**3GPP TSG- Meeting #**

**, FL, United States, 19 -**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.2* | | | | | | | | |
| **PSEUDO CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  |  | **CR** |  | **rev** |  | **Current version:** |  |  |
|  | | | | | | | | |
| *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:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** |  | | | | | | | | | |
| ***Source to TSG:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** |  | | | | |  | ***Date:*** | | |  |
|  |  | | | |  | |  | | |  |
| ***Category:*** |  |  | | | | | ***Release:*** | | |  |
|  | *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 can be 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:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | |  | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **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 is not yet completed, but will be extended with the detailed definition. It is also expected that test material will be provided for the operation points. | | | | | | | | |
|  | |  | | | | | | | | |
| ***This CR's revision history:*** | | |  |  |  |  | | --- | --- | --- | --- | | [**S4-241898**](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_130_Orlando/Docs/S4-241898.zip) | [FS\_Beyond2D] Updates to Stereoscopic Representation Formats | Qualcomm Incorporated | Thomas Stockhammer |   **Revisions**: none  **Online Discussion**:   * **Session 6: 11:00 - 12:30** * Thomas presents. * Madhukar: for user generated content the use of depth and alpha is not clear, why is it needed? Why do they need to be sent to the receiver? What UC are we targeting?   + Alexis: you can extract the background.   + Madhukar: yes you can do that at the source since the sender does it not the receiver.   + Alexis: at the receiver you can use alpha to include other objects in the same scene   + Madhukar: what is the value of sending depth? Capture at the sender is fine. * Gilles: note for clarification. * Gaelle: can we change the terminology for immersive video.   + Thomas: it was the terminology used by Apple. * Gaelle: what are you going to test?   + Thomas: the contribution is about representation formats   + Gaelle: why are we including this without tests? Are these 5 representation formats?   + Thomas: the first 3 are covered in VOPs * Serhan: why do we have so many names for stereoscopic video?   **Decision**:   * Session 6: parked | | | | | | | | |

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

# 2 References

The following documents contain provisions which, through reference in this text, constitute provisions of the present document.

- References are either specific (identified by date of publication, edition number, version number, etc.) or non‑specific.

- For a specific reference, subsequent revisions do not apply.

- For a non-specific reference, the latest version applies. In the case of a reference to a 3GPP document (including a GSM document), a non-specific reference implicitly refers to the latest version of that document *in the same Release as the present document*.

[1] 3GPP TR 21.905: "Vocabulary for 3GPP Specifications".

[2] Allied Market Research, “3D Technology Market Size, Share, Competitive Landscape and Trend Analysis Report by Product, Application : Global Opportunity Analysis and Industry Forecast, 2021-2030.”, [www.alliedmarketresearch.com/3d-technology-market.](http://www.alliedmarketresearch.com/3d-technology-market.)

[3] Mordor Intelligence, “Mobile 3D Market Size & Share Analysis - Growth Trends & Forecasts (2024 - 2029).”, <https://www.mordorintelligence.com/industry-reports/mobile-3d-market.>

[4] Grand View Research, “Immersive Technology Market Size, Share & Trends Analysis Report By Component (Hardware, Software, Services), By Technology, By Application, By Industry, By Region, And Segment Forecasts, 2023 - 2030.”, [https://www.grandviewresearch.com/industry-analysis/immersive-technology-market-report.](https://www.mordorintelligence.com/industry-reports/mobile-3d-market.)

[5] 3GPP TS 26.119: "Media Capabilities for Augmented Reality".

[6] 3GPP TS 26.118: "Virtual Reality (VR) profiles for streaming applications".

[7] 3GPP TS 26.143: "Messaging Media Profiles".

[8] 3GPP TS 26.511: "5G Media Streaming (5GMS); Profiles, codecs and formats".

[9] 3GPP TR 26.966: " Evaluation of new HEVC coding tools".

[10] 3GPP TS 26.265: "Media Delivery: Video Capabilities and Operating Points".

[11] 3GPP TR 26.955: "Video codec characteristics for 5G-based services and applications".

[S1] Apple HEVC Stereo Video - Interoperability Profile (Beta), Version 0.9, June 21, 2023, <https://developer.apple.com/av-foundation/HEVC-Stereo-Video-Profile.pdf>

[S2] Mike Swanson, "Spatial Video", March 7 2024, https://blog.mikeswanson.com/spatial-video/

[S3] Video Contour Map Payload, Version 0.9, June 21, 2023, <https://developer.apple.com/av-foundation/Video-Contour-Map-Metadata.pdf>

[S4] ITU-T H.273 (09/23), Coding-independent code points for video signal type identification

[S5] M. Satya, "3D Image Reconstruction From Multi-View Stereo", <https://medium.com/@satya15july_11937/3d-image-reconstruction-from-multi-view-stereo-782e6912435b>, March, 2023.

