This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Distortion Propagation Model-Based V-PCC Rate Control for 3D Point Cloud Broadcasting

Zhanyuan Cai, Wenxu Gao, Ge Li, *Member, IEEE*, and Wei Gao, *Senior Member, IEEE*

***Abstract*—For efficient point cloud broadcasting, point cloud compression technologies serve as the foundation, which plays a crucial role in immersive media communication and streaming. Video-based point cloud compression (V-PCC) is the recently developed standard by the Moving Picture Experts Group (MPEG) for dynamic point clouds. Its original fixed-ratio bit allocation (FR-BA) method in the unique all intra (AI) structure leads to a significant rate-distortion performance gap between the rate control manner and the fixed quantization parameters (FixedQP) scheme, as evidenced by significant increases in BD-Rate (Bjøntegaard Delta Rate) for both geometry and attribute. To address this issue, we propose a distortion propagation model-based frame-level bit allocation method that is specifically tailored for AI structure in V-PCC. First, the analysis is carried out for the distortion propagation model inside the group of pictures (GOP) for the AI configuration. Second, the skip ratio of 4x4 minimum coding units (CUs) is utilized to predict the distortion propagation factor. Third, the occupancy information is employed to refine the distortion propagation model and further enhance compression performance. Finally, experimental results demonstrate the effectiveness of the proposed distortion propagation model-based frame-level bit allocation method. Specifically, experimental results reveal that the proposed method achieves BD-Rate reductions of 0.92% and 4.85% in geometry and attribute, respectively, compared to the FR-BA method. Furthermore, with the introduction of distortion propagation factor prediction incorporating occupancy correction, the BD-Rate reductions are further extended to 2.16% and 6.13% in geometry and attribute, respectively.**

***Index Terms*—3D point cloud broadcasting, distortion propagation model, point cloud communication, rate control, streaming.**

Received 10 June 2024; revised 7 November 2024; accepted 25 November 2024. This work was supported in part by The Major Key Project of PCL under Grant PCL2024A02; in part by the Natural Science Foundation of China under Grant 62271013 and Grant 62031013; in part by the Guangdong Province Pearl River Talent Program under Grant 2021QN020708; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515010155; in part by the Shenzhen Science and Technology Program under Grant JCYJ20230807120808017; and in part by the Shenzhen Fundamental Research Program under Grant GXWD20201231165807007-20200806163656003. *(Zhanyuan Cai and Wenxu Gao contributed equally to this work.) (Corresponding author: Wei Gao.)*

Zhanyuan Cai and Ge Li are with the Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology, School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Shenzhen 518055, China (e-mail: 1901213111@pku.edu.cn; geli@pku.edu.cn).

Wenxu Gao and Wei Gao are the Guangdong Provincial Key Laboratory of Ultra High Definition Immersive Media Technology, School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Shenzhen 518055, China, also with Department of Media and Interaction, Peng Cheng Laboratory, Shenzhen 518066, China (e-mail: gaowx@stu.pku.edu.cn; gaowei262@pku.edu.cn).

Digital Object Identifier 10.1109/TBC.2024.3511950

I. Introduction

POINT cloud refers to a collection of massive points in 3D space, where each point is characterized by 3D position coordinates and a set of scalar or vector attributes. As an extensively adopted data format for representing 3D models, point cloud has garnered significant research attention from scholars in the field of 3D vision. Correspondingly advanced technologies are witnessing rapid growth [1], [2], [3], [4], [5], [6], [7]. Owing to the capability in precisely depicting the geometry structure of real-world objects or scenes, point clouds inherently offer advantages in 3D reconstruction while finding extensive practical applications including autonomous driving [8], [9], [10], immersive interaction [11], [12], [13], and digital twin cities [14], [15], [16]. However, in comparison to traditional video streaming, the direct transmission or storage of point clouds necessitates substantial bandwidth and storage capacity, thereby posing challenges for real-world deployment. In order to tackle this issue and facilitate immersive communication and broadcasting as well as efficient point cloud broadcasting, the Moving Picture Experts Group Immersive Media Working Group (MPEG-I) is currently engaged in developing point cloud compression techniques that encompass three distinct categories: static objects and scenes, dynamic objects, and dynamic acquisition. Following the release of Point Cloud Compression (PCC) 1.0 by Omnidirectional Media Format (OMAF), the Point Cloud Compression Working Group of MPEG-I issued a call for proposals for PCC 2.0 version in 2017 [17]. MPEG-I Phase 2a has established two technical pathways to address the requirements for 6DoF, i.e., geometry-based point cloud compression standard (G-PCC) for static objects/scenes and dynamic acquisition point clouds, and video-based point cloud compression standard (V-PCC) for dynamic point clouds [18]. Although V-PCC demonstrates commendable compression performance [19], there exists a dearth of research focusing on achieving high rate-distortion performance for compressing point cloud data under transmission bandwidth constraints. Rate control is a highly significant technical module in V-PCC encoder, as it ensures that the bitrate of the point cloud stream adapt to the channel bandwidth and optimizes coding performance (including coding efficiency and perceptual quality) as much as possible. Therefore, it is imperative to develop an effective rate control algorithm tailored specifically for V-PCC in 3D point cloud broadcasting.

The V-PCC framework consists of two primary stages, i.e., projection and 2D video coding. In the projection stage, the 3D point cloud data is projected onto 2D projection planes [20]

1557-9611 © 2024 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.
Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

to derive occupancy map, geometry map, and attribute map, along with generating other auxiliary information. The geometry map represents the depth information by reflecting the distance of each point from the projection plane. Meanwhile, the attribute map captures color information as an indication of various attributes associated with each point. In the 2D video coding stage, the traditional 2D video coding software is employed to encode the aforementioned types of videos. In the process of V-PCC and Visual Volumetric Video-based Coding (V3C) standardization, a reference software named Test Model Category2 (TMC2) [21] was developed to evaluate the coding performance of corresponding visual media data forms. The encoder of TMC2 incorporates various techniques, including patch packing, generation of occupancy map, geometry map, attribute map and auxiliary information, along with corresponding compression processes to enhance coding efficiency. It should be noted that by default, the TMC2 reference software utilizes traditional video coding standard such as High Efficiency Video Coding (HEVC) [22] for 2D video compression, allowing for flexible choice of video codec. There are two common coding structures implemented in the V-PCC standard for projected videos, including all intra (AI) and random access (RA) coding structures. To tackle the issue of self-occlusion in 3D space, the thickness of the point cloud is taken into consideration at the projection stage, resulting in the generation of two 2D patches with high content similarity by projecting a 3D patch twice. Consequently, a near frame and a far frame are generated for the geometry and attribute maps, respectively, establishing a group of pictures (GOP) structure, which is the exclusive "IPIP…" structural characteristic of AI coding configuration in V-PCC. On the other hand, the RA coding structure in the V-PCC standard aligns with that employed by Versatile Video Coding (VVC) [23], utilizing forward and backward reference technology to efficiently predict high-level frames from low-level ones and eliminate temporal redundancy.

The primary challenge in V-PCC rate control lies in the well-known "chicken and egg" dilemma, which refers to the interdependence between the rate-distortion optimization process and the quantization parameters (QPs) determination for rate control. To alleviate this problem, the two-pass encoding approach can be employed to acquire model parameters through precoding. The precoding strategy enables accurate rate control for projected videos derived from different point cloud sequences. However, when using a fixed-ratio bit allocation (FR-BA) method within the AI configuration, there still exists a significant gap in rate distortion performance compared to the fixed quantization parameters (FixedQP) coding scheme with common test conditions, reflected in terrible and pronounced BD-Rate (Bjøntegaard Delta Rate) increases for both geometry and attribute. In light of this concern, this paper proposes a distortion propagation model-based frame-level bit allocation method (DPM-BA) for V-PCC. In addition, considering the different rate-distortion characteristics between occupied coding units (CUs) and unoccupied CUs in projected videos of V-PCC, we propose to utilize occupancy information to correct the distortion propagation model and derive an adjusted distortion propagation model-based bit allocation method (ADPM-BA).

The main contributions of this paper are as follows:

- We propose a distortion propagation model-based frame-level bit allocation method for V-PCC, which fully considers the structural characteristics of projected videos under V-PCC's all intra configuration. Theoretical deduction is provided for the frame-level bit allocation method based on distortion propagation model inside the GOP. The proposed method combines the distortion propagation factor from intra frame to inter frame within the GOP and the Lagrange multiplier method to determine the target bits for each frame.
- We propose that the skip ratio of minimum CUs is introduced to predict distortion propagation factor. Based on the occupancy information, the unoccupied CUs are excluded in calculating skip ratio. In this manner, the more precise distortion propagation factor is predicted.
- Experimental results demonstrate that our proposed distortion propagation model-based frame-level bit allocation method for V-PCC exhibits substantial performance improvements in terms of geometry and attribute performance compared to the original fixed-ratio bit allocation approach. Moreover, incorporating our proposed distortion propagation factor prediction scheme with occupancy information correction further enhances these gains.

The remaining sections of this paper are structured as follows. Section II provides a comprehensive overview of the existing related works on rate control for 2D video coding and V-PCC. Section III presents the research motivation for the bit allocation task specific to the AI coding structure in V-PCC. In Section IV, the frame-level bit allocation scheme based on distortion propagation model is proposed. Detailed experimental results and analysis are presented in Section V. Finally, concluding remarks are provided in Section VI.

