**3GPP TSG-SA4 Meeting #Post 129-e *S4aV240057***

**Online, 08 Oct 2024**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *CR-Form-v12.2* | | | | | | | | |
| **PSEUDO CHANGE REQUEST** | | | | | | | | |
|  | | | | | | | | |
|  | **26.956** | **CR** | **pseudo** | **rev** | **1** | **Current version:** | **0.1.0** |  |
|  | | | | | | | | |
| *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 |  |

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | | | | | | | | | |
| ***Title:*** | [FS\_Beyond2D] Scenario Volumetric Video with single asset | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Source to WG:*** | InterDigital, Nokia, Philips, Sony | | | | | | | | | |
| ***Source to TSG:*** |  | | | | | | | | | |
|  |  | | | | | | | | | |
| ***Work item code:*** | FS\_Beyond2D | | | | |  | ***Date:*** | | | 2024-10-08 |
|  |  | | | |  | |  | | |  |
| ***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 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:*** | | PD 0.0.4 ([S4-241714](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_129-e/Docs/S4-241714.zip)) includes 5 scenarios including scenario “Streaming of Beyond 2D Produced VoD Content – Use Case Volumetric Video with single asset”, which relies on dense point clouds as representation format. TR 0.1.0 includes already dense point clouds as a representation format and therefore it is proposed to include the corresponding scenario. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Summary of change:*** | | This pCR adds elements for the scenario “Streaming of Beyond 2D Produced VoD Content – Use Case Volumetric Video with single asset” to the scenario clause 7.X in the TR. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Consequences if not approved:*** | | Without this scenario description the motivation of the dense point cloud representation format may not be well understood, which makes the FS\_Beyond2D study less comprehensive. | | | | | | | | |
|  | |  | | | | | | | | |
| ***Clauses affected:*** | | 4.3 | | | | | | | | |
|  | |  | | | | | | | | |
|  | | **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:*** | | |  |  |  |  | | --- | --- | --- | --- | |  |  |  |  | | | | | | | | | |

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

[X0] Broadpeak, V3C standardized content distribution at scale, <https://broadpeak.tv/newsroom/mpeg-v3c-standardized-content-distribution-at-scale/>

[X1] 5G-MAG, V3C Immersive platform, <https://5g-mag.github.io/Getting-Started/pages/v3c-immersive-platform/>

[X2] DVB project, DVB [Study Mission Report S101](https://dvb.org/wp-content/uploads/2024/02/S101_Study-Mission-on-Volumetric-Video_Feb-2024.pdf), [DVB Bluebook S101](https://dvb.org/wp-content/uploads/2024/02/S101_Study-Mission-on-Volumetric-Video_Feb-2024.pdf)

[X3] ISO/IEC 23090-5 Visual Volumetric Video-based Coding (V3C) and Video-based Point Cloud Compression (V-PCC) – 2nd edition

[X4] 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, "<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>"

[X5] ISO/IEC 23090-10 Carriage of visual volumetric video-based coding data – 1st edition

[X6] futuresource Consulting, Spotlight on HEVC, https://www.interdigital.com/white\_papers/spotlight-on-hevc-the-codec-of-choice-for-the-video-streaming-industry

[X7] MPEG WG7, Visual Volumetric Video-based Coding (V3C) and video-based Point Cloud Compression (V-PCC), [Subjective verification test report for V-PCC](https://www.mpeg.org/wp-content/uploads/mpeg_meetings/136_OnLine/w20992.zip)

[X8] SBTVD Forum, TV3.0, [Phase 2 test results video coding](https://forumsbtvd.org.br/wp-content/uploads/2021/12/SBTVD-TV_3_0-VC-Report.pdf)

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

### 7.X Streaming of Beyond 2D Produced VoD Content – Use Case “Volumetric Video with single asset”

### 7.X.1 Motivation for the scenario

This scenario handles streaming of produced Beyond 2D content providing experiences beyond what is achievable with 2D content. In this scenario, particularly in AR application, the user can watch the volumetric video asset as if it were naturally present. In AR and VR applications, the user can move smoothly around the asset or make the asset rotate.

Beyond 2D content can 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 has a colour that may depend on the viewing direction.Volumetric video has the potential to provide a more immersive and interactive experience than 2D content.

Several use cases for on-demand volumetric video streaming can be envisioned related to various domains including but not limited to entertainment, education and industrial monitoring. For example, in the entertainment domain users can stream a performance from their favorite singer or band to their living room and experience greater immersion potentially combined with spatial audio. Let’s imagine an education/training scenario, where a produced volumetric video of a fitness instructor shows how to perform an exercise that can help a 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 trainee 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.

