**Source: China Mobile Com. Corporation, ZTE**

**Title: [FS\_Beyond2D] Scenario: Glasses-free 3D Live Streaming**

**Agenda item: 9.9**

**Document for: Discussion and Agreement**

## 1 Introduction

This proposal provides a scenario on “Glasses-free 3D Live Streaming”.

## 2 Discussion

## Scenario 1: Glasses-free 3D Live Streaming

The following aspects are considered for a scenario:

1. **Scenario name**

Glasses-free 3D Live Streaming

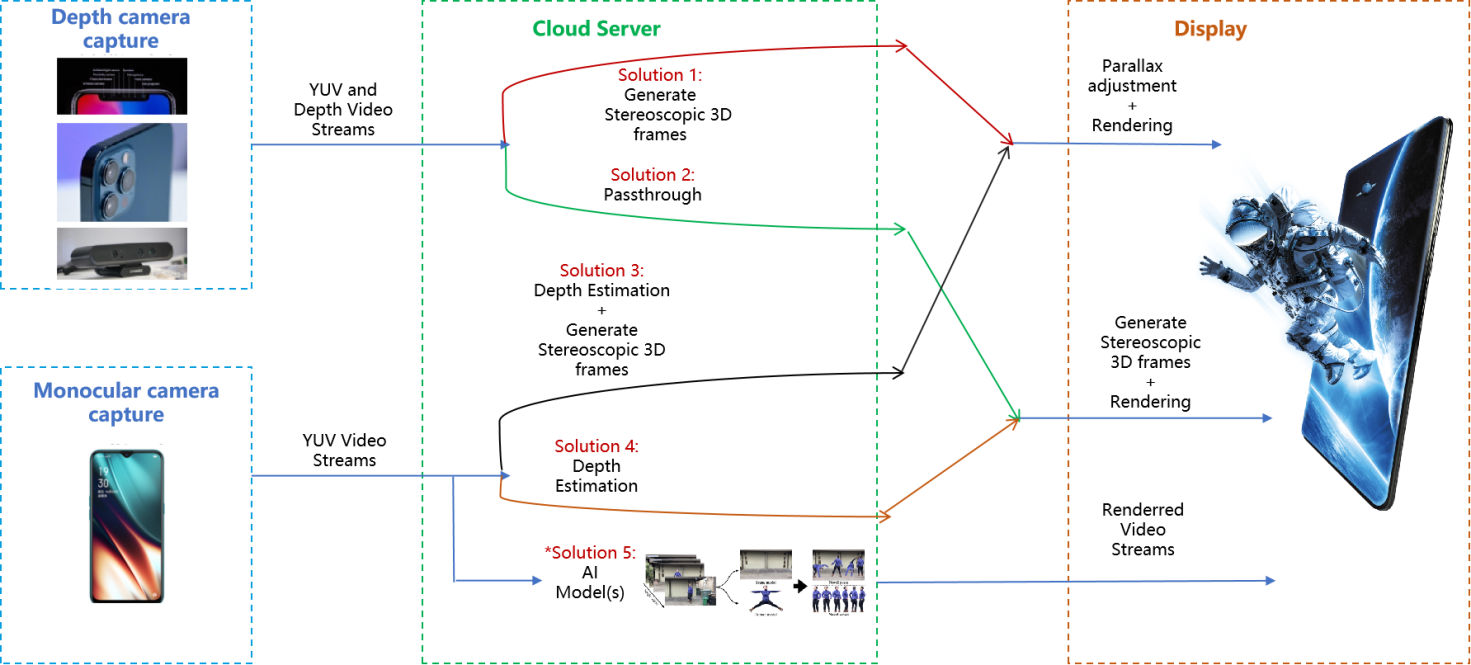
1. **Motivation for the scenario**

Live Streaming services can be deployed across various platforms, including social media platforms like YouTube Live, Facebook Live, and TikTok, as well as though e-commerce platforms such as eBay and Taobao [1]. It significantly impact marketing by providing a dynamic and interactive channel to directly connect markets and their target audiences in real time. In 2023, US livestream sales are projected to reach $50 billion, according to Coresight Research [2].

