POINT CLOUD ATTRIBUTE COMPRESSION VIA CHROMA SUBSAMPLING

Shashank N. Sridhara[⋆]*, Eduardo Pavez*[⋆]*, Antonio Ortega*[⋆]*, Ryosuke Watanabe*[†]*, Keisuke Nonaka*[†]

[⋆]University of Southern California, Los Angeles, CA, USA
[†]KDDI Research, Inc.

ABSTRACT

We introduce chroma subsampling for 3D point cloud attribute compression by proposing a novel technique to sample points irregularly placed in 3D space. While most current video compression standards use chroma subsampling, these chroma subsampling methods cannot be directly applied to 3D point clouds, given their irregularity and sparsity. In this work, we develop a framework to incorporate chroma subsampling into geometry-based point cloud encoders, such as region adaptive hierarchical transform (RAHT) and region adaptive graph Fourier transform (RAGFT). We propose different sampling patterns on a regular 3D grid to sample the points at different rates. We use a simple graph-based nearest neighbor interpolation technique to reconstruct the full resolution point cloud at the decoder end. Experimental results demonstrate that our proposed method provides significant coding gains with negligible impact on the reconstruction quality. For some sequences, we observe a bitrate reduction of 10-15% under the Bjontegaard metric. More generally, perceptual masking makes it possible to achieve larger bitrate reductions without visible changes in quality.

Index Terms— chroma subsampling, compression, interpolation, attributes.

1. INTRODUCTION

Thanks to advances in sensor technology and faster acquisition processes, point clouds are becoming increasingly popular means to represent 3D scenes. Therefore, compression of point clouds has received considerable attention and the Moving Picture Experts Group (MPEG) is currently developing compression standards specifically for point clouds [1]. A 3D point cloud is a list of unordered points in $\mathbb{R}^{N\times 3}$, $X = \{(x_i, y_i, z_i)\}$, along with their corresponding attributes $\{(r_i, g_i, b_i)\}$. Attributes can be color, intensity or surface normals. In order to effectively store and transmit point clouds, both geometry and attributes have to be compressed. In this work, we focus on compression of point cloud color attributes.

Geometry-based point cloud compression (G-PCC) and video-based point cloud compression (V-PCC) are the two approaches standarized by MPEG. G-PCC methods exploit geometry information to encode the associated attributes, while V-PCC is a projection-based method wherein 3D points are projected onto a 2D image so that the existing image/video encoding methods can be used [2]. The region adaptive hierarchical transform (RAHT) [3], a multi-resolution orthogonal transform, has been evaluated as a G-PCC codec by MPEG [1, 2]. In addition to RAHT, various alternative transforms have been proposed for attribute compression [4, 5, 6, 7].

For color attribute compression, the RGB values are first transformed to a YUV space where the Y, U and V components are encoded. Given that the human visual system is more sensitive to the changes in luminance (Y) than to the changes in chrominance (U and V) [8, 9], in image and video coding chroma components are often encoded at a lower resolution than luminance components, without significant perceptual degradation [10, 11, 12]. Most current video compression standards, including H.264/AVC and H.265/HEVC, provide different sampling rates for the luminance and chrominance components [13]. These perceptual observations are also valid for point cloud color attributes [14, 15].

On regular 2D grids, sampling at different rates can be performed by selecting pixels using specific regular patterns in both horizontal and vertical directions [16]. As a result, V-PCC methods can leverage directly the effectiveness of chroma subsampling for video by sampling pixels after they have been projected onto a 2D image. However, G-PCC methods compress the attributes by using the geometry information and the irregular placement of points in 3D space means that chroma subsampling is not as straightforward as for 2D images. In fact, to the best of our knowledge, none of the existing G-PCC methods use chroma subsampling.

In this paper, we propose a new approach to leverage the perceptual redundancies and smoothness of the chroma signals on the point cloud in order to improve the performance of G-PCC methods for attribute compression. We also propose low complexity sampling patterns for different sampling rates. Our approach is based on intersecting a regular sampling pattern on the complete 3D grid with the point cloud geometry. The proposed 3D sampling method is fast and can be implemented with few bitwise operators. We use a simple graph based filter for interpolation, which can be implemented by K-nearest neighbor graph construction and sparse matrix-vector product. Thus, the proposed sampling method has a low complexity overhead for G-PCC encoders while providing significant gains.

The rest of the paper is organized as follows. Section 2 describes proposed sampling and interpolation methods. Experimental results and complexity analysis are shown in Section 3 and we conclude the paper in Section 4.

