Geometry-Based Compression of Plenoptic Point Clouds

Davi R. Freitas
Inria - Rennes Bretagne-Atlantique
Rennes, France
rabbouni.davi@ieee.org

Gustavo L. Sandri
Instituto Federal de Brasília
Brasília, Brazil
gustavo.sandri@ieee.org

Ricardo L. de Queiroz
Universidade de Brasília
Brasília, Brazil
queiroz@ieee.org

***Abstract*—Plenoptic point clouds (PPC) are novel data structures that represent the light from different viewing directions in order to provide a higher degree of realism to regular point clouds. This is achieved by associating each point to multiple colors instead of a single one. Here, we present a method to efficiently compress the attributes of a PPC, consisting of a Karhunen-Loève transform over the color attributes followed by multiple attribute coders with intra prediction capability. This compression scheme can be incorporated within the MPEG's geometry-based PCC (G-PCC) standard, using any of G-PCC's existing solutions for attribute coding. Compression performance assessment using PPCs of different spatial resolutions reveals competitive results in comparison to existing methods, such as RAHT-based or video-based PCC solutions. We believe our coder to be the new state of the art.**

***Index Terms*—Point cloud compression, plenoptic point clouds**

I. Introduction

A point cloud (PC) is a data structure that represents objects and space in a three-dimensional (3D) coordinate system. Frequently used in applications of real-time rendering and capture of 3D objects [1], it is common to voxelize points by constraining them to a grid of $N \times N \times N$ voxels [2]. PCs have been favored over meshes due to the greater simplicity in its capture process. However, there is a huge amount of data required to represent these data structures. Therefore, the development of compression techniques in order to transmit and store this data is paramount. As a result, the Moving Picture Expert Group (MPEG) is among groups with ongoing standardization processes towards point cloud data compression [3].

The geometry-based point cloud compression (G-PCC) standard has been proposed by MPEG to compress high-detail static point clouds as efficiently as possible [3], [4], [5]. G-PCC offers two distinct algorithms for attribute coding: an encoder based on the region-adaptive hierarchical transform (RAHT) [6] and one based on level of detail (LoD).

The plenoptic function aims to represent the intensity of light at any given point by a 7-dimensional function

$$P(x, y, z, \theta, \phi, \lambda, t), \tag{1}$$

where (x, y, z) are space coordinates, (θ, ϕ) indicate the viewing direction, λ is the light wavelength and t is the time [7]. A plenoptic point cloud (PPC) can represent such a function by discretizing the parameters. The coordinates (x, y, z) are discretized into voxel positions, the wavelength λ is discretized into red-green-blue (RGB) color components and time t is discretized with the capture of successive frames. The azimuth and elevation angles (θ, ϕ) are naturally sampled by a finite number of camera rigs in the capture process, and the intermediate values for these parameters can be interpolated by the renderer. Hence, a dynamic point cloud with N_c color attributes per voxel — where N_c represents the number of camera rigs — can represent the plenoptic function and it is referred here as a PPC.

Point cloud compression is an active research field [3] and the representation of the plenoptic function has been the subject of MPEG and JPEG's ongoing standardization activities [8], [9]. Sandri et al. proposed the compression of PPCs by several approaches using the RAHT coder [10], [11]. The results have shown that the method using the Karhunen-Loève transform (KLT) over the voxel colors, followed by the use of RAHT over the transformed attributes, achieved the best performance at the time. Later on, Krivokuća et al. proposed subdividing the PPC into clusters of specular and diffuse components prior to KLT and RAHT [12]. Zhang et al. model the PPC from a different perspective, continuously representing the color function over (θ, ϕ) and compressing the spherical functions [13]. Naik et al. proposed a solution to handle PPC data with MPEG's video-based point cloud compression (V-PCC) standard [14]. Afterwards, they proposed optimizations to this scheme by discarding some of the views based on the voxels' specularity [15]. Li et al. proposed a video-based solution using V-PCC by compressing the multiple attributes via the Multiview High Efficiency Video Coding (MV-HEVC) [16]. Finally, Krivokuća et al. [17] introduces a representation of PPCs as sets of 6-D spatio-angular locations, providing an occlusion-aware compression method for the plenoptic attributes.