On the other hand, glasses-free 3D technology is relatively mature. It offers users a completely new level of visual realism and immersion. In 2024, the glasses-free 3D industry is anticipated for commercialization, with the global market valued at US$887.40 million in 2023 and projected to reach US$3.88 billion by 2033, expanding at phenomenal CAGR of 15.9% from 2023 to 2033[3].

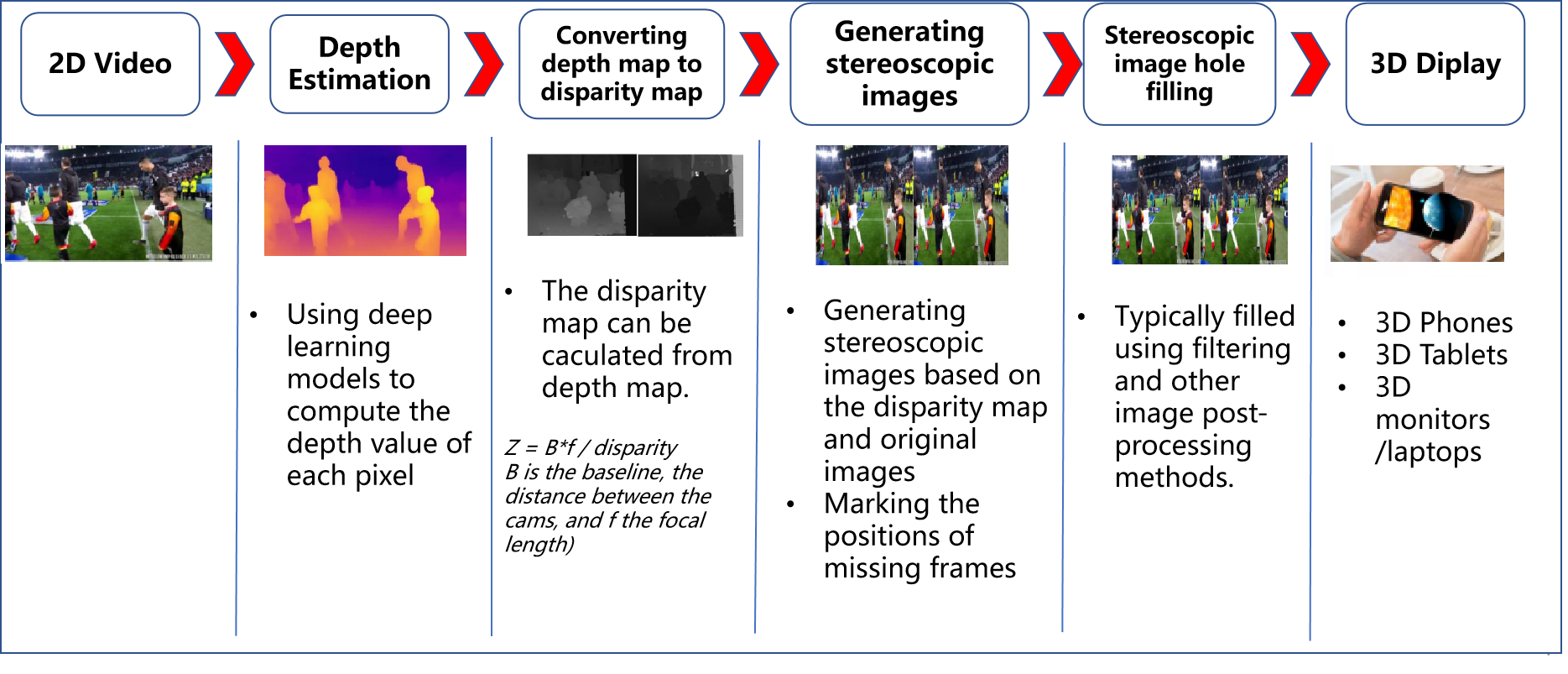
1. **Description of the scenario**

Generic end-to-end workflow, alignment with TS 26.501/506 is expected:



\*Parallax adjustment: track the viewer's eye position and adjust the 3D effect in real-time for single viewer applications.

