**Source:** **Deutsche Telekom, Fraunhofer HHI, Interdigital Europe, KDDI, Nokia, Philips, Samsung and Sony**

**Title:** **[Beyond2D] Scenario on Streaming of Beyond 2D Produced Content - Use Case “Volumetric Video with single asset”**

**Agenda Item:** **9.9**

**Document for:**  **Agreement**

# Introduction

A new study item FS\_Beyond2D ([SP-240479](https://www.3gpp.org/ftp/TSG_SA/TSG_SA/TSGS_103_Maastricht_2024-03/Docs/SP-240479.zip)) was approved at SA#103. One of the objectives of the study is:

2. Establish and document a set of beyond 2D video end-to-end reference scenarios, including real-time communication, streaming services, split rendering, and messaging and corresponding workflows (capturing, encoding, packaging, delivery, decoding, rendering, including general constraints on latency, as well as complexity) to support 3GPP network related delivery and devices leveraging the generation or display technologies. This includes identifying and defining relevant beyond 2D formats in the context of above workflows, and representation technologies to support delivery of these formats within 3GPP networks.

This document is an update to the scenario in the document [S4-240837](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_127-bis-e/Docs/S4-240837.zip+) presented in the SA4#127-bis-e meeting.

In this contribution, a draft scenario on Streaming of Beyond 2D Produced Content including a use case “Volumetric Video with single asset” is proposed for incorporation into FS\_Beyond2D TR 26.956 as basis for future work. The scenario is structured according to the template provided in TR 26.065 v.0.0.1 ([[S4-240825](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_127-bis-e/Docs/S4-240825.zip)](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_127-bis-e/Docs/S4-240825.zip)).

**========================= CHANGE 1 (all new) ==========================**

## 6.x Scenario #x: Streaming of Beyond 2D Produced Content – Use Case “Volumetric Video with single asset”

**6.x.1 Scenario name**

Streaming of Beyond 2D Produced Content - Use case “Volumetric Video with single asset”

**6.x.2 Motivation for the scenario**

*What is the market relevance of the proposed scenario within the next few years? Are there any commercially available or pre-released products or prototypes?*

This scenario handles the streaming of produced Beyond 2D content providing experiences beyond what is achievable with 2D content. “Beyond 2D” content may be in the form of volumetric video, which is a frame-based immersive experience whereby each frame represents a volumetric region in 3D space in which any point is either non-occupied or having a colour that may depend on the viewing direction.Volumetric video has the potential to provide a more immersive and interactive experience for use cases in diverse domains such as e.g. education, entertainment, and industrial monitoring.

Streaming of volumetric video has been previously considered in 3GPP in TR 26.928 (Cl. A.4 - Streaming of Immersive 6DoF, Cl 5.4 - XR Multimedia Streaming) [X2] and TR 26.998 (Cl. A.3 - Use Case 18: Streaming of volumetric video for glass-type MR devices) [X3].

On-demand volumetric video streaming enables distribution of high-quality, professionally captured and produced volumetric video content. Some aspects of production and capturing systems for volumetric representation formats such as point clouds and meshes are documented in TR 26.928, clause 4.6.7.

Several use cases of on-demand volumetric video streaming can be envisioned related to various domains including education, entertainment and industrial monitoring. For example, in an education/training scenario, a pre-recorded video of a fitness instructor showing how to perform an exercise can help the student to better understand how the exercise is done and thus replicate in a correct way. Another example in the education domain would be a mechanic giving a tutorial on how to assemble a mountain bike. The viewer can watch the movements of the mechanic from different angles and get an improved understanding of the different steps due to depth perception and different viewpoints. In the entertainment domain, users can stream a performance from their favorite band to their living room and experience greater immersion potentially together with spatial audio. For first implementations of relevant use cases the content can be quite simple without hindering the purpose, consisting of a camera captured 6 DoF person or object and 3D background or a background coming from an AR camera in the rendering device.

**Motivation for the use case:**

In the use case “Volumetric video with single asset”, and particularly in AR mode, the viewer can see the volumetric video asset, as if it were naturally in front of him. In AR and VR applications, the viewer can move smoothly around the asset or make the asset rotate.

