**Source: China Mobile Com. Corporation**

**Title: [FS\_Beyond2D] Scenario: Beyond2D Live Streaming**

**Agenda item: 3.8**

**Document for: Discussion and Agreement**

## 1 Introduction

This proposal provides a scenario on “Beyond2D Live Streaming”.

## 2 Discussion

## Scenario X: Beyond 2D Live Streaming

The following aspects are considered for a scenario:

1. **Scenario name**

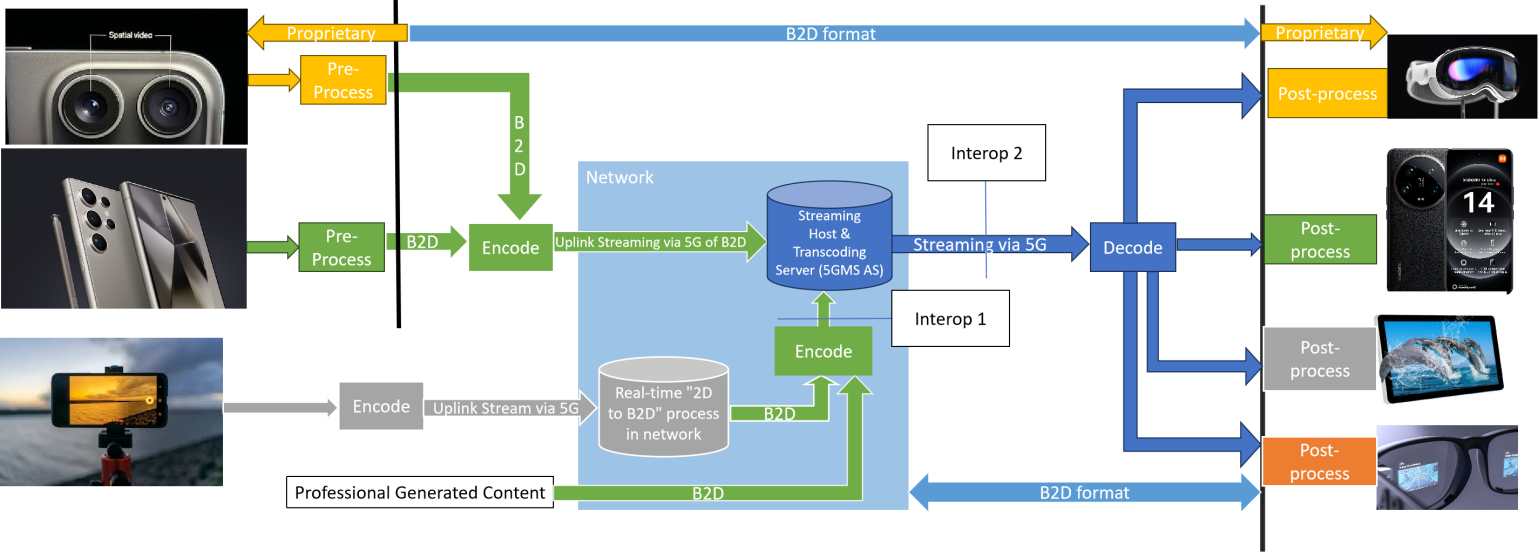
Beyond2D Live Streaming

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]. In 3GPP TR 26.955 [3] clause 6.2, Full HD Streaming has been introduced, complemented by 5G Media Streaming [4] and the TV Video Profiles [5] specifications. Additionally, TS 26.118 [6] defines VR profiles for streaming applications, focusing on the coded representation of 360 VR distribution signals. However, advancements in capturing devices (e.g., ToF cameras, phones equipped with depth sensors, spatial cameras) and displays technology (e.g., HMDs, AR Glasses, MR HMDs, glasses-free autostereoscopic displays, and multiscopic displays), are reshaping video services beyond traditional two-dimensional formats. The integration of beyond2D video into Live Streaming services can create a more life-like and immersive experience.

