Quality Assessment Challenges in MPEG's Current and Future Immersive Media Standards

Sebastian Schwarz and Sébastien Lasserre

Introduction

Recent years have shown significant advances in immersive media experiences. Three-dimensional representation formats allow for new forms of entertainment and communication. In this context, point cloud data has emerged as a promising enabler for such experiences. Because efficient enough point cloud compression technologies are still to be found, the Moving Picture Expert Group (MPEG) has just issued a Call for Proposals (CfP) on point cloud compression technologies. This letter will present the MPEG CfP evaluation procedure and try to anticipate some of the many challenges to be faced when assessing point cloud compression performance.

Dynamic point clouds have been identified as a promising format to code immersive worlds allowing free navigation to the user. The geometry-based description of data leads to new challenges, both compression technologies and quality assessment of the compressed immersive world.

MPEG Call for Point Cloud Compression Technology

There is now a huge interest from the Virtual Reality market in being able to represent the world in three dimensions, thus enabling the end-user to freely navigate in this world. MPEG has launched an ambitious roadmap including future coding technologies of 3D scenes. One of these technologies is Point Cloud Compression (PCC) and is expected to be delivered as an ISO standard in 2019/20. MPEG has issued a call for proposals on PCC technology and aims to evaluate submissions in October 2017 [1].

A point cloud is the given of a set of points, each defined by its 3D XYZ location and attributes, e.g. colour, reflectance, opacity. The MPEG call addresses various applications, resulting in several submission categories, i.e. static, dynamic, and dynamically acquired point clouds, and coding conditions, i.e. lossless geometry with lossy attributes and no/lossy geometry with lossy attributes. Testing material varies from huge, high-precision static point clouds, e.g. for map generation, to smaller but dynamic point clouds, thought as an input of a VR system. This article will focus on the lossy compression of the latter, as quality evaluation is considered the most critical for this scenario.

Figure 1. A three-dimensional object represented by a point cloud [3]. By the very nature of the point cloud format, free navigation is possible around the object.

Evaluation Anchors

For the compression of dynamic point cloud data, MPEG requests submissions based on a set of five test sequences, each with roughly one million points per "frame". An example for such a point cloud sequence is shown in Figure 1. Five target bit rates must be achieved for each sequence, ranging from 3 to up to 72 Mbit/s, to cover a wide range of use cases, for a total of 25 test points. At each test point, a proponent's decoded point cloud sequence will be evaluated against the competing submissions, as well as an anchor encoding generated with the provided experimental PCC software.

The anchor software relies on subsampling an octree representation for geometry and JPEG-based colour compression for attributes. The software allows for simple temporal prediction structures (IPIPIP), however, this feature is not used for the CfP. An example for the anchor compression result is shown in Figure 2. Due to the geometry subsampling, the decoded point cloud (middle) has fewer points than the original (left). This effect must be taken account in the objective and subjective quality assessment, for example by increasing the rendered point size for subjective viewing (right).

Objective Quality Assessment

Classically, encoding performance is assessed in a rate-distortion fashion, comparing the achieved bit rate with the introduced distortions. For 2D video, peak signal-to-noise ratio (PSNR), based on the mean squared error (MSE) between original and decoded pixel, is the most accepted distortion metric. While the PSNR does not necessarily fully represent all effects of the human visual system, it works well for typical 2D video coding artefacts, such as blocking and blurring.

Figure 2. Point cloud compression results using the anchor software [2] at 13 Mbit/s: Original, decoded point cloud (geometry subsampled), decoded point cloud rendered with larger point size.

For 3D point clouds, the relation between original and decoded point is not straight forward. As seen in Figure 2, the decoded point cloud might have less (or more) points than the reference point cloud. Furthermore, the decoded point has two kinds of distortion, geometry distortion and colour distortion.

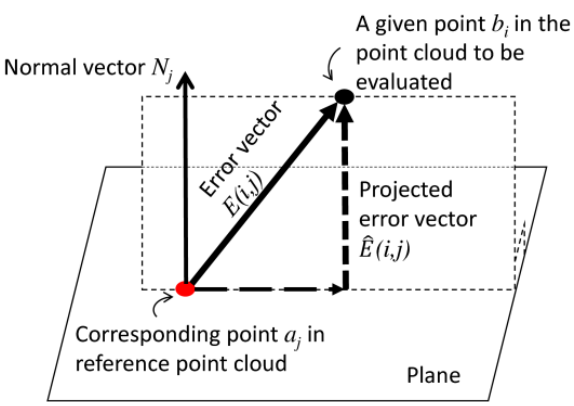


Figure 3. Illustration of the point-to-point error E (D1) and point-to-plane error Ê (D2).

