 ISO/IEC JTC 1/SC 29/WG 7 N513

**ISO/IEC JTC 1/SC 29/WG 7  
MPEG 3D Graphics and Haptics Coding   
Convenorship: AFNOR (France)**

**Document type:** Output Document

**Title:** Perf. Analysis of Currently AI-based Available Solutions for PCC

**Status:** Approved

**Date of document:** 2023-02-09

**Source:** ISO/IEC JTC 1/SC 29/WG 7

# Expected action: None

# Action due date: None

**No. of pages:** 14 (with cover page)

**Email of Convenor:** marius.preda @ imt . fr

**Committee URL:** [https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg7](https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3)

**INTERNATIONAL ORGANIZATION FOR STANDARDIZATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 7 MPEG Coding for 3D Graphics and haptics**

**ISO/IEC JTC 1/SC 29/WG 7 N513**

**January 2022, Online**

|  |  |
| --- | --- |
| **Title** | **Perform. Analysis of Currently AI-based Available Solutions for PCC** |
| **Author** | **Alexandre Zaghetto** |
| **Source** | **WG7, MPEG 3D Graphics Coding** |
| **Status** | **Approved** |
| **Serial Number** | **22430** |

Abstract

A baseline performance evaluation of different AI-based point cloud compression solutions was initiated after the 4th WG7 (135th MPEG) meeting in July 2021. Three AI-based solutions were investigated: (a) pcc\_geo\_cnn\_v2, (b) pcgcv2 and (c) adl-pcc. Since then, all results regarding these and new proposals were continuously being integrated meeting after meeting in subsequent versions of the Perform Analysis of Currently AI-based Available Solutions for PCC output documents. The version released in the 9th WG7 (140th MPEG) contains the full progression of results (w22092). In the current version of the document, only the performances of the latest active proposals being discussed in the group are reported.

# Introduction

After the 4th WG7 (135th MPEG), three AI-based solutions were initially evaluated as a baseline study, namely:

* **pcc\_geo\_cnn\_v2**: presented in m57301, introduces the discussion about Point Cloud Geometry Compression using the “Improved Deep Point Cloud Geometry Compression” architecture.
* **adl-pcc**: presented in m57556, aims at informing the MPEG community through SC29/WG7 on the public release of software for a Deep Learning (DL)-based Point Cloud (PC) geometry coding solution, as an effort to contribute to the MPEG Point Cloud Coding activity related to learning-based coding.
* **pcgcv2**: presented in m57453, proposes a geometry compression framework for AI-based PCC based on sparse convolution. In this method, the point cloud is represented by sparse tensor and processed by spatially sparse Convolution Neural Networks (CNN). More specifically, sparse CNNs are employed to exploit the spatial dependency between voxels and predict the occupancy probability, which will be used for entropy coding or binary classification of voxel occupancy symbols.

The details of the baseline study can be accessed in the document w22092 [1]. In the next section, performance results of active proposals being discussed in the group are presented.

# Performance Analysis

# EE 5.1 Deep Octree Coding

In this Exploration Experiment, the following techniques are being evaluated, m62102 [2]:

* G-PCC Octree
* SparseContextNet (SCN)
* VoxelContextNet (VCN)
* SparsePCGCv1
* SparsePCGCv2

Proposals target the *automotive-frame* use case. The training dataset includes Ford\_01\_q\_1mm and KITTI from 00 to 10 datasets. The test set is composed of Ford\_02\_q\_1mm, Ford\_03\_q\_1mm and KITTI 11,12 and 13 sequences. Average PSNR plots comparing the proposed SCN against G-PCC, VCN, SparsePCGCv1 and SparsePCGCv2 for Ford sequences are shown in Figure 1. Figure 2 shows results for the KITTI dataset.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Figure 1. Average R-D performance on Ford\_02 and Ford\_03.

Chart, line chart

Description automatically generated Chart, line chart

Description automatically generated

Figure 2. Average R-D performance on KITTI 11, 12 and 13.

BD-rate gains over G-PCC octree with angular mode enabled for Ford and with GPCC octree with angular mode disabled for KITTI data are as follows:

D1,D2

gain over GPCC SCN VCN SparsePCGCv1 SparsePCGCv2

Ford -8.6, -8.74 -1.91, -2.06 **-12.06,-12.12** **-14.31, -14.62**

KITTI -14.12, -14.13 -8.69, -8.7 **-19.37,-19.02** **-17.89, -17.89**

In m62101 (EE 5.1 summary report) [3], an additional test performed with SparsePCGCv1 identified as “new” is presented. The main difference between SparsePCGCv1 and SparsePCGCv1 “new” is in the convolution kernel size. Plots for Ford\_02\_q\_1mm and Ford\_03\_q\_1mm and are shown in Figure 3.