1. **Description of the scenario**

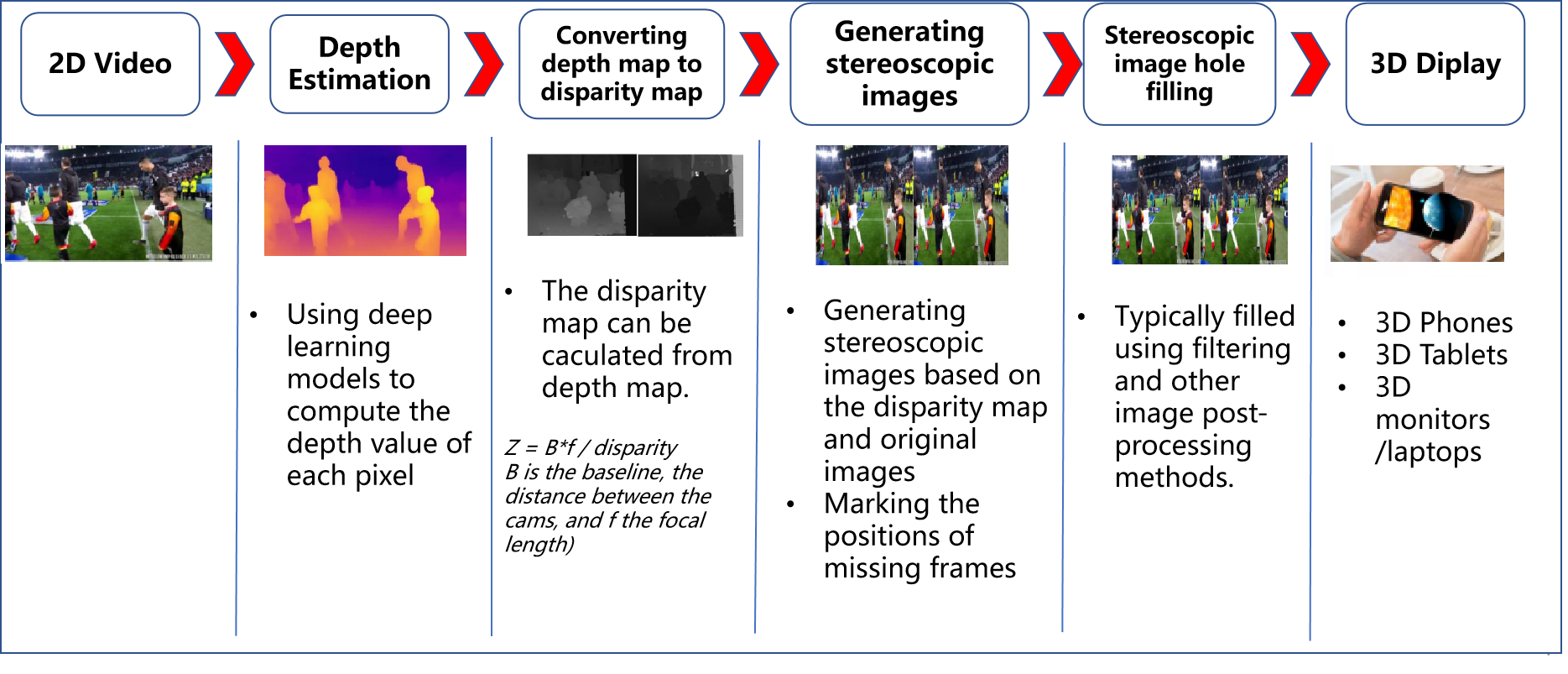
**Solution #1**:



For UE capable of directly capturing beyond 2D video on the device (e.g., 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 conversion (the process is described in the pipeline below). The cloud server then encodes the transcoded B2D video and streams it to the streaming server.

Pipeline for 2D-to-beyond 2D conversion:

