**3GPP TSG SA WG4 Meeting #130 *S4-241869***

**Orlando , US, 18th–22nd November 2024**

|  |
| --- |
| *CR-Form-v12.2* |
| **PSEUDO CHANGE REQUEST** |
|  |
|  | **26.956** | **CR** | **pseudo** | **rev** |  | **Current version:** | **0.1.1** |  |
|  |
| *For* ***[HE](http://www.3gpp.org/3G_Specs/CRs.htm%22%20%5Cl%20%22_blank)******[LP](http://www.3gpp.org/3G_Specs/CRs.htm%22%20%5Cl%20%22_blank)*** *on using this form: comprehensive instructions can be found at <http://www.3gpp.org/Change-Requests>.* |
|  |

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

|  |
| --- |
|  |
| ***Title:***  | [FS\_Beyond2D] Proposed Scenario: UE-to-UE Stereoscopic Video Live Streaming |
|  |  |
| ***Source to WG:*** | China Mobile Com. Corporation, ZTE |
| ***Source to TSG:*** |  |
|  |  |
| ***Work item code:*** | FS\_Beyond2D |  | ***Date:*** | 2024-11-11 |
|  |  |  |  |  |
| ***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 objectives:1. 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 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.

During previous SA4 meetings, a scenario on “UE-to-UE Beyond 2D Video Live Streaming” was defined and documented in PD (S4aV240022). We believe this scenario is ready for evaluation for the following reasons:1. This scenario uses a stereoscopic video format, which is well-defined in clause 4.3.2 of TR 26.956.
2. The test sequence has been proposed and documented in clause 2.5.9 of PD.
3. The evaluation metrics have also been proposed and documented in clause 4.2 of PD.
 |
|  |  |
| ***Summary of change:*** | This document provides a scenario on “UE-to-UE Stereoscopic Video Live Streaming”. |
|  |  |
| ***Consequences if not approved:*** | Incomplete TR. |
|  |  |
| ***Clauses affected:*** | 7 |
|  |  |
|  | **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)

[LS-1] Wang, Y., Lu, Z., Cao, P. et al. How Live Streaming Changes Shopping Decisions in E-commerce: A Study of Live Streaming Commerce. Comput Supported Coop Work 31, 701–729 (2022). <https://doi.org/10.1007/s10606-022-09439-2>

[LS-2] Xie, Junyuan et al. “Deep3D: Fully Automatic 2D-to-3D Video Conversion with Deep Convolutional Neural Networks.” European Conference on Computer Vision (2016).

[LS-3] Dumic, E. et al.. “Transmission of 3D Video Content. In: Assunção, P., Gotchev, A. (eds) 3D Visual Content Creation, Coding and Delivery.” Signals and Communication Technology (2019). Springer, Cham. https://doi.org/10.1007/978-3-319-77842-6\_8

[LS-4] Schierl, Thomas and Sam Narasimhan. “Transport and Storage Systems for 3-D Video Using MPEG-2 Systems, RTP, and ISO File Format.” Proceedings of the IEEE 99 (2011): 671-683.

[LS-5] 3GPP TR 26.905 V 18.0.0: "Mobile stereoscopic 3D video"

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

# 7 Considered Scenarios

## 7.1 Introduction

Editor’s note: This clause collects end-to-end scenarios and corresponding workflows for beyond 2D video, based on the template defined in Annex A. Alignment with the generalized media delivery architecture defined in TS 26.501/506 is expected, primarily addressing reference points M2 and M4.

## 7.x Scenario x: UE-to-UE Stereoscopic Video Live Streaming

### 7.x.1 Motivation

Live Streaming services can be deployed across various platforms, including social media platforms like YouTube Live™, Facebook Live™, and TikTokTM, as well as though e-commerce platforms such as eBayTM and TaobaoTM [LS-1]. It significantly impact marketing by providing a dynamic and interactive channel to directly connect markets and their target audiences in real time. To continue captivating users, it’s essential to explore a more immersive live streaming experience by incorporating beyond 2D video.

