Rate Control for Video-based Point Cloud Compression

Li Li, *Member, IEEE*, Zhu Li, *Senior Member, IEEE*, Shan Liu, Houqiang Li, *Senior Member, IEEE*

***Abstract*—Rate control is a necessary tool for the video-based point cloud compression (V-PCC). However, there is no solution specified on this topic yet. In this paper, we propose the first rate control algorithm for the V-PCC. Generally, a rate control algorithm is divided into two processes: bit allocation and bitrate control. In the V-PCC, the total bits are composed of three parts: the header information including auxiliary information and occupancy map, the geometry video, and the attribute video. The bit allocation aims to assign the total bits to these three parts. Since the auxiliary information and occupancy map are encoded losslessly, the bits cost of the header information is fixed. Therefore, we only need to assign bits between the geometry and attribute videos. Our first key contribution is the proposed bit allocation algorithm between the geometry and attribute videos in order to optimize the overall reconstructed point cloud quality. Then we assign the bits of geometry and attribute videos to each GOP, frame, and basic unit (BU) to finish the bit allocation process. Our second key contribution is that we assign zero bits to the BUs with only unoccupied pixels. The unoccupied pixels are useless for the reconstructed quality of the point cloud and therefore should be assigned zero bits. In the bitrate control process, the encoding parameters are determined and the model parameters are updated for each frame and BU to achieve the target bits. Our third key contribution is that we propose a basic unit level model updating scheme to handle the case where various patches may be put in different positions in neighboring frames. We use the auxiliary information to find the corresponding BU in the previous frame and use its model parameters for the current BU. The proposed algorithms are implemented in the V-PCC reference software and the corresponding High Efficiency Video Coding (HEVC) reference software. The experimental results show that the proposed rate control algorithm can achieve very small bits errors as well as quite good reconstructed point cloud quality.**

***Index Terms*—Bit allocation, High Efficiency Video Coding, Point cloud compression, Rate control, Video-based point cloud compression**

I. Introduction

A point cloud is a set of points in 3D space that can be used to represent a 3D surface. Not only the geometry information, but also each point contains some specific attributes, such as colors, material reflection, and so on. The capability of the point cloud to recover 3D objects makes it very promising for extended virtual reality applications such as 3D immersive telepresence [1] and virtual reality viewing with interactive parallax [2]. However, the high data rate of the point cloud is preventing the adoption of this media format. For example, for a typical dynamic point cloud captured by 8i with 30 frames per second, each frame usually has about one million points. If 30 and 24 bits are used to represent the geometry and attribute of the point cloud, the bitrate of the DPC can be as high as $180Mbytes$ per second without compression. The Moving Pictures Experts Group Immersive media working group (MPEG-I) is currently working on a video-based point cloud compression (V-PCC) standard [3] utilizing the existing video coding technologies to solve this problem.

Briefly speaking, the V-PCC projects the point cloud to the geometry and attribute videos and encodes these videos using video compression standard such as High Efficiency Video Coding [4]. Some header information also needs to be signaled in addition to the geometry and attribute videos. The V-PCC is already in the committee draft stage [5]. However, there is not any rate control scheme designed for the V-PCC yet. Rate control is a necessary tool for V-PCC for the point cloud transmission and storage. In this paper, we propose the first rate control scheme for the V-PCC. Generally, a rate control scheme can be divided into two processes: bit allocation and bitrate control. The bit allocation is responsible for assigning the total bits to each sub-unit such as video, frame, and basic unit (BU). The bitrate control aims to achieve the assigned bits for each sub-unit.

In the V-PCC, the total bits are composed of three parts: the header information including the auxiliary information and the occupancy map, the geometry video, and the attribute video. The auxiliary information that indicates the patch information is encoded losslessly. The occupancy map indicates whether the current pixel is occupied or not. It is first down-sampled and then encoded losslessly. Because of this, the bits of the header information is fixed. Therefore, the first step of the bit allocation focuses on the bit allocation among the geometry video and attribute video. This bit allocation step will be called as video level bit allocation in the following sections. The traditional bit allocation algorithms on video compression mainly focus on the picture level and BU level bit allocation [6] [7]. The picture level and BU level bit allocation algorithms take the inter-frame or inter-BU dependency into consideration to optimize the overall video or frame quality under the target bitrate. They have not considered the influence of the geometry and attribute videos on the reconstructed point cloud quality, which is the key of the video level bit allocation.

After the video level bit allocation, the video rate control

L. Li and Z. Li are with the Department of Computer Science and Electrical Engineering, University of Missouri-Kansas City, MO 64110, USA. L. Li is also with the CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China. Professor Zhu Li is the corresponding author (e-mail: lil1@umkc.edu; lizhu@umkc.edu).

S. Liu is with Tencent America, 661 Bryant St, Palo Alto, CA 94301 (e-mail: shanl@tencent.com).

H. Li is with the CAS Key Laboratory of Technology in Geo-Spatial Information Processing and Application System, University of Science and Technology of China, Hefei 230027, China (e-mail: lihq@ustc.edu.cn).

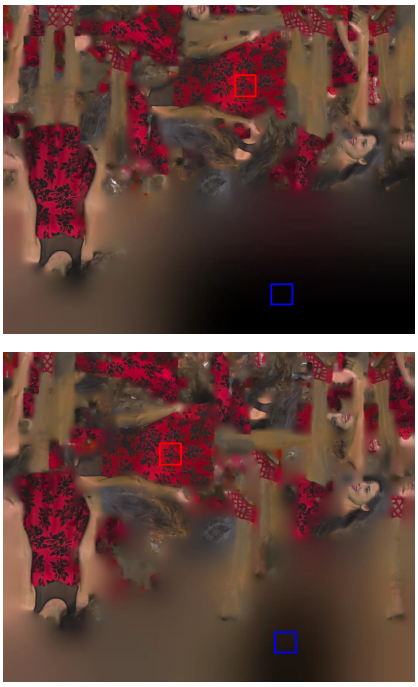


Fig. 1. Typical example of the projected neighboring attribute frames of the point cloud "RedAndBlack". The picture order counts of the top and bottom frames are 1450 and 1451, respectively.

algorithms will be used to control the bitrate of the geometry and attribute videos. The state-of-the-art video rate control algorithm is the λ-domain rate control algorithm [8]. The λ-domain rate control algorithm considers the Lagrangian Multiplier λ as the key factor to determine the bitrate and distortion. It has been adopted by both High Efficiency Video Coding (HEVC) [9] and Versatile Video Coding (VVC) [10] and integrated into the reference software. However, some new characteristics of the projected geometry and attribute videos make the current video rate control algorithms difficult to apply.

One typical example of the projected neighboring attribute frames of the point cloud "RedAndBlack" is shown in Fig. 1. We can see that some BUs indicated by the blue squares are unoccupied. These unoccupied BUs are useless for the reconstructed quality and should be assigned 0 bits. Therefore, the current BU level bit allocation scheme that treats all the BUs with equal importance becomes unsuitable. Additionally, we can see from Fig. 1 that the corresponding BUs in neighboring frames may be put in different positions as indicated by the red squares. This will have significant influences on the rate control model accuracy as we obtain the model parameters of the current BU from the co-located BU in the previous frame at the same hierarchical level. The inaccurate model parameters will lead to large bits errors and serious rate distortion (RD) performance losses.

In order to address the above problems, we propose the first rate control framework designed specified for V-PCC in this paper. The proposed framework mainly has the following key contributions.

- We propose a video level bit allocation algorithm between the geometry and attribute videos. To be more specific, we set the λ ratio between the geometry video and the attribute video inversely proportional to its influence on the reconstructed quality of the point cloud. The target bits of the geometry and attribute videos are closely related to their RD model parameters. Therefore, the proposed bit allocation algorithm is a content-related bit allocation algorithm.
- We propose assigning no bits to the unoccupied BUs that are useless for the reconstructed quality of the point cloud. Those BUs will be encoded using very large λs and QPs to minimize their bits costs. In addition, the RD characteristics of those BUs will not be included in RD characteristics of the picture level.
- We propose estimating the rate control model parameters of the current BU from the corresponding BU in the previous frame instead of the co-located BU. To be more specific, we use the auxiliary information to find the previous BU and obtain the corresponding model parameters. For the unoccupied BUs, we propose obtaining the model parameters from the unoccupied BUs in the previous frame.

The proposed algorithms are implemented in the V-PCC reference software [11] and the corresponding HEVC reference software (HM) [12]. The experimental results show that the proposed algorithm is able to achieve very small bits error as well as satisfactory RD performance.

The rest of this paper is organized as follows. In Section II, we will introduce the related works on rate control and point cloud compression. We will introduce the proposed rate control algorithm for V-PCC in Section III. In Section IV, the experimental results will be presented in detail. Section V will conclude the paper.

II. Related works

In this section, we will introduce some related works on bit allocation and rate control methods. We will also give a brief introduction on the V-PCC framework.

