LETTER

Improved Just Noticeable Difference Model Based Algorithm for Fast CU Partition in V-PCC*

Zhi LIU[†a)], *Member*, Heng WANG[†], Yuan LI[†], Hongyun LU[†], Hongyuan JING[††],
and Mengmeng ZHANG[†,††b)], *Nonmembers*

SUMMARY   In video-based point cloud compression (V-PCC), the partitioning of the Coding Unit (CU) has ultra-high computational complexity. Just Noticeable Difference Model (JND) is an effective metric to guide this process. However, in this paper, it is found that the performance of traditional JND model is degraded in V-PCC. For the attribute video, due to the pixel-filling operation, the capability of brightness perception is reduced for the JND model. For the geometric video, due to the depth filling operation, the capability of depth perception is degraded in the boundary area for depth based JND models (JNDD). In this paper, a joint JND model (J_JND) is proposed for the attribute video to improve the brightness perception capacity, and an occupancy map guided JNDD model (O_JNDD) is proposed for the geometric video to improve the depth difference estimation accuracy of the boundaries. Based on the two improved JND models, a fast V-PCC Coding Unit (CU) partitioning algorithm is proposed with adaptive CU depth prediction. The experimental results show that the proposed algorithm eliminates 27.46% of total coding time at the cost of only 0.36% and 0.75% Bjontegaard Delta rate increment under the geometry Point-to-Point (D1) error and attribute Luma Peak-signal-Noise-Ratio (PSNR), respectively.
key words: *V-PCC, JND, JNDD, partition*

1.   Introduction

In video-based point cloud compression (V-PCC) [1], the patch projection-based method produces a large number of empty pixels, and the far and near components are projected to different 2D images (video frames), respectively. Consequently, the projected video is always with high resolutions and double frame rates, which brings high computational complexity for coding in V-PCC.
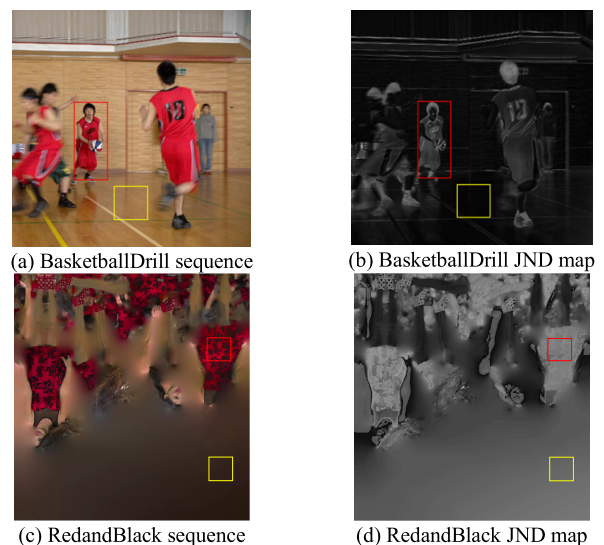
The partitioning of Coding Unit (CU) is an important task and has ultra-high computational complexity in V-PCC. In order to reduce the complexity, a few fast algorithms have been proposed recently [2]. In these proposed fast algorithms, cross information of attribute video, geometric video and occupancy maps was utilized to remove temporal and spatial redundancy in point cloud video. To reduce coding complexity, the characteristics of human visual system (HVS) are also adopted. In [3], [4], the Just Noticeable Difference (JND) model is used to categorize the coding units,

and to reduce the number of candidate modes. However, it is found in this paper that the traditional JND model [5] does not apply to projected video in V-PCC, due to the pixel and depth filling operation. Based on this observation, two improved JND models are designed, and a fast CU partition algorithm is proposed. The main contributions of this paper are as follows. Firstly, the performance of the typical JND model for the projected point cloud video is studied, and a joint JND (J_JND) model and an Occupancy Map-guided depth JND (O_JNDD) model are designed. Secondly, the coding tree units of attribute videos and geometric videos are classified using the improved JND model. Finally, combined with adaptive partitioning, the optimal partitioning range is determined for the coding tree units, and the process of partitioning is terminated in advance.