II. Proposed Framework

Consider a PPC where the n-th occupied voxel has a set of N_c RGB color values associated to the different camera viewpoints used in the capture process. Assume the RGB values are converted to YUV space, yielding plenoptic color vectors $[Y_n^1, U_n^1, V_n^1, \cdots, Y_n^{N_c}, U_n^{N_c}, V_n^{N_c}]$. This is due to the

Work partially supported by CNPq under grant 301647/2018-6

fact that the camera color variation is most prominent in the luminance component, which best reflects the changes in specularity. Thus, the smaller variance in the chrominance channels is a property that can be explored to enhance the compression. Hence, we apply a linear transform over each of the color components Y, U and V in order to take advantage of the correlation between the N_c plenoptic views. Consider C as one of the color channels and let

$$\mathbf{c}(n) = [C_n^1, C_n^2, C_n^3, \cdots, C_n^{N_c}]^T \tag{2}$$

represent the color component for the n-th voxel as viewed by the N_c camera rigs. For this approach, the KLT is applied over the $\mathbf{c}(n)$ vector signal as

$$\mathbf{s}(n) = \mathbf{H}_c \mathbf{c}(n), \tag{3}$$

where $\mathbf{s}(n) = [S_n^1, S_n^2, \cdots, S_n^{N_c}]^T$ is the vector with the transformed coefficients for the n-th voxel. $\mathbf{H}_c$ is the $N_c \times N_c$ KLT matrix of the color channel *C*, i.e. an orthogonal matrix made of the eigenvectors of the covariance matrix $\mathbf{K}_{cc}$ whose entries are $K_{ij} = \{E[(C_n^i - E[C_n^i])(C_n^j - E[C_n^j])]\}$. Each of the S_n^k is an attribute of the n-th transformed voxel and the set of $\{S_n^k\}$ is a PC to be transformed and encoded. The set $\{S_n^1\}$ is often called the DC coefficient PC. The others ($\{S_n^{k>1}\}$) are referred as AC PCs. The AC coefficients may contain negative values and, because of that, they are made positive by adding an offset. Moreover, since we allocate distortion instead of rate, all KLT channels for each of YUV are properly scaled and subject to the same QP. Therefore, the impact of our solution on the overall bit-rate consists of the transmission of both the DC and the AC coefficients, in addition to the point cloud's main color. The coefficients of $\mathbf{K}_{cc}$ can be incorporated in the high-level syntax of G-PCC's bitstream, as well as a single-bit flag signaling the encoding with the plenoptic enhancement. These coefficients can be conveyed in the bitstream using 32 bits floating-point numbers. Since the covariance matrix is symmetrical, $\frac{N_c(N_c+1)}{2}$ coefficients for each color channel have to be transmitted. According to our experiments, this side information corresponds to less than 0.07% of the total bit-rate.

We propose to encode the $\{S_n^k\}$, which are still redundant signals, using state-of-the-art attribute coding schemes such as RAHT or LoD, which are used in MPEG G-PCC. The operations and functionality of those point cloud compression (PCC) methods (RAHT and LoD) can be found elsewhere [3], [5], [18]. In summary, however, the main advantage of using these methods for encoding the coefficients over what is used in [11] stems from the ability to provide an intra-frame prediction feature that explores attribute correlation among neighboring voxels, achieving further bit-rate savings.

In essence, we run N_c attribute coders in parallel as depicted in Figs. 1 and 2. Figure 1 illustrates the encoder based on RAHT, wherein the N_c KLT output signals are fed to the multiple encoders. Note that it is important, however, that we use intra-frame-prediction within RAHT [19], [20] in order to further remove redundancy and to improve compression by exploiting the correlation between neighboring voxels. We have tested the use of intra-frame-prediction in both the DC and AC channels and the results consistently pointed to a superior performance when using the prediction, which is described in [19]. A very similar approach based on LoD is illustrated in Fig. 2 and is referred as Proposed Method 2. In Figs. 1 and 2, we used the G-PCC entropy coder.