2. PROPOSED METHOD

2.1. Chroma subsampling in Images/Videos

In chroma subsampling the chroma signals (U and V) are encoded at a lower resolution to reduce bitrate, which allows more bandwidth to be allocated to the luma signal. These approaches are motivated by perceptual studies showing that the effect of chroma subsampling is masked by the presence of full resolution luminance [14, 9].

Chroma subsampling is used in most video coding standards including the recent Versatile Video Coding (VVC) [13]. The main components of chroma subsampling are, sampling and interpolation. There are different schemes to sample the chroma components on a 2D grid as shown in Table 1. The most commonly used for-

This work was funded in part by KDDI Research, Inc. and by the National Science Foundation (NSF CNS-1956190).

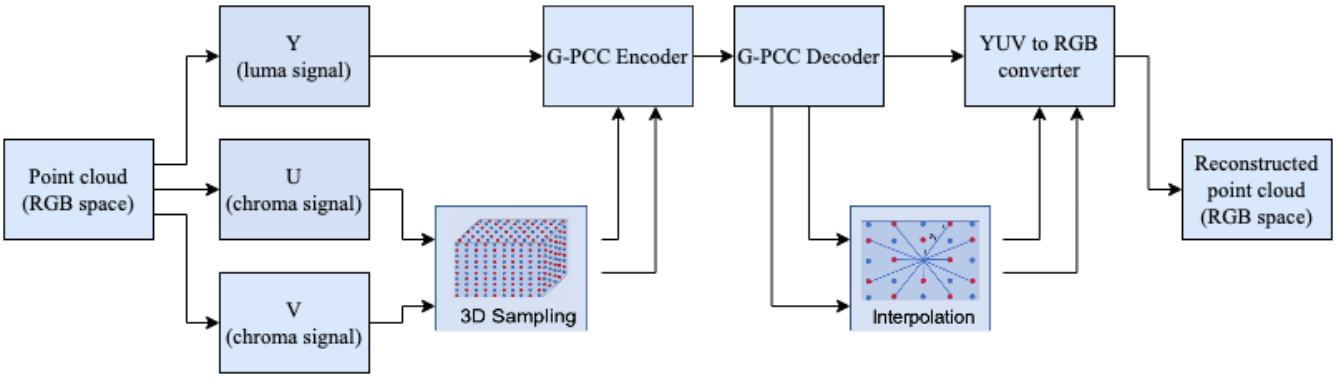


Fig. 1: Point cloud attribute compression pipeline using G-PCC encoders with chroma subsampling. The luminance signal (Y) will be directly passed to the G-PCC encoder, whereas chrominance signals (U and V) will be downsampled at a given sampling rate before encoding. At the decoder side, we interpolate the downsampled chrominance signal values and reconstruct the attributes of the full resolution point cloud.

mat is 4:2:0, where the horizontal and vertical resolutions of both the chroma components are halved. In all these sampling formats, 2×2 blocks are considered, where either the upper-left U and V component of the block are selected, or the sampled U and V component are obtained by averaging all the U,V components of a 2×2 block. Note that interpolation/upsampling is carried out by copying the same sampled value to all the 4 respective locations of a 2×2 U, V block. A more sophisticated and widely used method is bilinear interpolation, where upsampling is done by computing weighted average of nearest pixels.

subsampling scheme	luma resolution		chroma resolution	
	V	H	V	H
4:4:4	Full	Full	Full	Full
4:2:2	Full	Full	Full	1/2
4:2:0	Full	Full	1/2	1/2

Table 1: Different chroma subsampling schemes in images/videos, where V and H correspond to vertical and horizontal resolutions, respectively.

2.2. Challenges in point cloud chroma subsampling

Unlike in images, defining a sampling pattern and formats for different sampling rates is not straightforward for point clouds. The main challenges with point clouds are that 1) points are irregularly placed in 3D space, 2) there is a lack of spatial correlation, and 3) there is variation in spatial point density. Because of these challenges, classical image and video processing techniques cannot be applied directly to point clouds. In order to use chroma subsampling for G-PCC encoders, we have to design sampling and interpolation techniques for 3D points with desirable properties, including computational efficiency. The complete pipeline of chroma subsampling for G-PCC encoders is shown in Figure 1.