A. Rate Control

The rate control algorithms can be roughly divided into three groups: Q-domain rate control algorithm, ρ-domain rate control algorithm, and λ-domain rate control algorithm. The Q-domain rate control algorithm [13] [14] considers the quantization step Q as the key factor to determine the bitrate. The ρ-main rate control algorithm [15] [16] considers the percentage of zeros among the quantized coefficients ρ as the key factor to determine the bitrate. However, the Q or ρ can only determine the residue bitrate. Li *et al.* [8] pointed out that, along with the header bits increase in HEVC and VVC, the Q or ρ can no longer determine the overall bitrate. They also proposed that the Lagrange Multiplier λ is essentially the key factor to determine the overall bitrate. A corresponding hyperbolic $R-\lambda$ model based rate control algorithm is proposed and integrated into the HEVC and VVC reference software. Therefore, in this paper, we use the λ-domain rate control algorithm as the basis of the proposed rate control algorithm for V-PCC.

In addition to the rate control algorithms focusing on the bitrate control, there are also many works on the bit allocation. The bit allocation algorithms can be divided into two levels: picture level and BU level. The picture level bit allocation is closely related to the encoding structure. In the early stages of bit allocation algorithms, they are dealing with the simple IPPP structure. In the IPPP structure, each P frame will be referenced by the immediate subsequent frame. Therefore, all the P frames will be considered with equal importance. Those bit allocation algorithms consider the frame complexity to perform the picture level bit allocation. Jiang *et al.* [17] first introduced the mean absolute difference (MAD) ratio to characterize the frame complexity. They further proposed using the weighted sum of the MAD ratio and Peak-Signal-to-Noise-Ratio (PSNR) drop to enhance the frame complexity [18]. To overcome the scene change case, Zhou *et al.* [19] proposed using the histogram of difference frame (HOD) to describe the frame complexity.

Along with the introduction of the hierarchical-B coding structure to video coding framework [20], the reference relationships of various frames become much more complex. More and more works focus on characterizing the inter frame dependency among various frames. Hu *et al.* [21] proposed a linear model to characterize the quality dependency and deduced a Q-domain bit allocation algorithm for H.264/AVC [22]. Wang *et al.* [23] extended the linear model to GOP level distortion and rate models and derived a ρ-domain bit allocation algorithm for HEVC. Gao *et al.* [24] introduced a synthesized Laplacian model to describe the Discrete Cosine Transform (DCT) coefficient distribution and introduced a picture level bit allocation algorithm for HEVC. Under the λ-domain rate control algorithm, Li *et al.* [6] proved that the λ ratios of various pictures should be inversely proportional to their influences on the sequence to achieve the optimal performance. Gao *et al.* [25] [26] further carefully analyzed the distortion propagation in hierarchical-B coding structure and provided a content-related bit allocation scheme. These picture level bit allocation algorithms carefully investigated the distortion propagation influences of various frames on the sequence. However, the influences of the distortions of the geometry and attribute videos on the reconstructed point cloud have not been addressed until now.

In addition to the picture level bit allocation algorithms, there are many works focusing on the BU level bit allocation. Many works on the BU level bit allocation consider the BUs as independent since most BUs in inter frames obtain the prediction from the previously coded frames. The most straightforward method is using the complexity of a BU as the measure to determine its target bits. Seo *et al.* [27] introduced a combination of the variance of difference (VOD) and MAD as the complexity measure and performed BU level bit allocation based on it. However, the BU level bit allocation should be essentially determined by the RD characteristics instead of only the distortion. Yuan *et al.* [28] proposed a linear model between the distortion and Q, and derived the BU level bit allocation based on the model. He and Mitra [29] introduced a second order model between the distortion and ρ, and deducted the BU level bit allocation algorithm. In addition, Li *et al.* [7] proposed the λ-domain BU level bit allocation algorithm under the constraint of the picture level target bits. Guo *et al.* [30] provided a BU level rate control by considering the inter dependency directly in the block level.

There are also many works focusing on the dependent BU level bit allocation which considers the dependency among various BUs within one frame. The dependent BU level bit allocation is usually considered for the intra frame instead of the inter frame. Ferguson and Allinson [31] introduced the dependent quantization to video coding and proposed the modified steepest-descent algorithm to solve the problem. This algorithm showed significant gains on the I frames with strong dependency while only very small gains on the P frames with weak dependency. Lee and Song [32] proposed using the gradient as the complexity measure and introduced intra BU level bit allocation algorithm based on the complexity. Wang *et al.* [33] extended this idea for the λ-domain rate control algorithm and applied it to HEVC intra frame rate control. In addition, Gao *et al.* [34] formulated the bit allocation as a game theory problem and proposed optimizing the structure similarity (SSIM) [35] instead of PSNR. However, all these BU level bit allocation schemes are unable to solve the problem of unoccupied BUs in the V-PCC framework.

Except for the bitrate control and bit allocation algorithms, some other factors may have significant influences on the rate control such as the initial encoding parameter determination [36] [37] and the accuracy of the picture and BU level model parameters. For example, Li *et al.* [7] proposed calculating the model parameters from the RD characteristics from the co-located BU. Instead of estimating the model parameters from the previous BU, Chen and Pan [38] taken the distributions of the model parameters into consideration. In addition, Li *et al.* [39] proposed using the convolutional neural network (CNN) [40] to estimate the intra model parameters more accurately. However, these methods are unable to solve the problem that the current BU and its corresponding BU are not put the same position at all.

B. Video-based Point Cloud Compression

The conversion from a point cloud to videos can be roughly divided into three stages: clustering, packing, and padding. Initial clustering of the point cloud is first obtained by associating each point with the face having the most similar normal. The cluster is then refined by smoothing the generated patches. After patch generation, the patch-based projection method uses a simple packing strategy to organize the patches into videos. The patch location is determined through an exhaustive search in the raster scan order. The padding process then fills the empty space between the patches to make the generated frames more suitable for video coding. Note that the videos include geometry and attribute videos to record the geometry and attribute information.

To recover the point cloud from the encoded videos, some header information is signaled to the decoder. An illustration of the header information is shown in Fig. 2. In Fig. 2, we give an example of the patch projected to the yoz plane. For each patch, we need to signal the index of the projected

TABLE I
BRIEF INTRODUCTIONS OF THE HEADER INFORMATION

Auxiliary information	Notation	Explanation
Projection plane	n	Indication of the projected plane; value range: 0, 1, 2.
2D bounding box	$(u0, v0)$ $(u1, v1)$	Top left 2D position of the bounding box; bpr basis. Size of the 2D bounding box; bpr basis.
3D location	$\delta 0$ $s0$ $r0$	Minimum depth of the current patch. Tangential shift of top left position of the current patch. Bi-tangential shift of the top left position of the current patch.
Block to patch	bpr BP	Block resolution; typical value: 16. Block to patch array; bpr basis.
Occupancy map	opr OM	Occupancy map precision; typical value: 4 or 2. Occupancy map array; or basis.

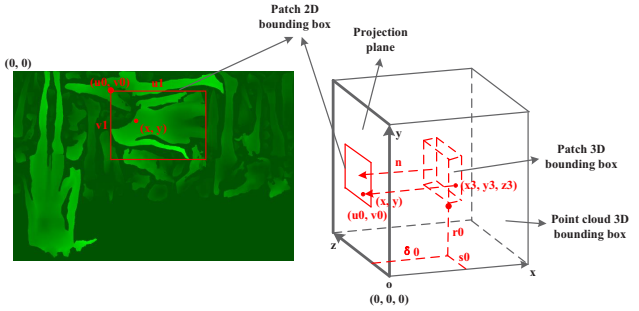


Fig. 2. Illustration of the header information.

TABLE II
PERFORMANCE OF DOWN-SAMPLING THE OCCUPANCY MAP BY 4 TIMES FOR ALL THE BITRATES COMPARED WITH THE V-PCC REFERENCE SOFTWARE.

Test point cloud	Geom.BD-GeomRate D1	Geom.BD-GeomRate D2	Attr.BD-AttrRate Luma	Attr.BD-AttrRate Cb	Attr.BD-AttrRate Cr
Loot	0.5%	0.8%	0.3%	0.0%	–0.3%
RedAndBlack	0.3%	0.9%	0.1%	0.4%	0.2%
Soldier	–0.2%	0.6%	0.8%	0.8%	0.9%
Queen	0.2%	0.8%	–0.5%	–1.1%	–1.1%
LongDress	0.1%	0.6%	0.4%	0.3%	0.2%
Avg.	**0.1%**	**0.7%**	**0.2%**	**0.1%**	**0.0%**

plane n, the 2D bounding box $(u0, v0, u1, v1)$, and the 3D location $(\delta 0, s0, r0)$. Additionally, for each block with size $bpr \times bpr$, the array BP indicates which patch it belongs to. for each block with size $opr \times opr$, the array OM indicates whether it is occupied or not. A detailed explanation of the header information is shown in Table I. Note that both the 2D bounding box and the block to patch information are signaled based on the bpr to save bits.