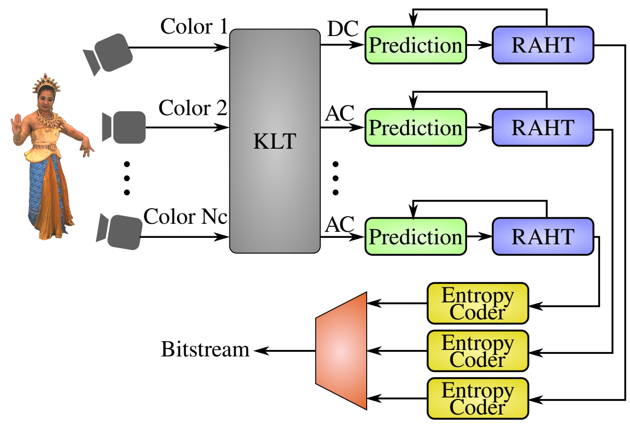

Fig. 1. Compression scheme for the intra-frame-predicted RAHT (proposed method 1).

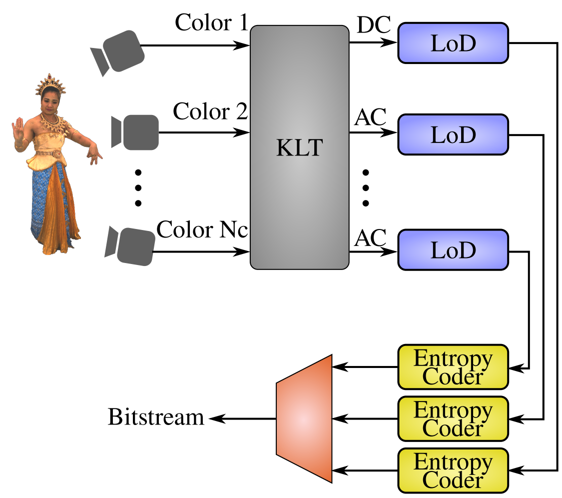

Fig. 2. Compression scheme using LoD (proposed method 2).

III. Evaluation

The experiments in this work were carried with version 10.0 of the G-PCC reference software (Test Model Categories 1 and 3 or TMC13) [5]. Both our proposed coding schemes were compared to the works of Sandri et al. [11] and Li et al. (MV-HEVC) [16]. We compared our solutions to [16], which is a state-of-the-art but not compliant with V-PCC. Other methods that were compliant with either V-PCC or G-PCC did not prove competitive. Ours, however, is competitive and compliant with MPEG's TMC13 (G-PCC).

We ran the experiments with the publicly available 8i Voxelized Surface Light Field Dataset (8iVSLF) [21]. These are the sequences being considered for the common test conditions (CTC) with respect to the MPEG efforts for plenoptic point cloud compression [22], and were the basis for the benchmark experiments due to compliance with MPEG's standards. The creation of new datasets which explore additional characteristics of the PPCs for a more thorough evaluation is something that we are aware of, and it is regarded as future work. Each PPC from this set has N_c RGB color values that are associated with each different camera viewpoint, and the geometry information of the voxels is constrained within a cube of $4096 \times 4096 \times 4096$ voxels, whose resolution is often referred to as a PC of "depth-12". However, downsampled versions of the PPCs from this dataset have also been generated which, among other advantages, provide reduced processing time and reduced memory consumption [23]. Therefore, tests for our proposed methods and the method in [11] were also performed with "depth-10" versions of those sequences. Details of the PPCs are provided in Table I.

TABLE I
CHARACTERISTICS OF THE 8IVSLF DATASET

Sequence	Voxels		N_c
	Depth 12	Depth 10	
Boxer	3493085	995099	13
Longdress	3096122	912518	12
Loot	3017285	869565	13
Redandblack	2770567	839315	12
Soldier	4001754	1193515	13
Thaidancer	3130215	283206	13

We calculate the distortion by considering the peak signal-to-noise ratio (PSNR) in between original and reconstructed Y channels for the N_c cameras either concatenated as a single signal or taking the average across all cameras, in accordance to [16]. We use the former method except when comparing to MV-HEVC, in which case we use the latter.

First, we compare our proposed methods 1 and 2. Every PPC from the 8iVSLF dataset comes with a main RGB channel in addition to the N_c colors from the different camera viewpoints, which is generated either by a weighted average of the N_c colors or by taking the frontal view of the PC [24]. Hence, we applied both coders over the non-transformed main color information in order to assess their performances in addition to our proposed compression schemes. In order to adequately compare the plenoptic results with the ones using the main color, both the DC and ACs coefficients in our proposed solution are scaled to fit an 8-bit representation. Table II presents the comparison results for the six depth-10 8iVSLF sequences. Bjøntegaard-delta PSNR (BD-PSNR) [25] results with the main color show an even comparison between coders. However, plenoptic results for the proposed method 1 present an average BD-PSNR value of 0.45 dB over the method 2.