2.3. Proposed sampling

We propose a computationally fast sampling method for irregular points in 3D. Let $V = \{v_i = (x_i, y_i, z_i)\}, \forall i \in \mathbb{R}^{N\times 3}$ represent the 3D coordinates of N points. We assume that the points are voxelized and represented as integer coordinates. We define a sampling pattern on a regular 3D grid for a particular sampling rate. Then, the defined sampling pattern is overlapped onto the irregular 3D points of the point cloud to generate the sampling set. For a sampling rate of 0.5, we sample every second point in the x, y and z directions. Similarly, for a sampling rate of 0.33 and 0.25, we sample every third and fourth point respectively in the x, y and z directions. The sampling patterns for 0.5, 033 and 0.25 sampling rates are shown in Figure 2a, 2b and 2c, respectively. To mathematically formulate the sampling rule, let us define a function $f_k(v_i)$ such that,

$$f_k(v_i) = \begin{cases} 1, & \text{if } x_i + y_i + z_i = 0 \ (\mathrm{mod} k) \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

The function $f_k(v_i)$ maps every point in V to $\{0, 1\}$ i.e., $f_k : V \rightarrow \{0, 1\}$. Therefore, the sampling set can be defined as,

$$S = \{i : f_k(v_i) = 1\}, \tag{2}$$

and the sampling rate for $f_k(\cdot)$ is $1/k$. The proposed sampling method has a low overhead on G-PCC encoders since it can be implemented with $\mathcal{O}(N)$ operations.

2.4. Proposed interpolation

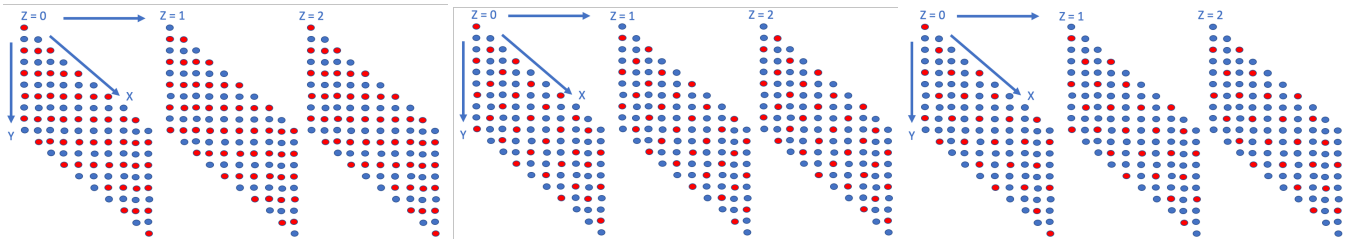
After the sampled points of chroma signal are encoded, we use interpolation to reconstruct chroma signal at the decoder end. Note that the full resolution geometry has to be encoded, since it is needed to represent the full resolution luma information. In our comparisons we assume that the point cloud geometry is available at both the encoder and decoder side and focus only on the cost of encoding attributes. We use graph based filters with K-nearest neighbor (K-NN) graph construction for interpolation.

Let $\hat{U}_s = \{u_i, \forall i \in S\}$ be the low resolution, reconstructed chroma signal and $U_{s^c} = \{\tilde{u_j}, \forall j \in S^c\}$ be the chroma values to be interpolated. The interpolated value at node $j \in S^c$ is,

$$\tilde{u_j} = \frac{1}{d_j} \sum_{i \in \mathbb{N}(j)} w_{ij} \hat{u_i}; \tag{3}$$

where the sum is over K-nearest neighbors of node j and $d_i = \sum_j w_{ij}$ is the degree of node j. We use a Gaussian weighting function for K-NN graph construction ,

$$w_{ij} = e^{-\frac{\|v_i - v_j\|_2^2}{2\sigma^2}}, \tag{4}$$



(a) 50% sampling (b) 33% sampling (c) 25% sampling

Fig. 2: Sampling pattern for different sampling rates on a 3D grid. Red points represents the sampling set. The attributes of blue points should be interpolated using the attributes of red point at the decoder end.

where $\sigma = (\sum_{i,j} \|v_i - v_j\|_2^2)/N$, $j \in S^c, i \in \mathbb{N}(j)$ and $\mathbb{N}(j) \subset S$, v_i, v_j are the 3D coordinates of points i and j respectively. As the distance between the points in S and S^c change for different sampling rates, we choose mean distance between the points in these two sets for the parameter σ. The proposed interpolation method can be implemented by a sparse matrix-vector product.