Based on the header information, from Fig. 2, we can derive the 3D to 2D correspondence from $(x3, y3, z3)$ to (x, y) as follows,

$$\begin{cases} x = (z3 - s0) + u0 \times bpr \\ y = (y3 - r0) + v0 \times bpr \\ h(x, y) = x3 - \delta 0, \end{cases} \tag{1}$$

where $h(x, y)$ is the pixel value of position (x, y) in the geometry frame. Also, we can derive the 2D to 3D correspondence as follows,

$$\begin{cases} x3 = \delta 0 + h(x, y) \\ y3 = y - v0 \times bpr + r0 \\ z3 = x - u0 \times bpr + s0. \end{cases} \tag{2}$$

These 3D to 2D relationships will be used to find the corresponding BU in the previous frame.

III. PROPOSED ALGORITHMS

As we have mentioned in Section I, a rate control algorithm can be divided into bit allocation and bitrate control. We will introduce the proposed bit allocation and bitrate control algorithms in Section III-A and Section III-B, respectively.

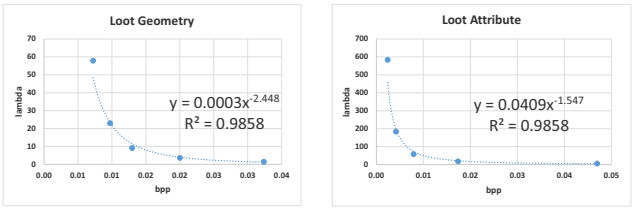
A. Bit allocation

The bit allocation includes four levels: video level, GOP level, picture level, and BU level.

1) Video level bit allocation: The video level bit allocation aims to assign the total bits to the header information including auxiliary information and occupancy map, the geometry video, and the attribute video. The auxiliary information is always encoded losslessly. The occupancy map is encoded losslessly after down-sampling. However, the occupancy map resolution opr that indicates the down-sampling ratio in both horizontal and vertical directions needs to be determined. In the current V-PCC reference software, the opr is related to the bitrate. It is set as 2 and 4 in the highest bitrate and all the other bitrates, respectively.

To find a good relationship between the target bits TR_T and opr, we compare the performance of setting the opr to 4 for all the bitrates with the V-PCC reference software default setting as shown in Table II. The experimental results show that setting the opr to 4 suffers an average of 0.1% and 0.7% losses for the geometry under D1 and D2 measurements, respectively. In terms of the attribute, it leads to 0.2%, 0.1%, and 0.0% RD performance losses for the Luma, Cb, and Cr components, respectively. The experimental results obviously demonstrate that setting the opr to 4 leads to almost no performance losses for all the target bitrates within a normal range. In the extreme high target bitrate case, the opr should be set as 2. Therefore, the opr is determined using the following equation according to the target bits per point $TBpp_T$,

$$\begin{cases} opr = 2, & TBpp_T > 1.0 \\ opr = 4, & TBpp_T \leq 1.0. \end{cases} \tag{3}$$



(a) Loot Geometry (b) Loot Attribute

(c) RedAndBlack Geometry (d) RedAndBlack Attribute

Fig. 3. Validation of the hyperbolic $R-\lambda$ relationship.

After the *opr* is determined, the actual bits of the header information AR_H are fixed. The total target bits of the geometry and attribute videos TR_V are calculated by subtracting AR_H from TR_T,

$$TR_V = TR_T - AR_H. \tag{4}$$

Before introducing the video level bit allocation to assign TR_V to the geometry and attribute videos, we first validate that the hyperbolic $R-\lambda$ model proposed in [8] is still valid for the geometry and attribute videos,

$$\lambda = \alpha R^{\beta}. \tag{5}$$

We count the R and λ of the geometry and attribute videos under various bitrates. We fit the data using the hyperbolic model as shown in Fig. 3. From Fig 3, we can see that the determination coefficients R^2 of all the fitted curves are larger than 0.98. The experimental results obviously demonstrate that the hyperbolic model is still valid for the geometry and attribute videos. Additionally, note that αs of the geometry and attribute videos show a large difference. The α of the attribute is 100 times larger than that of the geometry.

The target of the video level bit allocation is to minimize the distortion of the reconstructed point cloud under the constraint of TR_V. The distortion of the reconstructed point cloud is modeled as a weighted combination of the geometry and attribute distortions,

$$\min_{TR_G, TR_A} \omega TD_G + TD_A, \quad TR_G + TR_A \le TR_V. \tag{6}$$

where TD_G and TD_A are the distortions for the geometry and attribute videos, respectively. TR_G and TR_A are the target bits for the geometry and attribute videos, respectively. ω indicates the relative importance of the geometry and attribute videos. Generally, the geometry quality is more important than the attribute quality. Therefore, ω is always larger than 1.

The constrained problem is converted to an unconstrained problem by introducing a Lagrangian Multiplier λ,

$$\min_{TR_G, TR_A} \omega TD_G + TD_A + \lambda(TR_G + TR_A). \tag{7}$$

The unconstrained problem is solved using the Lagrangian method by setting the derivatives to TR_G and TR_A as 0. Since the geometry and attribute videos are encoded independently, the unconstrained problem is solved using the following two equations,

$$\omega \times \frac{\partial TD_G}{\partial TR_G} + \lambda = 0, \tag{8}$$

$$\frac{\partial TD_A}{\partial TR_A} + \lambda = 0. \tag{9}$$

As λ_G and λ_A are the slopes of the RD curves of the geometry and attribute videos,

$$\lambda_G = -\frac{\partial TD_G}{\partial TR_G}, \quad \lambda_A = -\frac{\partial TD_A}{\partial TR_A}. \tag{10}$$

Through substituting (10) into (8) and (9), we can derive the solution as follows,

$$\lambda_G = \frac{\lambda_A}{\omega}. \tag{11}$$

As we can see from (11), the larger the ω is, the smaller the λ of the geometry video is. This leads to more bits assigned to the geometry video while fewer bits assigned to the attribute video. Therefore, the larger the ω is, the better the RD performance of the geometry is. However, it will lead to worse RD performance for the attribute.

In addition to the constraints shown in (11), we have the TR_V as the other constraint,

$$TR_G + TR_A = TR_V. \tag{12}$$

Through substituting (12) and the hyperbolic model in (5) into (11), we can derive the following equation,

$$\omega \times \alpha_G TR_G^{\beta_G} - \alpha_A (TR_V - TR_G)^{\beta_A} = 0. \tag{13}$$

In (13), the only unknown parameters are α_G, β_G, α_A, β_A, and ω. The α_G, β_G, α_A, and β_A are sequence-related parameters. We obtain them through encoding the first 32 frames of the point cloud twice using different bitrates. We choose the bitrates $r2$ and $r4$ defined in the V-PCC common test condition (CTC) [41] to calculate the parameters. The ω is set as 8 in the overall experiment according to our experience. We will analyze the influence of ωs on the performance of video level bit allocation in the experimental results.

As shown in Fig. 3, the α_G and α_A are positive values and the β_G and β_A are negative values. Therefore, the left term of (13) is a monotone function of TR_G. We can solve it using the Bisection Method. After TR_G is determined, we encode the geometry video and obtain its actual bits AR_G. Then we calculate the target bits of the attribute video TR_A as follows,

$$TR_A = TR_V - AR_G. \tag{14}$$

The target bits will guide the encoding process of the attribute video.

TABLE III
Ω_{Pic_i} SETTING FOR THE GEOMETRY AND ATTRIBUTE VIDEOS

Frame level	Geometry	Attribute
0	1.0	1.0
1	1.0	1.5874
2	1.0	3.1748
3	1.0	5.0397
4	1.0	6.3496

2) GOP level bit allocation: The GOP level bit allocation has a simple aim in making the bit stream more adaptable to the bandwidth. We follow the current method in the HM to determine the GOP level target bits TR_{GOP},

$$TR_{GOP} = \frac{AVGR_{Pic} \times (N_{Coded} + SW) - R_{Coded}}{SW} \times N_{GOP}, \tag{15}$$

where $AVGR_{Pic}$ is the average bits per picture. The N_{Coded} and R_{Coded} are the number of coded pictures and spent bits, respectively. The N_{GOP} is the number of pictures in a GOP. The SW is the size of the sliding window that aims to make the bits adjustment more smoothly. The SW used in our experiments is set to 40.

