3D-Distortion Based Rate Distortion Optimization for Video-Based Point Cloud Compression

Yihao Fu[1], Liquan Shen[2*], and Tianyi Chen[3]
[1] School of Communication and Information Engineering, Shanghai University,
Shanghai 200072, China
[e-mail: yihao_fu97@163.com]
[2] Shanghai Institute for Advanced Communicationand Data Science, Shanghai University,
Shanghai 200072, China
[e-mail: jsslq@shu.edu.cn]
[3] School of Communication and Information Engineering, Shanghai University,
Shanghai 200072, China
[e-mail: chentianyi1995@163.com]
* corresponding author : Liquan Shen

Abstract

The state-of-the-art video-based point cloud compression(V-PCC) has a high efficiency of compressing 3D point cloud by projecting points onto 2D images. These images are then padded and compressed by High-Efficiency Video Coding(HEVC). Pixels in padded 2D images are classified into three groups including origin pixels, padded pixels and unoccupied pixels. Origin pixels are generated from projection of 3D point cloud. Padded pixels and unoccupied pixels are generated by copying values from origin pixels during image padding. For padded pixels, they are reconstructed to 3D space during geometry reconstruction as well as origin pixels. For unoccupied pixels, they are not reconstructed. The rate distortion optimization(RDO) used in HEVC is mainly aimed at keeping the balance between video distortion and video bitrates. However, traditional RDO is unreliable for padded pixels and unoccupied pixels, which leads to significant waste of bits in geometry reconstruction. In this paper, we propose a new RDO scheme which takes 3D-Distortion into account instead of traditional video distortion for padded pixels and unoccupied pixels. Firstly, these pixels are classified based on the occupancy map. Secondly, different strategies are applied to these pixels to calculate their 3D-Distortions. Finally, the obtained 3D-Distortions replace the sum square error(SSE) during the full RDO process in intra prediction and inter prediction. The proposed method is applied to geometry frames. Experimental results show that the proposed algorithm achieves an average of 31.41% and 6.14% bitrate saving for D1 metric in Random Access setting and All Intra setting on geometry videos compared with V-PCC anchor.

1. Introduction

With the rapid development of multi-media, more attention is paid to the high quality video, which should have high resolution, high dynamic range and more degrees of freedom. New emerging 3D videos are also entering people's daily lives. Many new downstream applications, such as Virtual Reality [1], Vector Map and Immersive Telepresence [2], [3] have made great progress in recent years. Point cloud [4] also plays an important role in these applications.

Point cloud is a set of 3D points which contain both geometry content and attribute content. Most point cloud used in applications is continuously moving, which is called dynamic point cloud(DPC) [5], [6]. Several DPC models are shown in **Fig. 1**. Although point cloud could be well used to construct high precision 3D models with the help of Point Cloud Library(PCL) [7], it always takes up a lot of storage space. For example, a 300-frames DPC with each frame of nearly one million points is usually larger than 2GBytes. Therefore, it is necessary to figure out an efficient strategy to compress DPC.

(a) *Longdress*   (b) *Loot*   (c) *Red and black*

(d) *Soldier*   (e) *Queen*   (f) *Basketball player*

Fig. 1. Examples of DPC models.

The Moving Picture Experts Group(MPEG) developed a new scheme of DPC compression in 2017, which is named video-based point cloud compression(V-PCC) [8], [9]. In particular, the compression model converts 3D point cloud to 2D images by projecting these points to its

bounding box [10]. These points are then allocated to different patches according to their normal. The generated patches are packed into regular 2D grids to generate geometry images and attribute images [11], which record the geometry content and attribute content respectively.

Auxiliary patch information is also generated during image packing. It consists of patch indices, patch view-ids, patch sizes, the start position of patches in 2D video and the relevant start position of 3D point cloud in 3D space. The view-id decides the projection parameters including projection axis and projection mode of each patch, which is indispensable in reconstruction. The relationship between projection parameters and view-ids is provided in **Table 1**. u, v are coordinates in 2D images and depth is the value of geometry images. X, Y, Z are coordinates in 3D space. It should be noted that we only provide the relationship of coordinates between 2D images and 3D point cloud. They are not equal in numerical value.

Table 1. The relationship between projection parameters and view-ids.