2.   Observation and Motivation

2.1   JND Model Mapping Analysis for Attribute Video

The JND model is used to delineate the sensitivity of the HVS toward changes in stimuli. For traditional video, the JND value in the visual sensitive area is distinct from that in nonvisual sensitive area. The sensitive area usually has relatively high JND value, while the nonvisual sensitive area has relatively low one, as the red box and the yellow box

(a) BasketballDrill sequence   (b) BasketballDrill JND map

(c) RedandBlack sequence   (d) RedandBlack JND map

Fig. 1   Ordinary and attribute video frame JND map.

shown in Figs. 1 (a) and (b), respectively. Sequence Basket-ballDrill and Redandblack in Fig. 1 are derived from [6] and [7], respectively.

In V-PCC, the 3D points of the point cloud are projected and packaged to form 2D attribute and geometric videos. Due to the mechanism of the packaging method [1], the output attribute video contains plenty of unoccupied areas. To improve the coding efficiency, the unoccupied areas of the attribute video are filled based on pixel fill operation. The yellow box in Fig. 1 (c) shows an example of the filled unoccupied area in attribute videos. While this method improves encoding efficiency, it inevitably results in relatively large man-made low-contrast areas. Our study found that, in these areas, the brightness perception sensitivity of the traditional JND model is degraded, as shown of the yellow and red boxes in Figs. 1 (c) and (d).

To confirm this observation, we carried out a series of experiments to study the performance of JND model in the filled unoccupied areas and the visually sensitive areas. The mean JND values are used as the metric. The results are illustrated in Fig. 2. It can be found from the figure that, the average JND values in filled unoccupied areas are comparable to that in sensitive areas.

2.2 JNDD Model Mapping Analysis for Geometry Video

The geometric video represents the depth information of the point cloud video, as an example shown in Fig. 3 (a). Like the attribute video, the output geometric video of the packaging process contains plenty of unoccupied areas, and the depth filling operation uses the residual values from the original and predicted blocks to fill these areas.

The HVS has a visual masking effect in depth due to the varying sensitivities of human eyes to changes in depth.
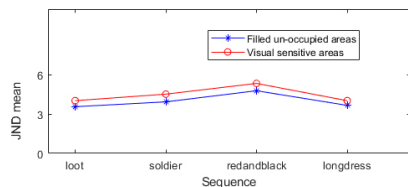
Fig. 2 Comparison of JND mean values between visual sensitive and nonvisual sensitive areas.

(a) Unoccupied geometric video   (b) Occupied geometric video   (c) JNDD map
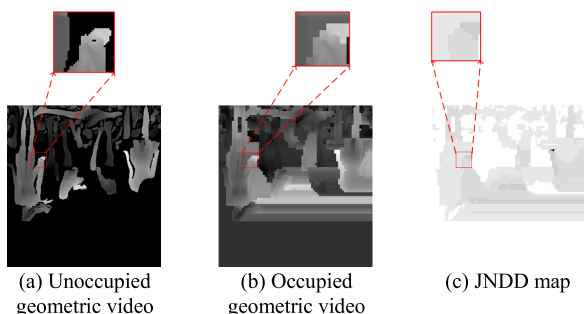
Fig. 3 Geometric image JNDD map.

The depth based JND (JNDD) model [8] is used to describe this feature. However, it is found in our study that, the aforementioned depth filling operation yields a decrease in geometric video's boundary depth difference, as shown in Fig. 3 (b), which degrades the differentiation of the JNDD model to the changes of the depth boundary, as shown in Fig. 3 (c).

2.3 Motivations

Since the area with large JND values represents the visually sensitive area, and it usually contains more information, for the CU partition process in video coding, to increase the prediction efficiency, this kind of area can be partitioned into small CU and has relatively large depth in the coding tree. While nonvisual sensitive area, it is always partitioned into large CU and has relatively small depth in the coding tree.

In the following sections, the observations stated in this section will be utilized to improve JND models and used to design a fast CU partition algorithm.

