**3GPP TSG SA WG4 Meeting 133-eS4-251383**

**online, , 18 - 25 July 2025 revision of S4-251319**

**Source: Qualcomm Incorporated, Tencent, CMCC**

**Title: Pseudo-CR on [FS\_Beyond2D] Draft Conclusions**

**Spec: 3GPP TR 26.956v1.0.0**

**Agenda item: 9.6**

**Document for: Decision**

**1. Introduction**

Study closes at this meeting

**2. Reason for Change**

No conclusions on recommendations yet available

**3. Conclusions**

They are ready.

**4. Proposal**

It is proposed to agree the following changes to 3GPP TR 26.956 v1.0.0 at SA4-133-e.

\* \* \* First Change \* \* \* \*

11 Conclusions and Recommendations

## 11.1 Summary and Conclusions

This technical report addresses the evolution of video services from traditional 2D formats to "beyond 2D" video, which includes immersive and interactive experiences based on stereoscopic 3D, multi-view plus depth, dense dynamic point clouds, dynamic meshes, and emerging research formats like Neural Radiance Fields (NeRF), light fields, and 3D Gaussian Splatting (3DGS). The report aims to evaluate the feasibility, performance, and interoperability of these formats and codecs within 3GPP services, considering implementation constraints and network requirements.

The report categorizes beyond 2D video formats as follows:

**- Stereoscopic 3D and Extensions:** Provides depth perception by presenting slightly different images to each eye. Widely supported by current devices and workflows, with extensions for higher resolutions and additional metadata (e.g., depth, alpha).

**- Multi-view Plus Depth:** Offers multiple synchronized camera views, optionally with depth maps, enabling free viewpoint navigation and immersive experiences.

**- Dense Dynamic Point Clouds:** Represents scenes or objects as high-density 3D points with attributes (color, normals, etc.), allowing detailed volumetric rendering.

**- Dynamic Meshes:** Uses vertices, edges, faces and attribute maps to define 3D geometry and texture, supporting animation and real-time rendering, commonly used for avatars and digital twins.

**- Light Fields, NeRF, and 3D Gaussian Splatting:** Advanced research formats that capture and render scenes with high realism and flexibility, though not yet standardized for commercial deployment.

A generic end-to-end reference model is introduced, covering content capture (via cameras or computer graphics), processing (conversion to beyond 2D formats), encoding (compression for efficient transmission), delivery (over 5G/6G networks), decoding, and rendering on various devices (smartphones, VR/AR headsets, autostereoscopic displays). The model emphasizes interoperability points and the need for systematic evaluation.

The report also defines a comprehensive evaluation and characterization framework, including:

**- Reference Scenarios:** Streaming of 1) UE-to-UE live stereoscopic video, 2) professionally produced volumetric video and 3) multi-view plus depth content, each with detailed workflows and constraints.

**- Performance Metrics:** Both objective (e.g. point-based PSNR and PCQM for point clouds, IV-SSIM for multi-view plus depth, HV3D for stereoscopic video) and subjective (user studies) metrics are used to assess quality, efficiency, and user experience. Subjective tests are enabled by the provision of videos allowing interested parties to conduct a formal subjective test, but 3GPP did not organize such a formal subjective test. External objective and subjective test reports are referenced where available.

**- Test Sequences:** A curated set of reference sequences (e.g., volumetric video represented as dense dynamic point cloud and dynamic mesh of people, multi-view scenes) are provided for benchmarking codecs and workflows.

Video codecs supported in existing 3GPP specifications (H.265/HEVC, MV-HEVC) are evaluated for their ability to support beyond 2D formats. The report identifies gaps in current capabilities, especially for new formats like point clouds and dynamic meshes. Network requirements such as latency, bandwidth, and real-time processing are also discussed, with an emphasis on leveraging 5G capabilities.

The report concludes that certain beyond 2D video formats are maturing and becoming market-relevant, driven by advances in capture, production, compression, and display technologies.

## 11.2 Recommendations

Based on the evaluation in this document, the following aspects are recommended:

- 3GPP TS 26.265 defines representation formats to support stereoscopic video. Based on the conclusions in this document, considering extensions for stereoscopic representation formats including intrinsic and extrinsic camera parameters, depth, alpha and possibly improved colour subsampling formats is worthwhile to study in more details including stereoscopic capturing with optical systems on typical UE form factors and then to:

- Identify relevant new representation formats not yet documented in TS 26.265 and provide the benefits in terms of user experience. The evaluation includes potential capturing and rendering of the formats. Candidates include support for alpha, support for depth together with stereo, additional color subsampling 4:2:2 or 4:4:4.