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Figure 3. Average R-D performance on Ford\_02 and Ford\_03.

BD-rate gains against G-PCC octree and number of model parameters is reported for SparsePCGCv1, VCN and SCN, as follows:

* SparsePCGCv1 -12.06% 5.6M parameters
* VCN -01.91% 0.9M parameters
* SCN -08.60% 9.2M parameters

In m62177 [4], a proposal for dynamic point cloud geometry compression for LiDAR point clouds with ego-motion compensation is presented. The training dataset is composed by sequences 00 to 05 of KITTI. KITTI 06 is used as the verification dataset, and KITTI 07 to 10 are used for testing. Gains against G-PCC Octree are reported in Table 1.

Table 1. Gains against G-PCC Octree.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Lossless (Bpp) | | | | Lossy (BD-Rate) | |
| point clouds | Number of frames | G-PCC | SparsePCGCv1 | Ours | CR Gain | D1 | D2 |
| KITTI\_07\_vox1mm | 1100 | 19.389 | 18.203 | **14.439** | **-25.53%** | **-51.94%** | **-51.97%** |
| KITTI\_08\_vox1mm | 4070 | 20.261 | 19.654 | **15.23** | **-24.83%** | **-49.49%** | **-49.53%** |
| KITTI\_09\_vox1mm | 1590 | 20.325 | 19.158 | **15.235** | **-25.04%** | **-50.39** | **-50.41%** |
| KITTI\_10\_vox1mm | 1200 | 19.669 | 18.702 | **14.797** | **-24.77%** | **-52.57%** | **-52.62%** |

An illustrative rate-distortion plot for KITTI 10 sequence, that includes comparison against G-PCC octree and SparsePCGCv1, is show in Figure 2.

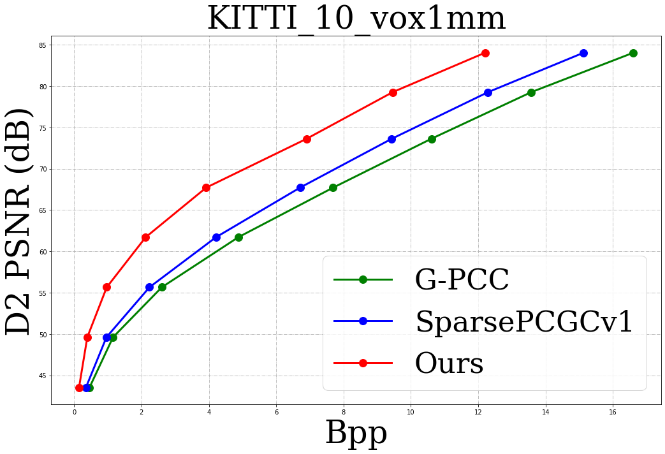
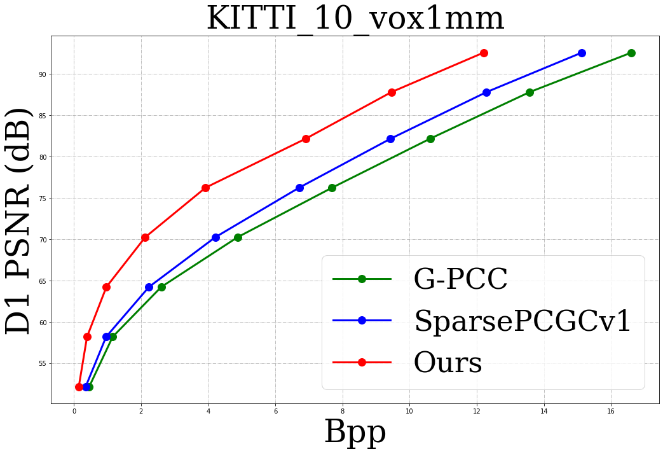


Figure 4. Average R-D performance on KITTI 10.