Most of the current Beyond 2D streaming services provided by network operators, services providers, and device manufacturers on the market are based on the stereoscopic video format, as defined in clause 4.3.2. In terms of distribution, existing stereoscopic 3D video formats, such as frame-compatible side-by-side and 2D video plus depth. Particular emphasis is given to the DVB systems [LS-3] and IP transport, focusing HTTP/TCP streaming, adaptive HTTP streaming, RTP/UDP streaming, P2P Networks, and Information-Centric Networking-ICN. Hybrid transport technologies, combining broadcast and broadband networks for video delivery are also addressed. The most important standards are MPEG-2 systems, which is used for digital broadcast and storage on Blu-ray discs, real-time transport protocol (RTP), which is used for real-time transmissions over the Internet, and the ISO base media file format, which can be used for progressive download in video-on-demand applications [LS-4].In cause 6.2 of TR 26.905 [LS-5], it provides a DASH-based streaming solution for streaming Stereoscopic 3D video.

### 7.x.2 Description of the Anticipated Application

#### 7.x.2.1 Overall Description

3GPP until now has very restricted set of services but based on the considerations in clause 7.x.1, the following encoding benchmark capabilities are considered for decoding:

- The capability of supporting up to two (N=2) concurrent decoder instances with the aggregate capabilities of H.265 (HEVC) YUV 4:2:0, 10 bit, Max Resolution 4096 x 2048..

The considered scenarios is low-latency streaming. Important aspects that are expected to be considered when evaluating a codec in the context of this UE-to-UE Stereoscopic Video Live Streaming scenario are:

- Quality and Coding Efficiency:

- High and uninterrupted visual quality, taking into account the services constraints.

- The ability to compress 2 or more B2D streams in real-time to minimize latency requirements.

- Any savings can provide significant benefits due to the expected large volume of the traffic either in quality or network utilization.

- Considered settings for encoding:

- Low-latency settings

- Encoding in this scenario is typically done as:

- Live and On-Demand distribution and encoding

- Sever and Cloud-based Encoding

#### 7.x.2.2 Capturing and processing

The existing and emerging capture methods include:

- **Stereoscopic camera**: a dual-lens camera which can directly capture stereo 3D video. For example, the SpatialLabs Eyes™, a stereoscopic camera cable of capturing at up to 8-MP (aka 4K) per eye at 30 fps or 2K per eye at 60 fps (https://www.tomshardware.com/cameras/3d-call-me-maybe-acers-new-spatiallabs-camera-live-streams-impressive-3d-video-in-8k-but-few-can-view-it). Another example is the ZTE Nubia Pad 3D II™, which can capture stereo 3D video at up to 13-MP per eye at 30 fps with the rear camera and 8-MP per eye at 30 fps with the selfie camera.

- **3D Camera Rig**: The cameras setup is shown in the figure below, which consists of two identical HD camcorders (Canon HG-20™) and an adjustable stereo mount. The mount ensures that optical axes of the cameras are parallel and supports the continuous adjustment of the camera distance in the range 7-50 cm. To ensure matching of the focal length the wide angle end of the zoom lens with a focal length of 43 mm has been used. In order to match the cameras with each other the focal length, white balance and shutter speed have been set manually. The synchronized operation of the two camcorder is ensured through the use of a single remote control. The camcorders support the capture of images with a resolution of 1920×1080 pixels and store them as high quality JPEG files.

Figure 7.x.1-1 Camera Rig for stereoscopic video capture



- **AI Based 2D-to-Stereo3D Conversion**: The AI-based conversion leverages deep neural networks to perform real-time, end-to-end conversion of 2D videos and images into stereoscopic 3D format [LS-2]. This technology is proving commercially viable and meets the growing demand for high-quality stereoscopic images, as demonstrated by commercial services.

For UE capable of directly capturing beyond 2D video on the device (e.g., UE with a stereoscopic camera**,**  UE equipped with ToF, LiDAR or Spatial camera), it pre-processes the captured video frames into a well-defined B2D format and sends them to the encoder as input. The encoded B2D video streams are then streamed to the streaming server within the network, where the server may transcode them into different bitrates and distribute them to various audiences. The receiving end decodes B2D video streams and perform post-processing to adapt to the rendering system.

For UE limited to capturing only 2D video (e.g., UE with a monocular camera), the UE initially encodes the regular 2D video and streams it to a cloud server capable of real-time 2D-to-beyond 2D transcoding (a generic pipeline for this transcoding process is described in Figure 7.x.1-2). The cloud server then encodes the transcoded B2D video and streams it to the streaming server.