II. Related Work

A. Rate Control for Video Coding

The rate control process can be divided into two main phases: bit allocation and bitrate control. The bit allocation phase employs the rate-distortion optimization criterion to allocate appropriate bits across various coding levels, including GOP level, frame level, and finer block level such as coding tree units (CTU), in order to achieve optimal overall coding performance. Once the bit allocation is completed at each coding level, the subsequent bitrate control phase aims to determine accurate QPs for quantization while regulating the actual encoded bits to closely match the target bitrate. The control model plays a crucial role in achieving effective and efficient bitrate control. For traditional 2D video compression, current rate control models can be categorized into three categories: Q domain [24], [25], [26], [27], ρ domain [28], [29], [30], [31], [32], and λ domain [33], [34], [35]. In earlier studies, the quantization step Q is considered as a pivotal factor in determining the rate [24], [25], [26], [27].

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

The ρ domain rate control considers that the proportion of non-zero terms ρ in transformation coefficients is inversely correlated with the rate [28], [29], [30], [31], [32]. However, both Q domain and ρ domain algorithms solely focus on residual bits while disregarding the significantly increased header bits introduced by HEVC and VVC. Consequently, in [33], a strong correlation is suggested between Lagrange multiplier λ and bitrate, resulting in a corresponding rate control algorithm based on the R-λ power function model integrated into reference softwares of HEVC and VVC. Numerous experiments have demonstrated that the R-λ model employed in traditional 2D video compression applications achieves accurate and stable rate control. By leveraging the power function relationship between rate and distortion, along with rate-distortion optimization theory, this model employs the Lagrange multiplier method to convert the constrained convex optimization problem into an unconstrained one while solving for Lagrange multipliers to guide bit allocation and determine QPs. In [34], the R-λ relationship is augmented with the inclusion of complexity, which is utilized for bit allocation at the CTU level. To address the issue of imprecise parameter updates in [33], a more intricate parameter updating strategy is employed in this proposal. In [35], an adaptive bit allocation method is introduced, applicable to both frame and CTU levels. Furthermore, improvements are made to intra-frame coding rate control to account for its typically higher bit overhead compared to inter-frame coding. Relevant studies conducted subsequently have predominantly focused on the R-λ model [36], [37], [38], [39], [40]. In [36], an improved rate control method for HEVC is introduced, which considers inter-block dependency to enhance the rate-distortion performance and reduce bitrate errors. In [37], a novel rate control scheme for HEVC screen content coding is presented, taking into consideration the distinctive characteristics of screen content videos in comparison to traditional nature videos. In [38], a high-performance rate control algorithm targeting the VVC standard is proposed, aiming to achieve superior coding efficiency by considering both spatial and temporal features complexity. In [39], a joint optimization scheme of structural similarity (SSIM) for video coding based on CTU-level bit allocation and rate-distortion optimization is proposed, which effectively improves the coding performance based on SSIM. In [40], an innovative bit allocation method for multi-view texture video coding is suggested, leveraging inter-view dependency and spatio-temporal correlation. Our proposed rate control algorithm in this study also falls within the λ domain category, aligning with these advancements. In addition, considering the initial quantization parameter, buffer state and maintaining video quality consistency hold significant research importance in the field of rate control. In [41], a learning-based initial quantization parameter method is proposed to enhance the rate control performance of HEVC. A benchmark label is constructed using a single rate-distortion pair, and it is suggested to utilize the target bits per pixel for all remaining frames. In [42], a novel CTU-level rate control method for intra frames in HEVC is proposed, ensuring consistent perceptual quality maintenance through a new perceptual hyperbolic rate-distortion optimization method and a CTU-level perceptual distortion model. In [43], a quality control method is designed as an alternative to bitrate control in video coding, which realizes the quality control with the minimum bitrate. The proposed algorithm demonstrates high accuracy in quality control while maintaining constant quality. In [44], a novel joint rate control scheme is presented, which is composed of satisfied-user-ratio (SUR)-based perception modeling and SUR-based perception-buffer rate control for HEVC. This scheme aims to maximize human visual perception quality while preventing buffer underflow and buffer overflow. To balance intra and inter coding, a learning-based method is proposed in [45] for determining intra features as well as an adaptive bitrate reduction method based on real-time bandwidth and encoded information, which can effectively improve the quality of compressed video while ensuring consistent video coding quality.

B. Rate Control for V-PCC

Unlike rate control for 2D video, the development of rate control for V-PCC remains limited. Existing rate control schemes in 2D video encoders do not account for the bit allocation between geometry and attribute videos at a video level. Additionally, in V-PCC, unoccupied units do not contribute to the quality of reconstructed point clouds; however, traditional rate control schemes fail to allocate zero bits to these units, resulting in extra bit overhead and degradation of rate-distortion performance. Furthermore, compared to natural scene videos, projected videos generated by V-PCC exhibit reduced temporal consistency, which leads to unreasonable bit allocation at basic unit level and subsequently impacts the overall rate-distortion performance. In [46], a pioneering dedicated rate control framework is proposed for V-PCC, tailored to the characteristics of projected videos. They demonstrate exceptional performance in terms of high rate-distortion performance and significant bitrate savings. In [47], a learning-based rate control method is introduced to enhance the rate-distortion performance of V-PCC. The approach reduces encoding delay by employing low-latency synchronous bit allocation and accurately predicts basic unit-level parameters through neural network. In [48], a novel V-PCC bit allocation method is proposed to establish a correlation between the number of coding bits for both geometry and attribute, and the distortion of reconstructed point clouds. This optimization ensures superior reconstructed quality while significantly reducing coding time. In [49], a region-based 3D point cloud compression technique is developed to map patches after V-PCC projection onto different regions. Video sequences are encoded within each region, and a coarse-to-fine rate control algorithm is proposed. In [50], a rate control algorithm is designed for all intra configuration in V-PCC, which developed an optimization citeria to allocate the target bits between the geometry video and the attribute video according to the quality dependency. A two-pass coding method is suggested to consider the characteristics of projected video and improve the accuracy of rate control. In addition, in terms of V-PCC low-complexity optimization, a fast rate-distortion-guided attribute coding unit partition method is

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

introduced using cross-projection information of V-PCC's all intra coding configuration in [51], along with a rate-distortion-guided learning approach to reduce the coding loss caused by the wrong prediction of coding unit partition. In [52], a low complexity coding unit decision algorithm for V-PCC is proposed, aiming to enhance predictive performance by leveraging the inherent correlation between 2D projected sequences and introducing cross-projection information. The aforementioned studies investigate rate-distortion optimization for V-PCC from various perspectives. However, the existing frame-level bit allocation algorithms fail to consider the substantial similarity between near and far frames under the all intra "IPIP…" coding structure of projected videos. Therefore, there exists significant potential to explore distortion propagation model within the GOP for V-PCC rate control.

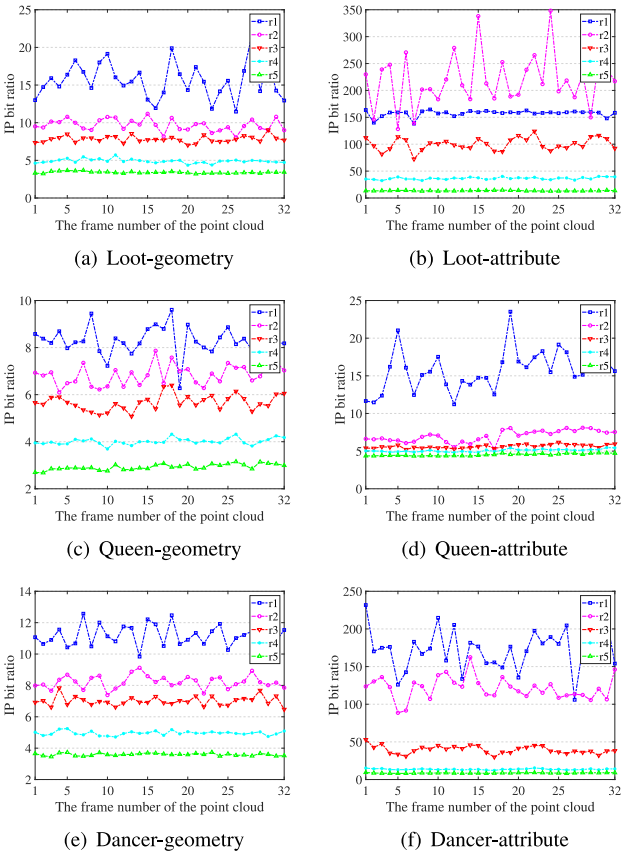
III. Motivation

The AI coding structure employed in V-PCC can be denoted as "IPIP…". A GOP is formed by combining one intra frame and one inter frame, which jointly impact the geometry or attribute quality of a point cloud frame. Hence, the bit allocation between intra frame and inter frame inside the GOP will significantly affect the compression performance of projected videos. The bit allocation task between intra frame and inter frame under this AI structure is simply referred to as "IP" frame-level bit allocation. In the original reference software, the "IP" frame-level target bit ratio is set to 5:1, which is known as the FR-BA method.

Note that the default bitrate control method in TMC2 is too imprecise to compare different bit allocation methods. Therefore, the generalized pre-encoding (GPE) method is adopted to achieve accurate R-λ model parameters and bitrate control. This approach encodes the current frame using five QPs and obtains rate-distortion information, which is then applied to fit the parameters of R-λ model. Importantly, only the frames of the first GOP are pre-encoded, while subsequent frames refer to the model parameters of this initial GOP. By avoiding interference from imprecise QP determination, the GPE method serves as a prerequisite for various bit allocation methods.