[D1] Greg Turk, The Polygon File Format, Stanford University, 1994.

[D2] Volumetric Format Association VFA, <https://www.volumetricformat.org/>

[D3] V-PCC, Visual volumetric video-based coding (V3C) and video-based point cloud compression (V-PCC), ISO/IEC 23090-5 2nd Ed, Nov 2023.

[D4] G-PCC, Geometry-based point cloud compression, ISO/IEC 23090-9, Mar 2023

[D5] Draco Bitstream Specification, <https://google.github.io/draco/spec/>

[N1] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2021. NeRF: representing scenes as neural radiance fields for view synthesis. Commun. ACM 65, 1 (January 2022), 99–106. https://doi.org/10.1145/3503250

[N2] Gao, Kyle et al. “NeRF: Neural Radiance Field in 3D Vision, A Comprehensive Review.” (2022).

[N3] Li, Sicheng et al. “NeRFCodec: Neural Feature Compression Meets Neural Radiance Fields for Memory-Efficient Scene Representation.” ArXiv abs/2404.02185 (2024): n. pag.

[N4] Dong-Ha Kim, Jun Young Jeong, Gwangsoon Lee, and Jae-Gon Kim "Compression method of NeRF model using NNC and VVC", Proc. SPIE 13164, International Workshop on Advanced Imaging Technology (IWAIT) 2024, 131642V (2 May 2024); https://doi.org/10.1117/12.3019533

[N5] G. Lafruit, Y. Liao, and G. Bang, “AhG on Implicit Neural Video Representations (INVR),” ISO/IEC JTC1/SC 29/WG04, M60641, Oct. 2022.G. Lafruit, Y. Liao, and G. Bang, “AhG on Implicit Neural Video Representations (INVR),” ISO/IEC JTC1/SC 29/WG04, M60641, Oct. 2022

[N6] RABBY, AKM SHAHARIAR AZAD and Chengcui Zhang. “BeyondPixels: A Comprehensive Review of the Evolution of Neural Radiance Fields.” ArXiv abs/2306.03000 (2023): n. pag.

[N7] Daniel Duckworth, Peter Hedman, Christian Reiser, Peter Zhizhin, Jean-François Thibert, Mario Lučić, Richard Szeliski, and Jonathan T. Barron. 2024. SMERF: Streamable Memory Efficient Radiance Fields for Real-Time Large-Scene Exploration. ACM Trans. Graph. 43, 4, Article 63 (July 2024), 13 pages. https://doi.org/10.1145/3658193

[N8] Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. ACM transactions on graphics (TOG), 41(4), 1-15..

[R1] OpenCV, <https://opencv.org>.

[R2] Colmap, <https://colmap.github.io/index.html>.

[R3] AliceVision Photogrammetric Computer Vision Framework, <https://alicevision.org>.

[R4] Open Multiple View Geometry (openMVG), <https://github.com/openMVG/openMVG>.

[R5] Higher-quality projected stereoscopic video Depth Estimation (IVDE), <https://gitlab.com/mpeg-i-visual/ivde>.

[R6] Test model for MPEG Higher-quality projected stereoscopic video , <https://gitlab.com/mpeg-i-visual/tmiv>.

[R7] Reference view synthesizer, <https://gitlab.com/mpeg-i-visual/rvs>.

[R8] Open Realtime Depth Image Based Renderer (OpenDIBR), <https://github.com/IDLabMedia/open-dibr>.

[R9] A. Dziembowski, D. Mieloch, J. Stankowski and A. Grzelka, "IV-PSNR – the objective quality metric for Higher-quality projected stereoscopic video applications," in IEEE Transactions on Circuits and Systems for Video Technology, doi: [10.1109/TCSVT.2022.3179575](https://doi.org/10.1109/TCSVT.2022.3179575), software: <https://gitlab.com/mpeg-i-visual/ivpsnr>

[R10] Quality Metrics for Higher-quality projected stereoscopic video (QMIV), <https://gitlab.com/mpeg-i-visual/ivpsnr>.

[3dtv] A. Quested and B. Zegel, "3D-TV production standards - first report of the ITU-R Rapporteurs", EBU Technical Review, 2011 Q2, <https://tech.ebu.ch/publications/trev_2011-Q2_3dtv_quested>

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

### 4.3.2 Extensions to Stereoscopic Video Representation Formats

Editor’s Note: Some additional references need to be extracted.

