**3GPP TSG-SA4 Meeting #129-e *S4-241481***

**online, , 19 - 23 Aug 2024 revision of S4aV240044**

|  |
| --- |
| *CR-Form-v12.2* |
| **PSEUDO CHANGE REQUEST** |
|  |
|  | **26.956** | **CR** | **pseudo** | **rev** | **1** | **Current version:** | **0.0.3** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | [FS\_Beyond2D] Representation Format - Stereoscopic Video with extensions |
|  |  |
| ***Source to WG:*** | Qualcomm Incorporated |
| ***Source to TSG:*** |  |
|  |  |
| ***Work item code:*** | FS\_Beyond2D |  | ***Date:*** | 2024-08-12 |
|  |  |  |  |  |
| ***Category:*** | **B** |  | ***Release:*** | Rel-19 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19)* |
|  |  |
| ***Reason for change:*** | The study item description in SP-240479 addresses the following objectives1. Identify and document beyond 2D formats, that are market-relevant within the next years, generated from established and emerging capturing systems (including cameras for spatial video capturing), contribution, and usable on display technologies (smartphones, VR HMDs, AR glasses, autostereoscopic and multiscopic displays).

During SA4#128, several scenarios were defined, that are considered to address the distribution scenarios and evaluation frameworks. However, some of the scenarios already assume a specific Representation Format that seems to be of less relevance initially.The evaluation framework is important, once Representation formats are defined. |
|  |  |
| ***Summary of change:*** | This document focusses on stereoscopic video with extensions. It is a starting point. |
|  |  |
| ***Consequences if not approved:*** |  |
|  |  |
| ***Clauses affected:*** | 4.3.2 |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |

|  |  |  |  |
| --- | --- | --- | --- |
| [S4aV240044](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV240044.zip) | [FS\_Beyond2D] Representation Format - Stereoscopic Video with extensions | Qualcomm Germany | Thomas Stockhammer |

**Revisions**: none**Presenter**: Thomas Stockhammer**Online Discussion**:* Thomas presents
* Serhan : Camera characteristics ?
	+ Thomas : will bring more information on that matter.
* Serhan : Quality observations are just statements from a single person. We need something more proven.
* Waqar : Offer to help on camera characteristics.
* Madhukar : What are these extensions to stereoscopic video?
	+ Thomas : anything not related to stereoscopic but still can be used : HDR, fisheye, …
* JiaYi : Signal properties : Is there any reference for them?
* Gaelle : need to better clarify what could go in TR and what would not. Format description and extensions should be clearly separated.
* Gilles : need to revise document to address these comments

**Decision**:[S4aV240044](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV240044.zip) is **noted**. This document addresses the above documents. |

## ===== CHANGE ===== (add to References)

[X] Apple HEVC Stereo Video - Interoperability Profile (Beta), Version 0.9, June 21, 2023, <https://developer.apple.com/av-foundation/HEVC-Stereo-Video-Profile.pdf>

[Y] Mike Swanson, "Spatial Video", March 7 2024, https://blog.mikeswanson.com/spatial-video/

[Z] Video Contour Map Payload, Version 0.9, June 21, 2023, <https://developer.apple.com/av-foundation/Video-Contour-Map-Metadata.pdf>

[A] ITU-T H.273 (09/23), Coding-independent code points for video signal type identification

[S] M. Satya, "3D Image Reconstruction From Multi-View Stereo", [https://medium.com/@satya15july\_11937/3d-image-reconstruction-from-multi-view-stereo-782e6912435b](https://medium.com/%40satya15july_11937/3d-image-reconstruction-from-multi-view-stereo-782e6912435b), March, 2023.

## ===== CHANGE =====

### 4.3.X Stereoscopic Video with Extensions

#### 4.3.X.1 Definition

Stereoscopic video presents one image to the user’s left eye and another image (typically correlated) to the user’s right eye to produce the stereopsis effect, defined as "the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision." [X].

Different type of Beyond 2D video using projections to left and right eye may be used [Y]:

- rectangular, traditional 3D

- spherically-projected 3D video as defined in TS 26.118.