* 1. **EE 5.2 Deep feature-based PC Coding**

In this Exploration Experiment, the following techniques are being evaluated:

* SparsePCGCv1; and
* GRASP-Net

The EE 5.2 summary report [5] presented in the 10th WG7 (141th MPEG) meeting compares the performance of SparsePCGCv1 and GRASP-Net for automotive-frame (LiDAR) and surface (solid, dense and sparse) point clouds. Two training configurations were evaluated, as described in Table 2.

Table 2. Training dataset configurations of SparsePCGC and GRASP-Net.

|  |  |  |
| --- | --- | --- |
| **Category** | **SparsePCGC (NJU/OPPO)** | **GRASP-Net (InterDigital)** |
| **Solid** | Modified ShapeNet | ModelNet40 |
| **Dense** | 4 dense MPEG sequences | Pretrain on ModelNet40 for 50 epochs, finetune on 4 dense MPEG sequences |
| **Sparse** | 4 dense MPEG sequences | ModelNet40 |
| **LiDAR** | Ford\_01 | Ford\_01 |

Results for Ford sequences and corresponding BD-rate gains are shown in Figure 5 and Table 3, respectively.

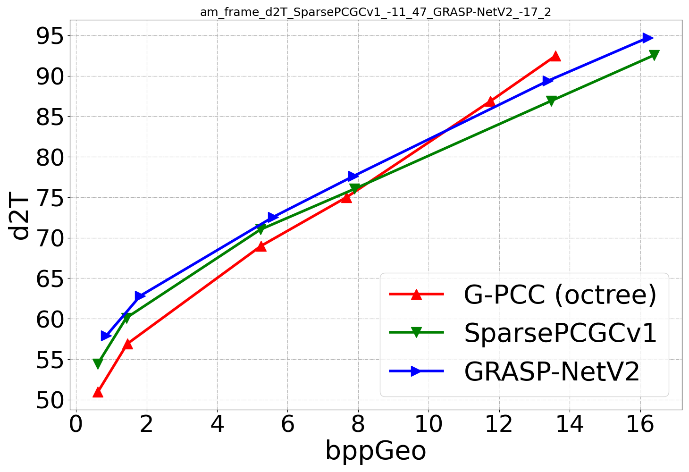
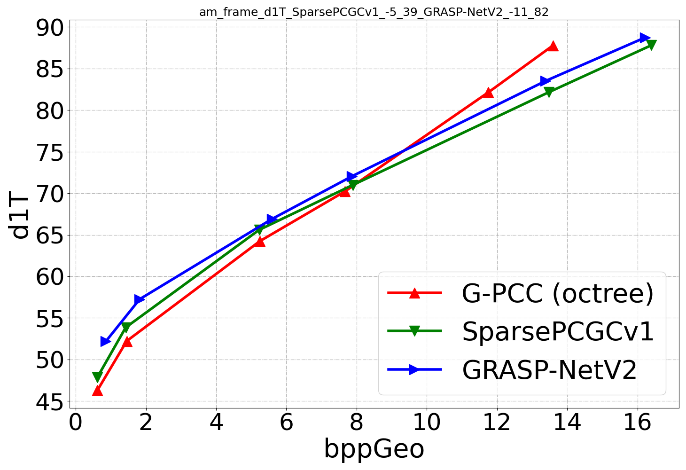


Figure 5. Average R-D performance on Ford\_02 and Ford\_03.

Table 3. Average BD-Rate gains (in %) with respect to G-PCC (octree).

|  |  |  |
| --- | --- | --- |
| Category | LiDAR | |
| Metric | D1 | D2 |
| SparsePCGCv1 | -5.39 | -11.47 |
| GRASP-NetV2 | **-11.82** | **-17.20** |

For surface point clouds, BD-rate gains are shown in Tables 4 and 5 under the two configurations described in Table 2.

Table 4. Average BD-Rate gains (in %) with respect to G-PCC (octree) under the training configuration of NJU/OPPO

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Category | Sparse | | Dense | | Solid | |
| Metric | D1 | D2 | D1 | D2 | D1 | D2 |
| SparsePCGCv1 | +14.96 | +6.18 | **-64.40** | **-80.27** | -92.10 | **-85.80** |
| GRASP-NetV2 | **-9.36** | **-13.78** | -60.54 | -71.55 | **-92.39** | -84.90 |

Table 5. Average BD-Rate gains (in %) with respect to G-PCC (octree) under the training configuration of IDC