Table I shows the rate-distortion performance of the FR-BA rate control method compared with the FixedQP scheme without rate control. BD-Rate is used to quantify the rate difference between the two methods at the same quality level [53]. Geometry distortion is assessed through point-to-point distance (D1) and point-to-point distance (D2), respectively. The final geometry BD-Rate is calculated as the average of the BD-rates under these two distortion metrics. The attribute distortion encompasses three attribute components (Luma, Cb, Cr), and the attribute BD-Rate is determined by weighing the BD-Rates of these components in a ratio of 6:1:1. It can be observed that the original "IP" frame-level bit allocation method exhibits a significant rate-distortion performance loss.

Fig. 1 collects the actual encoded bits of intra frames and inter frames in FixedQP method, which shows the "IP" frame-level actual bit ratios of geometry and attribute videos for 32 point cloud frames corresponding to three sequences ("Loot", "Queen" and "Dancer") under five rate points. The dataset in this paper is the common test condition for V-PCC from [54]. Fig. 2 further shows the geometry and attribute "IP" frame-level actual bit ratios of 32 point cloud frames corresponding to seven sequences under r1 bitrate scenario for the FixedQP and FR-BA rate control method, where the 32-frame data are averaged. According to Fig. 1, the following conclusions can be drawn. First, the actual "IP" frame-level bit ratio in different frames of the same projected video with the same bitrate has slight fluctuations, but is relatively close on the whole. Second, under the condition of the same projected video with different rate points, the actual "IP" frame-level bit ratio shows a downward trend with the increase of bitrate. Third, the actual "IP" frame-level bit ratio of the geometry video is lower than that of the attribute video for the same point cloud sequence. Finally, the actual "IP" frame-level bit ratios of different point cloud sequences vary significantly.



(a) Loot-geometry (b) Loot-attribute

(c) Queen-geometry (d) Queen-attribute

(e) Dancer-geometry (f) Dancer-attribute

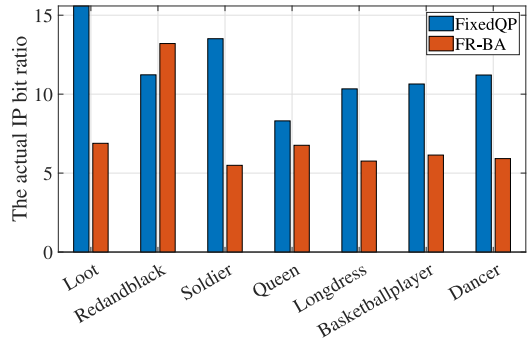
Fig. 1. Actual "IP" bit ratios under different bitrates in FixedQP scheme.

TABLE I
Comparison of BD-Rate Performance Between FR-BA Method and FixedQP Scheme Under All Intra Coding Structure

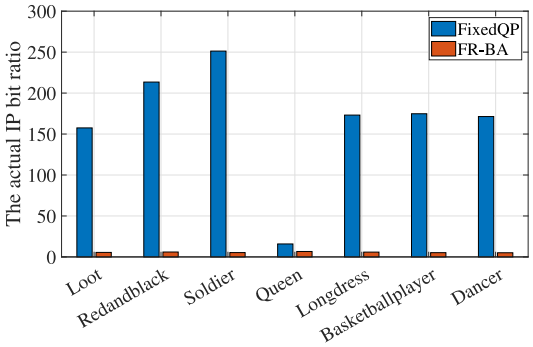
Point Cloud Sequence	Geometry BD-Rate		Attribute BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	4.28%	4.64%	14.78%	21.07%	23.80%
Redandblack	2.58%	2.17%	10.00%	17.64%	20.94%
Soldier	3.55%	3.91%	10.35%	14.37%	13.08%
Queen	2.36%	1.72%	6.23%	8.86%	10.80%
Longdress	3.38%	3.28%	8.34%	14.86%	15.39%
Basketballplayer	3.45%	3.47%	13.16%	18.56%	17.89%
Dancer	4.24%	4.37%	13.31%	19.53%	19.12%
Average	**3.41%**	**3.37%**	**10.88%**	**16.41%**	**17.29%**

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

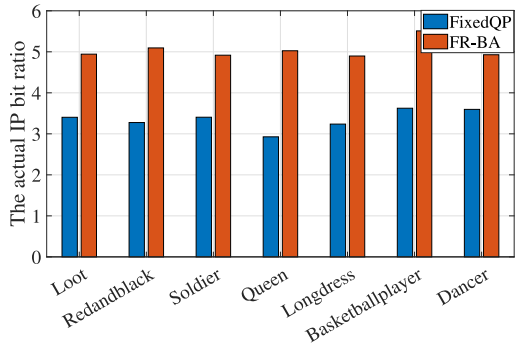


(a) Actual "IP" bit ratio for geometry video

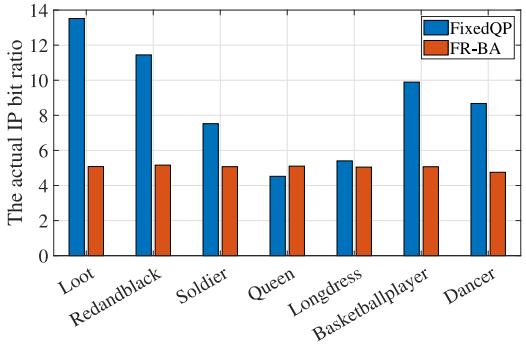


(b) Actual "IP" bit ratio for attribute video

Fig. 2. Comparison of "IP" bit ratio between FixedQP scheme and FR-BA method under r1 bitrate scenario.



(a) Actual "IP" bit ratio for geometry video



(b) Actual "IP" bit ratio for attribute video

Fig. 3. Comparison of "IP" bit ratio between FixedQP scheme and FR-BA method under r5 bitrate scenario.

Fig. 2 and Fig. 3 compare the actual "IP" bit ratio data of FixedQP scheme and FR-BA method at different rate points. The actual "IP" bit ratios of geometry and attribute videos at the lowest rate point r1 are presented in Fig. 2 (a) and (b) respectively. Fig. 3 (a) and (b) illustrate the actual "IP" bit-ratios of geometry and attribute videos at the highest rate point r5, respectively. It can be found that when the "IP" frame-level target bit ratio is set to a fixed value of 5:1, the actual bit ratio is fluctuated around 5 because it is difficult to achieve accurate bitrate control for inter frames. From these two figures, the following conclusions can be drawn. First, for both the geometry and attribute videos, the actual "IP" bit ratio of the FixedQP scheme decreases significantly with the increase of bitrate, which is consistent with Fig. 1. Second, the difference in actual "IP" bit ratio between the FR-BA method and the FixedQP scheme for attribute video is much larger than that for geometry video, which is one of the reasons for the significant attribute quality gap between the two methods in Table I. Third, the FR-BA method can not maintain the bit allocation method adaptive to the target bitrate at the lowest rate point and the highest rate point. Therefore, it is necessary to design an adaptive bit allocation method, which has great potential to improve the performance especially for attribute video.

To sum up, the inadequate "IP" frame-level target bit ratios of the FR-BA method depicted in Fig. 2 and Fig. 3 contribute to its subpar rate-distortion performance in Table I. Therefore, setting fixed "IP" frame-level target bit ratio for the V-PCC bit allocation task under the AI structure is an unreasonable bit allocation method. There arises a necessity to devise a bit allocation method that adaptively changes the target bit ratios at "IP" frame level in response to the change of target bitrates for the two projected videos.

IV. Methodology

For the unique AI structure in V-PCC, the distortion propagation model of "IP" structure has not been studied yet, where frame-level bit allocation task based on the distortion propagation model needs to consider different R - λ models for intra and iner frames. In this section, we theoretically derive a frame-level bit allocation method based on distortion propagation model inside the GOP under the AI coding structure. Then the skip ratio of CTUs is utilized to predict the distortion propagation factor. On this basis, the occupancy information is employed to correct the distortion propagation prediction model and further improve the coding performance.

A. Analysis of Distortion Propagation Model

The GOP in the AI structure consists of an intra frame and an inter frame, where the inter frame belongs to the forward reference frame, that is, the P-frame. The definition of the distortion minimization problem under AI structure can be expressed as:

$$\min_{\lambda_I,\lambda_P} \sum D_I + D_P, \text{s.t.} \sum (R_I + R_P) \leq R_T, \tag{1}$$

where $D_I, R_I, \lambda_I, D_P, R_P, \lambda_P$ and R_T represent distortion of intra frame, bits of intra frame, λ of intra frame, distortion of inter frame, bits of inter frame, λ of inter frame, target bits

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

of the GOP respectively. The distortion is measured using the mean square error (MSE) per pixel.