NOTE 1: This use case is part of the scenario “Streaming of Beyond 2D Produced Content”.

In recent years, several collaborations regarding on-demand volumetric video streaming were established between various mobile network operators, volumetric capture studios and technology providers. Some of these collaborations are listed here:

* Volucap, based in Potsdam, Germany:
  + Tagesschau: Volucap and the German news broadcaster Tagesschau collaborated to capture a volumetric representation of a news anchor: <https://volucap.com/portfolio-items/tagesschau-2025/>
  + Music Group: Volucap and the music group “Boss Hoss” prototyped the your favorite band in your living room: <https://volucap.com/portfolio-items/the-bosshoss-augmented-reality/>
  + Book enhanced with AR: Volucap enhanced a children song book with AR content on a smartphone: <https://volucap.com/portfolio-items/rolf-zuckowski/>
  + Sports training: Volucap and Deutsche Telekom produced a clip to learn cool dribbles and precise throws from the former basketball star Josh Mayo: <https://volucap.com/portfolio-items/meeting-josh/>
  + XR Fashion show: Volucap and Lana Mueller, a Berlin based designer, produced a fashion presentation in XR: <https://volucap.com/portfolio-items/lana-mueller-fashion/>
* Volograms, based in Dublin, Ireland
  + Provides professional volumetric content creation services to feed AR use cases such as augmented museum, training or fashion experiences: <https://www.volograms.com/made-with-volograms>
  + The company has also developed AI based solution to enable AR volumetric content from 2D single photo or video: <https://www.volograms.com/>
* [8i](https://8i.com/), [Mantis Vision](https://mantis-vision.com/), [Metastage](https://metastage.com/), [Volograms](https://www.volograms.com/), [XD Productions](https://www.xdprod.com/), etc. present volumetric capturing projects on their websites, similar to Volucap
* XD Productions and Volograms content (both professional and AI-based) has been showcased in public trade shows and conferences by InterDigital as part of MPEG-I V3C platform demonstration with the V-PCC player.
* Zerospace and Canon are collaborating to open a volumetric video capturing studio in Spring 2024. With over 100 Canon Cinema EOS cameras, it claims to offer unmatched capabilities. The website illustrates capture of sports content (e.g. basketball, Karate): <https://www.zerospace.co/studios/canon-volumetric-capture>
* Philips, InterDigital and Broadpeak are collaborating on an end-to-end implementation platform for packaging and delivery of volumetric video over content delivery network (CDN).
  + <https://broadpeak.tv/newsroom/mpeg-v3c-standardized-content-distribution-at-scale/>
  + <https://ir.interdigital.com/news-events/press-releases/news-details/2024/InterDigital-and-Broadpeak-Announce-Collaboration-on-MPEG-V3C-Standardized-Content-Distribution-At-Scale/default.aspx>
* The following video presents new video production technology used by Hollywood production studios and volumetric video is presented starting at minute 14: <https://www.youtube.com/watch?v=u7pu1cQBqtQ&t=373s>

NOTE 2: The examples are meant to provide motivation and demonstrate the market relevance of the scenario and use case and not to give detailed information on the capture setup, formats or other aspects of the workflow. For the workflow description see clause 3.

Volumetric video content can be consumed on devices such as smartphones, tablets, HMDs, TV sets, STBs and PCs, provided that a player for volumetric video is installed. The renderer in the player adapts the content to the specific display.

Brazilian SBTVD Forum has adopted volumetric video for inclusion in their [TV 3.0 standards](https://forumsbtvd.org.br/tv3_0/#panel-phase2) (support will not be mandatory in all receivers; focus on content distribution over the Internet and consumption on smartphones and HMDs). TV3.0 services are planned to be launched in 2025.