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Category | Sparse | | Dense | | Solid | |
| Metric | D1 | D2 | D1 | D2 | D1 | D2 |
| SparsePCGCv1 | +12.35 | -0.36 | Not available | Not available | -91.93 | **-85.65** |
| GRASP-NetV2 | **-15.29** | **-21.83** | -61.54 | -72.78 | **-92.51** | -85.27 |

The number of parameters of the neural network models used by SparsePCGCv1 and GRASP-Net are provided in Table 6. The model size of GRASP-NetV2 is much smaller than that of SparsePCGCv1. However, GRASP-NetV2 trains one model per rate-point, while SparsePCGCv1 trains one model and applies it for multiple (but not all) rate points

Table 6. Number of model parameters (in 103) when coding different types of PCs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| PC Category | LiDAR | Sparse | Dense | Solid |
| SparsePCGCv1 | 8279 | 916 to 5713 | 916 to 5713 | **210** to 1499 |
| GRASP-NetV2 | **290** | **290** to 292 | **211** to 290 | 293 to 318 |

A more detailed analysis of GRASP-NETv2 trained under the configurations described in Table 2 is presented in contribution m62096 [6]. Tables 7 summarizes the results.

Table 7. B-D rate savings (%) of GRASP-NetV2 on the original training settings (Config IDC) and on the training settings of SparsePCGC (Config NJU).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Training Config** | **Solid** | | **Dense** | | **Sparse** | |
| **D1** | **D2** | **D1** | **D2** | **D1** | **D2** |
| Config IDC | **-92.51** | **-85.27** | **-61.54** | **-72.78** | **-15.29** | **-21.83** |
| Config NJU | -92.39 | -84.90 | -60.54 | -71.55 | -9.36 | -13.78 |

**EE 5.3 AI-based Dynamic PC Coding.**

The EE 5.3 summary report [7] presented in the 10th WG7 (141th MPEG), four main technologies are listed:

* “DPCC-SC: Dynamic Point Cloud Geometry Compression using Sparse Convolutions”
* “SparsePCGCv3: Dynamic SparsePCGC with Inter Frame Prediction”
* “S-DPCC: Unified Intra/Inter Deep Dynamic Point Cloud Compression with Multiple Reference Frames and Rate Adaptation”
* “A Deep Dynamic Point Cloud Geometry Compression Framework for AI-based PCC”.

In m62076 [8] (and supporting m59617/m60307/m61201 document), a “Dynamic Point Cloud Geometry Compression using Sparse Convolutions” is discussed. Results against PCGCv2 and V-PCC (hm\_ld, hm\_ra and vvenc\_slow\_ra) are summarized in the Tables 8 and 9.

Table 8. Dynamic Point Cloud Geometry Compression using Sparse Convolutions: 96 Frames – 10bit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **hm\_ld (%)** | **hm-ra (%)** | **vvenc\_slow\_ra (%)** | **Intra (PCGCv2) (%)** |
| basketball\_player | -57.28 | -52.98 | -40.06 | -29.62 |
| dancer | -57.5 | -54.72 | -43.25 | -17.42 |
| exercise | -56.06 | -51.5 | -37.88 | -32.71 |
| model | -54.6 | -50.36 | -37.76 | -22.66 |
| **Average** | -56.52 | -52.61 | -40.09 | -25.67 |

Table 9. Dynamic Point Cloud Geometry Compression using Sparse Convolutions: 96 Frames – 11bit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **hm\_ld (%)** | **hm-ra (%)** | **vvenc\_slow\_ra (%)** | **Intra (PCGCv2) (%)** |
| basketball\_player | -54.11 | -50.44 | -35.35 | -38.49 |
| dancer | -56.78 | -54.02 | -40.46 | -34.7 |
| exercise | -52.89 | -48.39 | -33.08 | -39.35 |
| model | -55.99 | -52.86 | -38.88 | -36.45 |
| **Average** | -55.09 | -51.62 | -37.23 | -37.24 |

In m62181 [9] (and supporting documents m60354/m61006), the “Dynamic SparsePCGC” is discussed. Results are shown in Table 10.