In order to investigate where the gains come from, we broke down the results of the KLT-transformed color vectors by the rate-distortion (RD) performance of the individual DC and AC coefficient PCs. Some of the results are shown in Table II, in which the DC yields an average BD-PSNR gain of 0.11 dB

TABLE II
BD-PSNR (IN DB) FOR THE Y COMPONENT OF DEPTH-10 PCS, COMPARING PROPOSED METHOD 1 OVER 2.

PC	Main	Plenoptic	DC	AC 1	AC2
Boxer	−0.02	0.27	0.02	0.29	0.28
Longdress	0.10	0.72	0.26	0.26	0.33
Loot	−0.30	0.24	0.02	0.28	0.32
Redandblack	0.07	0.29	0.05	0.17	0.21
Soldier	0.02	0.34	−0.07	0.17	0.20
Thaidancer	0.04	0.81	0.37	0.21	0.22
Average	−0.02	0.45	0.11	0.23	0.26

for the method 1 over the proposed method 2, and the first and second ACs yield higher gains of 0.23 dB and 0.26 dB, respectively. We believe that the prediction model in predictive RAHT performs better than LoD's scheme for the transformed color information due to the selection of neighboring voxels. The LoD generation defines the order in which the colors are encoded. This order establishes which attribute values are available as references for prediction, which is based on the k-nearest-neighbors algorithm and uses point-to-point Euclidean distance thresholds. Therefore, the encoding order for this approach may not be optimal when using the transformed attributes. Hence, the lower correlation between neighbors for the AC-coefficient PCs may produce larger residuals in method 2 in comparison to method 1. This may lead to worse coding performance over the scaled plenoptic colors. The usage of unscaled coefficients for both methods further accentuates these differences, as can be seen in Table III.

Table III presents the BD-PSNR [25] comparing our proposed methods 1 and 2 and MV-HEVC, using [11] as an anchor. The BD metrics were computed using the points within the bit-rate range that comprises RAHT's quantization stepsizes from 12 to 300. The proposed method 1 significantly outperforms the others for most point clouds. Figs. 3 and 4 present RD curves comparing the four methods for the depth-12 *Thaidancer* and *Boxer* PCs. Results show greater rate-distortion gains for method 1 in comparison to MV-HEVC for medium and high bit-rate cases.

TABLE III
BD-PSNR(IN DB) COMPARISONS OF THE Y COMPONENT FOR DIFFERENT METHODS, USING [11] AS REFERENCE.

Sequence	Depth 12			Depth 10	
	MV-HEVC	Method 1	Method 2	Method 1	Method 2
Boxer	0.55	**1.78**	0.69	**1.83**	0.95
Longdress	3.40	**3.42**	2.17	**2.69**	1.79
Loot	1.88	**2.17**	1.11	**1.82**	0.99
Redandblack	1.66	**2.73**	1.66	**2.36**	1.58
Soldier	2.13	**2.53**	1.54	**1.99**	1.14
Thaidancer	**3.37**	3.33	1.86	**2.46**	1.27

Table III also presents the BD-PSNR results for both our proposed solutions compared to [11] for the depth-10 PPCs. As is the case with the depth-12 sequences, method 1 offers substantial coding gains in comparison to the method in [11]. Fig. 5 presents the RD performance for the *Thaidancer* sequence, which exhibits the same pattern found with the

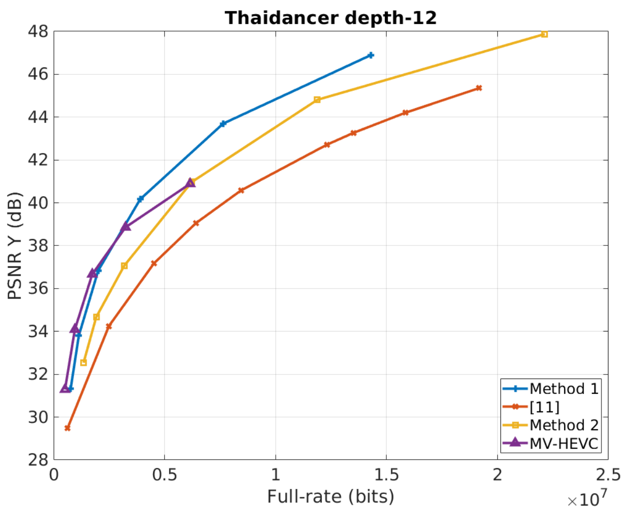

Fig. 3. Rate-distortion performance over the depth-12 *Thaidancer* for the four compression schemes considered.