$$\begin{pmatrix} \tilde{u_1} \\ \tilde{u_2} \\ \vdots \\ \tilde{u_{|S^c|}} \end{pmatrix} = \begin{pmatrix} w_{11} & \dots & w_{1|S|} \\ w_{21} & \dots & w_{2|S|} \\ \vdots & \dots & \vdots \\ w_{|S^c|1} & \dots & w_{|S^c||S|} \end{pmatrix} \begin{pmatrix} \hat{u_1} \\ \hat{u_2} \\ \vdots \\ \hat{u_{|S|}} \end{pmatrix}. \quad (5)$$

The chroma signal of the full resolution point cloud can be obtained by combining the interpolated and decoded chroma signal,

$$\hat{U} = \{\hat{u_i} \cup \tilde{u_j}, \forall i \in S \text{ and } j \in S^c\}. \quad (6)$$

The proposed interpolation method can be implemented with $\mathcal{O}(KNlog(N))$ operations for graph construction and $\mathcal{O}(KN)$ operations for matrix-vector multiplication.

3. EXPERIMENTS

In this section, we present an end-to-end evaluation of the effectiveness of chroma subsampling on attribute compression for two G-PCC encoders – RAHT and RAGFT. We evaluate the proposed method on the "8iVFBv2" point cloud dataset[1] [17]. The dataset consists of four sequences: *longdress, redandblack, soldier and loot.* We compare the performance of the proposed method against RAHT and RAGFT without chroma subsampling. We use the Bjontegaard metric [18] to compare the rate-distortion curves of attribute coding with and without chroma subsampling. We report results of chroma subsampling at a sampling rate of 50%. Even though the preliminary results for 25% and 33% sampling rates are promising, due to lack of space we are not including them, as they also require a more thorough evaluation of perceptual quality of the reconstructed point cloud, which is left for future work.

In all the experiments, we perform uniform quantization and entropy code the coefficients using the adaptive run-length Golomb-Rice algorithm (RLGR) [19]. For the RAGFT encoder, the block size of 16 was used in all the experiments [5]. We use $K = 10$ for K-NN interpolation. The distortion PSNR for the attributes is calculated in both the RGB and the YUV space. In the RGB space, we use pixel-wise PSNR which is given by,

$$\text{PSNR}_{rgb} = -10 \log_{10} (\text{MSE}_{rgb}) \quad (7)$$

[1] http://plenodb.jpeg.org/pc/8ilabs/

where,

$$\text{MSE}_{rgb} = (\|R - \hat{R}\|_2^2 + \|G - \hat{G}\|_2^2 + \|B - \hat{B}\|_2^2)/(3N255^2)$$

and N is the total number of points in the full resolution point cloud.

Along with the pixel-wise PSNR value in RGB space, we use color-sensitive-based combined PSNR (CS-PSNR), a video quality assessment metric which takes into account the sensitivity of the human visual system to different color components [20],

$$\text{CS-PSNR} = -10 \log_{10} (P_Y \text{MSE}_Y + P_U \text{MSE}_U + P_V \text{MSE}_V), \quad (8)$$

where $P_Y = 0.695, P_U = 0.130, P_V = 0.175$ are the weighting coefficients. The total rate is reported in bits per point [bpp] $B = (B_Y + B_U + B_V)/N$ where, B_Y, B_U and B_V represent the bits required to encode Y, U and V components respectively.

3.1. Results of attribute compression

The rate-distortion curves for four point cloud sequences in the RGB space and the YUV space are shown in Figure 3 and Figure 4 respectively. There is a considerable gain in attribute coding using chroma subsampling for both RAHT and RAGFT encoders. Tables 2 and 3 show average PSNR gain and percentage bitrate saving for RAHT and RAGFT respectively at a sampling rate of 50%. The bitrate savings for *longdress* and *redandblack* sequences is around 10-12%, and for *soldier* and *loot* sequences it is around 2-3%. From our own qualitative comparison of reconstructed point clouds with and without chroma subsampling in Figure 5, we have observed that these significant reductions in bitrate for attribute coding are achieved with little or no impact on perceived quality of the reconstructed point cloud.

sequence	RGB space		YUV space	
	avg. psnr gain(dB)	bitrate saving(bpp)	avg. psnr gain(dB)	bitrate saving(bpp)
longdress	0.59	9.61%	0.85	14.24%
redblack	0.56	11.09 %	0.81	17.12%
soldier	0.09	1.9%	0.15	3.0%
loot	0.1	3.13%	0.18	5.63%