The Lagrange multiplier method is employed to convert the above problem into an unconstrained optimization problem:

$$\min_{\lambda_I,\lambda_P} D_I + D_P + \mu(R_I + R_P), \tag{2}$$

where μ is the Lagrange multiplier. By setting the partial derivatives of the objective function in (2) equal to 0 with respect to both λ_I and λ_P, the unconstrained optimization problem can be solved by:

$$\frac{\partial D_I + \partial D_P}{\partial \lambda_I} + \mu \frac{\partial R_I}{\partial \lambda_I} = \frac{\partial D_I + \partial D_P}{\partial D_I}\frac{\partial D_I}{\partial \lambda_I} + \mu \frac{\partial R_I}{\partial \lambda_I} = 0, \tag{3}$$

$$\frac{\partial D_P}{\partial \lambda_P} + \mu \frac{\partial R_P}{\partial \lambda_P} = 0. \tag{4}$$

Convert the partial derivative of distortion D with respect to λ to the partial derivative of bitrate R with respect to λ:

$$\frac{\partial D}{\partial \lambda} = \frac{\partial D}{\partial R} \cdot \frac{\partial R}{\partial \lambda} = -\lambda \cdot \frac{\partial R}{\partial \lambda}. \tag{5}$$

Let $\frac{\partial D_P}{\partial D_I} = \gamma_P$, then equations (3) and (4) are converted into the following two equations:

$$\left[\mu - \lambda_I(1+\gamma_P)\right]\frac{\partial R_I}{\partial \lambda_I} = 0, \tag{6}$$

$$[\mu - \lambda_P]\frac{\partial R_P}{\partial \lambda_P} = 0. \tag{7}$$

It is known that the relations between the intra bits R_I, the inter bits R_P and the corresponding Lagrange multipliers λ_I, λ_P are as follows:

$$R_I = SATD^{\beta'} \cdot \left(\frac{\lambda_I}{\alpha_I}\right)^{-\frac{1}{\beta_I}}, \tag{8}$$

$$R_P = \left(\frac{\lambda_P}{\alpha_P}\right)^{\frac{1}{\beta_P}}, \tag{9}$$

where α_I, α_P, β_I and β_P are parameters related to the sequence, *SATD* is the sum of absolute transformed difference that reflects the frequency domain characteristics. The aforementioned two equations are derived from the VTM-11.0 reference software.

Thus, $\partial R_I/\partial \lambda_I$ and $\partial R_P/\partial \lambda_P$ can be easily obtained as:

$$\frac{\partial R_I}{\partial \lambda_I} = -\frac{1}{\beta_I} SATD^{\beta'} \cdot \left(\frac{\lambda_I}{\alpha_I}\right)^{-\frac{1}{\beta_I}-1}, \tag{10}$$

$$\frac{\partial R_P}{\partial \lambda_P} = \frac{1}{\beta_P}\left(\frac{\lambda_P}{\alpha_P}\right)^{\frac{1}{\beta_P}-1}. \tag{11}$$

Partial derivatives $\partial R_I/\partial \lambda_I$ and $\partial R_P/\partial \lambda_P$ are non-zero. Consequently, by utilizing equations (6) and (7), (12) can be derived, establishing the association between λ_I and λ_P:

$$\mu = \lambda_P = \lambda_I(1+\gamma_P). \tag{12}$$

According to the condition that the sum of intra and inter bits inside the GOP is equal to the target GOP bit R_T, μ can be determined by combining the aforementioned (8), (9), and (12):

$$\begin{aligned}\psi(\mu) &= R_I + R_P \\ &= SATD^{\beta'} \cdot \left(\frac{\mu}{\alpha_I(1+\gamma_P)}\right)^{-\frac{1}{\beta_I}} + \left(\frac{\mu}{\alpha_P}\right)^{\frac{1}{\beta_P}} \\ &= R_T.\end{aligned} \tag{13}$$

Considering the above equation, where $\alpha_I > 0, \alpha_P > 0, \beta_I > 0$ and $\beta_P < 0$, it follows that both R_I and R_P exhibit monotonically decreasing behavior with respect to the parameter range of interest $\mu \in (0, +\infty)$. Consequently, $\psi(\mu)$ also possesses this monotonic property. It means that employing the bisection method becomes a viable approach for obtaining a numerical solution to equation (13). The implementation details are as follows: the initial value of μ is set to 100, the solution interval is the closed interval $[0.1, 10000]$, the termination condition is $|\psi(\mu) - R_T| \leq 0.000001$, and the maximum number of iterations is 20.

After obtaining the numerical solution of μ, the corresponding λ_I and λ_P of intra frame and inter frame can be obtained from equation (12), and then the target bits of the two frames can be obtained from equation (8) and equation (9). In summary, based on the distortion propagation model, the frame-level bit allocation method of the non-first GOP in AI structure is implemented. For the first GOP, this work draws on the frame-level bit allocation method of RA structure in VTM-11.0 reference software, using a fixed "IP" bit ratio, e.g., 5.

B. Distortion Propagation Model-Based Frame-Level Bit Allocation

This subsection will build upon the theoretical analysis in Section IV-A to propose a novel frame-level bit allocation scheme for V-PCC, named Distortion Propagation Model-based Bit Allocation Method (DPM-BA).

In the previous subsection, the distortion propagation factor of intra frame to inter frame γ_P is introduced, and the procedure of obtaining γ_P is as follows: (a) Fix quantization parameters of inter frame as QP_{inter}. (b) Encode the frames by multiple QP_{intra} to record multiple groups of intra distortion and inter distortion $(D_{\text{intra}}, D_{\text{inter}})$. (c) Linear fitting is performed for $(D_{\text{intra}}, D_{\text{inter}})$. The slope is the distortion propagation factor γ_P, and the classical least squares fitting method is adopted. In this work, we unify the offset of QP_{intra} with respect to QP_{inter} as:

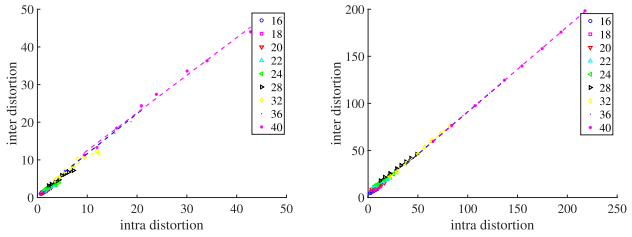
$$\{-10, -8, -6, -4, -3-2, -1, 0\}. \tag{14}$$

Based on the above rules, the values of the first four distortion propagation factors γ_P of two videos are obtained for seven point cloud sequences under the AI structure for analysis. Due to the specific "IPIP…" coding structure, the distortion propagation factor γ_P of geometry video, physically describes the distortion propagation relationship between two projected geometry frames (intra and inter frames, respectively) generated by a point cloud frame. In the same way, there are similar physical meanings in the attribute video.

Fig. 4 demonstrates the distortion propagation of intra frame to inter frame in two projected video of "Loot". We fix QPs of

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



(a) Geometry (b) Attribute

Fig. 4. Distortion propagation from intra frame to inter frame of "Loot".

inter frame within the set {16, 18, 20, 22, 24, 28, 32, 36, 40}. And based on (14), distortion pair ($D_{\text{intra}}, D_{\text{inter}}$) can be obtained by changing QP_{intra}, and represented by discrete points. They are linearly fitted by the classical least squares method and depicted by dashed line. The figure intuitively presents the linear distortion propagation relationship, thereby providing initial evidence for the research idea of introducing distortion propagation factor into frame-level bit allocation.

In addition, Table II further shows the linear slopes in the two subfigures, namely the distortion propagation factor γ_P, the fitting coefficient R^2, and the skip ratio θ_{skip} of minimum CUs. The following conclusions can be found. First, the linear distortion propagation from intra frame to inter frame of "Loot" is significant, and the fitting coefficients R^2 are very close to 1, which proves the reliability of introducing distortion propagation factor for frame-level bit allocation. Second, the distortion propagation factor γ_P of geometry video is significantly higher than that of attribute video. Thirdly, the skip ratio θ_{skip} of the minimum CUs is observed to be significantly lower in geometry videos as compared to attribute videos. However, this gap is not obvious when the QP of inter frame is large. For example, when QP_{inter} is greater than or equal to 32, the θ_{skip} for both the geometry and attribute videos approaches 1.

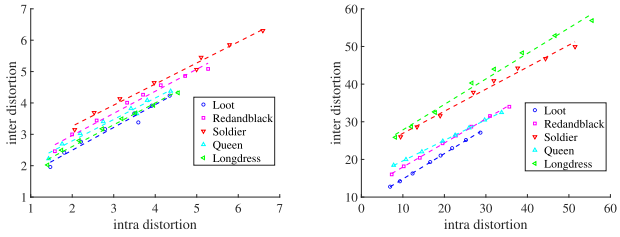
In order to further compare the distortion propagation factors of more sequences, Fig. 5 shows that when the QP of the first inter frame is fixed at 24, the distortion of the first inter frame is affected by the distortion of the first intra frame when the offset changes according to equation (14). It can be observed that the linear relationship is still valid for both kinds of two projected videos. Meanwhile, the slope of the linear relationship is the distortion propagation factor, and it is preliminarily proved that frame-level bit allocation can be carried out based on the distortion propagation model.

Table III presents the distortion propagation factors and linear fitting coefficients R^2 for different point cloud sequences. The phenomenon can be observed that the fitting coefficients R^2 of the linear distortion propagation model for the two projected videos are very close to 1, and the average values are 0.9806 and 0.9964, respectively.