3) Picture level bit allocation: The picture level bit allocation can be divided into intra frame bit allocation and inter frame bit allocation. We follow the current method in the HM to determine the bits of the intra frame. The intra frame target bits is proportional to the frame complexity that is measured using the sum of absolute transformed difference (SATD). The inter frame bit allocation is related to both the reference structures and video content. As proved in [6], the λ ratio of various pictures should be inversely proportional to their influences on the subsequent pictures,

$$\frac{\lambda_{Pic_i}}{\lambda_{Pic_j}} = \frac{\Omega_{Pic_j}}{\Omega_{Pic_i}}. \tag{16}$$

Additionally, the GOP level target bits are used as a constraint of the sum of the target bits of all the pictures in the GOP,

$$\sum_{i=1}^{N_{GOP}} TR_{Pic_i} = TR_{GOP}, \tag{17}$$

where TR_{Pic_i} is the picture level target bits of the ith picture in the GOP. Combining (5), (16), and (17), we can derive the target bits for each picture.

To solve the above equations, we still need to derive the unknown parameters including the hyperbolic model parameters and Ω_{Pic_i}. The determination of the model parameters for each picture α_{Pic_i} and β_{Pic_i} will be introduced in the bitrate control process. Then the only unknown parameter is the Ω_{Pic_i}. We followed the V-PCC CTC to set the Ω_{Pic_i}. It is different from the setting in HEVC due to the following two reasons. First, corresponding patches that put in different positions of neighboring frames make the inter correlations become less. Second, the geometry video has fewer correlations compared with the attribute video as the neighboring depth frames are less related. The detailed settings of the Ω_{Pic_i} for geometry and attribute videos are shown in Table III. Note that the pictures at the same hierarchical level have the same Ω_{Pic_i}. We test under the random access coding structure with GOP size 16. Therefore, there are five hierarchical levels. Note that the Ω_{Pic_i} of the pictures at level 0 is set as 1.0 and used as the basis for the other pictures.

4) Basic unit level bit allocation: The optimization target of the BU level bit allocation is to minimize the reconstructed quality of the point cloud frame under the constraint of TR_{Pic}. When performing the BU level bit allocation, it is difficult to accurately measure the influence of each BU on the reconstructed quality of the point cloud frame. For simplification, we consider the BUs containing occupied pixels have the same importance. The BUs containing only unoccupied pixels will have no influences on the reconstructed quality. Therefore, we formulate the BU level bit allocation problem as follows,

$$\min_{\lambda_{BU_j}} \sum_{i=1}^{N_{Pic}} \sum_{BU_i \in BU_O} TD_{BU_i}, \quad s.t. \sum_{i=1}^{N_{Pic}} TR_{BU_i} = TR_{Pic}, \tag{18}$$

where TD_{BU_i} and TR_{BU_i} are the target distortion and bits for the ith BU, respectively. N_{Pic} is the number of BUs in the picture. BU_O is the set of the occupied BUs in the picture. As shown in (18), we care about the distortion of the occupied PUs and ignore the distortions of the unoccupied PUs.

The unconstrained problem is converted to the following unconstrained problem by introducing the Lagrangian Multiplier λ,

$$\min_{\lambda_{BU_j}} \sum_{i=1}^{N_{Pic}} \sum_{BU_i \in BU_O} TD_{BU_i} + \lambda \sum_{i=1}^{N_{Pic}} TR_{BU_i}. \tag{19}$$

This unconstrained problem is solved by setting its derivative to λ_{BU_j} as 0,

$$\frac{\partial \sum_{i=1}^{N_{Pic}} \sum_{BU_i \in BU_O} TD_{BU_i}}{\partial \lambda_{BU_j}} + \lambda \frac{\partial \sum_{i=1}^{N_{Pic}} TR_{BU_i}}{\partial \lambda_{BU_j}} = 0. \tag{20}$$

If $BU_j \in BU_O$, (20) is converted to the following equation,

$$\frac{\partial TD_{BU_j}}{\partial \lambda_{BU_j}} + \lambda \frac{\partial TR_{BU_j}}{\partial \lambda_{BU_j}} = 0, \quad if\ BU_j \in BU_O. \tag{21}$$

Since λ_{BU_j} is the slope of the current BU, we can have

$$\lambda_{BU_j} = \lambda, \quad if\ BU_j \in BU_O. \tag{22}$$

Eq. (22) is in accordance with the conclusion in [6] and [7]. The λ_{BU_j} of each BU should be set as equal as possible to optimize the RD performance of the current frame.

If $BU_j \notin BU_O$, (20) is converted to the following equation,

$$\lambda \frac{\partial TR_{BU_j}}{\partial \lambda_{BU_j}} = 0, \quad if\ BU_j \notin BU_O. \tag{23}$$

From (5), we can have

$$\frac{\partial TR_{BU_j}}{\partial \lambda_{BU_j}} = \frac{1}{\beta}(\frac{1}{\alpha})^{\frac{1}{\beta}} \lambda_{BU_j}^{(\frac{1}{\beta}-1)} \triangleq \alpha_1 \lambda_{BU_j}^{\beta_1}. \tag{24}$$

As β_1 is negative, through substituting (24) into (23), we can have

$$\lambda_{BU_j} = +\infty, \quad if\ BU_j \notin BU_O. \tag{25}$$

According to (5), we can see that (25) indicates the unoccupied BUs should be assigned 0 bits. This conclusion is in accordance with our common sense that the unoccupied BUs that have no influences on the reconstructed quality. Additionally, the unoccupied BUs should be encoded using the infinity λ_{BU_j}. In our implementation, to avoid data overflow during the calculation of the RD cost, we set λ_{BU_j} as 160000.

In addition to the constraint shown in (22), the sum of the target bits of all the occupied BUs should follow the picture level target bits TR_{Pic},

$$\sum_{i=1}^{N_{Pic}} \sum_{BU_i \in BU_O} TD_{BU_i} = TR_{Pic}. \tag{26}$$

Combining (22), (5), and (26), we can derive the target bits TD_{BU_i} for each BU. Additionally, we modify two other aspects of the original BU level rate control algorithm in HM to make it in accordance with the proposed BU level bit allocation. First, the λ_{BU_j} is clipped in a limited range of the λ_{Pic}. The λ_{BU_j} of unoccupied BUs is not restricted within a limited range of λ_{Pic} in the proposed algorithm. Second, only the distortions of the occupied BUs are counted into the overall distortion of the current picture.

B. Bitrate control

The above bit allocation processes assign the bits to each picture and BU. In this section, we will determine the λ and quantization parameter QP for each picture and BU to finish the encoding process. We use (5) to determine the λ. Then the QP is determined using

$$QP = 4.3281 \times ln\lambda + 14.4329. \tag{27}$$

The only problem left in both the bit allocation and rate control processes is how to determine the model parameters α and β. In [7], it is proposed that the rate control model parameters are calculated using

$$\beta = -\frac{R\lambda}{D} - 1, \tag{28}$$

$$\alpha = \lambda R^{(1+\frac{R\lambda}{D})}. \tag{29}$$

During the rate control process, the current picture or BU has not been encoded yet. They use λ, R, and D of the previously encoded picture or the co-located BU at the same hierarchical level to estimate the current model parameters.

The above method works well for the update of the picture level model parameters under the V-PCC framework. However, for the BU level bitrate control, the corresponding BUs may be put in different positions. An occupied BU in the current frame may correspond to an unoccupied BU in the co-located position. Therefore, using the co-located BU to estimate the model parameters becomes not working at all. In this paper, we use the auxiliary information to find the corresponding BU in the previous frame to obtain more accurate model parameters for the current BU.

We first handle the case where some pixels are occupied in the current BU according to the occupancy map of the current picture OM_c. If the center pixel of the current BU

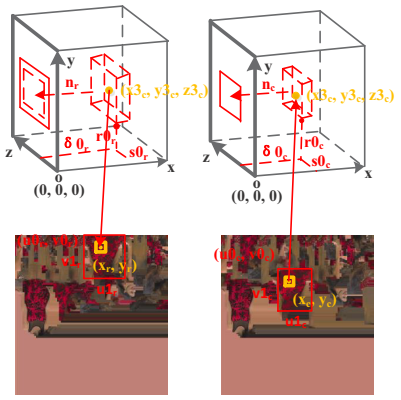


Fig. 4. Illustration of finding the corresponding BU.

is occupied, we will use it as the representative of the current BU. If the center pixel of the current BU is unoccupied, we will go through the current BU using the raster scan and use the first occupied pixel as the representative of the current BU. The representative pixel is denoted as (x_c, y_c). The idea of finding the corresponding BU (x_r, y_r) is shown in Fig. 4. Based on the block to patch information BP_c, we can find the current patch (x_c, y_c) belongs to. The patch can provides us with the patch projected plane n_c, the patch 2D bounding box $(u0_c, v0_c)$, the patch 3D location $(\delta 0_c, s0_c, r0_c)$. With all these information, we can calculate the corresponding 3D coordinate $(y3_c, z3_c)$ using (2).