View-id	u	v	Depth	Projection Mode
0	Z	Y	X	0
1	Z	X	Y	0
2	X	Y	Z	0
3	Z	Y	X	1
4	Z	X	Y	1
5	X	Y	Z	1

After packing, these images are padded to generate geometry videos and attribute videos, shown in **Fig. 2(a)** and **(b)**. Padded images are smoother, which are more suitable for video compression. Moreover, occupancy maps are also generated to indicate whether pixels are occupied [12], which are essential to the reconstruction of point cloud, shown in **Fig. 2(c)**. Particularly, occupancy maps in V-PCC are always down sampled for bitrate saving. Therefore, they are up-sampled during geometry reconstruction, which bring a lot of additional points to reconstructed point cloud. After padding, all the videos, maps and auxiliary information are compressed by mature video coding codecs, such as High-Efficiency Video Coding(HEVC) [13]. The whole V-PCC framework is shown in **Fig. 3**.

(a) Geometry Video   (b) Attribute Video   (c) Occupancy Map

Fig. 2. Video and map created from 3D point cloud.

All pixels in padded images could be classified into three groups including origin pixels, padded pixels and unoccupied pixels. Origin pixels are generated from projection of 3D point cloud. They are existent before image padding. Each origin pixel has a relevant point in the uncompressed 3D point cloud. Padded pixels and unoccupied pixels are both generated by

copying values from origin pixels during image padding. For padded pixels, although they have no correlation with the uncompressed 3D point cloud, they are reconstructed to 3D space during geometry reconstruction because of the down-up sampling of occupancy maps, shown in **Fig. 4**. For unoccupied pixels, they are not reconstructed because they are shown unoccupied in up-sampled occupancy maps. Current video coding codecs compress these pixels in the same method, which bring massive waste of bits for geometry videos.
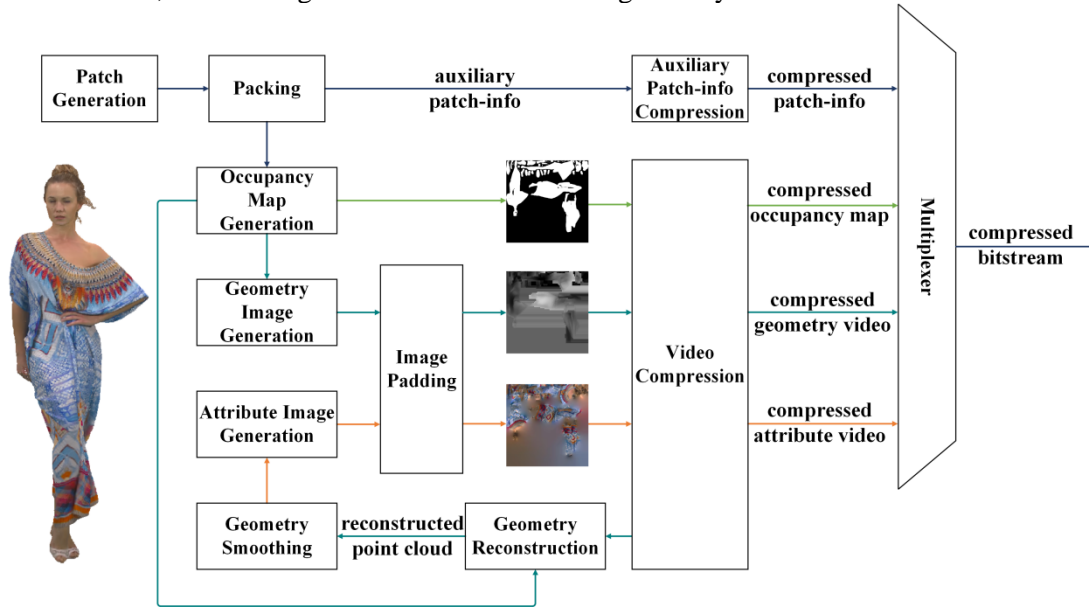

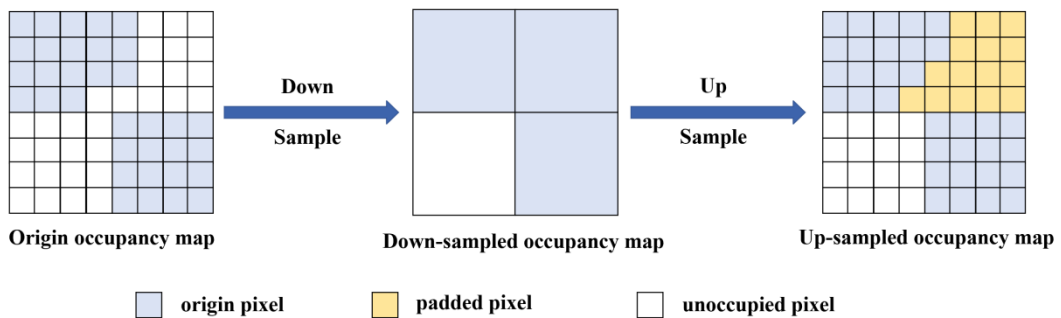**Fig. 3.** Overview of V-PCC framework.