- either of the two may be extended with additional depth data, also referred to as video contour maps [Z].

In addition, the detailed signal properties of the video each eye needs to be defined:

- Sample aspect ratio for each eye, defined according to the ITU-T H.273 [A], SampleAspectRatio. Typical parameters are 1:1 (value 1) or 4:3 (value 14).

- Picture aspect ratio for each eye. Typical parameters are 1:1 or 16:9.

- Resolutions per eye of left eye and right eye are

- for picture aspect ratio 1:1: 1080x1080, 1440x1440, 2160x2160, 4320x4320

- for picture aspect ratio 16:9: 1280x720, 1440x1080 (with sample aspect ratio 4:3), 3840x2160, 7680x4320

- Framerates for each eye are: 30 fps, 50fps, 60 fps, 90 fps, 120 fps, 144 fps and possibly fractional variants.

- Signal characteristics

- The video signal is YUV with 4:2:0 chroma subsampling.

- Bitdepth: 8 or 10 bits

- Colour primaries, defined according to the ITU-T H.273 [A], ColourPrimaries. Typical parameters are BT-709 (value 1), and BT-2020/BT-2100 (value 9).

- Transfer characteristics, defined according to the ITU-T H.273 [A], TransferCharacteristics. Typical parameters are BT-709 (value 1), BT-2020 (value 14), BT-2100 PQ (value 16) and BT-2100 PQ (value 18).

- Matrix coefficients, defined according to the ITU-T H.273 [A], MatrixCoefficients. Typical parameters are BT-709 (value 1), and BT-2020/BT-2100 non-constant luminance (value 9).

- Typical combined values are BT-709 SDR with (1,1,1), HDR PQ with (9,16,9) and HDR HLG with (9,18,9).

 - Projection parameters:

- Projection: rectilinear, fisheye, equirectangular

- Field-of-view and restricted coverage.

Additional metadata may be present as follows:

- hero eye: A value that indicates which eye is the primary eye when rendering in 2D.

- camera parameters: camera parameters are typically represented in a 3 × 4 projection matrix called the camera matrix. The extrinsic parameters define the camera pose (position and orientation) while the intrinsic parameters specify the camera image format, specifically:

- extrinsic parameters denote the coordinate system transformations from 3D world coordinates to 3D camera coordinates. For details see: https://en.wikipedia.org/wiki/Camera\_resectioning#Extrinsic\_parameters

- intrinsic parameters describe a specific camera model. These parameters encompass focal length, image sensor format, and camera principal point. For details see: https://en.wikipedia.org/wiki/Camera\_resectioning#Intrinsic\_parameters

- disparity adjustment:

- horizontal disparity adjustment, a value that indicates a relative shift of the left and right images, which changes the zero-parallax plane.

- Disparity/depth map: 10bit, same resolution as source content, monochrome, can possibly be sub-sampled

- Line time (per camera) – rolling shutter readout time, only relevant in poorer quality/reduced functionality camera pipelines typically used on HMD tracking cameras.

- Examples: https://github.com/MPEGGroup/FileFormatConformance/tree/m62054\_exintrinsics/data/file\_features/under\_consideration

#### 4.3.X.2 Production and Capturing Systems

The formats as defined in clause 4.3.X.1 may be captured at least with a reduced set of parameters by mobile devices and Head Mounted Displays (HMD) – for more details refer to the following information:

- <https://techcrunch.com/2023/12/11/apple-releases-spatial-video-recording-on-iphone-15-pro/>

- Spatial Video with 1080p at 30fps

- <https://9to5mac.com/2024/01/04/will-the-iphone-16-be-able-to-record-4k-spatial-video/>

- Spatial Video with 4K is expected to be available

- <https://appleinsider.com/articles/24/03/06/capturing-spatial-video-apple-vision-pro-vs-iphone-15-pro>

- The spatial video captured is in a square 1:1 format at 2200 pixels by 2200 pixels. It is a near-perfect recreation of the passthrough viewed by the user.