We then search all the patches in the previously coded frame at the same hierarchical level to find the corresponding patch. The patches should at least follow two constraints. First, the projected plane of the reference patch n_r should be the same as n_c. If the projected plane is different, the possibility to find the corresponding patch will be very small. Even if the current BU can find the corresponding BU in the previous frame, different projection planes will lead to various shape changes of the corresponding BUs. This will make the estimation of the model parameters inaccurate. Therefore, we bypass the reference patches with different projected planes from the current patch. Second, the $(y3_c, z3_c)$ should be within the range of the reference patch,

$$\begin{cases} s0_r \le z3_c \le s0_r + u1_r \times bpr - 1 \\ ro_r \le y3_c \le r0_r + v1_r \times bpr - 1. \end{cases} \tag{30}$$

If the reference patch does not contain $(y3_c, z3_c)$, it will be impossible for the reference patch to have the current BU.

There may be multiple patches following the above two constraints. For example, if one reference patch can satisfy the constraint, the patch in its opposite position in 3D space also has a good chance of satisfying the constraint. The differences between these patches are the values of the $\delta 0_r$. We will choose the patch with the smallest difference between $\delta 0_r$ and $\delta 0_c$ since the motion is usually not large between the current frame and the previously coded frame at the same level.

After the reference patch is found, based on the assumption that the 3D coordinates of the current pixel are the same as

TABLE IV
CHARACTERISTICS OF THE TEST DYNAMIC POINT CLOUD

Test point cloud	Frame rate	Number of points	Geometry precision	Attributes
Loot	30	~ 780000	10 bit	RGB
RedAndBlack	30	~ 700000	10 bit	RGB
Soldier	30	~ 1500000	10 bit	RGB
Queen	50	~ 1000000	10 bit	RGB
LongDress	30	~ 800000	10 bit	RGB

that of the reference pixel,

$$\begin{cases} s0_r + (x_r - u0_r \times bpr) = s0_c + (x_c - u0_c \times bpr) \\ r0_r + (y_r - v0_r \times bpr) = r0_c + (y_c - v0_c \times bpr). \end{cases} \tag{31}$$

The coordinate of the reference BU (x_r, y_r) can be calculated as

$$\begin{cases} x_r = x_c + (s0_c - s0_r) + (u0_r - u0_c) \times bpr \\ y_r = y_c + (r0_c - r0_r) + (v0_r - v0_c) \times bpr. \end{cases} \tag{32}$$

As we can see from (32), the right term is a combination of the 3D and 2D patch offsets. Therefore, this coordinate offset is essentially a global patch offset. We use the λ, R, and D of the BU containing (x_r, y_r) to update the model parameters of the current BU using (28).

For the case where no occupied pixels exist according to the occupancy map, these BUs are usually very smooth. They can share the model parameters. Therefore, we always obtain the model parameters from the previously coded BU at the same hierarchical level. Additionally, as we have mentioned in Fig. 3, the αs of the geometry and attribute videos differ significantly. Therefore, we set the initial α of the geometry as the default one in HM dividing 10. We set the initial α of the attribute as the default one in HM multiplying 10.

IV. EXPERIMENTAL RESULTS

The proposed algorithms are implemented in the V-PCC reference software TMC2-5.0 [11] and the corresponding HM [12] to compare with the V-PCC anchor without control. We also use the λ-domain rate control algorithm designed for the general videos as the anchor for comparison. Note that the differences between the proposed algorithm and the anchor are only the proposed BU level bit allocation and model updating algorithms. We set all the other parameters the same including the initial model parameters for a fair comparison. Note that we use the same video level bit allocation algorithm for the anchor and proposed rate control algorithm since there is no video level bit allocation algorithm yet designed for the V-PCC. For both the anchor and the proposed rate control algorithm, the target bits are generated as follows. We run the TMC2-5.0 anchor without rate control and count the number of bits ranging from the low bitrate r1 to the high bitrate r5 following the V-PCC CTC. The counted bits are then used as the target bits for both the anchor and the proposed rate control algorithm.

We test the lossy geometry, lossy attribute, random access (RA) case to demonstrate the effectiveness of the proposed algorithms. We perform the experiments on the five dynamic point clouds defined in the V-PCC CTC [41]. The detailed characteristics of the test point clouds are shown in Table IV. In the current V-PCC reference software, every 32 frame is encoded as independent sequences. Therefore, we test 32 frames as a good representative for the whole point cloud to verify the performance of the proposed algorithms. Since the bits generated by the anchor and the proposed algorithms are not the same, the Bjontegaard-Delta-rate (BD-rate) [42] is used to compare the respective RD performances.

TABLE V
OVERALL PERFORMANCE OF THE PROPOSED RATE CONTROL ALGORITHM COMPARED WITH THE ORIGINAL RATE CONTROL ALGORITHM

Test point cloud	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	–1.7%	–2.0%	–2.1%	–1.6%	–2.4%
RedAndBlack	–0.1%	–0.3%	–3.8%	–4.9%	–5.5%
Soldier	–2.1%	–2.9%	–0.1%	0.5%	1.2%
Queen	–2.7%	–2.5%	–1.5%	0.6%	0.5%
LongDress	–3.3%	–3.1%	–4.4%	–5.4%	–5.4%
Avg.	**–2.0%**	**–2.2%**	**–2.4%**	**–2.1%**	**–2.3%**

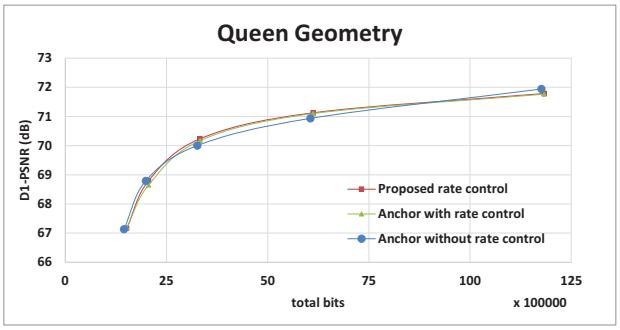
TABLE VI
OVERALL PERFORMANCE OF THE PROPOSED RATE CONTROL ALGORITHM COMPARED WITH THE TMC2-5.0 ANCHOR

Test point cloud	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	6.1%	4.5%	2.7%	0.2%	1.0%
RedAndBlack	12.5%	12.9%	2.3%	2.0%	2.5%
Soldier	–7.3%	–6.2%	1.0%	–2.4%	–1.3%
Queen	–2.3%	–1.9%	6.7%	0.2%	2.2%
LongDress	15.7%	11.1%	7.8%	–2.8%	0.6%
Avg.	**4.9%**	**4.1%**	**4.1%**	**–0.6%**	**1.0%**

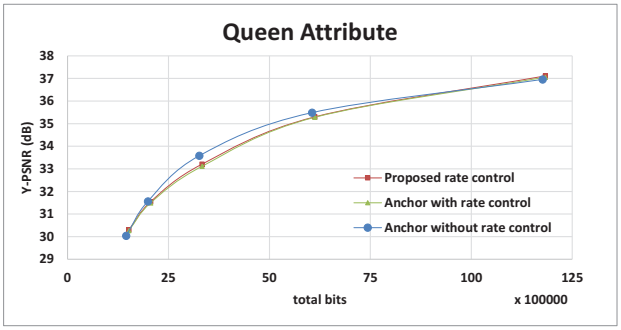
We show the benefits of the proposed algorithm in two aspects: the RD performance and the bits error between the target bits and actual bits. For the RD performance of the geometry, we report the BD-rates for both point-to-point PSNR (D1) and point-to-plane PSNR (D2) [41]. For the RD performance of the attribute, the BD-rates for the Luma, Cb, and Cr components are reported. In the following subsections, we first introduce the overall performance of the proposed rate control algorithm. Then we report the performance and analysis of the proposed algorithms individually.

A. Overall performance of the proposed rate control algorithm

Table V shows the overall RD performance of the proposed rate control algorithm compared to the anchor with rate control. Note that the ω is set as 8 in the overall experimental result. We will show the influences of different ωs in the next subsection. In addition, we only enable the proposed BU level bit allocation algorithm for the attribute instead of the geometry. This will also be explained in the following subsections. From Table V, we can see that the proposed algorithm can lead to 2.0% and 2.2% RD performance improvements on average for the geometry under the D1 and D2 measurements, respectively. The proposed algorithm can save an average of 2.4%, 2.1%, and 2.3% bits for the Luma, Cb, and Cr components, respectively. The proposed algorithms bring up to 3.3% and 3.1% RD performance improvements for



(a) Queen Geometry



(b) Queen Attribute

Fig. 5. Typical examples of the RD curves.

the geometry for the point cloud "LongDress". In addition, it brings 4.4%, 5.4%, and 5.4% bits savings for the attribute accordingly. The experimental results demonstrate that the proposed rate control algorithm can lead to obvious bits savings compared to the anchor with rate control.