3. Proposed Algorithm

3.1 Improvements of the JND Model

#### 3.1.1 Proposed Joint JND Model for Attribute Video

The traditional JND model [5] can be expressed as,

$$JND = LA + CM - C^{lc} * \min\{LA, CM\} \quad (1)$$

Where LA and CM represent the luminance adaptation, and contrast masking feature of HVS, respectively, and C^{lc} is a parameter.

As shown in Sect. 2.1, our study found that, for the attribute video in V-PCC, the JND model fails to describe the perceived brightness changes in the unoccupied areas, which means the LA module in (1) needs to be improved. In this subsection, a joint LA (J_LA) module is designed, as shown of the red area in Fig. 4, and the improved model is named as joint JND (J_JND) model.

The BartenCSF based LA [9] is a variation of LA. It takes a global perspective and directly converts the brightness value into perceived brightness, which can significantly improve the uniformity of brightness perception. In the proposed J_JND model, for the unoccupied area, the BartenCSF
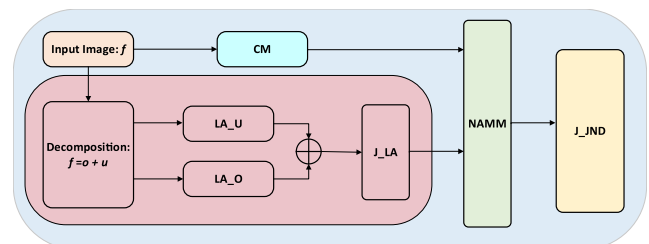
Fig. 4 Structure of the proposed J_JND model.

based *LA* module, denoted as *LA_U*, is used to substitute the original *LA* module, and is obtained using:

$$LA\_U = \begin{cases} \dfrac{lum(B(x,y)-62)-lum(B(x,y)-63)}{lum(B(x,y)-62)+lum(B(x,y)-63)} * w \\[2ex] \dfrac{lum(B(x,y)-63)-lum(B(x,y)-64)}{lum(B(x,y)-63)+lum(B(x,y)-64)} * w \end{cases} \tag{2}$$

where $B(x,y)$ represents the brightness at point (x,y), and *lum* is an inverse perceptual luminance converter, and w is set to 0.238 [9].

For the occupied area, the original *LA* module of JND is used, denoted as *LA_O*. The joint luminance adaptation (J_LA) is expressed as,

$$J\_LA = \alpha * LA\_O + (1-\alpha)LA\_U \tag{3}$$

where α is set to 1 for unoccupied areas, and 0 for occupied areas. The proposed J_JND model is expressed as,

$$J\_JND = J\_LA + CM - 0.3 * \min\{J\_LA, CM\} \tag{4}$$

3.1.2 Proposed Occupancy Map-Guided JNDD Model for Geometric Video

As has been stated in Sect. 2.2, the depth filling in geometric videos may blur the depth boundaries of geometric videos, and degrade the differentiation of the JNDD model to the changes of the depth boundary.

The occupancy map in V-PCC not only provides the information of occupied and unoccupied pixels, but also provides accurate boundary information [1]. To address the aforementioned issue in geometric video, the occupancy map is used as a visual masking factor to calibrate the JNDD model, denoted as Occupancy Map-Guided JNDD (O_JNDD) model, which is expressed as,

$$D_{O\_JNDD} = \lambda * JNDD \tag{5}$$

where λ is the visual masking factor, which is set to 1 when current pixel is in the occupied area, and 0 when it is not.

3.2 Fast CU Partition Algorithm Based on the Improved JND Model

The main idea of the fast CU partition algorithm based on improve JND model can be stated as follows. Since the HVS sensitivity of the area has correlation to the partition depth in the coding tree, the JND values are used to classify all the coding tree units (CTUs). For each kind of CTU, the optimal depth range is obtained through statistics method. Based on the optimal depth range, the encoder can skip the search of the depth out of the optimal range, and save coding time.