However, it is inadvisable to acquire the distortion propagation factor γ_P in practical coding scenarios using the aforementioned time-consuming coding method. Therefore, without the time-absorbing pre-coding strategy, and based on the reference characteristics of γ_P in adjacent GOPs, this paper uses the proportion θ_{skip} of all the 4×4 minimum CUs in inter frame whose coding mode is skip mode to predict γ_P. This

TABLE II
DISTORTION PROPAGATION FACTOR γ_P FROM INTRA FRAME TO INTER FRAME AND CORRELATION COEFFICIENTS OF LINEAR FITTING R^2 FOR DIFFERENT QP_{inter}

QP_{inter}	Geometry			Attribute		
	γ_P	R^2	θ_{skip}	γ_P	R^2	θ_{skip}
16	0.6684	0.9972	0.5490	0.6023	0.9932	0.7549
18	0.7574	0.9929	0.5268	0.5996	0.9942	0.7872
20	0.7826	0.9660	0.4906	0.5903	0.9962	0.8106
22	0.6927	0.9463	0.4815	0.6175	0.9985	0.8381
24	0.7261	0.9685	0.5116	0.6738	0.9991	0.8761
28	0.8775	0.9726	0.6655	0.7643	0.9996	0.9106
32	0.8586	0.9450	0.9382	0.9092	0.9989	0.9678
36	1.0651	0.9900	0.9983	0.8991	1.0000	0.9984
40	0.9988	0.9930	0.9991	0.9058	0.9999	0.9969
Average	**0.8252**	**0.9746**	**0.6845**	**0.7291**	**0.9977**	**0.8823**

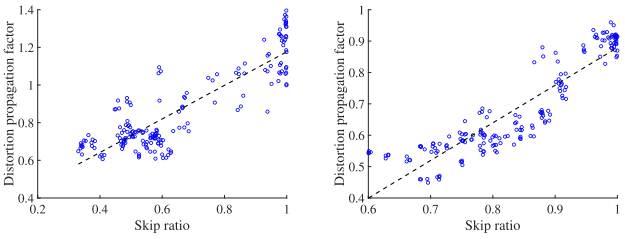


(a) Geometry (b) Attribute

Fig. 5. Distortion propagation from intra frame to inter frame for different point cloud sequences.

TABLE III
DISTORTION PROPAGATION FACTOR γ_P FROM INTRA FRAME TO INTER FRAME AND CORRELATION COEFFICIENTS OF LINEAR FITTING R^2 FOR DIFFERENT POINT CLOUD SEQUENCES

Point Cloud Sequence	Geometry		Attribute	
	γ_P	R^2	γ_P	R^2
Loot	0.7261	0.9685	0.6738	0.9991
Redandblack	0.7040	0.9808	0.6401	0.9988
Soldier	0.6713	0.9878	0.5859	0.9889
Queen	0.6832	0.9715	0.5453	0.9997
Longdress	0.6939	0.9870	0.6765	0.9905
Basketballplayer	0.7124	0.9762	0.6514	0.9990
Dancer	0.7565	0.9923	0.6488	0.9985
Average	**0.7068**	**0.9806**	**0.6317**	**0.9964**



(a) Geometry (b) Attribute

Fig. 6. Relationship between distortion propagation factor and skip ratio of two projected videos.

predicted value then serves as the distortion propagation factor in the next GOP.

Skip mode is a special prediction mode that consumes almost no bits in inter frames. According to [55], it is proved that the distortion propagation factor is approximately equal to the skip ratio θ_{skip} in the traditional 2D video coding environment HEVC. Reference [56] investigates the linear relationship

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Algorithm 1 Distortion Propagation Model-Based Frame-Level Bit Allocation

Input: The projected videos to be coded, target bits of the current GOP.

Output: target bits of intra frame and inter frame R_I, R_P.

1: Acquire the skip ratio θ_{skip} for inter-frame minimum CUs within the previous GOP;
2: Select the appropriate distortion propagation factor model according to the category of the projected video base on euqation (15);
3: Calculate the distortion propagation factor γ_P;
4: Solve the Lagrange multiplier μ in euqation (13) based on the bisection method;
5: Calculate the corresponding λ_I and λ_P in the rate-distortion optimization process base on euqation (12);
6: Calculate the target bits of intra frame and inter frame according to equations (8) and (9);
7: **return** target bits of intra frame and inter frame R_I, R_P.

between the distortion propagation factor and the skip ratio in VVC coding environment. In this study, we examine the relationship between the skip ratio θ_{skip} and the distortion propagation factor γ_P. Fig. 6 illustrates the linear fitting effects of these two variables for all test point cloud sequences. In the figure, the horizontal and vertical coordinates of the blue scatter points are skip ratio and distortion propagation factor respectively, while the black dashed lines depict fitting lines. The obtained fitting coefficients R^2 are 0.7498 and 0.8238 for geometry and attribute, respectively, which essentially confirm the linear relationship between these two factors.

Therefore, the two distortion propagation factor prediction models for projected video are obtained as:

$$\gamma_P = \begin{cases} 0.8502\theta_{\text{skip}} + 0.3172, \text{ geometry}, \\ 1.1636\theta_{\text{skip}} - 0.3006, \text{ attribute}. \end{cases} \tag{15}$$

Finally, Algorithm 1 describes the distortion propagation model-based frame-level bit allocation method in the form of pseudo-code. Utilizing the value of γ_P, the numerical solution of μ in equation (13) is solved. Consequently, we can calculate the target bits for both intra and inter frames by equations (8), (9) and (12).

C. Adjusted Distortion Propagation Model-Based Bit Allocation

To address the issue of inadequacy in the distortion propagation model used in traditional 2D video rate control for V-PCC, this study reconstructs the distortion propagation model by incorporating occupancy information and proposes a distortion propagation factor prediction approach based on occupation information correction. The resulting frame-level bit allocation approach is referred to as Adjusted Distortion Propagation Model-based Bit Allocation Method (ADPM-BA), as illustrated in Fig. 7.

There is a large amount of unoccupied information in the V-PCC's projected videos, which has minimal impact on the quality of reconstructed point clouds. It is also pointed out that the rate-distortion characteristics of unoccupied CUs in

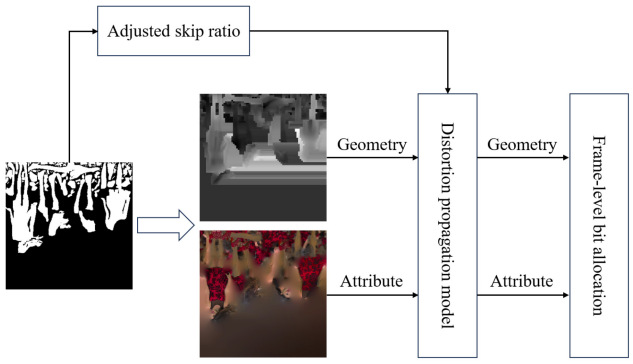
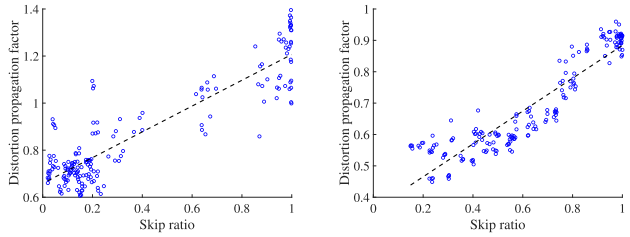


Fig. 7. Adjusted distortion propagation model-based bit allocation method.



(a) Geometry (b) Attribute

Fig. 8. Relationship between distortion propagation factor and skip ratio of two projected videos after the adjustment with occupancy information.

projected videos are quite different from those of occupied CUs. Therefore, this subsection aims to investigate the effect of disregarding unoccupied CUs on the distortion propagation factor prediction model.

For inter frames, we define the skip ratio of minimum CUs $\theta_{\text{skip}}^{\text{occu}}$ after introducing occupancy correction as:

$$\theta_{\text{skip}}^{\text{occu}} = \frac{N_{\text{occu}}^{\text{skip}}}{N_{\text{occu}}}, \tag{16}$$

where N_{occu} and N_{occu}^{skip} denote the number of occupied minimum CUs in the projected video and the number of occupied minimum CUs in which skip mode is used, respectively.

Fig. 8 presents the linear fitting effect between the skip ratio $\theta_{\text{skip}}^{\text{occu}}$ of the projected geometry and attribute videos, and the distortion propagation factor after the adjustment with occupancy information for point cloud sequences. The horizontal and vertical coordinates of the blue scatter points in the figure are the adjusted skip ratio and distortion propagation factor respectively, while the black dashed line is the fitting line. Notably, the fitting coefficients R^2 are 0.7935 and 0.8079 for geometry and attribute, respectively, which indicates better fitting effect on geometry video compared with the unadjusted skip ratio in the previous subsection. Consequently, the following two distortion propagation factor prediction models for the projected video are obtained:

$$\gamma_P^* = \begin{cases} 0.5330\theta_{\text{skip}}^{\text{occu}} + 0.6539, \text{ geometry}, \\ 0.5053\theta_{\text{skip}}^{\text{occu}} + 0.3531, \text{ attribute}. \end{cases} \tag{17}$$

Based on the above equations, we have developed an adjusted distortion propagation model to accommodate the novel characteristics of V-PCC projected videos. Algorithm 2

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE IV
BD-RATE COMPARISON OF DPM-BA WITH FR-BA UNDER ALL INTRA STRUCTURE

Point Cloud Sequence	Geometry BD-Rate		Attribute BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	−0.34%	−1.64%	−7.80%	−10.96%	−10.96%
Redandblack	0.43%	0.45%	−5.10%	−5.11%	−5.53%
Soldier	−0.95%	−3.13%	−5.87%	−6.10%	−5.48%
Queen	−0.49%	−0.14%	−0.21%	1.75%	0.08%
Longdress	−0.79%	−1.65%	−3.85%	−3.32%	−3.89%
Basketballplayer	−0.98%	−2.94%	−5.04%	−4.96%	−5.72%
Dancer	−0.08%	−0.71%	−5.70%	−4.31%	−5.50%
Average	**−0.46%**	**−1.39%**	**−4.80%**	**−4.72%**	**−5.29%**

TABLE V
BD-RATE COMPARISON OF ADPM-BA WITH FR-BA UNDER ALL INTRA STRUCTURE

Point Cloud Sequence	Geometry BD-Rate		Attribute BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	−1.55%	−3.93%	−8.76%	−12.94%	−12.75%
Redandblack	−0.62%	−1.87%	−6.19%	−7.53%	−7.27%
Soldier	−0.42%	−4.21%	−7.07%	−7.61%	−7.36%
Queen	−2.04%	−1.93%	0.32%	0.44%	−0.36%
Longdress	−1.26%	−3.35%	−5.35%	−6.73%	−7.07%
Basketballplayer	−0.77%	−4.36%	−6.46%	−8.02%	−8.37%
Dancer	−1.12%	−2.82%	−7.00%	−6.76%	−8.00%
Average	**−1.11%**	**−3.21%**	**−5.79%**	**−7.02%**	**−7.31%**

Algorithm 2 Adjusted Distortion Propagation Model-Based Bit Allocation Method

Input: The projected videos to be coded, target bits of the current GOP, occupancy information.