Table 10. Dynamic SparsePCGC:

96 Frames – 10bit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **hm\_ld (%)** | **hm-ra (%)** | **vvenc\_slow\_ra (%)** | **Intra (%)** |
| basketball\_player | -67.65 | -64.64 | -54.36 | -14.06 |
| dancer | -70.01 | -68.15 | -59.85 | -7.49 |
| exercise | -65.14 | -61.7 | -50.45 | -14.04 |
| model | -69.03 | -66.2 | -57.38 | -15.16 |
| **Average** | **-68.19** | **-65.46** | **-55.97** | -12.75 |

In m62066 [10], a “Unified Intra/Inter Deep Dynamic Point Cloud Compression with Multiple Reference Frames and Rate Adaptation”, is discussed. In addition to BD-rate gains observed against “Dynamic Point Cloud Geometry Compression using Sparse Convolutions” [7], shown in Table 11, architectural improvements on top of the reference technique are proposed, namely:

1. a **new** multi-scale spatio-temporal **predictor network**;
2. a **unified I- and P-frame compression** network
3. the use of a modulation network (gain and inverse gain units) that allows for **a single trained model to operate over a discrete set of rate points**; and
4. an interpolation strategy that uses the discrete gain and inverse gain units to achieve a **continuous rate adaptation**.

Table 11. Unified Intra/Inter Deep Dynamic Point Cloud Compression with Multiple Reference Frames and Rate Adaptation: 96 Frames – 10bit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| BD-Rate | | | | |
| Basketball player | Dancer | Exercise | Model | Average |
| -11.25 | -18.04 | -11.96 | -14.19 | -13.86 |

In m58780 [11] (and supporting documents m59685/m60267), “A Deep Dynamic Point Cloud Geometry Compression Framework for AI-based PCC” is proposed. No recent activity (two last meeting cycles) was identified for this proposal.

* 1. **EE 5.4 AI-based Attribute Coding**

In m59037 [12], a point cloud attribute compression (PCAC) method based on sparse tensor representation and corresponding sparse tensor network, like the geometry compression method in m57453 [13], is presented. Results are summarized in Table 12.

Table 12. BD-BR (%) and BD-PSNR (dB) comparisons of the proposed method with TMC13 on Y channel and YUV channels.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Point Cloud** | **TMC13v14 (RAHT)** | | | | **TMC13v6 (RAHT)** | | | |
| **BD-BR (%)** | | **BD-PSNR (dB)** | | **BD-BR (%)** | | **BD-PSNR (dB)** | |
| **Y** | **YUV** | **Y** | **YUV** | **Y** | **YUV** | **Y** | **YUV** |
| longdress | +27.3 | +26.9 | -0.84 | -0.79 | -31.3% | -32.2 | +1.24 | +1.23 |
| loot | +115.5 | +139.1 | -2.18 | -2.49 | 1.3% | +15.5 | +0.03 | -0.37 |
| red&black | +34.8 | +54.8 | -0.88 | -1.27 | -36.1% | -26.5 | +1.36 | +0.91 |
| soldier | +37.2 | +64.6 | -1.03 | -1.58 | -33.3% | -17.7 | +1.29 | +0.52 |
| **average** | **+53.7%** | **+71.4%** | **-1.23dB** | **-1.53dB** | **-24.8%** | **-15.2%** | **+0.98dB** | **+0.57dB** |

In m61007 [14], a lossless Point Cloud Attribute Compression (PCAC) method based on Multiscale Sparse Tensor (MST) representation is proposed. Results are shown in Tables 13 to 15.

Table 13. Gains over G-PCC on color attribute compression.

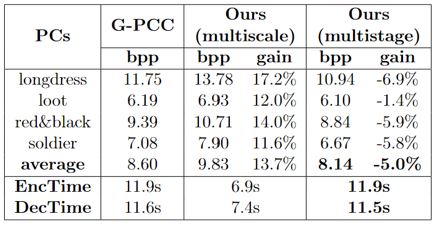


Table 14. Comparison using different training dataset.

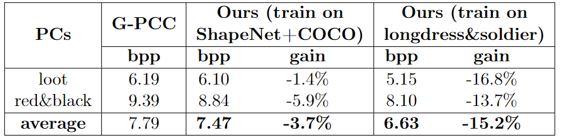
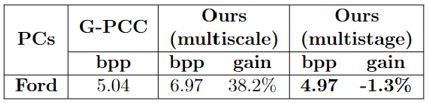


Table 15. Gains over G-PCC on reflectance attribute compression (Train on Ford\_01\_q1mm and test on Ford\_02\_q1mm and Ford\_03\_q1mm.)



In m62176 [15], a study about the impact of training strategies on techniques described in m59037 [12] and m61007 [14] is presented. Datasets are described in Table 16, and results are summarized in Figure 6 and Table 17.