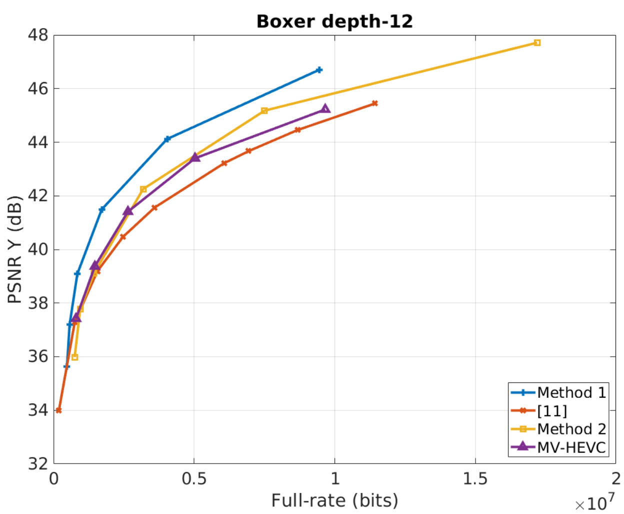

Fig. 4. Rate-distortion performance over the depth-12 *Boxer* for the four compression schemes considered.

other PPCs tested. Please note that MV-HEVC results are only available for depth-12 PPCs.

IV. Conclusions

This work proposes the incorporation of plenoptic capability for MPEG's geometry-based encoder (G-PCC). This is achieved by compressing plenoptic point clouds using a combination of the KLT over the color vector of the different camera viewpoints followed by the usage of G-PCC's attribute coders based on RAHT and LoD. Results show that the RAHT-based coding scheme (proposed method 1) provides substantial coding gains in comparison to competing approaches for both depth-12 and depth-10 PPCs. It even surpasses a V-PCC-based approach, making this proposed solution the state of the art in the compression of PPCs.

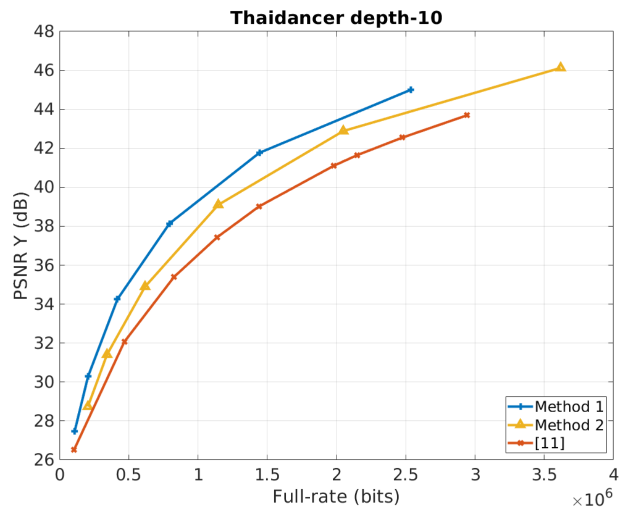

Fig. 5. Rate-distortion performance over the depth-10 *Thaidancer* for the methods 1 and 2 compared to [11].

References

[1] C. Tulvan, R. Mekuria, Z. Li, and S. Laserre, "Use Cases for Point Cloud Compression (PCC)," ISO/IEC JTC1/SC29/WG11 MPEG, document N16331, Jun. 2016.

[2] R. L. de Queiroz and P. A. Chou, "Motion-Compensated Compression of Dynamic Voxelized Point Clouds," *IEEE Transactions on Image Processing*, vol. 26, no. 8, pp. 3886–3895, 2017.

[3] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li, J. Llach, K. Mammou, R. Mekuria, O. Nakagami, E. Siahaan, A. Tabatabai, A. M. Tourapis, and V. Zakharchenko, "Emerging MPEG Standards for Point Cloud Compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2019.

[4] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: video-based (v-pcc) and geometry-based (g-pcc)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, pp. e13, 2020.