Output: target bits of intra frame and inter frame R_I, R_P.

1: Introduce occupancy information in the projected video coding stage;
2: Acquire the skip ratio $\theta_{\text{skip}}^{\text{occu}}$ for inter-frame occupied minimum CUs within the previous GOP;
3: Select the appropriate distortion propagation factor model according to the category of the projected video base on euqation (17);
4: Calculate the adjusted distortion propagation factor γ_P^*;
5: Solve the Lagrange multiplier μ in euqation (13) based on the bisection method;
6: Calculate the corresponding λ_I and λ_P in the rate-distortion optimization process base on euqation (12);
7: Calculate the target bits of intra frame and inter frame according to equations (8) and (9);
8: **return** target bits of intra frame and inter frame R_I, R_P.

outlines the pseudo-code for the procedure of proposed adjusted distortion propagation model-based bit allocation method.

V. EXPERIMENTS

In this section, the two frame-level bit allocation methods based on the previously proposed distortion propagation model are experimentally evaluated. It is important to note that maintaining fairness and consistency with the experimental contrast is crucial. Therefore, subsequent bit allocation method are implemented on the basis of the GPE scheme, including FR-BA benchmark, DPM-BA and ADPM-BA. The configuration parameters and operations of the these three methods remain consistent with each other except for the different implementation of frame-level bit allocation. The rate-distortion performance evaluation indexes align with Table I, and more detailed objective indicators are provided.

A. Rate-Distortion Performance

Table IV presents the BD-Rate performance comparison between DPM-BA method and FR-BA method under the AI structure. It can be observed that, by employing the distortion propagation model for frame-level bit allocation, DPM-BA achieves average BD-Rate reductions of 0.46% and 1.39% on D1 and D2, respectively. These improvements are relatively modest due to the similarity between the actual "IP" frame-level bit ratio of FR-BA method on geometry video and that of FixedQP scheme, as depicted in Fig. 2. For attribute videos, DPM-BA exhibits a more significant performance improvement compared to FR-BA, with average BD-Rate reductions of 4.80%, 4.72% and 5.29% for Luma, Cb, and Cr channels respectively. This pronounced enhancement can be attributed to the small actual "IP" frame-level bit ratio of the FR-BA method on the attribute video. Using the weighting technique described in Section III, the geometry and attribute BD-Rate are −0.92% and −4.85%, respectively.

Table V shows the BD-Rate performance comparison between ADPM-BA method and FR-BA method under the AI structure. It can be noted that after incorporating the distortion propagation factor prediction model adjusted by occupancy information, ADPM-BA achieves average BD-Rate reductions of 1.11% and 3.21% on D1 and D2, respectively, compared to FR-BA. For attribute indicators Luma, Cb and Cr, the average BD-Rate reductions of ADPM-BA compared with FR-BA are 5.79%, 7.02% and 7.31%, respectively. By employing the weighting technique, the geometry and attribute

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VI
BD-RATE COMPARISON OF ADPM-BA WITH FIXEDQP UNDER ALL INTRA STRUCTURE

Point Cloud Sequence	Geometry BD-Rate		Attribute BD-Rate		
	D1	D2	Luma	Cb	Cr
Loot	2.57%	0.56%	4.74%	5.31%	7.92%
Redandblack	1.93%	0.29%	3.23%	8.86%	12.32%
Soldier	3.08%	−0.46%	2.59%	5.83%	4.88%
Queen	0.01%	−0.21%	6.57%	9.46%	10.42%
Longdress	2.03%	−0.25%	2.64%	7.32%	7.42%
Basketballplayer	2.50%	−1.06%	6.07%	9.11%	8.34%
Dancer	3.08%	1.50%	5.55%	11.53%	9.76%
Average	**2.17%**	**0.05%**	**4.49%**	**8.20%**	**8.72%**

BD-Rates are -2.16% and -6.13%, respectively. In conclusion, our results highlight that introducing occupancy information to adjust the distortion propagation factor prediction model further enhances the rate-distortion performance of ADPM-BA in both geometry and attribute aspects.

Finally, Table VI demonstrates the BD-Rate performance of ADPM-BA in comparison with FixedQP scheme under the AI structure, clearly illustrating the noticeable rate-distortion performance gap between these two methods. The weighted geometry and attribute BD-Rates are calculated to be 1.11% and 5.48%, respectively. Consequently, it can be inferred that the rate-distortion performance of the ADPM-BA rate control method closely approximates that of the FixedQP scheme without rate control. In general, the rate-distortion performance of FixedQP scheme without rate control is optimal when compared to other rate control methods. However, in the field of real-time communication, the FixedQP scheme proves to be impractical as it fails to adapt to bandwidth limit and network fluctuations.

In contrast to the earliest research on rate control for V-PCC [46], they failed to address the unique "IP. IP" all intra structure being studied, that is, the content quality dependency of near and far frames was not taken into account. Furthermore, although [50] has investigated the quality dependence between the geometry video and the color video within the all intra structure of V-PCC, the influence of the unoccupied CUs is still not considered, which results in the characteristics of the projected video not being fully mined and optimized. All these have been solved in this work to optimize the rate-distortion performance of rate control.

B. Bit Error

The bit error is another crucial metric for evaluating rate control algorithms, as it indicates the relative deviation between the actual coded bitrate and the target bitrate, thereby intuitively reflecting the accuracy of rate control. In this subsection, the overall sequence-level bit errors e_{seq} of the three rate control methods are compared, and the calculation method is shown as:

$$e_{seq} = \frac{|R_{act} - R_{tar}|}{R_{tar}} \times 100\%, \tag{18}$$

where e_{seq} is the sequence-level bit error, R_{act} and R_{tar} are the actual bitrate and the target bitrate, respectively. Table VII gives detailed data of overall sequence-level bit errors. It can be observed that all three bit allocation methods maintain very small sequence-level bit errors for AI-structured projected videos. For geometry videos, the average bit errors of the three bit allocation methods are less than or equal to 0.10%. For attribute videos, the average bit errors are equal to 0.05%. These results illustrate that the proposed methods maintains a high rate control accuracy.

C. Quality Consistency

In order to quantify the quality consistency of point clouds under different bit allocation methods, this subsection calculates the standard deviation of objective quality for the whole point cloud sequence in two dimensions: geometry and attribute. The mean standard deviation values of FR-BA method for all sequences in geometry and attribute quality under AI structure are 0.21 and 0.17 respectively, as presented in Table VIII. Here, the standard deviation values of each sequence represent the mean values of the standard deviation values obtained at five target rate points. Additionally, both DPMBA and ADPM-BA methods exhibit slightly higher quality standard deviation values compared to the FR-BA method, with a marginal increase of 0.01 and 0.03 in geometry and attribute dimensions respectively.

Due to the excellent quality consistency among the three rate control methods under the AI structure, it becomes challenging to compare their performance directly. Hence, Fig. 9 shows the attribute quality consistency comparison of the four sequences ("Loot", "Redandblack", "Basketballplayer", and "Dancer") using the three bit allocation methods at the lowest rate point r1. As shown in Fig. 9, the quality of initial point cloud frame is identical for all three methods. This uniformity is because both the DPM-BA and ADPM-BA methods employ a fixed "IP" bit ratio during encoding the initial frame. In addition, Table IX provides specific standard deviation values. It is observed that although the quality consistency of DPM-BA and ADPM-BA methods is slightly lower than that of FR-BA methods, this can be ascribed to the initial point cloud frame's low quality. However, for long point cloud videos, the influence of initial frame quality is insignificant. This is because the DPM-BA and ADPM-BA approaches substantially enhance the attribute quality of the subsequent point cloud frames. Notably, when encoding the 5th and subsequent point cloud frames, reconstructed point clouds using DPM-BA and ADPM-BA methods achieve more than 0.2dB improvement in attribute quality compared to FR-BA. This explains the substantial gain in rate-distortion performance achieved by our proposed methods. Consequently, adopting a more reasonable "IP" target bit ratio for initial point cloud frames enables DPM-BA and ADPM-BA to attain better quality consistency as well as improved rate-distortion performance.

