-4.2%	-4.7%	-10.2%	-4.4%	-4.9%	-10.3%	-4.0%	-4.6%
RSP	-10.3%	-5.2%	-5.5%	-9.5%	-0.4%	0.3%	-9.6%	-0.4%	0.2%	-10.4%	-5.3%	-5.6%

Table 3 –Results for all projections using HM-16.15

In Table 4, results of RSP for individual sequences are shown.

	SPSNR-NN-E2E			SPSNR-I-E2E			CPP-PSNR-E2E			WS-PSNR-E2E		
RSP	Y	U	V	Y	U	V	Y	U	V	Y	U	V
Trolley	-8.4%	2.5%	-2.3%	-8.1%	7.2%	3.8%	-8.2%	7.3%	3.8%	-8.5%	2.4%	-2.5%
GasLamp	-3.9%	3.2%	7.7%	-2.8%	8.8%	13.6%	-3.0%	8.7%	13.7%	-4.1%	2.9%	7.5%
Skateboarding_in_lot	-17.4%	-16.5%	-17.7%	-17.5%	-16.6%	-17.9%	-17.5%	-16.7%	-18.0%	-17.4%	-16.6%	-17.9%
Chairlift	-23.5%	-16.8%	-17.3%	-23.5%	-16.9%	-17.3%	-23.7%	-16.9%	-17.4%	-23.6%	-16.9%	-17.4%
KiteFlite	-9.5%	-3.2%	-4.0%	-9.1%	0.9%	1.4%	-9.0%	0.8%	1.5%	-9.4%	-3.3%	-4.1%
Harbor	-8.2%	-1.9%	-1.6%	-7.3%	4.5%	7.0%	-7.9%	4.5%	7.0%	-8.6%	-2.2%	-1.7%
PoleVault	-8.8%	-9.6%	-9.3%	-7.3%	-7.0%	-5.2%	-7.4%	-6.9%	-5.2%	-8.8%	-9.7%	-9.3%
AerialCity	-1.4%	0.4%	0.7%	0.3%	8.3%	8.2%	0.3%	8.2%	7.9%	-1.4%	0.3%	0.6%
DrivingInCity	-2.6%	7.3%	6.3%	-0.9%	17.0%	15.6%	-0.8%	17.0%	15.6%	-2.6%	7.3%	6.3%
DrivingInCountry	-19.8%	-17.2%	-17.8%	-18.6%	-9.9%	-6.7%	-18.6%	-10.0%	-6.7%	-19.8%	-17.2%	-17.8%
Overall	-10.3%	-5.2%	-5.5%	-9.5%	-0.4%	0.3%	-9.6%	-0.4%	0.2%	-10.4%	-5.3%	-5.6%

Table 4 – RSP results for all sequences

5 Conclusion

Rotated sphere projection (RSP) is proposed. RSP uses two symmetric faces to represent omnidirectional video. It has a simple 3:2 aspect ratio like CMP, proximity to sphere like ISP and SSP, high continuity like SSP and a very simple mathematical representation that is based on ERP. Hence, it is recommended to adopt and encourage further study of this projection format for 360 coding and transport standards.

6 References

[1] J. Boyce, E. Alshina, A. Abbas, Y. Ye, "JVET common test conditions and evaluation procedures for 360° video", Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E1030, 5th Meeting, Jan. 2017.

[2] Y. He, X. Xiu, Y. Ye, V. Zakharchenko, A. Singh, A. Dsouza, C.-C. Huang, J.-L. Lin, Y. Sun, A. Lu, L. Yu, G. Van der Auwera, C. Zhang, Y. Lu, “AHG8: Algorithm description of projection format conversion in 360Lib”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0084, 5th Meeting, Jan. 2017.
[3] V. Zakharchenko, E. Alshina, K.P. Choi, A. Singh, A. Dsouza,” AhG8: Icosahedral projection for 360-degree video content”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-D0028, 4th Meeting, Oct. 2016.
[4] S. N. Akula, A. Singh, A. Dsouza, R.K. K, C. Pujara, R.N. Gadde, V. Zakharchenko, E. Alshina, K.P.Choi, “AHG8: Efficient Frame Packing for Icosahedral Projection”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0029, 5th Meeting, Jan. 2017.
[5] H.-C. Lin, C.-C. Huang, C.-Y. Li, Y.-H. Lee, J.-L. Lin, S.-K. Chang, “AHG8: An improvement on the compact OHP layout” Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0056, 5th Meeting, Jan. 2017.
[6] C. Zhang, Y. Lu, J. Li, Z. Wen, “AHG8: Segmented Sphere Projection for 360-degree video” Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0025, 5th Meeting, Jan. 2017.
[7] K. Kammachi-Sreedhar, M. M. Hannuksela, “AHG8: Nested polygonal chain packing of 360-degree ERP pictures”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0090, 5th Meeting, Jan. 2017.
[8] Facebook Cube Map Implementation: https://code.facebook.com/posts/1638767863078802/under-the-hood-building-360-video/
[9] Equirectangular Projection https://en.wikipedia.org/wiki/Equirectangular_projection
[10] X. Xiu, Y. Ye, P. Hanhart, Y. He (InterDigital), “AHG8: projection format conversion only results using 360Lib”, Joint Video Exploration Team of ITU-T SG16 WP3 and ISO/IEC JTC1/SC29/WG11, JVET-E0134, 5th Meeting, Jan. 2017.
[11] 360Lib-2.1: https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-2.1/

7 Patent rights declaration(s)

GoPro may have current or pending patent rights relating to the technology described in this contribution and, conditioned on reciprocity, is prepared to grant licenses under reasonable and non-discriminatory terms as necessary for implementation of the resulting ITU-T Recommendation | ISO/IEC International Standard (per box 2 of the ITU-T/ITU-R/ISO/IEC patent statement and licensing declaration form).