Table V shows the overall RD performance of the proposed rate control algorithm compared to the anchor without rate control. First, we can see that the proposed algorithm leads to about 4% to 5% bits increase compared to the anchor without rate control. As the intra frame costs much more bits than the inter frames, the immediate inter frames after the intra frame are assigned much fewer bits compared with the anchor to keep the balance of the buffer. Those frames will suffer some quality degradations that will propagate to the subsequent frames. That is why we suffer some performance losses on average. Second, the performance of each individual point cloud varies significantly. This is because the anchor without rate control and the proposed rate control algorithm are setting different ωs. The anchor without rate control is using larger ωs for lower bitrate while smaller ωs for higher bitrate. Additionally, we show the geometry and attribute RD curves for "Queen" for validation in Fig. 5. The proposed rate control algorithm shows better RD performance compared with the anchor with rate control. However, it leads to some improvements for some target bits while some losses for the other target bits compared to the anchor without rate control.

Table VII shows the comparison of the bits error between the anchor and the proposed rate control algorithm. We can see that the proposed rate control algorithm achieves smaller average bits error as well as maximum bits error compared with the anchor. The smaller bits error mainly comes from

TABLE VII
BITS ERRORS COMPARISON BETWEEN THE ANCHOR AND THE PROPOSED RATE CONTROL ALGORITHM

Test case	Point cloud	r1	r2	r3	r4	r5
Anchor	Loot	2.65%	2.16%	1.41%	0.88%	0.40%
	RedAndBlack	6.23%	4.09%	2.93%	2.07%	0.95%
	Soldier	4.15%	3.21%	2.32%	1.29%	0.46%
	Queen	4.73%	3.45%	2.11%	1.13%	0.58%
	LongDress	4.04%	2.91%	2.04%	1.67%	0.46%
	Avg.			**2.33%**		
	Max			**6.23%**		
Proposed	Loot	2.57%	2.18%	1.42%	0.81%	0.40%
	RedAndBlack	5.36%	4.00%	2.82%	1.72%	0.83%
	Soldier	4.16%	3.34%	2.35%	1.29%	0.56%
	Queen	4.73%	3.44%	2.10%	1.13%	0.58%
	LongDress	3.79%	2.40%	1.72%	0.75%	0.22%
	Avg.			**2.19%**		
	Max			**5.36%**		

TABLE VIII
PERFORMANCE OF THE PROPOSED RATE CONTROL ALGORITHM WITH ω SET AS 6 COMPARED WITH THAT WITH ω SET AS 8

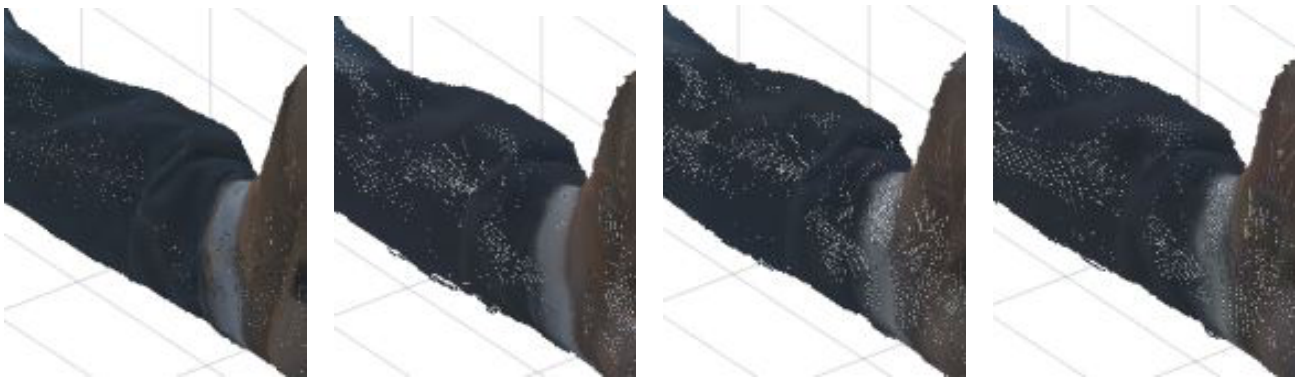
Test point cloud	Geom.BD-TotalRate D1	D2	Attr.BD-TotalRate Luma	Cb	Cr
Loot	10.1%	10.7%	–6.5%	–5.1%	–6.7%
RedAndBlack	10.4%	10.5%	–5.2%	–6.0%	–5.6%
Soldier	14.1%	14.6%	–4.1%	–2.4%	–3.8%
Queen	20.2%	20.4%	–0.6%	–2.6%	–3.9%
LongDress	24.4%	23.7%	–1.6%	–3.5%	–3.2%
Avg.	**15.8%**	**15.9%**	**–3.6%**	**–3.9%**	**–4.6%**

a more reasonable bit allocation and a better model updating scheme. Additionally, we can see that both the anchor and the proposed algorithm lead to much more serious bits errors in low bitrate case compared with those in high bitrate case. The portion of the intra bits in low bitrate case is higher than that in high bitrate case. The larger portion of intra bits leads to fewer bits available for the subsequent frames and brings larger errors.

B. Performance of the video level bit allocation

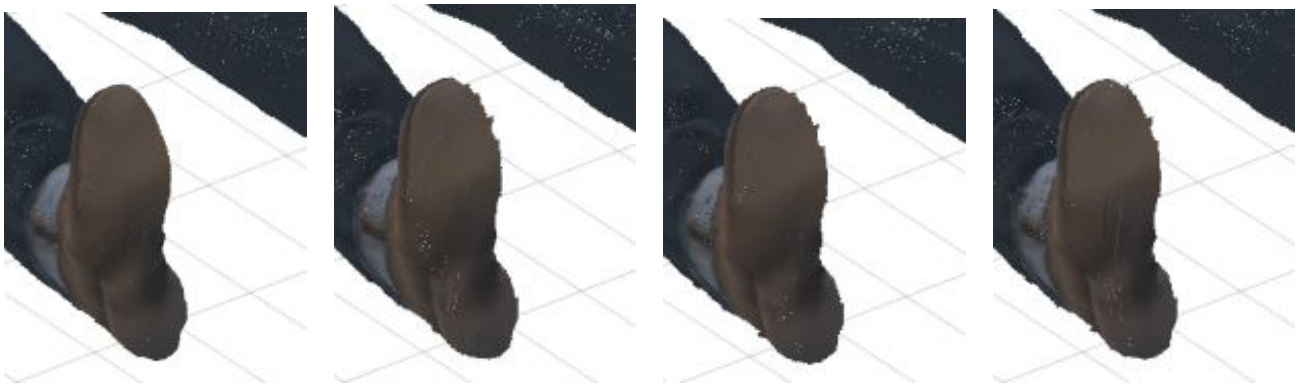
In this section, we will analyze the performance of the video level bit allocation algorithm under different ωs. Table VIII and Table IX show the performance of the proposed algorithm with ωs set as 6 and 10 compared to ω set as 8, respectively. We can see that the larger the ω is, the better RD performance we can achieve for the geometry. However, the worse RD performance we will obtain for the attribute. This is in accordance with our analysis in Section III-A that larger λ corresponds to better RD performance for the geometry while worse RD performance for the attribute. However, we are still confused about how to select a better ω for the overall reconstructed point cloud quality.

Essentially, to obtain the optimal ω, we need a full experimental result on the subjective quality of reconstructed point cloud under different ωs. However, this is not the main focus of this work. We just show a few examples of subjective qualities in high and low bitrates to find a suitable ω as shown in Fig. 6 and Fig. 7. We can see very obvious geometry distortions in the low bitrate case if we set the ω as 6. In addition, very



(a) original (b) $\omega = 6$ (c) $\omega = 8$ (d) $\omega = 10$

Fig. 6. Subjective qualities of setting different ωs in low bitrate case. The example cropped from "Loot" with picture order count 1001. The target bits are 1310816



(a) original (b) $\omega = 6$ (c) $\omega = 8$ (d) $\omega = 10$

Fig. 7. Subjective qualities of setting different ωs in high bitrate case. The example cropped from "Loot" with picture order count 1001. The target bits are 8097008.