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 a 3D background or a background coming from an AR camera in the rendering device.

*a. Technology evaluation on the market*

Volucap, based in Potsdam, Germany:

* Tagesschau: Volucap and the German news broadcaster Tagesschau collaborated to capture a volumetric representation of a news : <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/>

*b. Industry activities*

Several industry activities regarding on-demand volumetric video streaming have been established between various mobile network operators, volumetric capture studios and technology providers. Some of these industry activities are listed here:

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

*c. Production tools/companies*

The following is a non-exhaustive list of companies providing tools, equipment or services to produce volumetric video content:

* 4D Views : https://www.4dviews.com/
* Arcturus: <https://arcturus.studio/>
* Dimension Studio: <https://dimensionstudio.co/>
* Evercoast : <https://evercoast.com/>
* Mantis Vision : https://mantis-vision.com/
* Metastage : https://metastage.com/
* Volucap: <https://volucap.com/>
* XD Prod : <https://www.xdprod.com/>

*d. Delivery solutions*

An end-to-end implementation of a platform for packaging and delivery of volumetric video over content delivery network (CDN) is available [X0].

*e. Content decoding and rendering*

5G-MAG hosts a V3C Immersive Platform [X1]. It provides a Unity package to decode, render and play V3C content in Unity using the V3C Immersive Platform – Decoder Plugin.

### 7.X.2 Description of the scenario

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

A computer screen shot of a diagram

Description automatically generated

**Figure x1: Streaming of Beyond 2D produced VoD content (content courtesy XD Productions)**

Editor: This figure is a “specialized figure” of a generic streaming scenario. More information to be provided at SA4-130

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. The number of cameras per layer depend again on the designer of the rig and in the year 2024 there are typically between 30 and 60 cameras per layer. Cameras can be equipped with depth sensors. Hardware such as cameras and depth sensors are mostly 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 dense dynamic point cloud or a dynamic mesh. The use case “Volumetric Video with single asset” presented here is entirely based on dense point clouds as the representation format.

As an example, in the following the production pipeline of the company XD Productions is illustrated.

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**

The 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. 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 at <https://www.youtube.com/watch?v=xX4SJTE3hmQ&t=41s>

There are several companies that provide volumetric video capturing technology or entire volumetric video capturing studios. More detailed information can be found in chapter 5.4.4 of 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).[X2]

1. *Encoding*

The representation format dense 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. [X3]

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

A diagram of a company

Description automatically generated with medium confidence

**Figure x5 V-PCC encoder**

For coding of geometry, texture and occupancy map V-PCC relies on 2D video codecs. Due to its efficiency and market penetration HEVC is selected for the present scenario.

An example of point cloud data encoding processing is described in [X4].

1. *Packing 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 [X5].

1. *Decoding*

The decoder makes use of hardware HEVC video decoders capabilities for all pixel data, and a small number of metadata is decoded by a CPU. No dedicated hardware is required for V-PCC real-time decoding. The following figure shows the architecture of an MPEG V-PCC decoder.

A screenshot of a computer

Description automatically generated

**Figure x6 V-PCC decoder**

Decoding of the dense 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 dense point cloud in 3D space and render it to the display of a consumer device.

An example of point cloud data decoding processing has been described in [X4].

Futuresource estimates that by the end of 2023 there are 4.1 billion Smartphones globally in the field with the capacity to decode HEVC video [X6].

1. *Post-processing*

Editor: Information to be provided at SA4-130

1. *Rendering*

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

Editor: Information to be provided at SA4-130

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.

Editor: Information to be provided at SA4-130

### 7.X.3 Supporting companies and 3GPP members

Deutsche Telekom, Fraunhofer HHI, Huawei, Interdigital, KDDI, Nokia, Philips, Samsung and Sony are supporting the scenario.

Editor: This information is requested by the template, but this information may not go to the TR.

### 7.X.4 Source format properties

Editor: Information to be provided at SA4-130

### 7.X.5 Encoding and decoding constraints and settings

Editor: Information to be provided at SA4-130

### 7.X.6 Performance Metrics and Requirements

Editor: Information to be provided at SA4-130

### 7.X.7 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 [X5].

1. *RTP based delivery*

RTP is not proposed for this scenario.

### 7.X.8 Test Sequences

Editor: Information be provided at SA4-130.

### 7.X.8 Detailed Test conditions

Editor: Information be provided at SA4-130

### 7.X.9 External Performance data

The subjective verification test report for MPEG V-PCC can be downloaded from the public MPEG website [X7].

The Brazilian SBTVD Forum performed objective tests with MPEG 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> [X8]

### 7.X.9 Additional Information

Editor: Information be provided at SA4-130