Table 16. Dataset Details.

|  |  |  |
| --- | --- | --- |
|  | **Training Datasets** | **Test Datasets** |
| **Colorized ShapeNet** | 6000 frames | 32 frames |
| **8iVFB & Owlii** | longdress\_vox10, loot\_vox10, red&black\_vox10 (3\*300 frames)  dancer\_vox11, model\_vox11, exercise\_vox11 (3\*64 frames) | soldier\_vox10\_0690  basketball\_player\_vox11\_0200 |

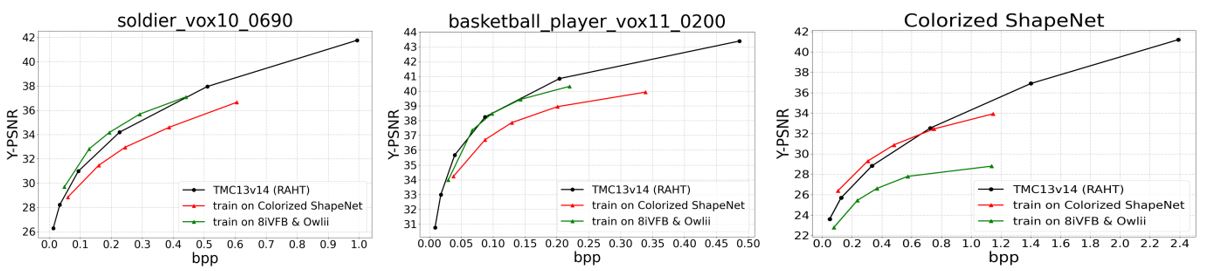


Figure 6. Lossy PCAC m59037 [10] performance comparison using rate-distortion curves.

Table 17. Lossless PCAC m61007 [12] performance comparison using bpp & gains over G-PCC.

|  |  |  |  |
| --- | --- | --- | --- |
| Test data | G-PCC | Train on Colorized ShapeNet | Train on 8iVFB & Owlii |
| soldier\_vox10 | 7.080 | 6.665 (-5.9%) | 5.621 (-20.6%) |
| basketball\_player\_vox11 | 7.720 | 7.762 (0.5%) | 7.036 (-8.9%) |
| Colorized ShapeNet | 11.000 | 8.966 (-18.5%) | 10.096 (-8.2%) |

In m61142 [16], a deep point cloud attribute compression scheme with normalizing flow is proposed. Results are shown in Figure 7.

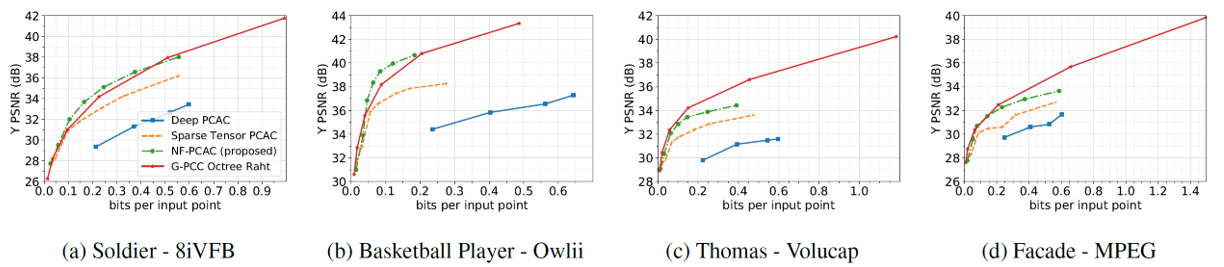


Figure 7. Y-PSNR plots against TMC13v14.

* 1. **General Explorations**

In m61313 [17] and m62180 [18], authors propose a flexible configuration of the deep learning-based framework where rather than having a joint geometry and attribute compression scheme, we proposed to use a deep learning-based geometry compression along with G-PCC’s attribute compression. Results are show in Figure 8.