Fig. 4. Down-up sampling of occupancy map.

Many works have been proposed to optimize this compression model in motion prediction, rate control scheme and RDO process. Li et al. proposed an advanced 3D motion prediction to utilize auxiliary information and accurate 3D reconstructed geometry to estimate 2D motion vector used in both geometry videos and attribute videos [14]. The estimated 2D motion vector could well deal with the patch inconsistency problem. Liu et al. proposed a model-based joint bit allocation between geometry videos and attribute videos to reduce the complexity of bit allocation by building a distortion prediction model and a rate prediction model [15]. Xiong et al. proposed an efficient geometry surface coding by estimating the point normal [16]. Herglotz et al. proposed a rate-distortion optimized signal extrapolation algorithm to select better transform coefficients for unoccupied regions [17]. Li et al. proposed an occupancy-

map-based RDO to consider only rate of unoccupied pixels during RDO process [18]. Although these works could achieve bitrate savings in geometry videos or attribute videos, they ignore that the traditional RDO process is not suitable for both padded pixels and unoccupied pixels. Therefore, origin pixels, padded pixels and unoccupied pixels should be treated respectively during the full RDO process.

In this paper, we propose a new RDO process scheme which uses 3D-Distortions instead of traditional video distortions for padded pixels and unoccupied pixels. Firstly, the up-sampled occupancy map is used to classify origin pixels, padded pixels and unoccupied pixels. Secondly, different RDO strategies are applied to these pixels. For origin pixels, the sum square error(SSE) is still used as their distortions since their SSE has a strong correlation with their geometric distortions. For padded pixels, they are reconstructed to 3D space. The K-Dimensional tree [19] constructed from the uncompressed 3D point cloud is used to search their nearest points. The square of the geometric distance between reconstructed points and nearest points are regarded as their 3D-Distortions. The 3D-Distortion is more reliable than SSE for padded pixels because it is based on existent points in the uncompressed 3D point cloud. For unoccupied pixels, their 3D-Distortions are set to "0" because they will not be reconstructed during geometry reconstruction. Finally, the new defined 3D-Distortions replace SSE for padded pixels and unoccupied pixels during full RDO process in intra prediction and inter prediction. As far as we can see, it is the first work to distinguish padded pixels from origin pixels during RDO process.

The rest of this paper is organized as follows. In Section 2, the investigations of current RDO processes are provide at first. Then, the deficiencies of these RDO processes are analyzed. In Section 3, the definition and calculation processes of 3D-Distortions for unoccupied pixels and padded pixels are given. Furthermore, the new 3D-Distortion based RDO and its implementations are explained in detail. In Section 4, experimental setting and results are presented. Finally, concluding remarks and future work are given in Section 5.

2. Investigations of Current RDO Processes

In this section, we first investigate the difference between the traditional full RDO process and the occupancy-map-based RDO process in [18]. Then, we analyze the irrationality of existing RDO processes for padded pixels and unoccupied pixels.

2.1 Investigations of different full RDO processes

In existing video codecs, RDO plays an important role in mode decisions. For example, HEVC provides lots of prediction modes and transform unit sizes in both intra prediction and inter prediction. Full RDO process is always used to decide the best mode of current coding unit. In HEVC, it is calculated by using Lagrange multiplier λ,

$$\min_{P} J = \sum_{i=1}^{N} D_{sse}(i) + \lambda R, \tag{1}$$

where P is encoding parameter, J is the RD cost of current mode, N is the number of pixels in current block, $D_{sse}(i)$ is the square error between the uncompressed pixel and the reconstructed pixel in position i, R is the bitrate of current block including residual bitrate and flag bitrate. The full RDO process used in HEVC could keep the balance of video distortion and video bitrate. However, SSE is not suitable for padded pixels and unoccupied pixels in V-PCC.

The occupancy-map-based RDO process distinguished unoccupied pixels from other pixels, which adds a mask to (1)

$$\min_{P} J = \sum_{i=1}^{N} D_{sse}(i) \times M_i + \lambda R, \tag{2}$$

where M_i indicates whether the pixel in position i is occupied or not. In case of unoccupied pixels, only bitrate is considered when calculating their RD cost. This method is only applied to full RDO process, excluding sum of absolute transformed difference(SATD) [20] and sum of absolute difference(SAD) used in rough intra prediction and motion estimation (ME). The mask is also used in sample-adaptive offset(SAO) process [21], which only considers the weight of occupied pixels when choosing the filter type.