[DVB](https://dvb.org/) is running a study mission on volumetric video and the first results are published in [Study Mission Report S101](https://dvb.org/wp-content/uploads/2024/02/S101_Study-Mission-on-Volumetric-Video_Feb-2024.pdf). [X4]

**6.x.3 Description of the scenario**

*This provides a description of beyond 2D video end-to-end workflows, which includes identifying and defining beyond 2D formats being used in the context and representation technologies to delivery these formats. The following aspects may be considered for each workflow:*

This scenario considers streaming of Beyond 2D produced content, with the use case "Volumetric Video with single asset”.

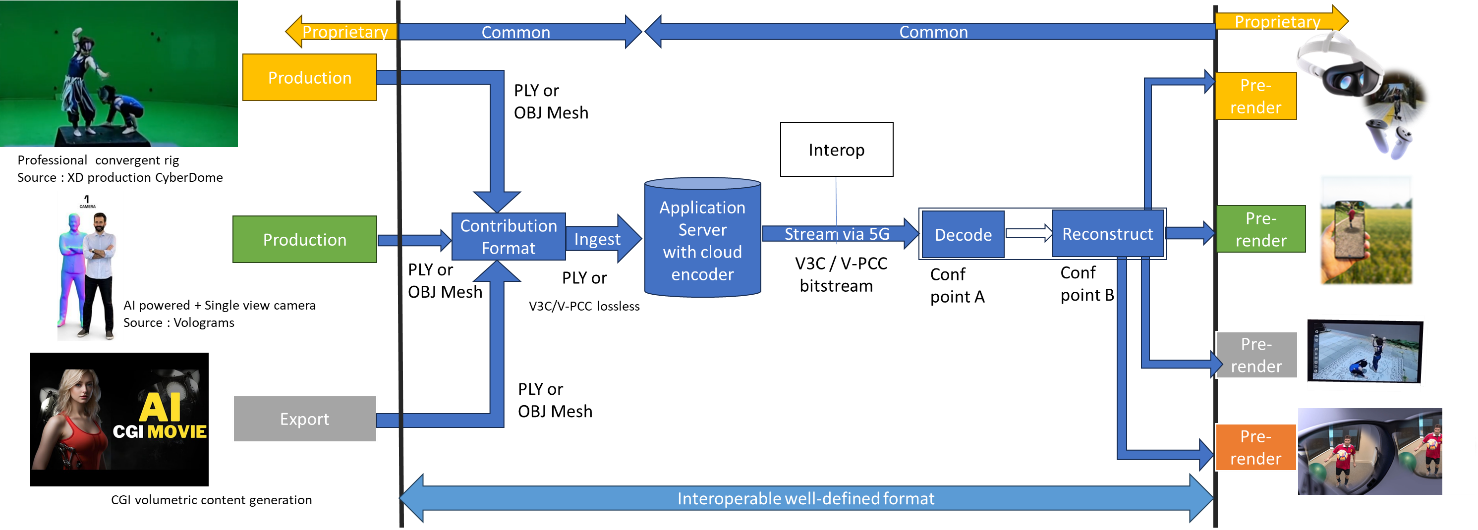


Figure x1: Streaming of Beyond 2D produced content

1. *Capturing and processing*

Capturing of high-quality 6 DoF assets as a volumetric video is typically done with a rig of cameras aligned on a circle around the asset(s) to be captured. Depending on the rig, there can be one or more layers of cameras at different height positions, with each layer consisting of up to 60 cameras. Cameras can be equipped with depth sensors. Hardware such as cameras and depth sensors are off the shelf equipment, but the assembly in the rig is vendor dependent and proprietary.

The various camera and depth sensor signals are fed into a production pipeline that produces the volumetric video. Production includes stitching the various signals, filling holes, correcting occlusions, etc. Persons or physical objects (e.g. a ball or an instrument) can be combined in an asset or separate assets can be used for each person or object. For simplification and not hindering the purpose, the use case described in this document is limited to a single asset. The representation format of a produced asset is typically a dynamic point cloud or a dynamic mesh. The use case “Volumetric Video with single asset” described in this document is entirely based on point clouds as the representation format.

As an example, in the following we describe the production pipeline of the company XD Productions.

The figures below show the XD Productions CYBERDOME capture rig and associated real time viewing to control acquisition.