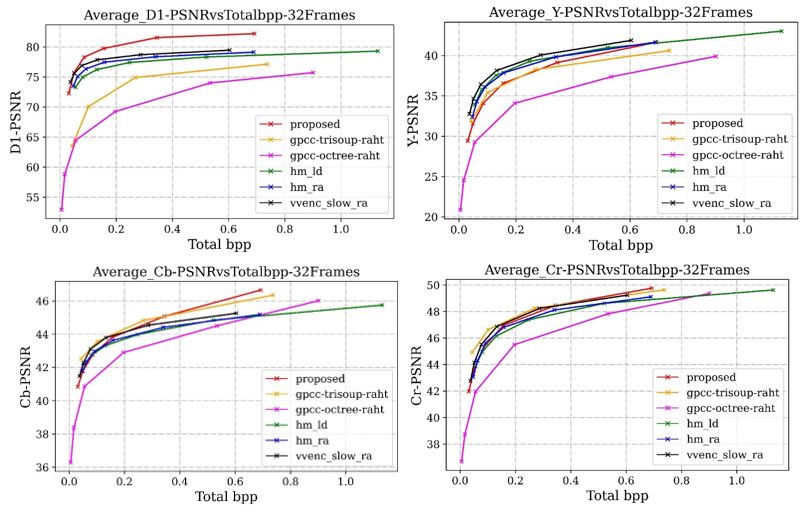


Figure 8. Geometry and attribute PSNR plots.

# Conclusion

In this document, the performances of active proposals discussed in the group were presented. Main use cases addressed are automotive (LiDAR frame) and surface (solid, dense and sparse, up to 13 bits geometry) static and dynamic point clouds. Full alignment with the “Guidelines for conducting AI exploration experiments for PCC” [19] and the “Preliminary data set collection for AI experiments” [20] is strongly recommended. Common training and test conditions for AI-based Attribute Coding needs to be defined.

# Participants

|  |  |  |  |
| --- | --- | --- | --- |
| **Contact** | **Affiliation** | **Email** | **Role** |
| **Alexandre Zaghetto** | Sony | alexandre.zaghetto@sony.com | Coordinator |
| **Jianqiang Wang** | Nanjing University | wangjq@smail.nju.edu.cn | Collaborator |
| **Jiahao Pang** | IDC | jiahao.pang@interdigital.com | Collaborator |
| **Vladyslav Zakharchenko** | OPPO | vladyslav.zakharchenko@oppo.com | Collaborator |
| **Anique Akhtar** | Qualcomm | aniquea@qti.qualcomm.com | Collaborator |

# References

[1] “*Performance analysis of currently AI-based available solutions for PCC,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document w22092, Mainz, October 2022

[2] “*Point cloud geometry compression with hierarchical feature propagation (update),”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62102, Online, January 2023.

[3] “*Summary on EE5.1 on Deep Octree Coding,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62101, Online, January 2023.

[4] “*Dynamic Point Cloud Geometry Compression for LiDAR Point Cloud with Ego-Motion Compensation,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62177, Online, January 2023.

[5] “*Summary Report of EE 5.2 on End-to-end AI PC Coding,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62093, Online, January 2023.

[6] “*On Training Datasets for EE 5.2,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62096, Online, January 2023.

[7] “*Summary Report AI-Based Dynamic PC Coding,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62077, Online, January 2023.

[8] “*Results dynamic point cloud compression,*” ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62076, Online, January 2023.

[9] “*Test 3* *Cross-check results,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62181, Online, January 2023.

[10] “*Unified Intra/Inter Deep Dynamic Point Cloud Compression with Multiple Reference Frames and Rate Adaptation,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62066, Online, January 2023.

[11] “*A Deep Dynamic Point Cloud Geometry Compression Framework for AI-based PCC,*” ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m58780, Online, January 2022.

[12] “*Point Cloud Attribute Compression using Sparse Tensor-Representation,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m59037, Online, January 2022.

[13] “*A Geometry Compression Framework for AI-based PCC via Sparse Convolution,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m57453, Online, July 2021.

[14] “*Lossless SparsePCAC: Multiscale Sparse Representation for Lossless Point Cloud Attribute Compression,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m61007, Mainz, October 2022.

[15] “*On the Training Datasets for Attribute Compression,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62176, Online, January 2023.

[16] “*NF-PCAC: deep point cloud attribute compression with normalizing flow,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m61142 , Mainz, October 2022.

[17] “*Baseline attribute compression for ML-based PCC,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m61313 , Mainz, October 2022.

[18] “*Update on baseline attribute compression for ML-based PCC,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document m62180 , Online, January 2023.

[19] “*Guidelines for conducting AI exploration experiments for PCC,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document w22429, Online, January 2023.

[20] “*Preliminary data set collection for AI experiments,”* ISO/IEC JTC1/SC29 WG7 (MPEG 3D Graphics Coding) Document w22431, Online, January 2023.