2.2 The deficiencies of current RDO processes

The traditional RDO process for HEVC compresses all pixels in the same quality. It inevitably wastes a lot of bits on unoccupied pixels, which are not reconstructed during geometry reconstruction. The occupancy-map-based RDO makes up for this shortcoming by not considering the distortion of unoccupied pixels. However, it still calculates the distortion of padded pixels by SSE as well as origin pixels. Although both origin pixels and padded pixels are reconstructed to 3D space during geometry reconstruction, they are different in essence.

All origin pixels and padded pixels have relevant 3D points, which are determined by their position and value. Since the positions of pixels in 2D videos are fixed, for reconstructed 3D points, only one of the coordinates is changed according to their relevant 3D points by quantization noise. Specially, SSE in 2D videos is equal to the distortions between the reconstructed 3D points and the relevant 3D points for origin pixels and padded pixels. However, the quality of geometry reconstruction is based on the difference between the reconstructed point cloud and the uncompressed point cloud, not their relevant 3D points. For origin pixels which are generated from projection, their relevant 3D points always belong to the uncompressed 3D point cloud. So, their SSE has a strong correlation with their geometric distortions. Therefore, it is suitable to apply SSE to origin pixels. But for padded pixels which are generated by copying values from origin pixels, their relevant 3D points do not belong to the uncompressed point cloud. For this reason, SSE for padded pixels has no correlation with their actual distortions in geometry reconstruction. In another word, SSE is not able to measure the geometric distortions for padded pixels because their relevant 3D points are not existent in the uncompressed point cloud.

Thus, it is meaningful to figure out a dependable measure for distortion of padded pixels, which is proposed in this paper, named as 3D-Distortion. The proposed 3D-Distortion is based on the minimum Euclidean distance between the reconstructed 3D points and the uncompressed 3D point cloud, which has practical significance for padded pixels. Furthermore, the 3D-Distortions for padded pixels are the same as their point-to-point error, which is regarded as the geometry metric D1 in V-PCC common test model. By keeping consistent with the geometry metric in V-PCC, the 3D-Distortion could better reflect the distortion characteristics for padded pixels, which leads to more bitrate saving than using SSE on padded pixels.

3. Proposed Method

In this section, the definition and calculation processes of 3D-Distortions on unoccupied pixels and padded pixels are firstly provided. Then, the 3D-Distortion based RDO and its implementations are explained in detail. The framework of the 3D-Distortion based RDO is shown in **Fig. 5**. Specifically, the 3D-Distortion is actually the square of geometric distance

between reconstructed points and uncompressed points in the 3D point cloud. For unoccupied pixels, they are not reconstructed and their 3D-Distortions are set to "0". For padded pixels, they are firstly reconstructed to 3D space. A K-Dimensional tree constructed with the uncompressed 3D point cloud is used to find their nearest points. Finally, the square of geometric distance between reconstructed points and nearest points is regarded as the 3D-Distortions for padded pixels.
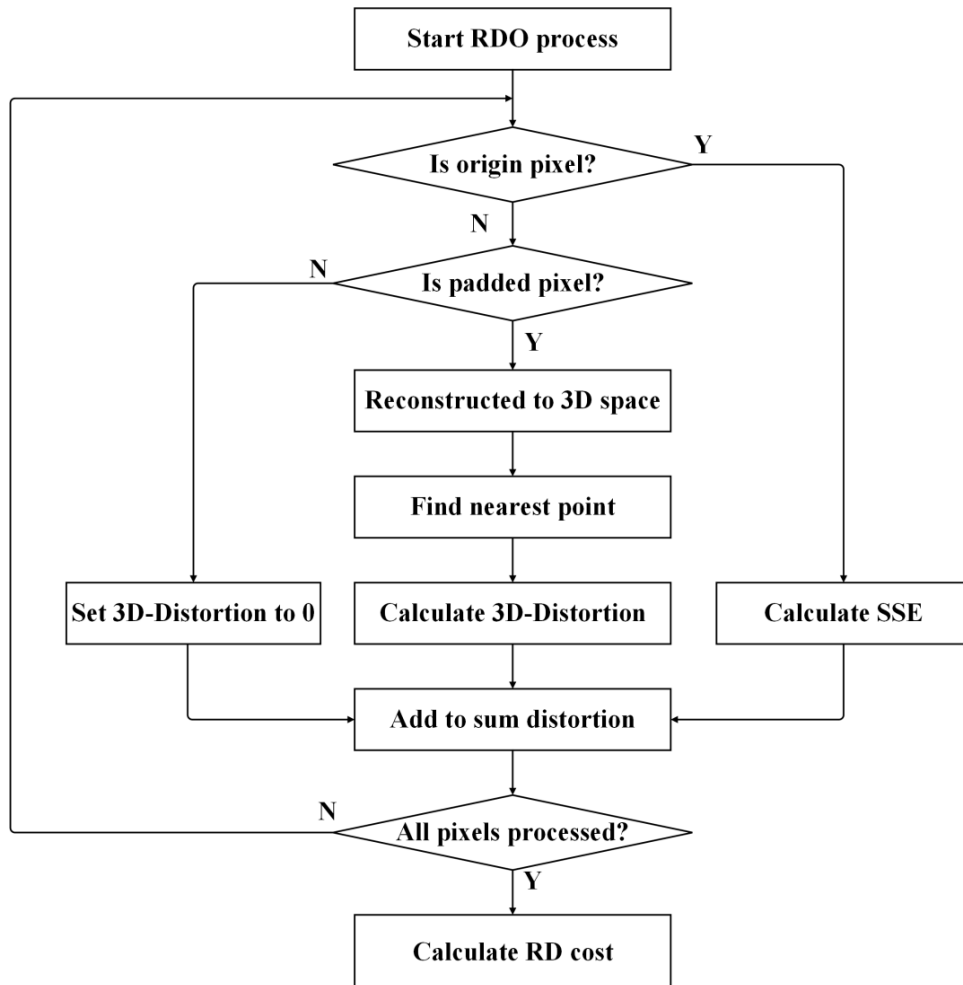
Fig. 5. The framework of 3D-Distortion based RDO.