#### 4.3.2.1 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." [S1].

#### 4.3.2.2 Stereoscopic Video format description according to TS 26.265

Stereoscopic video is defined in TS 26.265 [26265]. The focus on the definition in TS 26.265 is on providing a consistent description of stereoscopic video based on professional production, for example Hollywood movies. In this case, a view is available to be presented to the left eye and another view is available to be presented simultaneously to the right eye. The presentation of both the left and right views allows for an effect known as stereopsis which can be defined as "the perception of depth produced by the reception in the brain of visual stimuli from both eyes in combination; binocular vision." For signal representations, [3dtv] recommends that and Left and Right eyes comply to regular image formats auch as Rec ITU-R BT.709 or Rec ITU-T BT.2100 and any necessary 3D-specific metadata is incorporated with the data. Hence, for stereoscopic video, two synchronized video signals are available, each with identical parameters. The baseline content is described in TS 26.265 as part of the 3GPP Stereoscopic 3D TV Format in clause 4.4.3.4 as follows:

The stereoscopic 3D Cinema format uses two signals, one for the left eye and another view for the right eye as defined in Table 4.4.2-1. The components for each eye closely follow the specifications of the 3GPP HDR signals, but there are some restrictions and extensions, namely:

- Frame rates include high frame rate for movies, namely 48 fps.

- the spatial resolution is restricted to 4K

An informative summary of the parameters of a 3GPP Stereoscopic 3D TV format based on the parameters defined in Table 4.4.2-1 of TS 26.265 is provided in Table 4.3.2.1-1.

Table 4.3.2.1-1 Video Signal Parameters for 3GPP Stereoscopic 3D Cinema format

|  |  |
| --- | --- |
| Parameter | Restrictions |
| Picture aspect ratio | 16:9 |
| Spatial Resolution width x height | 3 840 × 2 160, 1 920 × 1 080  NOTE: For 1080, typically the encoded signal has 1088 lines and cropping is applied to remove spatial samples that are not presented. |
| Scan Type | the source scan type of the pictures as defined in clause 7.3 of Rec. ITU-T H.273 is progressive |
| Chroma format indicator | The chroma format indicator is 4:2:0. |
| Colour primaries  Transfer Characteristics  Matrix Coefficients | Only the following value combinations are permitted: (1, 1, 1), (9, 16, 9), and (9, 18, 9) for SDR, HDR PQ and HDR HLG, respectively. |
| Bit depth | The permitted value is 10 bit. |
| Colour primaries | Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Transfer Characteristics | Only the value 16 (for PQ) as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Matrix Coefficients | Only the value 9 as defined in clause 8.2 of Rec. ITU-T H.273 is permitted. |
| Frame rates | The permitted values are 60, 60/1.001, 48, 48/1.001, 50, 30, 30/1.001, 25, 24, 24/1.001 fps. |
| Frame packing | No frame packing is applied. |
| Projection | No projection is used. |
| Sample aspect ratio | The pixel aspect ratio is 1 (square pixel), i.e. only the value 1 as defined in clause 7.3 of Rec. ITU-T H.273 is permitted. |
| Chroma sample location type | For SDR, the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 0 (Chroma samples are colocated with the luma samples at the top-left corner).  For HDR PQ and HLG, the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 2 (chroma samples are centered horizontally between two luma samples). |
| Chroma sample location type | the location of chroma samples relative to the luma samples for progressive frames as defined in Rec. ITU-T H.273, clause 8.7 is set to 2 (chroma samples are centered horizontally between two luma samples). |
| Range | Restricted video range is used. |
| Stereoscopic Video | A signal for the Left and for the Right Eye is provided whereby the signals have the identical parameters as above and are timely synchronized. |

The format focuses on existing professionally generated movie content. This study will not evaluate further this format.

#### 4.3.2.3 Extensions to Stereoscopic Video Representation formats

Extensions to the above content format result from different new use cases:

1) User generated stereoscopic content from modern devices that allow to capture beyond 2D video formats.

2) Higher-quality projected stereoscopic video that includes higher resolution and projections for professional content

The above use cases require extensions to the Stereoscopic 3D Cinema format defined in TS 26.265.

For user-generated stereoscopic content production systems as introduced in clause 4.3.2.2, extensions to the above basic signals are to be considered. The major extensions are more flexible spatial resolutions beyond 16:9, as well as additional metadata, including alpha and depth information. Offline postprocess can be used to acquire accompanying depth and such information is beneficial in the rendering to reduce parallax effects.

Stereoscopic video for user-generated content may use projections to left and right eye as follows [S2]:

- rectangular, traditional 3D

- extensions with additional depth data, also referred to as video contour maps [S3].

- extensions with additional alpha maps.

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 [S4], 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

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

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

- Signal characteristics

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

- Bit depth: 8 or 10 bits

- Colour primaries, defined according to the ITU-T H.273 [S4], 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 [S4], TransferCharacteristics. Typical parameters are BT-709 (value 1), BT-2020 (value 14), BT-2100 PQ (value 16) and BT-2100 HLG (value 18).

- Matrix coefficients, defined according to the ITU-T H.273 [S4], 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) and HDR PQ with (9,16,9).