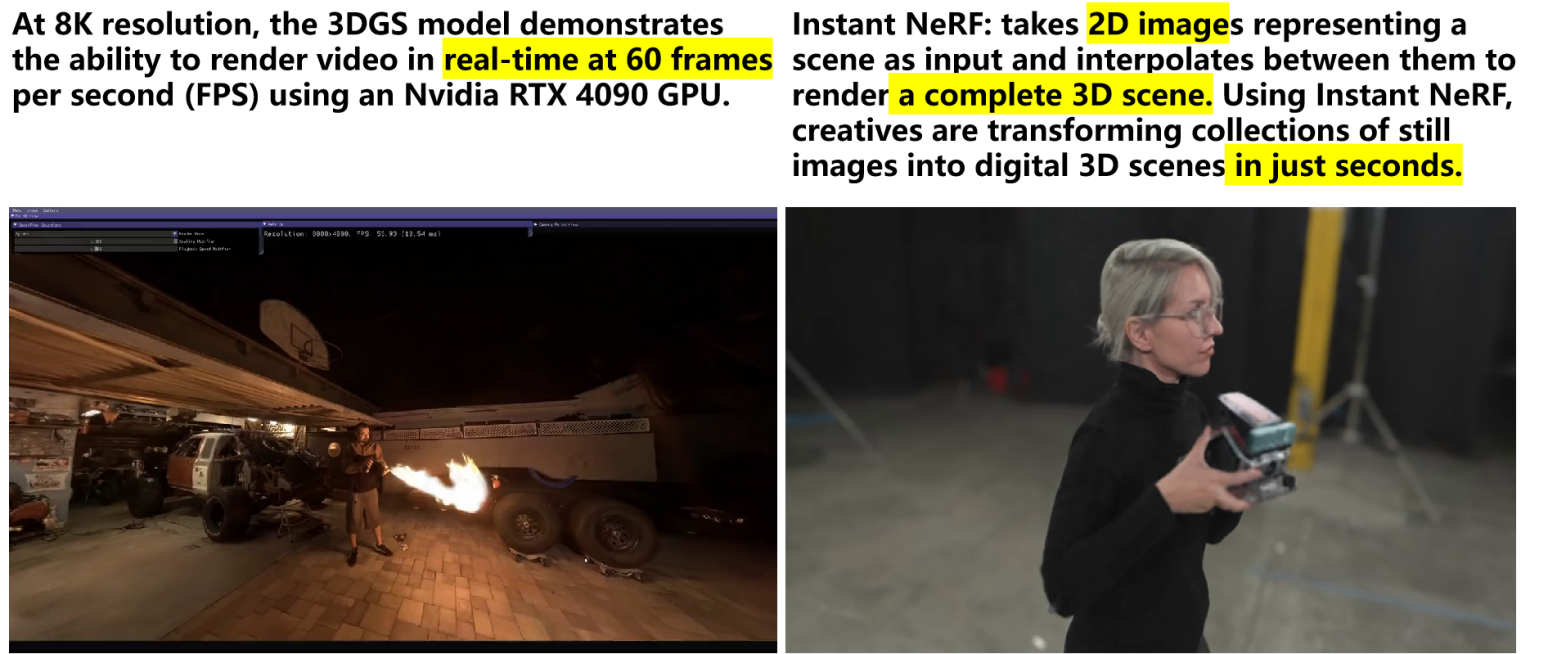


**Solution #2:**

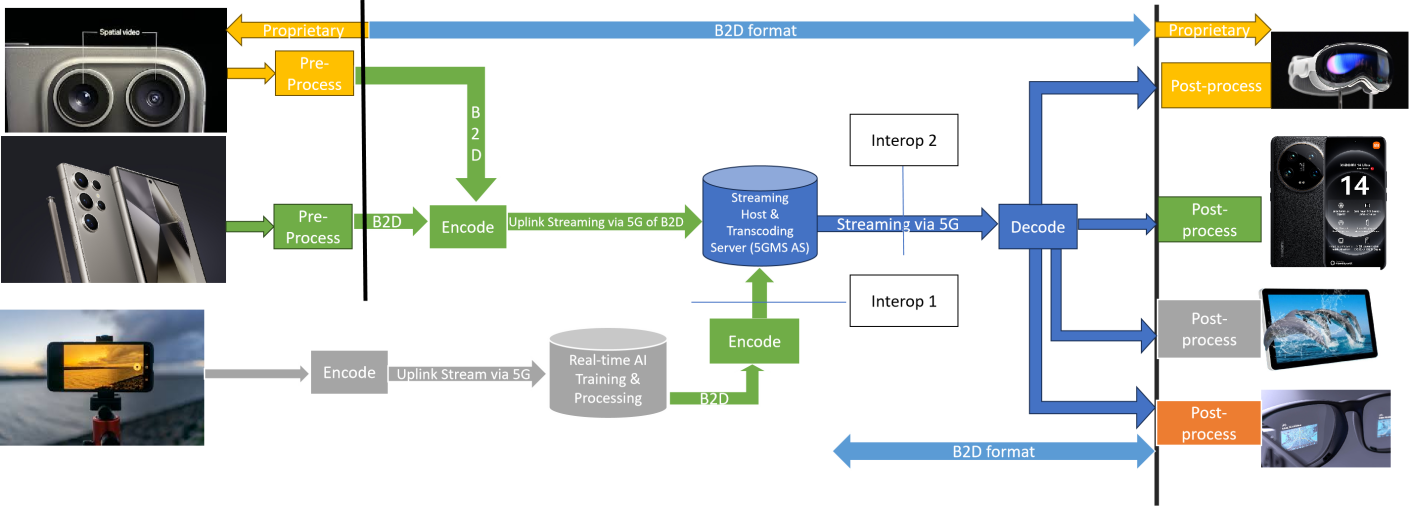
The usage of AI-generated representations (such as NeRF/3DGS) for live streaming is currently being explored, with these technologies showing great promise in achieving high-resolution photorealistic results, real-time rendering, and bandwidth reduction.

Observation: the latest 3DGS model can perform Fast-on-the-fly per-frame reconstruction within 12 seconds and real-time rendering at 200 FPS.

