IEICE TRANS. FUNDAMENTALS, VOL.E107–A, NO.8 AUGUST 2024

LETTER

CTU-Level Adaptive QP Offset Algorithm for V-PCC Using JND and Spatial Complexity

Mengmeng ZHANG[†,††a)], **Zeliang ZHANG**[†], **Yuan LI**[†], **Ran CHENG**[†], **Hongyuan JING**[††], ***Nonmembers***, ***and*** **Zhi LIU**[†b)], ***Member***

SUMMARY Point cloud video contains not only color information but also spatial position information and usually has large volume of data. Typical rate distortion optimization algorithms based on Human Visual System only consider the color information, which limit the coding performance. In this paper, a Coding Tree Unit (CTU) level quantization parameter (QP) adjustment algorithm based on JND and spatial complexity is proposed to improve the subjective and objective quality of Video-Based Point Cloud Compression (V-PCC). Firstly, it is found that the JND model is degraded at CTU level for attribute video due to the pixel filling strategy of V-PCC, and an improved JND model is designed using the occupancy map. Secondly, a spatial complexity detection metric is designed to measure the visual importance of each CTU. Finally, a CTU-level QP adjustment scheme based on both JND levels and visual importance is proposed for geometry and attribute video. The experimental results show that, compared with the latest V-PCC (TMC2-18.0) anchors, the BD-rate is reduced by −2.8% and −3.2% for D1 and D2 metrics, respectively, and the subjective quality is improved significantly.

key words: *V-PCC, adaptive QP, JND, spatial complexity, subjective quality*

1. Introduction

Point cloud video is one of the main representations to describe the three-dimension space. It usually contains large volume of data [1]. Therefore, the Moving Picture Experts Group (MPEG) developed a standard named Video-based Point Cloud Compression (V-PCC) to reduce storage and transmission requirements.

In V-PCC, 3D point cloud is projected onto 2D planes, and three types of videos are generated, including the occupancy map video, geometry video, and attribute video. For these projected videos, traditional video coding framework, such as High Efficiency Video Coding (HEVC) is utilized for compression [2]. In the field of video compression, the characteristics of the Human Visual System (HVS) play an important role.

Quantization Parameter (QP) adjustment is one of the main approaches for rate distortion optimization [3]. In this regard, the research based on HVS can be mainly divided into two categories: Just Noticeable Difference (JND)-based algorithms and visual attention-based algorithms [4]–[7]. To balance the bitrates between salient areas and less salient areas, Cui et al. [4] proposed a low-complexity Coding Tree Unit (CTU) layer saliency detection scheme, and applied it to adjust the QP in CTU level. Shen et al. [5] employed a pixel-wise pattern based JND model to determine a perceptually lossless distortion threshold for each CTU to decide the appropriate QP. Nami et al. [6] proposed a JND-based perceptual coding scheme, where CNN-based JND prediction and visual importance from visual attention models were used to adjust QP for each block.

In point cloud video, each point contains not only color information but also spatial position information [8]. HVS is both sensitive to color distortion and spatial position changes. However, the existing studies based on HVS did not consider the spatial position information. In this paper, a CTU-level adaptive QP offset adjust algorithm is proposed by jointly considering JND and spatial complexity. It includes three parts. Firstly, it is found that the averaged JND value in a CTU is influence by the unoccupied pixels, and an improved JND model is designed using occupancy map. Secondly, a spatial complexity metric is designed to measure visual importance in geometry video. Finally, a CTU-based QP adjust scheme is proposed based on both JND levels and visual importance.

In this paper, the CTU with occupied pixels is referred to as an occupied CTU, while the CTU where all pixels are unoccupied is referred to as an unoccupied CTU. The proposed algorithm is applied to occupied blocks. For the unoccupied blocks, it has been found that unoccupied pixels have no impact on the reconstruction of point cloud [9]. In this paper, the QP of this area is set to the maximum to save as much bit as possible.

The remainder of the paper is organized as follows. The proposed algorithm is described in Sect. 2. The experimental results are presented in Sect. 3. Finally, Sect. 4 concludes the paper.

2. Proposed CTU-Level Adaptive QP Offset Algorithm

2.1 JND-Based Perception Description for Attribute Video

JND is always used to measure the limit of distortion level that cannot be perceived by human visual system. In this study, the pixel domain JND model is adopted for the attribute video in V-PCC, which considers the joint effects of luminance adaptation (LA) and contrast masking (CM), as shown in (1).