Figure 7.x.1-2 Pipeline for 2D-to-Stereo3D Conversion



#### 7.x.2.3 Encoding

The following solutions can be used to realize this scenario:

- Concurrent H.265/HEVC

- MV-HEVC

#### 7.x.2.4 Packing and Delivery

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

#### 7.x.2.5 Decoding

The following solutions can be used to realize this scenario:

- Concurrent H.265/HEVC decoding capabilities

- MV-HEVC

#### 7.x.2.6 Rendering

Rendering can be on:

- Backward-compatible to 2D presentation, e.g., a mobile phone, but the stereoscopic effect is lost in this case.

- A device for 3D presentation, e.g., autostereoscopic displays, VR headset, and AR glasses, these devices can track the viewer's eye position and adjusts the 3D effect in real-time for single viewer applications (parallax adjustment) and rendering.

### 7.x.3 Sourcer Format Properties

Table 7.x.3-1 provides an overview of the different source signal properties for UE-to-UE Stereoscopic Video Live Streaming. This information is used to select proper test sequences.

Table 7.x.3-1 UE-to-UE Stereoscopic Video Live Streaming Source Properties

|  |  |
| --- | --- |
| Source format properties | B2D Live Streaming |
| Number of views | 2 |
| Spatial resolution for each view | For each view:1920 x 10802560 x 1600 |
| Chroma format | Y’CbCr, RGB |
| Chroma subsampling | 4:2:0 |
| Picture aspect ratio | 32:9 16:9 16:10 |
| Frame rates | 25, 30, 60, 90,120 Hz  |
| Bit depth | 8, 10 |

### 7.x.4 Encoding and Decoding Constraints

Table 7.x.4-1 provides an overview of encoding and decoding constraints for UE-to-UE Stereoscopic Video Live Streaming scenario using H.265/HEVC. This information supports the definition of detailed anchor conditions.

Table 7.x.4-1 Encoding and Decoding Configurations

|  |  |
| --- | --- |
| Encoding and Decoding Constraints | H.265/HEVC |
| Relevant Codec and Codec Profile/Levels | H.265/HEVC Main 10 Profile Level 4.1, 5.1 |
| Random access frequency | 1 seconds |
| Bit rates and quality configuration | Fixed QP: [17, 22, 27, 32, 37] CBRHalf Width/Height: 5-8MbpsFull Width/Height: 8-16MbpsCapped-VBR |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | Covering a range of relevant bitrates and qualities |
| Latency requirements and specific encoding settings | Low latency requirements |
| Encoding complexity context  | Real-time encoding, Cloud-based encoding |
| Required decoding capabilities | H.265/HEVC Main 10 Profile Level 4.1, 5.1 |

### 7.x.5 Performance Metrics

[PSNR, SSIM, SIM, BD-Rate, signal-to-noise ratio (SNR)]

<Editor’s Note: Clause 4.2 of the permanent document of TR 26.956 (S4-241868) provides verified Objective Metrics (e.g., HV3D Quality Metrics) and Subjective Assessment methods for stereoscopic video. This content is expected to be discussed and approved by the SA4 group before being added to the TR.>

### 7.x.6 Interoperability Consideration

For UE-to-UE Stereoscopic Live Streaming, DASH-based solutions are expected.

### 7.x.7 Reference Sequences

Table 7.x.7-1 provides the selected reference sequences for this scenario. Keys are identified to refer to the sequences in the context of the scenario. The sequences are named and a reference to the details of the sequence is provided. A justification is provided, why this sequence is selected.

<Editor’s Note: Clause 2.5.9 of the permanent document of TR 26.956 (S4-241868) lists public datasets, generation software, and capturing tools for creating source sequences for UE-to-UE Stereoscopic live streaming. This content is expected to be reviewed and approved by the SA4 group before being added to the TR.>

Table 7.x.7-1 Reference Sequences for UE-to-UE Stereoscopic Live Streaming

|  |  |  |  |
| --- | --- | --- | --- |
| Key | Name | Reference | Justification/Comment |
| SX-R01 |  |  |  |
| SX-R02 |  |  |  |
| SX-R03 |  |  |  |
| SX-R04 |  |  |  |

### 7.x.8 Anchor Definition

<Editor’s Note: This part of content will be addressed in a separate contribution.>

### 7.x.9 Anchor Result

<Editor’s Note: This part of content will be addressed in a separate contribution>