**3GPP TSG-SA4 Meeting #130S4-241988**

**Orlando, 18th - 22nd November 2024**

**Source: Apple Inc., Qualcomm Incorporated**

**Title: [FS\_Beyond2D] On Point Cloud format for Beyond 2D video**

**Agenda item: 9.8**

**Document for: Agreement**

# 1 Introduction

The FS\_Beyond2D permanent document and TR 26.956 [1] are gathering a possible set of 3D visual formats thought to be relevant to the *Beyond 2D video* 3GPP study. This includes point cloud representations in clause 4.3.3. During the recent sub-working group calls, contributions were provided on source material based on this format [2]. This discussion reviews the available information on this format and proposes a way forward to progress on this topic.

# 2 Background

Point cloud representations have been used since a long time, with detailed documentation available from 1998 [3]. The format gained favour due to its straightforward relation with capturing and rendering systems. However, already since that time, is was realized that this format requires a significant amount of data to be stored and transmitted for accurate representation. The evidence of this need can already be seen from the work that ensued in the following years (e.g. from 2001 [4]) to improve on performance. This work continues to date with approaches like neural radiance fields (NeRFs) [5] (already documented in the *Beyond 2D Video* 3GPP Technical Report (TR 26.956) [1]), Point-NeRF [6], 3D Gaussian splatting [7], [8] etc. Interested readers can refer to the literature. A main focus of these methods has been to reduce the amount of data needed to be captured (and potentially also stored) before it is actually rendered (and is possibly eventually compressed), for further optimization of quality, storage, and transmission.

The TR, at this point in time, also touches upon these needs, albeit a bit inconsistently. For example, the number of points needed for the point cloud representation format has been added as a necessary parameter for documentation of the available content. While in the introductory clause, a number of more than 500,000 points has been suggested, some included estimates appear to suggest that point clouds with ~10 mega points are needed for accurate human representation when viewed at a distance of 1.5 m. These numbers were provided without any significant documentation of the references. It is also evident that the number of points needed for accurate representation of an object really depend on aspects such as: the application, the type of the content, the rendering technique, and (more specifically in relationship to SA4) the device capabilities in terms of memory, processing resources, and device power consumption. Hence it would be hard to specify a single number for all types of content.

Finally, although some benefits of the technology have been documented, the work on its possible limitations needs to still be conducted. We feel that since the work on addressing some of the known limitations has continued over the past 25 years, such aspects need to be further elaborated. In addition, the essential aspects of the impact of this technology on mobile devices of various resource capabilities, such as in terms of display, processing requirements (e.g. GPU/CPU use, memory throughput), battery/power consumption, etc. would need to be better assessed. Clause 4.3.3.3 on Rendering and Display Systems adds some renderer-based optimizations, which may be a good starting point to document any relevant optimizations.

# 3 Proposal

Given that the work on alleviating limitations of point clouds has been pursued over a very long time and it is still in process, it is proposed that the beyond 2D study should (in some order of priority):

1. Pursue documentation of metrics that would determine an appropriate number of points for a given content type, depending upon the application.
2. Document further relevant optimization techniques and any alternate formats.
3. Evaluate the visual performance of such content on mobile devices. For example, a different set of objective or subjective metrics may be considered more suitable for mobile devices.
4. Preliminary evaluation of the resource consumption on mobile platforms for decompression and rendering of such content.

It should be noted that items 3 and 4 above apply in general to other such visual formats as well.

# References

1. [S4-241721](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/TSGS4_129-e/Docs/S4-241721.zip), [FS\_Beyond2D] TR 26.956 v0.1.0
2. [S4aV240072](https://www.3gpp.org/ftp/TSG_SA/WG4_CODEC/3GPP_SA4_AHOC_MTGs/SA4_VIDEO/Docs/S4aV240072.zip), [FS\_Beyond2D] Test sequences for “streaming single asset” scenario
3. Grossman, Jeffrey P., and William J. Dally. "Point sample rendering." In Rendering Techniques’ 98: Proceedings of the Eurographics Workshop in Vienna, Austria, June 29—July 1, 1998 9, pp. 181-192. Springer Vienna, 1998.
4. Zwicker, Matthias, Hanspeter Pfister, Jeroen Van Baar, and Markus Gross. "Surface splatting." In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 371-378. 2001.
5. Mildenhall, B., P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng Nerf. "Representing scenes as neural radiance fields for view synthesis.,” 2021.
6. Xu, Qiangeng, Zexiang Xu, Julien Philip, Sai Bi, Zhixin Shu, Kalyan Sunkavalli, and Ulrich Neumann. "Point-nerf: Point-based neural radiance fields." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5438-5448. 2022
7. Kerbl, Bernhard, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. "3D Gaussian Splatting for Real-Time Radiance Field Rendering." ACM Trans. Graph. 42, no. 4 (2023): 139-1.
8. Wu, Tong, Yu-Jie Yuan, Ling-Xiao Zhang, Jie Yang, Yan-Pei Cao, Ling-Qi Yan, and Lin Gao. "Recent advances in 3d gaussian splatting." Computational Visual Media 10, no. 4 (2024): 613-642.