Manuscript received February 7, 2024.
Manuscript publicized March 6, 2024.
[†]North China University of Technology, Beijing, China.
[††]Beijing Union University, Beijing, China.
a) E-mail: muchmeng@126.com (Corresponding author)
b) E-mail: lzliu@ncut.edu.cn (Corresponding author)
DOI: 10.1587/transfun.2024EAL2021

(a) attribute image (b) without mask (c) with mask

Fig. 1 Visualization of JND image at CTU level without mask and with mask for attribute image of sequence Longdress.

$$JND(x,y) = LA(x,y) + CM(x,y) - 0.03 * \min\{LA(x,y), CM(x,y)\}, \quad (1)$$

In each CTU, the obtained JND value is averaged and denoted as JND_{CTU}. In this study, it is found that although the pixels in unoccupied area do not affect the reconstruction quality, however, the existence of unoccupied pixels may influence the accuracy of the JND_{CTU}. This is due to that the filling strategy for attribute image are likely to create flat texture in unoccupied area, which reduces the threshold of the human eye to detect distortion. Therefore, the JND_{CTU} obtained in this kind of CTU is smaller than it ought to be. In the proposed algorithm, the occupancy map is employed as a mask for the calculation of the JND_{CTU} to exclude unoccupied pixels from the JND model, as shown in (2).

$$JND_{CTU} = \frac{1}{N}\sum_{x=0}^{63}\sum_{y=0}^{63} JND(x,y) \times M, \quad (2)$$

where $M = 1$ if the pixel at position (x, y) is occupied, otherwise $M = 0$. N represents the number of occupied pixels in a CTU.

Figure 1 shows the comparison of JND image at CTU level for attribute image of the sequence Longdress, without mask and with mask. It is evident that the mask can help to distinguish salient CTU.

In order to describe the visual sensitivity for attribute image, a parameter named Attribute Visual Sensitivity (AVS) is defined for each occupied CTU, as shown in (3),

$$AVS_{CTU} = \frac{\max JND_{CTU} - JND_{CTU}}{\max JND_{CTU} - \min JND_{CTU}}, \quad (3)$$

where $maxJND_{CTU}$ and $minJND_{CTU}$ represent the maximum and minimum JND levels of occupied CTUs in a frame, respectively. The smaller the JND value of the current occupied CTU, the larger the AVS value will be, indicating that the human eye is more sensitive to the current CTU.

2.2 Spatial Complexity Description for Geometry Video

In point cloud video, areas with significant spatial variation often reflect high contrast of depth, which conforms with the fact that depth edges have higher visual sensitivity and attention [10]. The projected depth information in point cloud video is stored in the luminance channel of the geometry image. Therefore, the spatial positional changes in the point cloud are represented by calculating the complexity of the luminance channel within each CTU. Areas with high complexity usually have high visual attention.

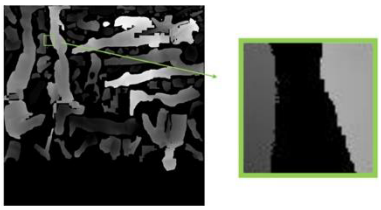


Fig. 2 An example where occupied pixels in a CTU do not belong to the same patch in geometry image.

Various methods exist for measuring complexity. Among them, variance is the popular one. However, the occupied pixels within a CTU may belong to multiple patches, as shown in Fig. 2. Different patches may correspond to different projection planes, and their projected depths may also vary significantly. Therefore, the spatial variations of the area may not be represented by the variance of occupied pixels in a CTU.

In this paper, a parameter named Geometry Visual Sensitivity (GVS) is designed to measure the visual importance of CTU, as shown in (4). GVS is obtained by scaling the variances using the average variance of all occupied CTUs in the frame. The variances of different patches within a CTU are calculated separately, and the maximum variance is selected to represent the complexity of the CTU.

$$GVS_{CTU} = \frac{s * Var_{CTU} + meanVar}{Var_{CTU} + s*meanVar}, \quad (4)$$

where $meanVar$ represents the average variance of all occupied CTUs in the frame, and Var_{CTU} represents the variance of the current CTU, and s is the intensity factor. In this paper, s is set to 2. The larger the GVS, the more sensitive for human is to the area.