Table 2: Average PSNR gain and percentage bitrate savings in RGB and YUV space using the RAHT encoder at a sampling rate of 50%

3.2. Complexity analysis

We compare the run-time of both RAHT and RAGFT, with and without chroma subsampling. We performed experiments on the first

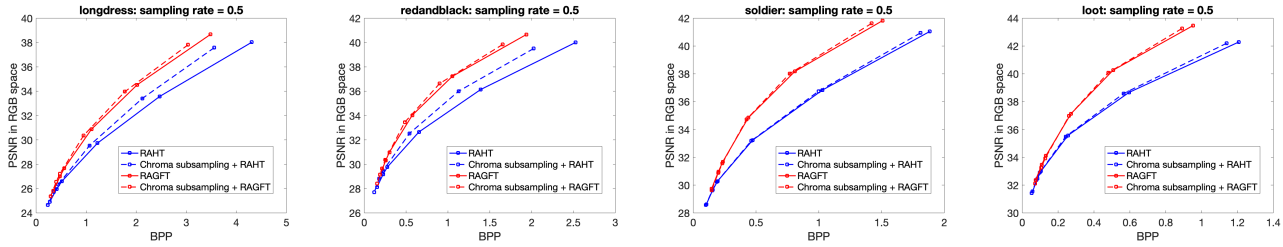


Fig. 3: Distortion rate curves for color attribute compression in RGB space

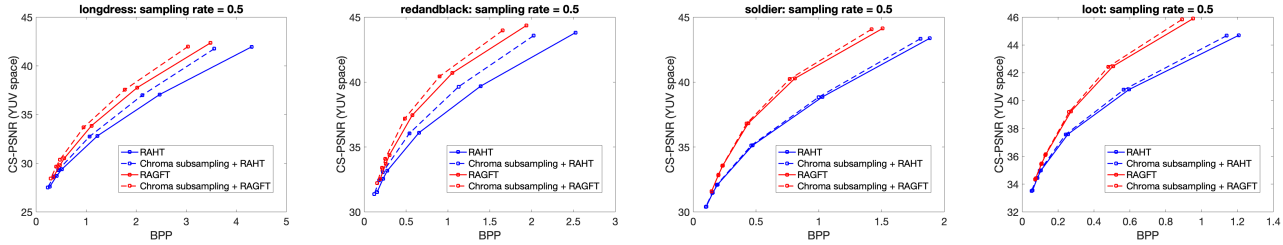


Fig. 4: Distortion rate curves for color attribute compression in YUV space using CS-PSNR

(a) Original (b) RAHT: 2.46bpp (c) RAGFT: 2.01bpp (d) S+RAHT: 2.12bpp (e) S+RAGFT: 1.77bpp

Fig. 5: Qualitative comparison of reconstructed point clouds. Figures 5b and 5c shows the reconstructed point clouds from RAHT and RAGFT without chroma subsampling. Figures 5d and 5e shows the reconstructed point cloud from RAHT and RAGFT respectively with chroma subsampling at a sampling rate of 50%.

sequence	RGB space		YUV space	
	avg. psnr gain(dB)	bitrate saving(bpp)	avg. psnr gain(dB)	bitrate saving(bpp)
longdress	0.23	4.23%	0.63	10.39%
redblack	0.13	2.64 %	0.53	10.70%
soldier	0.12	1.03%	0.26	3.25%
loot	0.05	1.45 %	0.21	2.74%

Table 3: Average PSNR gain and percentage bitrate savings in RGB and YUV space using the RAGFT encoder at a sampling rate of 50%

20 point cloud frames of *longdress* sequence and compute encoder and decoder run-times. From the results in Table 4 we observe a very small overhead at the encoder but considerable overhead at the decoder. This can be attributed to the interpolation process, which involves graph construction and filtering, whereas the sampling algorithm is implemented in $\mathcal{O}(N)$ complexity. However, we expect that the complexity can be reduced by constructing sparser graphs by limiting the number of nearest neighbors. We will address the issue of interpolation complexity in our future work.

time/frame (in seconds)	RAHT	RAHT + sampling	RAGFT	RAGFT+ sampling
encoding time	0.47	0.72	27.99	34.26
decoding time	0.42	2.53	26.75	35.61

Table 4: Comparison of encoding and decoding time (per frame) for RAHT and RAGFT encoders on 20 *longdress* point cloud frames.