* KIRI Engine - 3D Gaussian Splatting on Android and iOS mobile devices: <https://rebusfarm.net/news/kiri-engine-3d-gaussian-splatting-on-android-and-ios>
* Nividia Instant NeRF: https://www.nvidia.com/en-us/research/nerf-vr-contest/



UE can transmit 2D video to a cloud server, which employs on-the-fly AI training to generate B2D video sequences in real-time. These B2D video frames are then encoded and streamed to the streaming server for distribution to the receiving ends.



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

China Mobile

ZTE

1. **Source format properties**

Table.1 provides an overview of typical beyond 2D source signal properties for UE-to-UE Live Streaming. This information is used to select proper test sequences.

Table 1. Beyond 2D Source Format Properties for UE-to-UE Live Streaming

|  |  |
| --- | --- |
| Source format properties | Social Sharing |
| Number of views | 2 |
| Spatial resolution for each view | Full Width:  3840 x 1080  Full Height:  1920 x 2160  2560 x 3200  Half Width/Height:  1920 x 1080  2560 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 |
| Colour space formats | BT.709,BT.2020 |
| Transfer characteristics | BT.709,BT.2020 |

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

Table 2 provides an overview of encoding and decoding constraints for H.264/AVC and H.265/HEVC for UE-to-UE Live Streaming scenario. This information supports the definition of detailed anchor conditions.

Table 2. Encoding and Decoding Configurations

|  |  |  |
| --- | --- | --- |
| Encoding and Decoding Constraints | H.264/AVC | H.265/HEVC |
| Relevant Codec and Codec Profile/Levels | H.264/AVC Progressive  High Profile 4.2 | H.265/HEVC Main 10 Profile  Level 4.1, 5.1 |
| Random access frequency | 1 second | 1 seconds |
| Bit rates and quality configuration | Fixed QP  CBR  Half Width/Height: 4-8Mbps  Full Width/Height: 8-16Mbps  Capped-VBR | Fixed QP  CBR  Half Width/Height: 5-8Mbps  Full Width/Height: 8-16Mbps  Capped-VBR |
| Bit rate parameters (CBR, VBR, CAE, HRD parameters) | Covering a range of relevant bitrates and qualities | Covering a range of relevant bitrates and qualities |
| Latency requirements and specific encoding settings | Low latency requirements | Low latency requirements |
| Encoding complexity context | Real-time encoding, Cloud-based encoding | Real-time encoding, Cloud-based encoding |
| Required decoding capabilities | H.264/AVC Progressive  High Profile 4.2 | H.265/HEVC Main 10 Profile  Level 4.1, 5.1 |

QP: [17, 22, 27, 32]

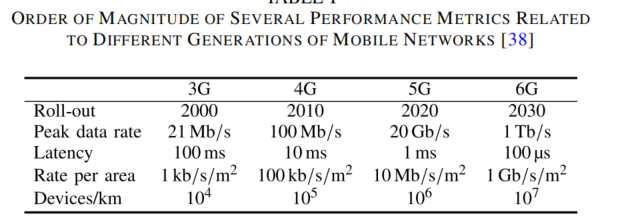
Other potential codecs:

1. MV-HEVC
2. MIV
3. **Performance Metrics and Requirements**

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

[Latency: 1s<FFS>

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

1. **Test Sequences**

<TBD>

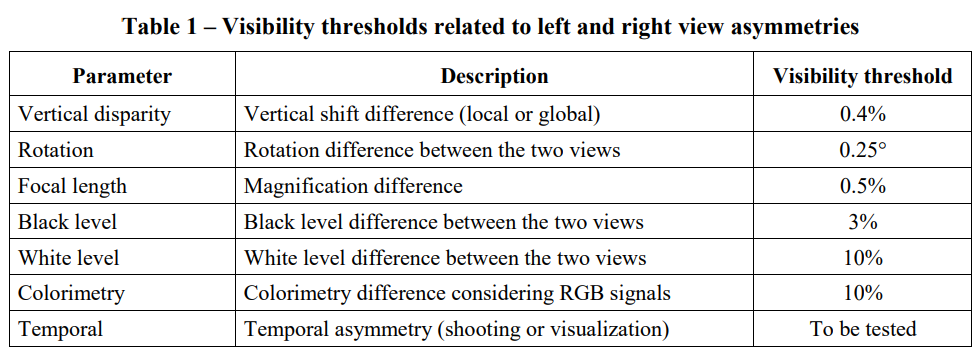
1. **Detailed test conditions**

<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.
3. **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] 3GPP TR 26.955: “5G Video Codec Characteristics”

[4] 3GPP TS 26.511: “5G Media Streaming (5GMS); Profiles, codecs and formats”

[5] 3GPP TS 26.116: “Television (TV) over 3GPP services; Video profiles”

[6] 3GPP TS 26.118: “3GPP Virtual reality profiles for streaming applications”