2.3 CTU-Level Adaptive QP Offset Algorithm

Based on the calculated AVS and GVS, the visual sensitivity factor (VS_{CTU}) for both geometry and attribute video of each CTU is defined, and is utilized to adjust the QP offset for each CTU. For geometry video, the VS_{CTU} is defined using the GVS obtained in (4). For attribute video, the GVS is used to decrease the bit-rate requirement by prioritizing the visual importance of different areas. This means that the VS_{CTU} of attribute video is obtained by weighted combining the GVS and AVS.

$$VS_{CTU} = \begin{cases} GVS_{CTU}, & if\ Type = GV \\ \beta_1 * AVS_{CTU} + \beta_2 * GVS_{CTU}, & if\ Type = AV \end{cases}, \quad (5)$$

where β_1 and β_2 are empirically selected from experimental results ($\beta_1 = 1.5$ and $\beta_2 = 0.25$). The parameter $Type$ indicates the type of the video, with GV representing geometry video and AV representing attribute video. If VS_{CTU} is

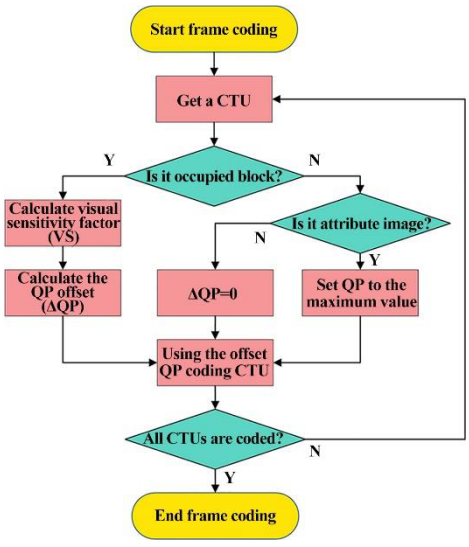


Fig. 3 The overall flowchart of the proposed adaptive QP offset algorithm.

greater than 1, it indicates that the occupied CTU belongs to a sensitive area. Conversely, it indicates that the occupied CTU belongs to a non-sensitive area.

The QP offset of each occupied CTU in the geometry and attribute image is calculated as follows:

$$\Delta QP_{CTU_i} = \begin{cases} 1, & if\ 0 \le \mathrm{VS}_{CTU_i} < \theta \\ 0, & if\ \theta \le \mathrm{VS}_{CTU_i} \le 1 \\ \left\lfloor -\alpha * \log_s^{VS_{CTU_i}} \right\rfloor, & if\ \mathrm{VS}_{CTU_i} > 1 \end{cases}, \tag{6}$$

where α is a constant, representing the maximum QP offset range in geometry and attribute video. In this paper, α is set to 2 and 6 respectively. $\lfloor\ \rfloor$ represents rounding down. The threshold θ is used to indicate the degree of visual sensitivity. When the VS_{CTU} is smaller than θ, the current occupied CTU is not visual sensitivity. In this paper, θ is set to 0.6.

For unoccupied blocks, QP is adjusted using two schemes. For attribute video, since unoccupied pixels have no impact on the reconstruction of point cloud, therefore, QP is set to the maximum value to decrease the bit-rate requirement. For geometry video, it had been proved that unoccupied blocks only require a very small amount of bit consumption [9]. Therefore, QP is not adjusted for this area.

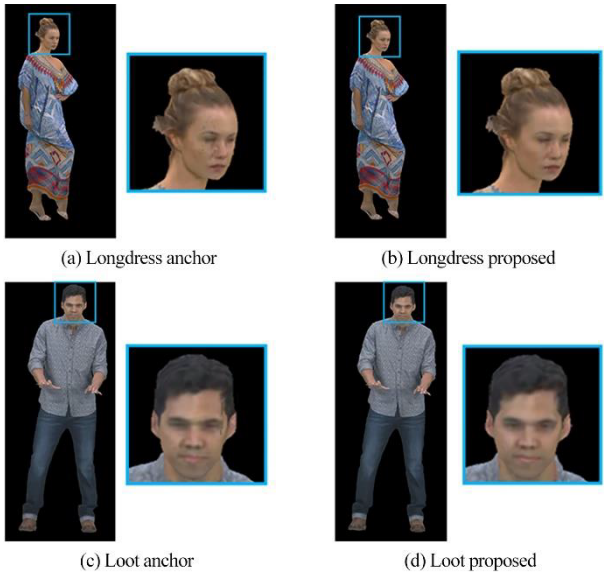
The flowchart for the algorithm is shown in Fig. 3.