- Once stereo is captured on supporting phones, offline postprocess can be used to acquire accompanying depth (using for example Depth-Anything <https://github.com/DepthAnything/Depth-Anything-V2/tree/main> and [ZoeDepth](https://github.com/isl-org/ZoeDepth%22%20%5Co%20%22https%3A//github.com/isl-org/zoedepth%22%20%5Ct%20%22_blank) https://github.com/isl-org/ZoeDepth or similar).

- Meta Quest™ can record spatial video: <https://360rumors.com/quest-3-3d-videos/>

- After recording, the video or photo is captured in side-by-side format, with a square aspect ratio. Photos will also be side-by-side but they are stretched vertically, and need to be edited to fix that.

- <https://deovr.com/blog/84-record-vr-footage-on-the-meta-quest-3>

- The Meta Quest 3™ features two cameras that deliver full-color passthrough, allowing users to record content in 4K (2k per eye), using the Meta Quest Developer HUB (https://developer.oculus.com/documentation/unity/ts-odh).

- The Quest 3's passthrough cameras record footage that is flat 120-100 (possibly 90) degrees.

Beyond user-generated content, an ecosystem is developing around this format including movie production, documentaries and live sports. Examples are mentioned here:

- <https://www.apple.com/newsroom/2024/02/2024-mls-season-kicks-off-today-exclusively-on-mls-season-pass-on-apple-tv/>

- <https://www.apple.com/newsroom/2024/01/apple-previews-new-entertainment-experiences-launching-with-apple-vision-pro/>

- <https://www.macrumors.com/2024/01/08/vision-pro-movies-games/>

#### 4.3.X.3 Rendering and Display Systems

Spatial video with the above parameters can be viewed on different rendering and display systems, including

- Backward-compatible to 2D (just view one eye), hence can be viewed on regular phones

- Apple Vision Pro ™

- Meta Quest ™: <https://techcrunch.com/2024/02/01/meta-quest-adds-support-for-apples-spatial-video-ahead-of-vision-pro-launch/>

In addition, OpenXR and WebXR defines APIs to render stereoscopic video with additional metadata.

- OpenXR APIs exist

- WebXR APIs exist

For rendering multi-view stereo video, including 3D reconstruction, refer to [S]. It is shown, how additional metadata can be used to improve rendering.

#### 4.3.X.4 Supporting Information

The baseline video can be encoded using HEVC-based encoding tools:

- framepacking (see for example TS 26.118)

- MV-HEVC (see TR 26.966)

The content can be delivered using regular ISO BMFF based distribution, including streaming with DASH/HLS/CMAF.

Editor’s Note

- Typical quality criteria for evaluating the format

 - Existing test and reference sequences

 - Conversion from other formats (lossless, lossy)

 - Uncompressed data size

- Extensibility of the format

#### 4.3.X.5 Benefits and Limitations

##### 4.3.X.5.1 Benefits

The extended stereoscopic video format has the following benefits:

- Simplicity: The technology is supported by existing content production workflows

- Device Support: The technology is supported by emerging devices on the market

- In device decoding and rendering: The technology generally allows that decoding and rendering can be done in the device, which makes it robust against impaired or lossy network connections.

- Content Industry starts to embrace the format, for details see clause 4.3.X.2

- The format is extensible to add additional metadata, for details see clause 4.3.X.1

- User-generated content production workflows exist.

- Backward-compatible rendering. The content can be rendered on 2D displays without quality degradation.

- It promises very good user experience: "It isn’t too hyperbolic to say that immersive video — when done right — makes you feel like you’ve been teleported to a new location. While you might “have seen” a place in a traditional flat video, with immersive video, you’ve “been there.” If you haven’t experienced video like this, I’m jealous, because there’s nothing like the first time. If you own a Vision Pro and haven’t watched these yet, stop reading and do it now!" [Y]

##### 4.3.X.5.2 Limitations

The format is primarily used to support lean-back and seated experiences, typically head movements with 3DOF and 3DOF+ can be supported, but may be extended in the future to address additional degrees freedom.