The results depicted in Fig. 9 demonstrate that, for the majority of point cloud frames, the attribute quality achieved by ADPM-BA surpasses or significantly outperforms that of DPM-BA. Particularly noteworthy is the substantial attribute quality improvement observed in the 10th to 20th point cloud frames of the "Dancer" when employing the ADPM-BA method compared to its counterpart, DPM-BA. These findings

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

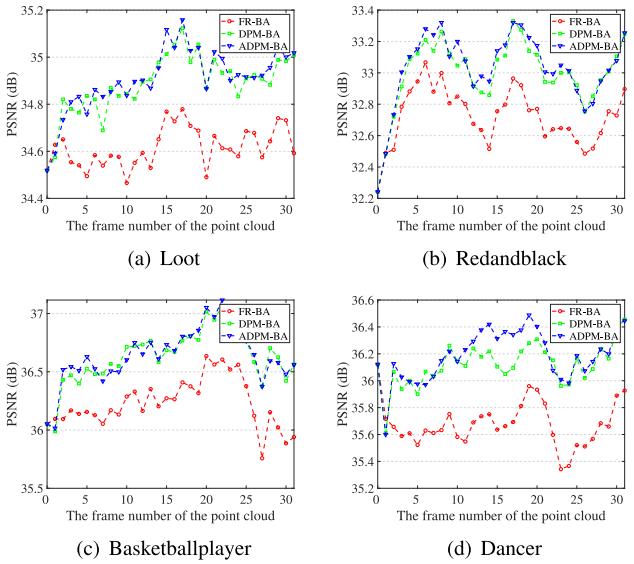
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE VII
COMPARISON OF BIT ERROR FOR THE THREE BIT ALLOCATION METHODS UNDER ALL INTRA STRUCTURE

Point Cloud Sequence	FR-BA		DPM-BA (Proposed)		ADPM-BA (Proposed)	
	Geometry	Attribute	Geometry	Attribute	Geometry	Attribute
Loot	0.17%	0.10%	0.20%	0.10%	0.15%	0.07%
Redandblack	0.11%	0.05%	0.10%	0.03%	0.09%	0.03%
Soldier	0.09%	0.04%	0.07%	0.04%	0.06%	0.03%
Queen	0.10%	0.06%	0.09%	0.08%	0.08%	0.11%
Longdress	0.13%	0.03%	0.13%	0.06%	0.11%	0.05%
Basketballplayer	0.06%	0.05%	0.03%	0.04%	0.03%	0.04%
Dancer	0.05%	0.05%	0.10%	0.05%	0.08%	0.04%
Average	**0.10%**	**0.05%**	**0.10%**	**0.05%**	**0.09%**	**0.05%**

TABLE VIII
QUALITY STANDARD DEVIATION VALUES (DB) OF GEOMETRY AND ATTRIBUTE FOR THE THREE BIT ALLOCATION METHODS UNDER ALL INTRA STRUCTURE

Point Cloud Sequence	FR-BA		DPM-BA (Proposed)		ADPM-BA (Proposed)	
	Geometry	Attribute	Geometry	Attribute	Geometry	Attribute
Loot	0.14	0.12	0.18	0.15	0.20	0.16
Redandblack	0.49	0.30	0.48	0.33	0.49	0.33
Soldier	0.07	0.05	0.12	0.10	0.13	0.12
Queen	0.11	0.14	0.16	0.15	0.16	0.16
Longdress	0.28	0.22	0.26	0.23	0.25	0.25
Basketballplayer	0.17	0.23	0.17	0.24	0.15	0.25
Dancer	0.18	0.16	0.18	0.15	0.18	0.16
Average	**0.21**	**0.17**	**0.22**	**0.20**	**0.22**	**0.20**



(a) Loot (b) Redandblack

(c) Basketballplayer (d) Dancer

Fig. 9. Comparison of attribute quality consistency for the three bit allocation methods under r1 bitrate scenario.

provide compelling evidence supporting the effectiveness of utilizing an adjusted distortion propagation model-based bit allocation approach.

VI. CONCLUSION

For V-PCC, the fixed-ratio bit allocation method under the AI coding structure, as implemented in the reference software, results in a considerable rate-distortion performance disparity between the rate control manner and the FixedQP scheme. Therefore, this paper proposes a distortion propagation model-based bit allocation method called DPM-BA for the specific AI configuration in V-PCC. On the basis of this foundation,

TABLE IX
QUALITY STANDARD DEVIATION VALUES (DB) OF ATTRIBUTE FOR THE THREE BIT ALLOCATION METHODS IN FIG. 9

Point Cloud Sequence	FR-BA	DPM-BA (Proposed)	ADPM-BA (Proposed)
Loot	0.08	0.13	0.13
Redandblack	0.18	0.22	0.24
Basketballplayer	0.21	0.25	0.25
Dancer	0.17	0.16	0.19
Average	**0.16**	**0.19**	**0.20**

an adjusted distortion propagation model-based bit allocation method, dubbed ADPM-BA, is introduced, which leverages the occupancy information to refine the distortion propagation model and further improve the coding performance. Experimental results validate the effectiveness of our proposed distortion propagation model-based frame-level bit allocation method, with ADPM-BA achieving a rate-distortion performance that closely approaches the upper bound of rate control while maintaining satisfactory quality consistency. Considering its potential applications in immersive communication scenarios and video streaming transmission under bandwidth-limited conditions, our proposed scheme holds promise for improving point cloud broadcasting efficiency.

REFERENCES

[1] X. Sheng, L. Li, D. Liu, and Z. Xiong, "Attribute artifacts removal for geometry-based point cloud compression," *IEEE Trans. Image Process.*, vol. 31, pp. 3399–3413, 2022.

[2] E. S. Jang et al., "Video-based point-cloud-compression standard in MPEG: From evidence collection to committee draft [standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 36, no. 3, pp. 118–123, May 2019.

[3] Y. Xu, Q. Yang, L. Yang, and J.-N. Hwang, "EPES: Point cloud quality modeling using elastic potential energy similarity," *IEEE Trans. Broadcast.*, vol. 68, no. 1, pp. 33–42, Mar. 2022.

[4] F. Nardo, D. Peressoni, P. Testolina, M. Giordani, and A. Zanella, "Point cloud compression for efficient data broadcasting: A performance comparison," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2022, pp. 2732–2737.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[5] L. Xie, W. Gao, H. Zheng, and G. Li, "ROI-guided point cloud geometry compression towards human and machine vision," in *Proc. 32nd ACM Int. Conf. Multimedia*, 2024, pp. 3741–3750.

[6] L. Xie and W. Gao, "LearningPCC: A PyTorch library for learning-based point cloud compression," in *Proc. 32nd ACM Int. Conf. Multimedia*, 2024, pp. 11234–11238.

[7] L. Xie and W. Gao, "PCHMVision: An open-source library of point cloud compression for human and machine vision," in *Proc. 32nd ACM Int. Conf. Multimedia*, 2024, pp. 11239–11243.

[8] S. Chen, B. Liu, C. Feng, C. Vallespi-Gonzalez, and C. Wellington, "3D point cloud processing and learning for autonomous driving: Impacting map creation, localization, and perception," *IEEE Signal Process. Mag.*, vol. 38, no. 1, pp. 68–86, Jan. 2021.

[9] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1907–1915.

[10] Q. Zhu, L. Chen, Q. Li, M. Li, A. Nüchter, and J. Wang, "3D LIDAR point cloud based intersection recognition for autonomous driving," in *Proc. IEEE Intell. Veh. Symp.*, 2012, pp. 456–461.

[11] M.-L. Champel, R. Doré, and N. Mollet, "Key factors for a high-quality VR experience," in *Proc. Appl. Digit. Image Process.*, 2017, pp. 183–194.

[12] D. Garrido, R. Rodrigues, A. A. Sousa, J. Jacob, and D. C. Silva, "Point cloud interaction and manipulation in virtual reality," in *Proc. 5th Int. Conf. Artif. Intell. Virtual Real. (AIVR)*, 2021, pp. 15–20.

[13] X. Wu, Y. Zhang, C. Fan, J. Hou, and S. Kwong, "Subjective quality database and objective study of compressed point clouds with 6DoF head-mounted display," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4630–4644, Dec. 2021.

[14] F. Xue, W. Lu, Z. Chen, and C. J. Webster, "From LiDAR point cloud towards digital twin city: Clustering city objects based on Gestalt principles," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 418–431, Sep. 2020.

[15] T. Deng, K. Zhang, and Z.-J. M. Shen, "A systematic review of a digital twin city: A new pattern of urban governance toward smart cities," *J. Manag. Sci. Eng.*, vol. 6, no. 2, pp. 125–134, 2021.

[16] V. V. Lehtola et al., "Digital twin of a city: Review of technology serving city needs," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 114, Nov. 2022, Art. no. 102915.

[17] "Call for proposals for point cloud coding V2," ISO/IEC JTC1/SC29/WG11 MPEG, Denver, CO, USA, document N16763, Apr. 2017.

[18] S. Schwarz et al., "Emerging MPEG standards for point cloud compression," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 9, no. 1, pp. 133–148, Mar. 2019.

[19] H. Liu, H. Yuan, Q. Liu, J. Hou, and J. Liu, "A comprehensive study and comparison of core technologies for MPEG 3-D point cloud compression," *IEEE Trans. Broadcast.*, vol. 66, no. 3, pp. 701–717, Sep. 2020.

[20] E. Lopes, J. Ascenso, C. Brites, and F. Pereira, "Adaptive plane projection for video-based point cloud coding," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, 2019, pp. 49–54.

[21] MPEG 3D Graphics Coding, "V-PCC codec description," ISO/IEC JTC1/SC29/WG7 MPEG, Denver, CO, USA, document N00012, Jan. 2021.

[22] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[23] B. Bross et al., "Overview of the versatile video coding (VVC) standard and its applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 10, pp. 3736–3764, Oct. 2021.

[24] J. Katto and M. Ohta, "Mathematical analysis of MPEG compression capability and its application to rate control," in *Proc., Int. Conf. Image Process.*, 1995, pp. 555–558.

[25] T. Chiang and Y.-Q. Zhang, "A new rate control scheme using quadratic rate distortion model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[26] D.-K. Kwon, M.-Y. Shen, and C.-C. J. Kuo, "Rate control for H.264 video with enhanced rate and distortion models," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 5, pp. 517–529, May 2007.