3. Experimental Results

The proposed algorithm is implemented in the V-PCC reference software TMC2-18.0 to compare with V-PCC anchor. The test point cloud sequences named "Loot", "Redandblack", "Soldier", and "Longdress" are encoded following the Common Test Conditions (CTC) of V-PCC. In the experiments, all-intra (AI) configuration is used. The first 32 frames of each point cloud sequence were tested to validate the performance of the proposed algorithm in terms of the Bjøntegaard-delta (BD)-rate and the subjective results.

Table 1 Overall performance of the proposed algorithm compared with TMC2-18.0 anchor.

Point Cloud	Geom.BD-TotalRate		Attr.BD-TotalRate		
	D1	D2	Luma	Cb	Cr
Loot	-2.8%	-3.4%	-0.5%	-0.4%	-0.9%
Redandblack	-2.5%	-3.2%	-0.5%	-1.2%	-0.5%
Soldier	-1.5%	-1.7%	-0.2%	-0.1%	0.1%
Longdress	-4.2%	-4.6%	-0.7%	-0.7%	-1.0%
Average	-2.8%	-3.2%	-0.5%	-0.6%	-0.6%
Enc.time	100.09%				



(a) Longdress anchor (b) Longdress proposed

(c) Loot anchor (d) Loot proposed

Fig. 4 Comparison of subjective quality for sequences "Longdress" and "Loot".

3.1 Overall Performance of the Proposed Algorithm

Table 1 shows the overall RD performance of the proposed algorithm. It can be observed that the proposed algorithm achieves 2.8% and 3.2% Geom.BD-TotalRate gains on average in D1 and D2. In terms of attribute video, the proposed algorithm can achieve average performance gains of 0.5%, 0.6%, and 0.6% for Luma, Cb, and Cr components respectively. In addition, the encoding time of the algorithm remains almost unchanged compared to the V-PCC anchor.

Figure 4 shows subjective quality comparison of the reconstructed point clouds for the "Longdress" and "Loot" at the R3 rate. Figure 4(a) and Fig. 4(c) demonstrate the reconstructed point clouds using the V-PCC anchor, while Fig. 4(b) and Fig. 4(d) demonstrate the reconstructed point clouds using the proposed algorithm. From Fig. 4(a), noticeable distortion can be observed in the relatively flat area, such as the facial area. For the proposed algorithm, the subjective results are obviously better than the V-PCC anchor, as evident in Fig. 4(b). Annoying cracks are found in Fig. 4(c), which greatly impair the subjective experience. As a comparison, our method does not exhibit noticeable cracks. In summary, the subjective quality is enhanced for the proposed algorithm when subject to bit rate limitations.

Table 2 Performance of geometry video with only GVS is enabled, compared with TMC2-18.0 anchor.

Point Cloud	Geom.BD-TotalRate		Attr.BD-TotalRate		
	D1	D2	Luma	Cb	Cr
Loot	-1.7%	-2.3%	0.5%	0.4%	-0.3%
Redandblack	-1.4%	-2.0%	0.3%	0.6%	0.3%
Soldier	-1.4%	-1.5%	0.2%	0.3%	-0.1%
Longdress	-2.7%	-3.1%	0.0%	0.2%	0.2%
Average	-1.8%	-2.2%	0.2%	0.4%	0.0%

Table 3 Performance of attribute video by only adjusting QP in occupied CTUs (with only AVS enabled, and with both AVS and GVS enabled), compared with TMC2-18.0 anchor.

Point Cloud	BD-AttrRate Only AVS			BD-AttrRate AVS + GVS		
	Luma	Cb	Cr	Luma	Cb	Cr
Loot	3.5%	4.2%	2.4%	0.2%	0.0%	-1.0%
Redandblack	4.2%	-2.7%	13.5%	0.2%	-0.6%	0.5%
Soldier	18.8%	11.1%	12.0%	0.2%	-0.4%	-0.4%
Longdress	4.7%	3.9%	2.0%	0.2%	-0.2%	-0.3%
Average	7.8%	4.1%	7.5%	0.2%	-0.3%	-0.3%

Table 4 Performance of attribute video by only adjusting QP in unoccupied CTUs, and the overall performance of attribute video, compared with TMC2-18.0 anchor.

Point Cloud	BD-AttrRate Unoccupied CTUs			BD-AttrRate Overall		
	Luma	Cb	Cr	Luma	Cb	Cr
Loot	-1.8%	-0.9%	-1.8%	-1.6%	-1.2%	-1.8%
Redandblack	-1.5%	-1.8%	-1.8%	-1.4%	-2.2%	-1.2%
Soldier	-0.6%	0.3%	0.2%	-0.4%	0.5%	-0.2%
Longdress	-1.0%	-0.9%	-1.0%	-0.9%	-1.2%	-1.3%
Average	-1.2%	-0.8%	-1.1%	-1.1%	-1.0%	-1.1%

3.2 Ablation Study

Table 2 shows the performance of geometry video with only GVS enabled. It achieves 2.8% and 3.2% Geom.BD-TotalRate gains on average in D1 and D2, with only 0.2% and 0.4% Color.BD-TotalRate loss on average for Luma and Cb, respectively.