3.2.1 Classification of CTUs Based on JND Values

Based on the JND values, CTUs in a video frame are classified into three kinds: Visual Sensitive (VS) CTUs, Medium Visual Sensitive (MVS) CTUs, and Non-Visual Sensitive
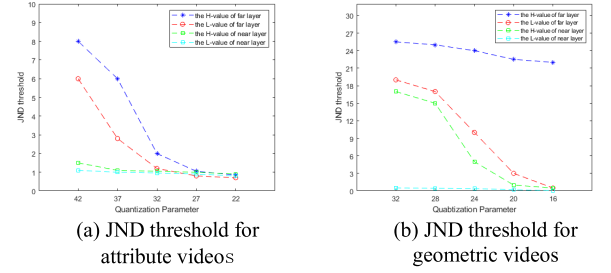

(a) JND threshold for attribute videos (b) JND threshold for geometric videos

Fig. 5 JND thresholds for different QPs.

Table 1 The optimal depth level for attribute videos.

CTU Type	Candidate depth	Depth ranges
NVS CTU	0	[0,0]
MVS CTU	0, 1, 2	[0,2]
VS CTU	0, 1, 2, 3	[0,3]

Table 2 The optimal depth level of near layer for geometric videos.

CTU Type	Candidate depth	Depth ranges
NVS CTU	0	[0,0]
MVS CTU	0, 1, 2	[0,2]
VS CTU	1, 2, 3	[1,3]

Table 3 The optimal depth level of far layer for geometric videos.

CTU Type	Candidate depth	Depth ranges
NVS CTU	0, 1	[0,1]
VS CTU	0, 1, 2	[0,2]

(NVS) CTUs. The classifier is expressed as,

$$\begin{cases} \text{NVS } CTU, & JND \le L \\ \text{MVS } CTU, & L < JND \le H \\ \text{VS } CTU, & JND > H \end{cases} \tag{6}$$

where L and H are thresholds and are obtained through statistics.

Figure 5 shows the values of L and H for far layer and near layer with different quantization parameters (QP). Due to the unique dual frame rate structure of V-PCC, one frame in point cloud video will generate one near layer frame and one far layer frame. There is a significant difference in the partition depth between the near and far frame, therefore, the CTUs in near and far frame are classified independently.

3.2.2 Optimal Depth Range for CU Partition

Based on extensive statistical analysis, the optimal depth range of CTU is obtained and shown as follows.

For attribute video, it is found that the optimal depth range of near layer frame and far layer frame is the same, and is shown in Table 1.

For geometric video, the optimal depth range for near layer frame is shown in Table 2. The optimal depth range for NVS CTUs is all 0, while the optimal depth range for VS CTUs is [1, 3].

Table 3 shows the optimal depth range for far layer frame

Table 4 Performance of V-PCC fast algorithm.

Sequences	Geom.BD-GeomRate [%]		End-to-End BD-AttrRate [%]			Geom.BD-TotalRate [%]		End-to-End BD TotalRate [%]			Δ *time* [%]			[2]'s Δ *time* [%]
	D1	D2	Luma	Cb	Cr	D1	D2	Luma	Cb	Cr	Δ*att*	Δ*geot*	Δ*Total*	Δ*Total*
Longdress	0.47	0.56	0.35	0.13	0.19	0.46	0.75	0.37	0.16	0.21	41.75	50.42	31.77	18.04
Loot	0.30	0.50	0.43	-1.53	-0.79	0.31	0.72	0.41	-1.07	-0.53	33.24	42.69	18.95	19.66
Queen	0.46	0.49	2.12	0.74	1.97	0.34	0.49	1.74	0.73	1.56	40.79	44.84	24.62	16.21
Basketball-player	0.32	0.36	1.24	-1.13	-1.06	0.40	0.49	1.02	-0.71	-0.67	45.85	46.04	29.53	10.19
Dancer	0.55	0.66	1.08	0.38	-0.49	0.47	0.71	1.00	0.41	-0.23	46.38	46.85	31.49	10.42
Redandblack	0.18	0.33	0.25	-0.30	0.15	0.14	0.38	0.25	-0.16	0.18	40.62	**51.27**	28.93	19.35
Soldier	**0.49**	0.48	0.37	-0.31	-0.29	0.36	0.35	0.44	-0.03	0.00	38.55	47.16	26.91	12.97
Average	**0.40**	**0.48**	**0.84**	**-0.29**	**-0.05**	**0.36**	**0.56**	**0.75**	**-0.09**	**0.08**	**41.03**	**47.04**	**27.46**	**15.26**

in geometric videos. Due to the presence of large flat areas in geometric video and the low probability of partitioning for far layer, there are only two kinds of CTUs: NVS CTU and VS CTU.