[5] "G-PCC Codec Description v10," ISO/IEC JTC1/SC29/WG11 MPEG, document N19331, Jun. 2020.

[6] R. L. de Queiroz and P. A. Chou, "Compression of 3D Point Clouds Using a Region-Adaptive Hierarchical Transform," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3947–3956, 2016.

[7] M. Landy and J. A. Movshon, "The plenoptic function and the elements of early vision," in *Computational Models of Visual Processing*, pp. 3–20. 1991.

[8] T. Ebrahimi, S. Foessel, F. Pereira, and P. Schelkens, "JPEG Pleno: Toward an Efficient Representation of Visual Reality," *IEEE MultiMedia*, vol. 23, no. 4, pp. 14–20, 2016.

[9] M. Domański, O. Stankiewicz, K. Wegner, and T. Grajek, "Immersive visual media — MPEG-I: 360 video, virtual navigation and beyond," in *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2017, pp. 1–9.

[10] G. Sandri, R. De Queiroz, and P. A. Chou, "Compression of plenoptic point clouds using the region-adaptive hierarchical transform," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 1153–1157.

[11] G. Sandri, R. L. de Queiroz, and P. A. Chou, "Compression of Plenoptic Point Clouds," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1419–1427, 2019.

[12] M. Krivokuća and C. Guillemot, "Colour compression of plenoptic point clouds using raht-klt with prior colour clustering and specular/diffuse component separation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 1978–1982.

[13] X. Zhang, P. A. Chou, M. Sun, M. Tang, S. Wang, S. Ma, and W. Gao, "A Framework for Surface Light Field Compression," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 2595–2599.
[14] D. Naik and S. Schwarz, "[V-PCC] CE2.15 report on Attribute Coding (SLF)," ISO/IEC JTC1/SC29/WG11 MPEG, input document M49123, Jul. 2019.
[15] D. Naik, S. Schwarz, V. K. M. Vadakita, and K. Roimela, "Surface Lightfield Support in Video-based Point Cloud Coding," in *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, 2020, pp. 1–6.
[16] L. Li, Z. Li, S. Liu, and H. Li, "Video-Based Compression for Plenoptic Point Clouds," in *2020 Data Compression Conference (DCC)*, 2020, pp. 378–378.
[17] M. Krivokuca, E. Miandji, C. Guillemot, and P. Chou, "Compression of Plenoptic Point Cloud Attributes Using 6-D Point Clouds and 6-D Transforms," *IEEE Transactions on Multimedia*, pp. 1–1, 2021.
[18] K. Mammou, A. Tourapis, J. Kim, F. Robinet, V. Valentin, and Y. Su, "Lifting Scheme for Lossy Attribute Encoding in TMC1," ISO/IEC JTC1/SC29/WG11 MPEG, input document M42640, Apr. 2018.
[19] S. Lasserre and D. Flynn, "[G-PCC] On an improvement of RAHT to exploit attribute correlation," ISO/IEC JTC1/SC29/WG11 MPEG, input document M47378, Jul. 2019.
[20] A. L. Souto and R. L. de Queiroz, "On predictive raht for dynamic point cloud coding," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 2701–2705.
[21] "JPEG Pleno Database," https://jpeg.org/plenodb/, [Online].
[22] R. L. de Queiroz, C. Dorea, D. C. Garcia, R. U. Ferreira, D. Freitas, R. Higa, I. Seidel, and V. Testoni, "Discussion on Common Test Conditions for plenoptic V-PCC ," ISO/IEC JTC1/SC29 Joint WG11/WG7 (MPEG/JPEG), input document WG7M55146, Oct. 2020.
[23] C. Dorea, D. C. Garcia, R. U. Ferreira, D. Freitas, R. Higa, R. L. de Queiroz, I. Seidel, and V. Testoni, "Downsampled 8iVSLF Dataset," ISO/IEC JTC1/SC29 Joint WG11/WG7 (MPEG/JPEG), input document WG7M55148, Oct. 2020.
[24] M. Krivokuća, P. A. Chou, and P. Savill, "8i Voxelized Surface Light Field (8iVSLF) Dataset," ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG), input document M42914, Jul. 2018.
[25] G. Bjøntegaard, "Improvements of the BD-PSNR Model," Tech. Rep., VCEG-AI11, ITU-T SG16/Q6, Berlin, Germany, Jul. 2008.