Table 3 shows the results of attribute video with only AVS enabled and with both AVS and GVS enabled. As has stated in Sect. 2, they are only applied to occupied CTUs. It can be found that, the objective loss is 7.8% for Luma, when only AVS is enabled. With both AVS and GVS enabled, the loss is only 0.2% in Luma. This is due to that the spatial complexity (GVS) obtained from geometry video can help to reduce the bitrate and improve PSNR of attribute video.

Table 4 shows the result of QP adjustment for unoccupied CTUs. The average gains are 1.2%, 0.8%, and 1.1% for Luma, Cb, and Cr, respectively. It shows that by adjusting the QP of unoccupied CTUs to the maximum can improve the objective performance. This is due to that it can save a significant amount of bits consumption without compromising the quality of the reconstructed video. Table 4 also show the overall performance of attribute video, with both AVS and GVS enabled, and the QPs are adjusted in both occupied and unoccupied CTUs. The average gains are 1.1%, 1.0%, and 1.1% for Luma, Cb, and Cr, respectively.

4. Conclusion

In this paper, a CTU-level adaptive QP offset algorithm based on visual sensitivity is proposed for geometry video and attribute video in V-PCC, to improve the subjective and objective quality of reconstructed point cloud. The visual sensitivity of each occupied CTU in the geometry video and attribute video is obtained, and is jointly applied to guide the QP adjustment. The experimental results show that, the proposed algorithm can significantly improve the subjective and objective quality of the reconstructed point cloud.

Acknowledgments

This work is supported by MOE Planned Project of Humanities and Social Sciences (No.20YJA870014).

References

[1] A. Chen, S. Mao, Z. Li, M. Xu, H. Zhang, D. Niyato, and Z. Han, "An introduction to point cloud compression standards," GetMobile: Mobile Computing and Communications, vol.27, no.1, pp.11–17, 2023.

[2] C. Cao, M. Preda, V. Zakharchenko, E.S. Jang, and T. Zaharia, "Compression of sparse and dense dynamic point clouds–methods and standards," Proc. IEEE, vol.109, no.9, pp.1537–1558, 2021.

[3] G. Xiang, H. Jia, M. Yang, Y. Li, and X. Xie, "A novel adaptive quantization method for video coding," Multimed. Tools Appl., vol.77, no.12, pp.14817–14840, 2017.

[4] Z. Cui, M. Zhang, X. Jiang, Z. Gan, G. Tang, and F. Liu, "Improving HEVC coding perceptual quality using saliency guided CTU layer QP adjustment," IEEE Chinese Automation Congress (CAC), pp.5524–5529, 2019.

[5] X. Shen, X. Zhang, S. Wang, S. Kwong, and G. Zhu, "Just noticeable distortion based perceptually lossless intra coding," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.2058–2062, 2020.

[6] S. Nami, F. Pakdaman, M.R. Hashemi, and S. Shirmohammadi, "BL-JUNIPER: A CNN-assisted framework for perceptual video coding leveraging block-level JND," IEEE Trans. Multimedia, vol.25, pp.5077–5092, 2023.

[7] D. Zhang, F. Li, M. Liu, R. Cong, H. Bai, M. Wang, and Y. Zhao, "Exploring resolution fields for scalable image compression with uncertainty guidance," IEEE Trans. Circuits Syst. Video Technol., vol.34, no.4, pp.2934–2948, 2024, DOI: 10.1109/TCSVT.2023.3307438.

[8] A. Tabatabai, T. Suzuki, A. Zaghetto, S. Kuma, O. Nakagami, and D. Graziosi, "An overview of ongoing point cloud compression standardization activities: Video-based (V-PCC) and geometry-based (G-PCC)," APSIPA Transactions on Signal and Information Processing, vol.9, no.1, e13, 2020.

[9] 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.

[10] Y. Zhang, H. Zhang, M. Yu, S. Kwong, and Y.-S. Ho, "Sparse representation-based video quality assessment for synthesized 3D videos," IEEE Trans. Image Process., vol.29, pp.509–524, 2019.