TABLE IX
PERFORMANCE OF THE PROPOSED RATE CONTROL ALGORITHM WITH ω SET AS 10 COMPARED WITH THAT WITH ω SET AS 8

Test point cloud	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	–7.0%	–6.6%	7.2%	9.9%	7.9%
RedAndBlack	–8.3%	–7.9%	7.0%	7.9%	7.0%
Soldier	–11.5%	–11.9%	5.4%	9.1%	7.6%
Queen	–12.4%	–12.8%	2.7%	3.2%	2.7%
LongDress	–15.8%	–15.1%	2.6%	3.9%	3.6%
Avg.	**–11.0%**	**–10.9%**	**5.0%**	**6.8%**	**5.8%**

TABLE XI
PERFORMANCE OF THE PROPOSED MODEL UPDATING ALGORITHM COMPARED TO THE ORIGINAL MODEL UPDATING ALGORITHM

Test point cloud	Geom.BD-TotalRate D1	Geom.BD-TotalRate D2	Attr.BD-TotalRate Luma	Attr.BD-TotalRate Cb	Attr.BD-TotalRate Cr
Loot	–1.7%	–2.0%	0.8%	1.5%	0.4%
RedAndBlack	0.0%	–0.1%	–0.6%	–1.1%	–1.3%
Soldier	–2.2%	–2.9%	0.9%	0.9%	2.1%
Queen	–2.7%	–2.5%	0.0%	1.6%	1.0%
LongDress	–3.2%	–3.0%	–2.5%	–4.0%	–3.6%
Avg.	**–1.9%**	**–2.1%**	**–0.3%**	**–0.2%**	**0.3%**

TABLE X
BITS ERRORS COMPARISON BETWEEN DIFFERENT ωS OF THE PROPOSED RATE CONTROL ALGORITHM

ω	Average bits error	Maximum bits error
6	2.17%	5.42%
8	2.19%	5.36%
10	2.19%	5.55%

obvious attribute distortions are observed in the high bitrate case if we set the ω as 10. The optimal ωs may vary under different target bits. In this paper, we choose ω as 8 to get a better performance balance in the low and high bitrate cases.

In addition to the RD performance, we show the bits errors of setting different ωs in Table X. We can see that the proposed algorithms have very similar average and maximum bits errors under various ωs. The different settings of ωs only have very small influences on the bits errors.

C. Performance of the proposed model parameter estimation

Table XI shows the performance of the proposed model updating algorithm compared with the original model updating algorithm. We can see that the proposed model updating algorithm leads to 1.9% and 2.1% bits savings on average under D1 and D2 quality measurements, respectively. It can bring an average of 0.3%, 0.2% and 0.3% performance improvements for the Luma, Cb, and Cr components, respectively.

TABLE XII
BITS ERRORS COMPARISON BETWEEN THE PROPOSED MODEL UPDATING ALGORITHM COMPARED TO THE ANCHOR WITH RATE CONTROL

	Average bits error	Maximum bits error
Anchor	2.33%	6.23%
Proposed	2.27%	5.73%

TABLE XIII
PERFORMANCE OF THE BU LEVEL BIT ALLOCATION ALGORITHM FOR ATTRIBUTE COMPARED WITH THE ORIGINAL BU LEVEL BIT ALLOCATION ALGORITHM

Test point cloud	Geom.BD-TotalRate D1	D2	Attr.BD-TotalRate Luma	Cb	Cr
Loot	2.8%	5.2%	–0.2%	1.0%	–2.0%
RedAndBlack	0.4%	1.8%	–1.1%	–1.0%	–1.1%
Soldier	2.2%	4.1%	1.1%	1.6%	–1.6%
Queen	5.7%	6.6%	1.2%	–1.3%	0.1%
LongDress	3.2%	5.0%	0.0%	0.2%	0.1%
Avg.	**2.8%**	**4.5%**	**0.2%**	**0.1%**	**–0.9%**

The experimental results show the proposed model updating algorithm can bring a better RD performance for the geometry compared with attribute. As we have shown in Section III-B, we use the center pixel to represent the current BU to find the corresponding pixel. However, the corresponding pixel may not be the center pixel of the corresponding BU. Therefore, the corresponding BUs are usually not with exactly the same content. This situation will be more serious for the attribute video whose texture is less smooth compared with that of the geometry video.

Table XII shows the average and maximum bits error comparison between the proposed model updating algorithm and the original model updating algorithm. We can see that the proposed model updating algorithm brings both smaller average and maximum bits errors. As most BUs can obtain more accurate model parameters, both smaller average and maximum bits errors are achieved. The experimental results demonstrate that the proposed model updating algorithm can bring better RD performance as well as smaller bits errors compared with the original model updating algorithm.

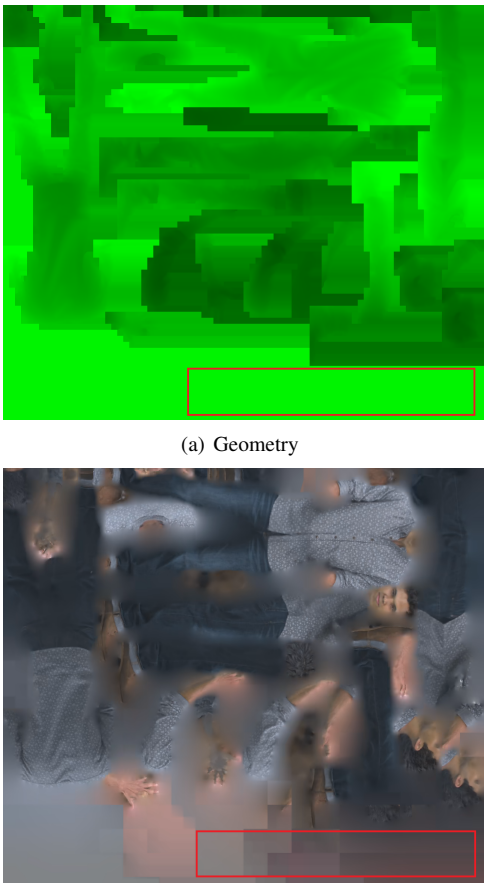
D. Performance of the proposed BU level bit allocation

Table XIII and Table XIV show the performance of the BU level bit allocation algorithm for the geometry and attribute compared with the original BU level bit allocation, respectively. We can see that the proposed BU level bit allocation algorithm leads to obvious RD performance improvements for the attribute. However, it brings some RD performance losses for the geometry. That is why we disabled the BU level bit allocation algorithm in the overall experimental results.

The different padding algorithms for the geometry and attribute are the key reasons why the proposed BU level bit allocation leads to different performances. An unoccupied geometry BU is padded using the horizontal extension from the left border pixels or the vertical extension from the top border pixels if the left border pixels are unavailable. The unoccupied geometry BU will always not cost many bits through intra horizontal prediction or vertical prediction even if the proposed BU level bit allocation is not used. However, an unoccupied attribute BU is padded using a pull-push algorithm to minimize the overall bitrate for both the completely unoccupied BU and partially unoccupied BU [43]. Therefore, the proposed BU level bit allocation has a larger chance to save some bits and improve the overall performance. To better illustrate this problem, we show an example of the reconstructed frames of the geometry and attribute videos in Fig 8. We can see very obvious blocking artifacts for the unoccupied BUs as indicated by the red triangles in the attribute frame. This is due to the bits savings for the residue of the unoccupied BUs. However, we cannot see similar artifacts in the geometry frame as the unoccupied BUs in the geometry frame is very smooth.

TABLE XIV
PERFORMANCE OF THE BU LEVEL BIT ALLOCATION ALGORITHM FOR GEOMETRY COMPARED WITH THE ORIGINAL BU LEVEL BIT ALLOCATION ALGORITHM

Test point cloud	Geom.BD-TotalRate D1	D2	Attr.BD-TotalRate Luma	Cb	Cr
Loot	0.0%	0.0%	–2.9%	–3.2%	–2.8%
RedAndBlack	–0.2%	–0.2%	–3.2%	–3.8%	–4.2%
Soldier	0.0%	0.0%	–1.0%	–0.3%	–0.9%
Queen	0.0%	0.0%	–1.5%	–0.9%	–0.5%
LongDress	–0.1%	–0.1%	–1.9%	–1.5%	–1.9%
Avg.	**–0.1%**	**–0.1%**	**–2.1%**	**–1.9%**	**–2.1%**



(a) Geometry

(b) Attribute

Fig. 8. Typical example of the reconstructed frame of the geometry and attribute. The example is from picture order count 5 of "Loot" with target bits 8097008.

TABLE XV
BITS ERRORS COMPARISON BETWEEN THE PROPOSED BIT ALLOCATION ALGORITHMS AND THE ORIGINAL BIT ALLOCATION ALGORITHM

	Average bits error	Maximum bits error
Original	2.27%	5.73%
Geometry	2.26%	6.30%
Attribute	2.19%	5.36%

Table XV shows the bits errors comparison between the proposed bit allocation algorithms and the original bit allocation algorithm. We can see that the proposed BU level bit allocation for the geometry leads to similar average bits error and a little bit higher maximum bits error compared with the original bit allocation algorithm. Since the overall bits errors are mainly determined by the rate control accuracy of the attribute video, these two algorithms show similar bits errors. The proposed BU level bit allocation for the attribute leads to smaller average and maximum bits errors due to a more reasonable bit allocation.