A group of people standing in a green room

Description automatically generated

Figure x2: XD Productions capture rig (<https://www.xdprod.com/services/studio/studio-virtuel/>)

A computer with people on it

Description automatically generated

Figure x3: XD Production real time virtual production (<https://www.xdprod.com/services/studio/studio-virtuel/>)

The figures below show CYBERDOME acquisitions covering single or multiple characters in dynamic scenes.

A collage of people playing football

Description automatically generated

Figure x4: XD Productions contents screenshots, from top right to left: Acrobat01, Soccer Blue, Soccer Red, Dancer01 and Acrobat Duo

Acquisition processing pipeline includes a rig of about sixty 4K cameras, arranged in hemispheres around the scene to be captured. The set is 15-meter in diameter for a 7-meter diameter capture area. Two types of lenses are simultaneously used, with variable focal lengths, which allows to adapt the size of the capture area, and to mix wide shots and close-ups on the same captures to improve the quality of the textures. Each content item is then converted into point cloud frames contained in a 10bit-sized bounding box with integer, positive coordinates. The processing output is shared in the PLY format.

Another example on how single asset volumetric video is produced is shown in a [video](https://www.youtube.com/watch?v=xX4SJTE3hmQ) by Metastage.

There are a number of companies that provide volumetric video capturing technology or entire volumetric video capturing studios. For a more detailed list we refer to chapter 5.4.4 in the [DVB Study Mission report S101 on Volumetric Video](https://dvb.org/wp-content/uploads/2024/02/S101_Study-Mission-on-Volumetric-Video_Feb-2024.pdf).[X4]

1. *Encoding*

The representation format(s) point cloud is encoded with MPEG V-PCC standardized in ISO/IEC 23090-5 Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC) – 2nd edition. [X5]

The following diagram shows the V-PCC encoder main steps.

A diagram of a company

Description automatically generated with medium confidence

For coding of geometry, texture and occupancy map, V-PCC relies on 2D video encoders such as HEVC or VVC.

An example of point cloud data encoding processing has been described in [X1]

1. *Packaging and delivery*

Storage and delivery of MPEG V-PCC is standardized in ISO/IEC 23090-10 Carriage of visual volumetric video-based coding data – 1st edition. [X6]

1. *Decoding*

The decoder(s) make use of hardware video decoders capabilities for all pixel data, and a small amount of metadata is decoded by a CPU. No dedicated hardware is needed for real-time decoding. The following diagram shows the architecture of MPEG V-PCC.

A screenshot of a computer

Description automatically generated

Decoding of the point cloud is terminated at the output of the video decoders, but these images are just intermediate results and do not represent a useable image for the human eye. Additional stages are needed to reconstruct the point cloud in 3D space and render it to the display of the consumer device.

An example of point cloud data decoding processing has been described in [X1].

By the end of 2023, Futuresource estimates that there are 4.1 billion Smartphones globally in the field with the capacity to decode HEVC video, therefore we propose using HEVC as the underlying 2D video codec.

1. *\*Post-processing*
2. *Rendering*

Rendering is implementation dependent, but it is typically performed on a GPU without dedicated V-PCC hardware.

1. *General constraints on latency, bandwidth, reliability and complexity*

For delivery, the volumetric frames are organized using a random-access reference frame structure.

All decoder and renderer processes are real-time and may have a latency in the order of a few frames.

An example of point cloud data rendering has been described in the paper “[Efficient Delivery and Rendering on Client Devices via MPEG-I Standards for Emerging Volumetric Video Experiences](mailto:https://www.ibc.org/technical-papers/ibc2023-tech-papers-efficient-delivery-and-rendering-on-client-devices-via-mpeg-i-standards-for-emerging-volumetric-video-experiences/10277.article)”

**6.x.4 Supporting companies and 3GPP members**

1. *This documents the 3GPP members that support this scenario in terms of providing the information, test material, test requirements and the characterization for the tests. For each of the identified necessities, a tick box is created in the template.*
2. *Preferably several 3GPP members are included in the support, and in addition a video service provider may be included (not necessarily a 3GPP member).*
3. *Cross-verification is preferably done by the supporters of the scenario*