Additional metadata may be present, either on a static or per frame basis, 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

- alpha maps: 8 bit, same resolution as source content

- 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>

For higher-quality projected stereoscopic video, the following are the core extensions beyond the TS 26.265 3D TV Format:

- projected video

- higher resolutions up to 8K

- additional depth and alpha data

For higher-quality projected stereoscopic video extensions content may use projections to left and right eye as follows [S2]:

- spherically-projected 3D video as defined in TS 26.118 [6].

- extended with additional depth data, also referred to as video contour maps [S3].

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 [S4], 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: 2160x2160, 4320x4320

- for picture aspect ratio 16:9: 3840x2160, 7680x4320

NOTE: 8K resolution is supported in TS 26.118 [6], and also supported in terms of decoding on modern mobile systems-on-chip. Whether 8K is supported in a full end-to-end workflow is application dependent, but withr appropriate capability negotation, a suitable resolution can be determined.

- Framerates for each eye are: 24 fps, 30 fps, 48fps, 50fps, 60 fps, 90 fps, 120 fps, 144 fps and possibly fractional variants.

NOTE: 120 and 144 fps are supported in terms of decoding on modern mobile systems-on-chip. Whether such high-frame rates supported in a full end-to-end workflow is application dependent, but withr appropriate capability negotation, a suitable resolution can be determined.

- Signal characteristics

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

- Bit depth: 10 bits

- Colour primaries, defined according to the ITU-T H.273 [S4], ColourPrimaries being BT-2100 (value 9).

- Transfer characteristics, defined according to the ITU-T H.273 [S4], TransferCharacteristics. being BT-2100 PQ (value 16).

- Matrix coefficients, defined according to the ITU-T H.273 [S4], MatrixCoefficients being BT-2100 non-constant luminance (value 9).

- The core presentation format is HDR PQ with (9,16,9).

- Projection parameters:

- Projection: fisheye, equirectangular

- Field-of-view and restricted coverage, typically 180 degree.

NOTE: The parameters may be aligned with TS 26.118 [6]

Additional metadata may be present, either on a static or per frame basis, as follows:

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

- 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

#### 4.3.2.4 Production and Capturing Systems

The formats as defined in clause 4.3.2.1 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) 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.

NOTE: In TV productions it was known that there were issues with visual fatigue, nausea due to bad content production. Guidelines that professional producers can take into account have been provided which minimize these effects. Indications whether this also is an issue for user generated content is for further study.

Beyond user-generated stereoscopic 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/>

Latest information on content production can for example also be found here: https://www.provideocoalition.com/creating-stereoscopic-video-for-the-apple-vision-pro/.

#### 4.3.2.5 Rendering and Display Systems

Stereoscopic 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. The stereoscopic effect is lost in this case.

- 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 define 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 [S5]. It is shown, how additional metadata as defined in clause 4.3.2.1 can be used to improve rendering.

#### 4.3.2.6 Supporting Information

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

- framepacking (see for example TS 26.118 [6])

- MV-HEVC (see TR 26.966 and TS 26.265)

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

Uncompressed data rate can be computed as 2 x height x width x (1.5 + depthflag + alphaflag) \* framerate \* bitdepth. Some examples are provided in Table 4.3.2.6-1.

Table 4.3.2.6-1 Uncompressed data rate examples

|  |  |
| --- | --- |
| Signal | Data rate |
| Stereoscopic 3D TV Format HD | 2 x 1080 x 1920 x (1.5 + 0 + 0) \* 24 \* 10 = 1.39 Gbit/s |
| Stereoscopic 3D TV Format UHD | 2 x 2160 x 3840 x (1.5 + 0 + 0) \* 24 \* 10 = 5.56 Gbit/s |
| User Generated Stereoscopic Content HD | 2 x 1080 x 1080 x (1.5 + 0 + 0) \* 30 \* 8 = 800 Mbit/s |
| User Generated Stereoscopic Content UHD with depth | 2 x 2160 x 2160 x (1.5 + 1 + 0) \* 30 \* 10 = 6.52 Gbit/s |
| Higher-quality projected stereoscopic video at 8K with alpha and depth | 2 x 4320 x7680 x (1.5 + 1 + 1) \* 60 \* 10 = 129.77 Gbit/s |

Editor’s Note

- Typical quality criteria for evaluating the format

- Existing test and reference sequences

- Conversion from other formats (lossless, lossy)

- Extensibility of the format

#### 4.3.2.7 Benefits and Limitations

##### 4.3.2.7.1 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 4.3.2.2

- The format is extensible to add additional metadata, for details see clause 4.3.2.1

- User-generated content production workflows exist.

- Backward-compatible rendering. The content can be rendered on 2D displays.

- Very good B2D user experiences have been reported, when the content is properly produced and suitable devices for playback and rendering are used [S2].

##### 4.3.2.7.2 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.

Editor’s Note: More Benefits and limitations will be added over time