Pipeline for 2D-to-Stereo 3D conversion:



Beyond 2D video formats:

* RGBD
* RGB
* Stereoscopic 3D frames
* \*AI-Generated representations: NeRF/3DGS (Under Investigation)

**Pre-knowledge:**

The commercialized glasses-free 3D monitors and tablets takes stereo 3D video frames (side-by-side) as input, combined with eye-tracking technology on the terminal side to render 3D effects. Based on this, we have listed potential solutions:

**For UE equipped with a depth camera (e.g., ToF or LiDAR):**

Solution 1: The UE sends captured YUV and depth video streams to a cloud server. The cloud server first converts YUV video frames to RGB video frames, also performs the depth to disparity map conversion, and it then generates stereoscopic 3D video frames, sending the converted frames to G3D displays. The G3D display tracks the viewer's eye position and adjusts the 3D effect in real-time for single viewer applications (parallax adjustment) and rendering.

Solution 2: The UE sends captured YUV and depth video streams to a cloud server. The cloud server forwards video streams to G3D displays. The G3D display converts the depth map to disparity map and generate stereoscopic 3D video frames. The G3D display then perform parallax adjustment and rendering.

**For UE equipped with a monocular camera:**

Solution 3: The UE sends captured YUV video streams to a cloud server. The cloud server estimates depth values and follows the procedures listed in Solution 1 for further processing.

Solution 4: The UE sends captured YUV video streams to a cloud server. The cloud server estimates depth values and follows the procedures listed in Solution 2 for further processing.

\*Solution 5: The usage of AI-generated representations (such as NeRF/3DGS) for live streaming is currently under investigation.

1. **Supporting companies and 3GPP members**

China Mobile

ZTE

1. **Source format properties**

[

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Input stream (2D)** | **Output stream (3D)** |
| Spatial resolution | 1920\*1080  2560\*1600 | Full Width: 3840\*1080  5120\*1600  Full Height: 1920\*2160  2560\*3200  Half Width/Height: 1920\*1080  2560\*1600 |
| Chroma Format | Y’CbCr, RGB | Y’CbCr, RGB |
| Chroma Subsampling | 4:2:0 | 4:2:0 |
| Aspect ratios | 16:9  16:10 | 32:9  32:10  16:9  16:10 |
| Frame rates | 25, 30, 60,90,120 | 25, 30, 60,90,120 |
| Colour space formats | BT.709, BT.2020 | BT.709,BT.2020 |
| Transfer Characteristics | BT.709, BT.2020 | BT.709, BT.2020 |
| Bit depth | 8 bit | 8 bit, 10 bit |
| Viewpoints | 1 | 2 views for mobile phone  <tbd> |
| Other signal properties |  |  |

]

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

[

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Input streams (2D)** | **Output streams (3D)** |
| Relevant Codec and Codec Profile/Levels according to 3GPP TS (e.g., TS 26.119) | H.264/AVC Progressive  High Profile 4.2 | H.265/HEVC Main 10 Profile |
| Random access frequency | 1 second | 1 second |
| Error resiliency requirements | N/A | N/A |
| Bitrates and quality requirements | Constant bitrates ={4,8,12} Mbps | Constant bitrates ={4,8,12,24} Mbps |
| Bitrate parameters (CBR, VBR, CAE, HRD parameters) | 4Mbps ~ 12Mbps CBR | Half Width/Height: 4Mbps ~12Mbps  Full Width/Height: 8Mbps ~ 24Mbps  CBR |
| ABR encoding requirements (switching frequency, etc.) | N/A | N/A |
| Latency requirements and specific encoding settings | Low latency requirements | Low latency requirements |
| Encoding context: real-time encoding, on device encoding, cloud-based encoding, offline encoding, etc | Real-time encoding | Real-time encoding  Cloud-based encoding |
| Required decoding capabilities | H.264/AVC Progressive  High Profile 4.2 | H.265/HEVC Main 10 Profile |
| Synchronization requirements | N/A | <tbd> |