- Study the feasibility of generating video signals following these representation formats on typical UE form factors, in particular smartphones based on existing and emerging optical systems

- Identify compression options for the representation formats based on existing 3GPP codecs, in particular HEVC and MV-HEVC.

- Identify the opportunities and needs to integrate the representation formats into different transport systems, including messaging, real-time communication, split rendering and streaming.

- Define the expected traffic characteristics for new representation formats to meet certain quality thresholds.

- Define a conformance environment, including hosting, tooling and process, as well as conforming test vectors to support operation points.

- Identify gaps in existing specifications and provide guidance for potential normative work.

- For other representation formats, in particular dense dynamic point clouds and dynamic meshes, it is recommended to continue monitoring the broader adoption of these formats. Once there is sufficient market traction, the baseline established in the study may be used to define the exact representation formats. This report also provides a good indication of how different HEVC profiles and features may be used to compress such formats. While these formats may provide some new experiences, there is no immediate necessity to add them to 3GPP specifications. Instead, the market should be monitored for traction of these technologies, particularly in content generation and broader availability of content.

Further 3GPP provided a study on 6G use cases and services requirements in TR 22.870 [Vol-36]. Clause 9.12 of this report describes a use case on personalized interactive immersive guided tour, where assets represented as volumetric video are part of the scene.

As MPEG V-DMC [DM-20] was finalized by MPEG at the closure date for the first version of this technical report, the evaluation of dynamic mesh and the codec MPEG V-DMC is not completed. Therefore, it is recommended:

- Evaluate the dynamic mesh representation format with MPEG V-DMC and HEVC as underlying video codec by providing objective test results and by delivering videos enabling subjective testing.

- Study dynamic mesh content generation for offline productions in prosumer case (e.g. social media) and for real-time applications.

- Immersive communication is one of the important scenarios of 6G, which requires beyond 2D modalities like vision, hearing, and touch to achieve a truly immersive experience. Therefore, it is recommended that potential formats and implementation strategies for B2D video in 6G scenarios be addressed in subsequent studies.

I - A particular format of interest is 3D Gaussian splats (3DGS) as introduced in clause 4.3.6.3. This format represents a 3D scene with a 3D Gaussian primitive, an anisotropic Gaussian ellipsoid. The rendering process is simple and can be executed in real time by projecting the sorted 3D Gaussian splats onto a screen and rendering them in a photorealistic manner. This results in real scenes rendered in real time with lighting and reflection effects, enhancing the realism of the rendered image. In addition, 3DGS has the potential to be generated with commonly available optical systems on existing and emerging smartphones, supported by AI-based workflows. All of this makes the formats an attractive candidate, possibly in the 6G era, and further detailed study of this format is recommended including:

- Identification of the use cases for mobile devices that demonstrate the practical applications of 3DGS contents.

- A full definition and analysis of the 3DGS representation format including the relevance and complexity of the parameters of the primitives. This includes the size of the 3D model and the associated processing requirements, as well as quality and complexity aspects.

- Identify the opportunities to generated 3DGS content, in particular with existing optical systems and the ability to integrate AI-based workflows on device and/or in the network.

- Study the integration of such formats into 3GPP services (e.g. messaging), expected traffic characteristics as well as other aspects related to provide fully interoperable solutions.

- Develop an end-to-end reference implementation for content delivery, covering the entire pipeline from content creation on a server or capture on the UE and, through compression, transmission, to rendering on a mobile device platform. C

Current evaluations are limited to single-user cases, and some test sequences contain only a single asset (excluding complex scenes or multiple assets). In actual deployment, however, multi-user high-concurrency cases are anticipated, which means even compressed B2D video may demand more network resources than traditional 2D video to satisfy latency and bandwidth requirements. Therefore, studying efficient network solutions and bandwidth optimization is crucial to enable real-time B2D video delivery across a wide viewing range without compromising perceptual immersion. Furthermore, B2D-related features, such as AI-based stereoscopic video generation, require substantial computing power, which may exceed UE capabilities. To address this point, it is beneficial to investigate the feasibility of implementing these features (fully or partially) at the network level, thereby reducing computing latency and improving energy efficiency.

\* \* \* Next Change \* \* \* \*

# 2 References

[Vol-36] 3GPP TR 22.870 V0.3.1, Study on 6G Use Cases and Service Requirements

\* \* \* End of Changes \* \* \* \*