4. Experimental Results

4.1 Configurations and Settings

The proposed algorithm is implemented in V-PCC reference software TMC2-18.0 [10] to verify performance. All the coding experiments follow the V-PCC common test conditions (CTCs) [10]. The coding configuration is set to all intra, and the recommended five bitrate sets {R1, R2, R3, R4, R5} are used. Time saving ratio is obtained using:

$$\Delta T = (T_{org} - T_{fast})/T_{org} \qquad (7)$$

where T_{org} and T_{fast} are the coding time of TMC2-18.0 and the proposed fast algorithm, respectively.

4.2 Total Performance of the Proposed Algorithm

Table 4 shows the overall performance of the proposed fast algorithm. Compared with TMC2-18.0, under all intra configuration, the average saving of total coding time, attribute video coding time, and geometric video coding is 27.46%, 41.03%, and 47.04%, respectively.

In addition, Table 4 shows that the average BD-rate increment is very small for the proposed algorithm. For attribute videos, the average BD-rate changes are 0.84%, −0.29%, and −0.05% for Luma, Cb, and Cr content, respectively. For geometric videos, the average BD-rate increment is 0.36% and 0.56% for D1 and D2 metrics, respectively. For the total bitrate, the results show that the BD-rate increment is small for both geometric and attribute videos, with a maximum value of 0.75%.

The algorithm proposed in [2] was implemented in the experimental condition of this paper, and the results are shown in Table 4. The results show that the time saving of the proposed algorithm is 12.20% higher than [2] with

less coding loss.

5. Conclusion

In this paper, the degradation of the traditional JND model in V-PCC is observed, and improved JND models are designed. Based on the improved JND model, a fast CU partition algorithm is proposed. The experimental results show that the proposed algorithm can significantly reduce coding time.

References

[1] E.S. Jang, M. Preda, K. Mammou, A.M. Tourapis, J. Kim, D.B. Graziosi, S. Rhyu, and M. Budagavi, "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, 2019.

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

[3] M. Liu, C. Zhang, H. Bai, R. Zhang, and Y. Zhao, "Cross-part learning for fine-grained image classification," IEEE Trans. Image Process., vol.31, pp.748–758, 2022.

[4] J.-R. Lin, M.-J. Chen, C.-H. Yeh, Y.-C. Chen, L.-J. Kau, C.-Y. Chang, and M.-H. Lin, "Visual perception based algorithm for fast depth intra coding of 3D-HEVC," IEEE Trans. Multimed., vol.24, pp.1707–1720, 2022.

[5] X.K. Yang, W.S. Ling, Z.K. Lu, E.P. Ong, and S.S. Yao, "Just noticeable distortion model and its applications in video coding," Signal Process. Image Commun., vol.20, no.7, pp.662–680, 2005.

[6] J. Boyce, K. Suehring, X. Li, and V. Seregin, JVET-J1010: JVET common test conditions and software reference configurations, 2018.

[7] M. Krivokuća, P.A. Chou, and P. Savill, "8i voxelized surface light field (8iVSLF) dataset," ISO/IEC JTC1/SC29 WG11 (MPEG) input document m42914, 2018.

[8] C. Chen, G. Jiang, and M. Yu, "Depth-perception based geometry compression method of dynamic point clouds," Proc. 2021 5th Int. Conf. Video Image Process., pp.56–61, 2021.

[9] C. Jung, Q. Lin, and S. Yu, "HEVC encoder optimization for HDR video coding based on perceptual block merging," 2016 Visual Communications and Image Processing (VCIP), pp.1–4, 2016.

[10] ISO/IEC JTC 1/SC 29/WG 11, Common Test Conditions for V3C and V-PCC, July 2020.