]

1. **Performance Metrics and Requirements**

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

[Latency: 1s

Bandwidth: 100Mbit/s～1Gbit/s (evaluate the impact of the beyond 2D video representations on the bandwidth requirements of streaming with minuscule loss probabilities)]

1. **Interoperability Considerations for the application**

DASH, LL-HLS

RTP

WebRTC

1. **Test Sequences**

<TBD>

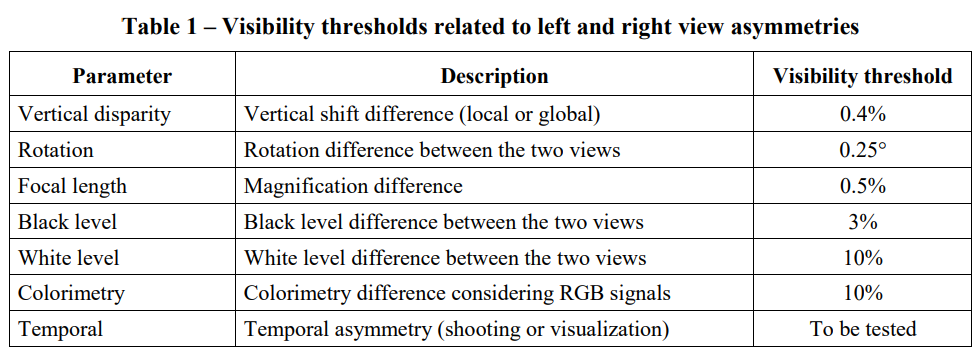
1. **Detailed test conditions**

<TBD>

1. **External Performance data**
2. Subjective Assessment Methods for 3D Video Quality, document ITU-T P.3D-sam, International Telecommunication Union, Geneva,Switzerland, Jul. 2015.

*“In stereo 3D systems, a binocular 3D image is formed by presenting the left and right image to their respective eye. If discrepancies arise between these two images, they can cause psychophysical stress, and in some cases 3D viewing can fail. For example, when shooting and displaying stereoscopic 3DTV programmes, there may be geometrical, optical, electrical or temporal asymmetries, such as size inconsistency, vertical shift, rotation error, and luminance or colour levels between the left and right images. For the production of natural scene content using two independent video cameras, the main issue is to guarantee that the asymmetries of the views are under perceptual limits.”*

*Table 1 illustrates visibility thresholds obtained from subjective experiments using an impairment scale and for a viewing distance of 4.5 times the display height.*



1. Assessment Methods of Visual Fatigue and Safety Guideline for 3D Video, document ITU-T J.3D-fatigue, International Telecommunication Union, Geneva, Switzerland, 2015.
2. Display Requirements for 3D Video Quality Assessment, document ITU-T J.3D-disp-req, International Telecommunication Union, Geneva, Switzerland, 2015.

<FFS>

1. **Additional Information**

<TBD>

## 3 Proposal

It is proposed to document this scenario to permanent document of TR 26.956.

## 4 References

[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>

[2] Coresight Research, “Inside the $50 billion shopping spree happening on livestreams.”, https://www.cnbc.com/video/2023/06/09/inside-the-50-billion-shopping-spree-happening-on-livestreams.html

[3] Fact. MR, “Glass-Free HD 3D Display Market”, https://www.factmr.com/report/286/glass-free-hd-3d-display-market