Therefore, the CfP specifies three distortion metrics. The first metric, D1, calculates the MSE between a position of a point and the position of its closest neighbour in a reference point cloud (point-to-point). D2 calculates the MSE between the position of a point and its projection onto a given reference plane, representing the surface of the point cloud model (point-to-plane). The difference between these two error calculations is illustrated in Figure 3. Finally, colour distortion is calculated in YUV domain, between the current point and its closest reference neighbour (D1).

To take the possibility of a largely varying number of original and decoded points into account, both metrics are run twice.

First, comparing all decoded points to the original, then comparing the original to the decoded points, and symmetric results of both runs are reported.

Subjective Quality Assessment

It is apparent that purely objective quality evaluation for point cloud data suffers from similar problems as 2D video. In addition, with no research on the relation of D1 and D2 to the actual perceived quality, and no knowledge of the different effects and relationship between geometry and colour distortion on subjective quality, subjective quality assessment becomes a key necessity for reliably assessing coding performance. However, there is no standard procedure for assessing visual quality of dynamic point cloud data. In order to establish some kind of standardised procedure, two key aspects need to be addressed: First, the rendering of the points, and second, how to ensure stable viewing between different test subjects and test points.

Regarding the rendering, a point has no size and is not supposed to be visible. Consequently, it must be visualised by something with some shape and size to make an object represented by a point cloud viewable. Given that points are located on a three-dimensional uniform integer-based grid, the minimum shape that fills the space between adjacent points without overlapping is a cube of size unity.

However, due to possible geometry sampling induced by the compression system, the optimal cube size may not be unity. The visual effect of this size is shown in Figure 3 and 4. A powerful rendering scheme would be to allow a local cube size depending on the location of the neighboring points, but this would interfere in some uncontrolled way on the compression scheme and may hide compression artifacts. Since the goal of MPEG is to provide a compression system and not a renderer, it has been decided to impose a uniform shape (cube) for all points and allow the proponents to provide a given point size for their decoded content to be rendered at.

Concerning the methodology for subjective quality evaluation and ensuring stable viewing conditions, it has been decided to not let the participants navigate freely around the object. Instead, 2D video based on a fixed path around the object will be rendered. This path is unknown to the proponents beforehand. In addition to ensuring stable viewing conditions for all test subjects, this approach has the benefit to keep participants focused on the quality evaluation and not distract them with the navigation controls. Video quality can then be evaluated using standardised methods for assessing subjective quality.

Figure 4. Effect of the rendered point size on the visual quality of the object. A too small size leads to a ghostly rendering (left) but, on the other hand, a too big size masks texture fine details (right). Determining the optimal point size (center) is one of the biggest challenges behind the subjective quality assessment.

Perspective and Future Work

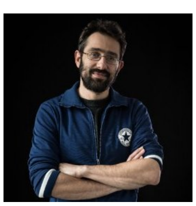
Looking at the presented objective and subjective quality assessment for point cloud compression technology, it becomes apparent that this field is far less researched than 2D video quality assessment. MPEG is aware that the chosen approaches do not necessarily present the final word in assessing point cloud compression quality. Nonetheless, they represent the current state of research at the time of issuing the CfP and should allow for an initial assessment.

Nonetheless, any help on refining point cloud compression quality assessment is more than welcome. Details on the currently chosen approaches are available in [1]. Interested VQEG experts are invited to contribute to this activity. In particular, inputs on the following problems are of high interest to the authors:

- What extensions or improvements to the current distortion metrics could be considered? Are there any other reliable metrics other than D1 and D2?
- How do the metrics for geometry distortion relate to perceived visual distortion?
- What is the relationship between geometry and colour distortion when it comes to visual quality?
- How to reliably assess the effects of geometry sub-/ over-sampling on objective and subjective quality?
- What are the effects of temporal geometry distortions on the visual quality. How to assess them?
- How to standardise visually quality assessment for (dynamic) point cloud data.

As for future work, the MPEG CfP has been issued and proponents are invited to submit their solutions. The above-described objective and subjective quality assessment will be carried out in October 2017 and results should be available by the 120th MPEG meeting. We intend to publish a follow-up article discussing the outcome and faced challenges during this evaluation in a later edition of this VQEG eLetter.

Sebastian Schwarz
Nokia Technologies, Finland



Sébastien Lasserre
Technicolor, France

References

[1] Call for Proposals for Point Cloud Compression V2. ISO/IEC JTC1/SC29/WG11 Doc. N16763, Hobart, Australia, Apr. 2017.

[2] R. Mekuria, K. Blom, P. Cesar, "Design, Implementation and Evaluation of a Point Cloud Codec for Tele-Immersive Video", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828-842, Apr. 2017.

[3] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i Voxelized Full Bodies – A Voxelized Point Cloud Dataset." ISO/IEC JTC1/SC29/WG11 Doc. M40059, Geneva, CH, Jan. 2017.