V. CONCLUSION

In this paper, we propose the first rate control algorithm for the video-based point cloud compression (V-PCC) framework. We mainly have the following key contributions to address the new features brought by the V-PCC framework. First, we propose a video level bit allocation algorithm to assign the bits between the geometry and attribute videos. Second, we introduce a basic unit (BU) level bit allocation algorithm assigning zero bits to the unoccupied BUs by ignoring their distortions. Third, we propose a more accurate model updating scheme by finding the corresponding BUs with similar video content in the previous frame at the same hierarchical level. The proposed rate control algorithms are implemented in the V-PCC and the corresponding High Efficiency Video Coding (HEVC) reference software. The experimental results show that the proposed algorithms can lead to very obvious rate distortion performance improvements as well as smaller bits errors. We will investigate the optimal video, picture level bit allocation algorithms in our future work to achieve even better performance.

REFERENCES

[1] H. Fuchs, A. State, and J. Bazin, "Immersive 3D Telepresence," *Computer*, vol. 47, no. 7, pp. 46–52, Jul. 2014.

[2] M.-L. Champel, R. Doré, and N. Mollet, "Key Factors for a High-Quality VR Experience," in *2017 SPIE Optical Engineering and applications*, vol. 10396, 2017.

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

[4] G. J. Sullivan, J. Ohm, W. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec. 2012.

[5] 3DG, "Continuous Improvement of Study Text of ISO/IEC CD 23090-5 Video-based Point Cloud Compression," Document ISO/IEC JTC1/SC29/WG11 w18479, Geneva, Switzerland, Mar. 2019.

[6] L. Li, B. Li, H. Li, and C. W. Chen, "λ-Domain Optimal Bit Allocation Algorithm for High Efficiency Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 130–142, Jan. 2018.

[7] S. Li, M. Xu, Z. Wang, and X. Sun, "Optimal Bit Allocation for CTU Level Rate Control in HEVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 11, pp. 2409–2424, Nov. 2017.

[8] B. Li, H. Li, L. Li, and J. Zhang, "λ Domain Rate Control Algorithm for High Efficiency Video Coding," *IEEE Transactions on Image Processing*, vol. 23, no. 9, pp. 3841–3854, Sept. 2014.

[9] B. Li, H. Li, L. Li, and J. Zhang, "Rate Control by R-lambda Model for HEVC," Document ISO/IEC JTC1/SC29/WG11 JCTVC-K0103, Shanghai, CN, Oct. 2012.

[10] Y. Li, Z. Chen, X. Li, and S. Liu, "Rate Control for VVC," Document ISO/IEC JTC1/SC29/WG11 JVET-K0390, Ljubljana, SI, Jul. 2018.

[11] "Point Cloud Compression Category 2 Reference Software, TMC2-5.0," http://mpegx.int-evry.fr/software/MPEG/PCC/TM/mpeg-pcc-tmc2.git, accessed: 2019.

[12] "High Efficiency Video Coding Test Model, HM-16.18+SCM8.7," https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/, accessed: 2019.

[13] T. Chiang and Y.-Q. Zhang, "A New Rate Control Scheme Using Quadratic Rate Distortion Model," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 1, pp. 246–250, Feb. 1997.

[14] S. Ma, Wen Gao, and Yan Lu, "Rate-distortion Analysis for H.264/AVC Video Coding and Its Application to Rate Control," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.

[15] Z. He, Y. K. Kim, and S. K. Mitra, "Low-delay Rate Control for DCT Video Coding via ρ-domain Source Modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 8, pp. 928–940, Aug 2001.

[16] Z. He and S. K. Mitra, "A Linear Source Model and a Unified Rate Control Algorithm for DCT Video Coding," *IEEE transactions on Circuits and Systems for Video Technology*, vol. 12, no. 11, pp. 970–982, 2002.

[17] M. Jiang, X. Yi, and N. Ling, "Frame Layer Bit Allocation Scheme for Constant Quality Video," in *2004 IEEE International Conference on Multimedia and Expo (ICME)*, vol. 2, Jun. 2004, pp. 1055–1058 Vol.2.

[18] M. Jiang and N. Ling, "Low-delay Rate Control for Real-time H.264/AVC Video Coding," *IEEE Transactions on*

Multimedia, vol. 8, no. 3, pp. 467–477, Jun. 2006.

[19] S. Zhou, J. Li, J. Fei, and Y. Zhang, "Improvement on Rate-Distortion Performance of H.264 Rate Control in Low Bit Rate," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 8, pp. 996–1006, Aug. 2007.

[20] H. Schwarz, D. Marpe, and T. Wiegand, "Analysis of Hierarchical B Pictures and MCTF," in *2006 IEEE International Conference on Multimedia and Expo (ICME)*, 2006, pp. 1929–1932.

[21] S. Hu, H. Wang, S. Kwong, T. Zhao, and C. . J. Kuo, "Rate Control Optimization for Temporal-Layer Scalable Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 8, pp. 1152–1162, Aug. 2011.

[22] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[23] S. Wang, S. Ma, S. Wang, D. Zhao, and W. Gao, "Rate-GOP Based Rate Control for High Efficiency Video Coding," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1101–1111, Dec. 2013.

[24] 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 Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 139–153, Jan 2016.

[25] Y. Gao, C. Zhu, S. Li, and T. Yang, "Source Distortion Temporal Propagation Analysis for Random-Access Hierarchical Video Coding Optimization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 2, pp. 546–559, Feb. 2019.

[26] T. Yang, C. Zhu, X. Fan, and Q. Peng, "Source Distortion Temporal Propagation Model for Motion Compensated Video Coding Optimization," in *2012 IEEE International Conference on Multimedia and Expo*, Jul. 2012, pp. 85–90.

[27] C. Seo, J. W. Kang, J. Han, and T. Q. Nguyen, "Efficient Bit Allocation and Rate Control Algorithms for Hierarchical Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 9, pp. 1210–1223, Sept. 2010.

[28] W. Yuan, S. Lin, Y. Zhang, W. Yuan, and H. Luo, "Optimum Bit Allocation and Rate Control for H.264/AVC," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 6, pp. 705–715, Jun. 2006.

[29] Z. He and S. K. Mitra, "Optimum Bit Allocation and Accurate Rate Control for Video Coding via ρ-domain Source Modeling," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 10, pp. 840–849, Oct. 2002.

[30] H. Guo, C. Zhu, M. Xu, and S. Li, "Inter-Block Dependency-Based CTU Level Rate Control for HEVC," *IEEE Transactions on Broadcasting*, pp. 1–14, 2019.

[31] K. L. Ferguson and N. M. Allinson, "Modified Steepest-Descent for Bit Allocation in Strongly Dependent Video Coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 1057–1062, Jul. 2009.

[32] Y. G. Lee and B. C. Song, "An Intra-Frame Rate Control Algorithm for Ultralow Delay H.264/Advanced Video Coding (AVC)," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 5, pp. 747–752, May 2009.

[33] M. Wang, K. N. Ngan, and H. Li, "An Efficient Frame-Content Based Intra Frame Rate Control for High Efficiency Video Coding," *IEEE Signal Processing Letters*, vol. 22, no. 7, pp. 896–900, July 2015.

[34] W. Gao, S. Kwong, Y. Zhou, and H. Yuan, "SSIM-Based Game Theory Approach for Rate-Distortion Optimized Intra Frame CTU-Level Bit Allocation," *IEEE Transactions on Multimedia*, vol. 18, no. 6, pp. 988–999, June 2016.

[35] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image Quality Assessment: from Error Visibility to Structural Similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[36] H. Wang and S. Kwong, "Rate-Distortion Optimization of Rate Control for H.264 With Adaptive Initial Quantization Parameter Determination," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 140–144, Jan. 2008.

[37] L. Li, B. Li, D. Liu, and H. Li, "λ-Domain Rate Control Algorithm for HEVC Scalable Extension," *IEEE Transactions on Multimedia*, vol. 18, no. 10, pp. 2023–2039, Oct 2016.

[38] Z. Chen and X. Pan, "An Optimized Rate Control for Low-delay H.265/HEVC," *IEEE Transactions on Image Processing*, pp. 1–1, 2019.

[39] Y. Li, B. Li, D. Liu, and Z. Chen, "A Convolutional Neural Network-based Approach to Rate Control in HEVC Intra Coding," in *2017 IEEE Visual Communications and Image Processing (VCIP)*. IEEE, 2017, pp. 1–4.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[41] S. Schwarz, G. Martin-Cocher, D. Flynn, and M. Budagavi, "Common Test Conditions for Point Cloud Compression," Document ISO/IEC JTC1/SC29/WG11 w17766, Ljubljana, Slovenia, Jul. 2018.

[42] G. Bjontegaard, "Calculation of Average PSNR Differences between RD-Curves," Document VCEG-M33, Austin, Texas, USA, Apr. 2001.

[43] D. Graziosi, "[V-PCC] TMC2 Optimal Texture Packing," Document ISO/IEC JTC1/SC29/WG11 m43681, Ljubljana, SI, Jul. 2018.