[27] S. Ma, W. Gao, and Y. Lu, "Rate-distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[28] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay rate control for DCT video coding via /spl rho/-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 8, pp. 928–940, Aug. 2001.

[29] M. Liu, Y. Guo, H. Li, and C. W. Chen, "Low-complexity rate control based on ρ-domain model for scalable video coding," in *Proc. IEEE Int. Conf. Image Process.*, 2010, pp. 1277–1280.

[30] Z. He and S. K. Mitra, "Optimum bit allocation and accurate rate control for video coding via /spl rho/-domain source modeling," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 10, pp. 840–849, Oct. 2002.

[31] Y. Pitrey, Y. Serrand, M. Babel, and O. Déforges, "Rho-domain for low-complexity rate control on MPEG-4 scalable video coding," in *Proc. 10th IEEE Int. Symp. Multimedia*, 2008, pp. 89–96.

[32] Y. Pitrey, M. Babel, and O. Déforges, "One-pass bitrate control for MPEG-4 scalable video coding using ρ-domain," in *Proc. IEEE Int. Symp. Broadband Multimedia Syst. Broadcast.*, 2009, pp. 1–5.

[33] B. Li, H. Li, L. Li, and J. Zhang, "λ domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, pp. 3841–3854, 2014.

[34] M. Karczewicz and X. Wang, "Intra frame rate control based on SATD," JCTVC ITU-T SG1 6WP3 and ISO/IEC JTC1/SC, Geneva, Switzerland, document M0257, Apr. 2013.

[35] B. Li, H. Li, and L. Li, "Adaptive bit allocation for R-lambda model rate control in HM," JCTVC ITU-T SG1 6WP3 and ISO/IEC JTC1/SC, Geneva, Switzerland, document M0036, Apr. 2013.

[36] H. Guo, C. Zhu, M. Xu, and S. Li, "Inter-block dependency-based CTU level rate control for HEVC," *IEEE Trans. Broadcast.*, vol. 66, no. 1, pp. 113–126, Mar. 2020.

[37] H. Yang, L. Shen, Y. Yang, and W. Lin, "A novel rate control scheme for video coding in HEVC-SCC," *IEEE Trans. Broadcast.*, vol. 66, no. 2, pp. 333–345, Jun. 2020.

[38] Z. Zhao, X. He, S. Xiong, L. He, H. Chen, and R. E. Sheriff, "A high-performance rate control algorithm in versatile video coding based on spatial and temporal feature complexity," *IEEE Trans. Broadcast.*, vol. 69, no. 3, pp. 753–766, Sep. 2023.

[39] Y. Li and X. Mou, "Joint optimization for SSIM-based CTU-level bit allocation and rate distortion optimization," *IEEE Trans. Broadcast.*, vol. 67, no. 2, pp. 500–511, Jun. 2021.

[40] T. Li, L. Yu, H. Wang, and Z. Kuang, "A bit allocation method based on inter-view dependency and spatio-temporal correlation for multi-view texture video coding," *IEEE Trans. Broadcast.*, vol. 67, no. 1, pp. 159–173, Mar. 2021.

[41] W. Gao, S. Kwong, Q. Jiang, C.-K. Fong, P. H. Wong, and W. Y. Yuen, "Data-driven rate control for rate-distortion optimization in HEVC based on simplified effective initial QP learning," *IEEE Trans. Broadcast.*, vol. 65, no. 1, pp. 94–108, Mar. 2019.

[42] G. Xiang et al., "Perceptual quality consistency oriented CTU level rate control for HEVC intra coding," *IEEE Trans. Broadcast.*, vol. 68, no. 1, pp. 69–82, Mar. 2022.

[43] M. Zhou, X. Wei, C. Ji, T. Xiang, and B. Fang, "Optimum quality control algorithm for versatile video coding," *IEEE Trans. Broadcast.*, vol. 68, no. 3, pp. 582–593, Sep. 2022.

[44] Z. Yang, W. Gao, G. Li, and Y. Yan, "SUR-driven video coding rate control for jointly optimizing perceptual quality and buffer control," *IEEE Trans. Image Process.*, vol. 32, pp. 5451–5464, 2023.

[45] W. Gao, Q. Jiang, R. Wang, S. Ma, G. Li, and S. Kwong, "Consistent quality oriented rate control in hevc via balancing intra and inter frame coding," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1594–1604, Mar. 2022.

[46] L. Li, Z. Li, S. Liu, and H. Li, "Rate control for video-based point cloud compression," *IEEE Trans. Image Process.*, vol. 29, pp. 6237–6250, 2020.

[47] T. Wang, F. Li, and P. C. Cosman, "Learning-based rate control for video-based point cloud compression," *IEEE Trans. Image Process.*, vol. 31, pp. 2175–2189, 2022.

[48] Q. Liu, H. Yuan, J. Hou, R. Hamzaoui, and H. Su, "Model-based joint bit allocation between geometry and color for video-based 3D point cloud compression," *IEEE Trans. Multimedia*, vol. 23, pp. 3278–3291, 2021.

[49] Q. Liu, H. Yuan, R. Hamzaoui, and H. Su, "Coarse to fine rate control for region-based 3D point cloud compression," in *Proc. IEEE Int. Conf. Multimedia & Expo Workshops (ICMEW)*, 2020, pp. 1–6.

[50] F. Shen and W. Gao, "A rate control algorithm for video-based point cloud compression," in *Proc. Int. Conf. Visual Commun. Image Process. (VCIP)*, 2021, pp. 1–5.

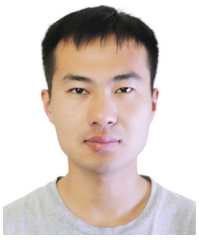
[51] H. Yuan, W. Gao, G. Li, and Z. Li, "Rate-distortion-guided learning approach with cross-projection information for V-PCC fast CU decision," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 3085–3093.

[52] W. Gao, H. Yuan, G. Li, Z. Li, and H. Yuan, "Low complexity coding unit decision for video-based point cloud compression," *IEEE Trans. Image Process.*, vol. 33, pp. 149–162, 2024.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[53] G. Bjøntegaard, "Calculation of average PSNR difference between RD curves," ITU Telecommun. Stand. Sector, VCEG, Austin, TX, USA, document VCEG-M33, Apr. 2001.
[54] MPEG 3D Graphics Coding, "Common test conditions for V3C and V-PCC," ISO/IEC JTC1/SC29/WG7, MPEG, Denver, CO, USA, document N00038, Nov. 2020.
[55] W. Gao, S. Kwong, H. Yuan, and X. Wang, "DCT coefficient distribution modeling and quality dependency analysis based frame-level bit allocation for HEVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 26, no. 1, pp. 139–153, Jan. 2016.
[56] M. Bichon, J. Le Tanou, M. Ropert, W. Hamidouche, and L. Morin, "Optimal adaptive quantization based on temporal distortion propagation model for HEVC," *IEEE Trans. Image Process.*, vol. 28, pp. 5419–5434, 2019.



Zhanyuan Cai received the B.S. degree from NorthWestern Polytechnical University, Xi'an, China, in 2019, and the M.S. degree from the School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, China, in 2022. His research interests include video/point cloud coding and image processing.

Wenxu Gao received the B.S. and M.S. degrees from Sun Yat-sen University, Guangzhou, China, in 2020 and 2022, respectively. He is currently pursuing the Ph.D. degree with the School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, China. His research interests include 3D visual perception modeling, point cloud quality assessment, and point cloud compression.

Ge Li (Member, IEEE) received the Ph.D. degree from the Department of Electrical Engineering, Auburn University, AL, USA, in 1999. He was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of California at Davis, from 2003 to 2004. He was a Summer Engineer with the CSG Research Laboratories, Motorola Inc., Libertyville, IL, USA. After several years of research work in industry, he joined as a Full Professor with the School of Electronic and Computer Engineering, Peking University, China, in 2014. His general research interests include image/video processing and analysis, machine learning, and signal processing.

Wei Gao (Senior Member, IEEE) received the Ph.D. degree in computer science from the City University of Hong Kong, Kowloon, Hong Kong, in February 2017. In 2016, he was a Visiting Scholar with the University of California at Los Angeles, CA, USA. From 2017 to 2019, he worked with the City University of Hong Kong, Hong Kong, and Nanyang Technological University, Singapore. Since 2019, he has been an Assistant Professor with the School of Electronic and Computer Engineering, Shenzhen Graduate School, Peking University, Shenzhen, China. He was an Organizer of workshops and special sessions, and was the Tutorial Speaker at several international conferences. He is also the Leader to establish several open source projects, including OpenPointCloud, OpenAICoding, and OpenDatasets. He authored or coauthored two books, including "Deep Learning for 3D Point Clouds", and "Point Cloud Compression: Technologies and Standardization" published by Springer Nature. His research interests include multimedia coding, multimedia processing, and artificial intelligence, especially 3D point cloud compression and processing. He is an Elected Member of IEEE Visual Signal Processing and Communications Technical Committee, and APSIPA Image, Video, and Multimedia Technical Committee. He is/was an Editorial Board Member of *Elsevier Signal Processing*, *Neural Processing Letter*, *IET Image Processing*, *IET Electronic Letters*, and *APSIPA Transactions on Signal and Information Processing*.

Authorized licensed use limited to: University Town Library of Shenzhen. Downloaded on December 26,2024 at 09:09:13 UTC from IEEE Xplore. Restrictions apply.