Deutsche Telekom, Fraunhofer HHI, Interdigital, KDDI, Nokia, Philips, Samsung and Sony

**6.x.5 Source format properties**

*This defines a clear range of the considered and relevant source formats, including the signal properties, but also the characteristics of the content. As an example, the texture and depth format properties of the source may be used which include:*

Table xxx1 lists Beyond 2D Source Format Properties for Volumetric Video with single asset streaming scenario.

|  |  |
| --- | --- |
| Source format properties | Volumetric Video with single asset streaming scenario |
| Number of points /Spatial Resolution | Up to 1 million points per frame |
| Chroma format | RGB |
| Chroma subsampling | Not Applicable |
| Picture aspect ratio | Not Applicable |
| Frame rates | 25, 30 Hz |
| Bit depth | 8 and 10 |
| Colour space formats | RGB 444 nonlinear, BT.709 |
| Transfer characteristics | BT.709 |
| Viewpoints | All assets can be viewed from all directions |

Table x1

**6.x.6 Encoding and decoding constraints and settings**

Table xx2 provides an overview of encoding and decoding constraints for V-PCC with H.265/HEVC for Volumetric Video with single asset streaming scenario. This information supports the definition of detailed anchor condition.

|  |  |
| --- | --- |
| Encoding and Decoding Constraints | H.265/HEVC |
| Relevant Codec and Codec Profile/Levels | H.265/HEVC Main 10 Profile  Level 4.1, 5.1  Metadata stream parsing |
| Random access frequency | 1 seconds |
| Bit rates and quality configuration | Fixed QP Geometry: [32;28;24;20;16]  Fixed QP Texture: [42;37;32;27;27]  bitrates [1;50 Mbps] |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | Covering a range of relevant bitrates and qualities |
| Latency requirements and specific encoding settings | No specific latency requirement |
| Encoding complexity context | Cloud-based encoding, offline encoding |
| Required decoding capabilities | 3 decoder instantiations of H.265/HEVC Main 10 Profile  Level 4.1, 5.1 for (occupancy, geometry and color)  One synchronised metadata bitstream (Atlas) |

**6.x.7 Performance Metrics and Requirements**

1. *A clear definition on how the performance needs to be evaluated including metrics, etc addressing the main KPIs of the scenario.*
2. *Objective measures such as PSNR, VMAF, etc, may be used.*
3. *Justification on whether objective metrics are sufficient and representative of the subjective performance.*

For objective tests it is proposed to use the point-based metric described in annex B1, B3 and B4 of the following document: <https://www.mpeg.org/wp-content/uploads/mpeg_meetings/136_OnLine/w21000.zip>

The point-based metric has been used by MPEG throughout the complete development process of V-PCC. On case-by-case basis and complementary to the point-based metric, MPEG produced snapshots of frames and studied subjectively the impact of a tool before selection.

**6.x.8 Interoperability Considerations for the application**

1. *Streaming with DASH/HLS/CMAF/QUIC*

MPEG-DASH is used with ISO/IEC 23090-10 Carriage of visual volumetric video-based coding data – 1st edition.[X6]

1. *RTP based delivery*

RTP is not proposed for this use case.

**6.x.9 Test Sequences**

*A set of selected test sequences that are provided by the proponents in order to do the evaluation. They should cover a set of source format properties*

* Longdress, Soldier, Loot, Red and Black from 8i as shown in Figure x5.
* Mitch, Thomas from Volucap as shown in Figure x6
* Red Soccer Player, Two\_Acrobats from XD Productions as shown in Figure x7

A group of people in military uniforms

Description automatically generated

Figure x5 - 8i content sequences, first frame, from left to right, Longdress, Soldier, Loot, Red and Black.

A person standing in front of a black background

Description automatically generated

Figure x6 - Volucap sequences, first frame, from left to right, Mitch, Thomas.

****

Figure x7 - XD Productions sequences, first frame, from left to right, Red Soccer Player, Two acrobats.

