**Source: Nokia Corporation1, Interdigital, Philips**

**Title: [FS\_Beyond2D] Scenario on on-demand volumetric video streaming**

**Agenda Item: 9.9**

**Document for: Discussion and 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.

In this contribution, a draft scenario on On-demand volumetric video streaming 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 [S4aV240003](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Inbox/Drafts/S4aV240003-onlineEdits.docx) including the online edits agreed at the VIDEO SWG telco on 26 March.

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

## 6.x Scenario #x: On-demand volumetric video streaming

1. **Scenario name**

On-demand volumetric video streaming

1. **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?*

Volumetric video is a technology, which enables capture of an object or scene in three dimensions and its playback independent from the original capture position(s) or orientation(s). This has the potential to provide a more immersive and interactive experience for use cases in diverse domains such as e.g. industrial monitoring, education and entertainment.

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) and TR 26.998 (Cl. A.3 - Use Case 18: Streaming of volumetric video for glass-type MR devices).

On-demand volumetric video streaming allows to provide high-quality, professionally captured volumetric video content. Several use cases of on-demand volumetric video streaming can be envisioned to related to various domains including education, professional training or entertainment. 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 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 jazz band to their living room and experience greater immersion potentially together with spatial audio.

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 below:

1. **NTT DOCOMO and Arcturus:**
	* Arcturus, the developer behind the post-production platform [HoloSuite](https://arcturus.studio/holosuite/), has partnered with **NTT DOCOMO**: <https://www.digitalmediaworld.tv/animation/arcturus-and-docomo-telecom-bring-volumetric-video-to-5g-mobile>
	* The goal of this collaboration is to bring **volumetric video** to **5G mobile devices**.
	* Arcturus specializes in creating, editing, and distributing volumetric video of digital humans using their software tools.
	* Their platform includes two critical applications:
		+ **HoloEdit**: Designed for editing and compressing volumetric video, including fixing multiple meshes for characters and cleaning textures.
		+ **HoloStream**: Adaptive bitrate streaming software that delivers volumetric video directly to mobile devices.
	* The partnership aims to make it easier for producers to reach their audiences by streaming uninterrupted volumetric experiences over a mobile network.
2. **SK Telecom and Jump Studio**
	* SK Telecom, a major South Korean telecom operator, has expanded and relocated **Jump Studio**, Asia’s first **Mixed Reality Capture Studio**.
	* Jump Studio uses **Microsoft’s volumetric video capture technology**.
	* The studio is now located at SK Telecom’s headquarters (SKT-Tower).
	* SK Telecom plans to promote **5G content business** by leveraging volumetric video capabilities: <https://www.sktelecom.com/en/press/press_detail.do?idx=1487>
3. **Deutsche Telekom and Volucap**
* Actively engaged in the development of augmented reality (AR) and virtual reality (VR) technologies in collaboration with partners.
* One notable collaboration involves [Volucap](https://volucap.com/), a volumetric capture studio in Germany.
* Their volumetric capture process captures digital actors in three dimensions, creating more lifelike and immersive experiences.
1. **Volucap and 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/>
2. **Sony Electronics and Levan Center of Innovation:**
	* The **Levan Center of Innovation**, in collaboration with **Sony Electronics** and **GeniusXR**, has opened a cutting-edge **Volumetric Capture Studio (VCS)**.<https://www.ravepubs.com/sony-electronics-genius-xr-levan-center-of-innovation/>
	* Sony provides the backbone of the VCS with its high-end cameras and machine learning software, essential for producing top-notch content.
3. **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://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>

NOTE 1: The examples are meant to provide motivation and demonstrate the market relevance of the scenario and not to give detailed information on the capture setup, formats or other aspects of the workflow. For the workflow description see the clause 3.

1. **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:*

1. *Capturing and processing*

This scenario considers volumetric video that is the result of a production pipeline that in general may be a combination of captured video, ray-traced computer graphics, neural rendering and compositing thereof. If video was captured, then zero or more of those cameras may be range-sensing cameras and one or more of these cameras have colour sensors. In the case of two or more cameras that are not rigidly connected, camera extrinsics are online calibrated.

The created content is offline converted to a representation format that is suitable for encoding. The parameters of the source cameras (if any) may or may not be part of the presentation.

1. *Encoding*

The representation format(s) are encoded by using conventional 2D video codec(s) and metadata stream(s).

1. *Packaging and delivery*

The multiple streams may be multiplexed or provided as separate tracks.

1. *Decoding*

The decoder(s) will 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.

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

Rendering is typically performed on a GPU without dedicated hardware.

When a viewing space is used, then:

* What is rendered is one or two viewports with perspective projection and with 6 degrees of freedom (3-D position and 3-D orientation).
* The pose of the viewport is within a viewing space that can be signaled or implicitly determined from a decoded frame. A viewing space can limit both position, orientation or both in combination. For instance, it is generally not intended for a viewport to intersect with scene elements.
* When a viewport would be rendered that is outside of the viewing space, then the renderer has to perform a mitigation to avoid a viewing experience that is not intended by the content provider.
1. *General constraints on latency, bandwidth, reliability and complexity*

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 seconds.

1. **Supporting companies and 3GPP members**
2. *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.*
3. *Preferably several 3GPP members are included in the support, and in addition a video service provider may be included (not necessarily a 3GPP member).*
4. *Cross-verification is preferably done by the supporters of the scenario*

Nokia, Interdigital, Philips

1. **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:*

1. *Spatial resolutions*

**TBD**

1. *Chroma Format*

**TBD**

1. *Chroma Subsampling*

**TBD**

1. *Aspect ratios*

**TBD**

1. *Frame rates*

**TBD**

1. *Colour space formats*

YCbCr 4:2:0

1. *Transfer Characteristics*

Video carrying texture information will have a BT.709 transfer function. All other video will have a linear transfer.

1. *Bit depth*

Colour will be 8 or 10 bit. Effective geometry bit depth may be higher or implicitly represented (not as video samples).

1. *Viewpoints*

The viewpoints are within a viewing space that can be signalled or implicitly determined from a decoded frame.

1. *Other signal properties*

**TBD**

1. **Encoding and decoding constraints and settings**

*Typical encoding constraints and settings such as*

1. *Relevant Codec and Codec Profile/Levels according to TS26.119*
2. *Random access frequency*
3. *Error resiliency requirements*
4. *Bitrates and quality requirements*
5. *Bitrate parameters (CBR, VBR, CAE, HRD parameters)*
6. *ABR encoding requirements (switching frequency, etc.)*
7. *Latency requirements and specific encoding settings*
8. *Encoding context: real-time encoding, on device encoding, cloud-based encoding, offline encoding, etc.*
9. *Required decoding capabilities*
10. *Synchronization requirements*

**TBD**

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

**TBD**

1. **Interoperability Considerations for the application**
2. *Streaming with DASH/HLS/CMAF/QUIC*
3. *RTP based delivery*

**TBD**

1. **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*

**TBD**

1. **Detailed test conditions**

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

**TBD**

1. **External Performance data**

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

**TBD**

1. **Additional Information**