4. CONCLUSION

In this work, we have presented chroma subsampling for 3D point cloud attribute compression in G-PCC based encoders. We have proposed a novel sampling method for 3D points at different sampling rates. Graph-based filters with K-nearest neighbors graph is used for interpolation. We evaluated, both qualitatively and quantitatively, the effectiveness of chroma subsampling on RAHT and RAGFT encoders. We observed significant bitrate savings with little or no impact on the perceived quality of reconstructed point cloud. We also performed complexity analysis and observed that the proposed sampling algorithm is fast and incurs low complexity overhead compared to interpolation at the decoder end as it involves graph construction.

5. REFERENCES

[1] Sebastian Schwarz, Marius Preda, Vittorio Baroncini, Madhukar Budagavi, Pablo Cesar, Philip A. Chou, Robert A. Cohen, Maja Krivokuća, Sébastien Lasserre, Zhu Li, Joan Llach, Khaled Mammou, Rufael Mekuria, Ohji Nakagami, Ernestasia Siahaan, Ali Tabatabai, Alexis M. Tourapis, and Vladyslav 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.

[2] 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.

[3] Ricardo L. de Queiroz and Philip 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.

[4] Cha Zhang, Dinei Florêncio, and Charles Loop, "Point cloud attribute compression with graph transform," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2066–2070.

[5] Eduardo Pavez, Benjamin Girault, Antonio Ortega, and Philip A. Chou, "Region adaptive graph fourier transform for 3d point clouds," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 2726–2730.

[6] Eduardo Pavez, André L. Souto, Ricardo L. De Queiroz, and Antonio Ortega, "Multi-resolution intra-predictive coding of 3d point cloud attributes," in *2021 IEEE International Conference on Image Processing (ICIP)*, 2021, pp. 3393–3397.

[7] Philip A Chou, Maxim Koroteev, and Maja Krivokuća, "A volumetric approach to point cloud compression—part I: Attribute compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 2203–2216, 2019.

[8] Li-Heng Chen, Christos G. Bampis, Zhi Li, Joel Sole, and Alan C. Bovik, "Perceptual video quality prediction emphasizing chroma distortions," *IEEE Transactions on Image Processing*, vol. 30, pp. 1408–1422, 2021.

[9] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[10] Douglas Issue, "Chrominance Subsampling in Digital Images," *The Pumpkin*, vol. 1, 2009.

[11] Emil Dumic, Mario Mustra, Sonja Grgic, and Goran Gvozden, "Image quality of 4:2:2 and 4:2:0 chroma subsampling formats," in *2009 International Symposium ELMAR*, 2009, pp. 19–24.

[12] Yongbing Zhang, Debin Zhao, Jian Zhang, Ruiqin Xiong, and Wen Gao, "Interpolation-dependent image downsampling," *IEEE Transactions on Image Processing*, vol. 20, no. 11, pp. 3291–3296, 2011.

[13] Benjamin Bross, Jianle Chen, Jens-Rainer Ohm, Gary J. Sullivan, and Ye-Kui Wang, "Developments in international video coding standardization after AVC, with an overview of versatile video coding (VVC)," *Proceedings of the IEEE*, vol. 109, no. 9, pp. 1463–1493, 2021.

[14] Ingmar Lissner, Jens Preiss, Philipp Urban, Matthias Scheller Lichtenauer, and Peter Zolliker, "Image-difference prediction: From grayscale to color," *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 435–446, 2013.

[15] Gabriel Meynet, Yana Nehmé, Julie Digne, and Guillaume Lavoué, "Pcqm: A full-reference quality metric for colored 3d point clouds," in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, 2020, pp. 1–6.

[16] Charles Poynton, "Chroma subsampling notation," *Retrieved June*, vol. 19, pp. 2004, 2002.

[17] Eugene d'Eon, Bob Harrison, Taos Myers, and Philip A Chou, "8i voxelized full bodies, version 2–a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m40059 M*, vol. 74006, 2017.

[18] Gisle Bjontegaard, "Calculation of average psnr differences between RD-curves," *VCEG-M33*, 2001.

[19] Henrique S. Malvar, "Adaptive run-length / golomb-rice encoding of quantized generalized gaussian sources with unknown statistics," in *Proceedings of the Data Compression Conference*, USA, 2006, DCC '06, p. 23–32, IEEE Computer Society.

[20] Xiwu Shang, Jie Liang, Guozhong Wang, Haiwu Zhao, Chengjia Wu, and Chang Lin, "Color-sensitivity-based combined PSNR for objective video quality assessment," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 5, pp. 1239–1250, 2019.