Figure x8 gives characteristics of the 3D point cloud sequences.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Sequence** | **Provider** | **FPS** | **Number of frames** | **Duration (s)** | **~ Point number per frame** | **Geometry precision** | **Attributes** |
| Longdress | 8i | 30 | 300 | 10,0 | 800000 | 10 bits | R,G,B |
| Soldier | 8i | 30 | 300 | 10,0 | 1000000 | 10 bits | R,G,B |
| Loot | 8i | 30 | 300 | 10,0 | 780000 | 10 bits | R,G,B |
| Red and black | 8i | 30 | 300 | 10,0 | 700000 | 10 bits | R,G,B |
| Mitch | Volucap | 30 | 475 | 15,8 | 860000 | 10 bits | R,G,B |
| Thomas | Volucap | 30 | 748 | 24,9 | 790000 | 10 bits | R,G,B |
| Red Soccer player | XD Productions | 25 | 125 | 5,0 | 950000 | 10 bits | R,G,B |
| Two Acrobats | XD Productions | 30 | 211 | 7,0 | 790000 | 10 bits | R,G,B |

Figure x8 - Test material datasets

**6.x.10 Detailed test conditions**

*Provides a proposal for detailed test conditions, for example based on a reference software together with the sequences and configuration parameters.*

It is suggested to use test model V24 for V-PCC which is the latest when drafting this scenario. Version V24 can be downloaded from the following public reference: <https://github.com/MPEGGroup/mpeg-pcc-tmc2>

The site includes information on how to build the codec for various platforms such as Windows, Linux or OSX. As we suggest using HEVC as the underlying video codec, HM shall be selected for the 2D video codec test model. The site also includes information on how to run the test model for the V-PCC encoder and decoder. A CTC fixes the detailed test conditions. For more details, a software manual can be consulted.

**6.x.11 External Performance data**

*References to external performance data that can be added, for example other SDOs, public documents and so on.*

The subjective verification test report for V-PCC can be downloaded from the public MPEG website: <https://www.mpeg.org/wp-content/uploads/mpeg_meetings/136_OnLine/w20992.zip>

The Brazilian SBTVD Forum performed objective tests with V-PCC. Full results are available in chapter 6.10 (Candidate Technology I), 6.10.3.2 and 6.10.4 of the following document: <https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-VC-Report.pdf>

**6.x.12 Additional Information**

Sequences can be decoded and visualized in real time using a 3D background or in Augmented Reality on a smartphone, tablet, head-mounted display using DASH streaming mode or local file system.

Nokia’s real-time V-PCC decoder implementation that was released as open source: <https://github.com/nokiatech/vpcc>

A simple scene description could be added to enable the placement of the asset in the scene (position, orientation, scale…) but is outside the scope of this document, which is focused on the format and codec evaluation.

**========================= CHANGE 2 ==========================**

# 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".

[X1] "[[Efficient Delivery and Rendering on Client Devices MPEG-I Standards for Emerging Volumetric Video Experiences,](https://www.ibc.org/technical-papers/ibc2023-tech-papers-efficient-delivery-and-rendering-on-client-devices-via-mpeg-i-standards-for-emerging-volumetric-video-experiences/10277.article)](https://www.ibc.org/technical-papers/ibc2023-tech-papers-efficient-delivery-and-rendering-on-client-devices-via-mpeg-i-standards-for-emerging-volumetric-video-experiences/10277.article)" By C. Guede, P. Fontaine, J. Mulard, B. Leroy, C. Quinquis, R. Gendrot, S. Gudumasu, V. Allié, B. Kroon, B. Sonneveldt, R. Schimanofsky, IBC 2023 Tech Papers

[X2] TR 26.928 Extended Reality (XR) in 5G

[X3] TR 26.998 Support of 5G glass-type Augmented Reality / Mixed Reality (AR/MR) devices

[X4] DVB [Study Mission Report S101](https://dvb.org/wp-content/uploads/2024/02/S101_Study-Mission-on-Volumetric-Video_Feb-2024.pdf)

[X5] ISO/IEC 23090-5 Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC) – 2nd edition

[X6] ISO/IEC 23090-10 Carriage of visual volumetric video-based coding data – 1st edition