3.1 3D-Distortions for padded pixels

The calculation processes of 3D-Distortions for padded pixels are divided into three steps: reconstruction, searching nearest points and calculating 3D-Distortions.

    Firstly, padded pixels are reconstructed to 3D space. Their relevant auxiliary patch information is found by their patch index. To each patch, their start position $(u0, v0)$ in 2D images is corresponding to their start position $(x0, y0, z0)$ in 3D space. Particularly, the offset of coordinates in 2D images equals to the offset of coordinates in 3D space in two dimensions. Furthermore, $u0$ and $v0$ are always divided by occupancy resolution for bitrate saving, and thus they have to multiply by occupancy resolution during reconstruction. For example, when the view-id of pixel (u, v) is "0", its 3D point coordinates (x, y, z) are calculated as follows,

$$\begin{cases} x = x0 + value(u,v) \\ y = y0 + (v - v0 \times occupancy\_resolution) \\ z = z0 + (u - u0 \times occupancy\_resolution). \end{cases} \quad (3)$$

It should be noted that the specific formulas for 3D coordinates in (3) are changed according to the view-id of padded pixels.

After reconstruction, a K-Dimensional tree is used to find the nearest points of the reconstructed points. A K-Dimensional tree is a space-partitioning data structure for organizing points in a K-Dimensional space. It is always utilized to find neighbor points during point cloud processing by nearest neighbor search(NNS) [22]. NNS is also a form of proximity search, which is based on Euclidean distance. The K-Dimensional tree used in the proposed method is constructed with the uncompressed 3D point cloud before patch segmentation. To each reconstructed point (x, y, z), one nearest point (x_o, y_o, z_o) could be found based on the minimum Euclidean distance.

Finally, the 3D-Distortion for padded pixels is calculated. It is based on the square of geometric distance between reconstructed points and nearest points as follows,

$$D = (x - x_o)^2 + (y - y_o)^2 + (z - z_o)^2. \quad (4)$$

3.2 3D-Distortion based RDO process

The 3D-Distortion based RDO process treated origin pixels, padded pixels and unoccupied pixels with different strategies. For origin pixels, their SSE is calculated as their distortions. For padded pixels and unoccupied pixels, their SSE is replaced by obtained 3D-Distortions. The RD cost of full RDO process is calculated as follows,

$$\min_P J = \sum_{i=0}^{N} D_{sse}(i) + \sum_{j=0}^{K} D_{3D}(j) + \lambda R, \quad (5)$$

where N is the number of origin pixels and K is the number of padded pixels.

This method is applied to full RDO process in HEVC/VVC, including precise intra prediction, inter prediction and residue tree split. The occupancy map mask proposed in [18] is also applied to our SAO process to avoid significant loss caused by unoccupied pixels. The 3D-Distortion based RDO process is not utilized in rough RDO process in HEVC/VVC, such as SATD in rough intra prediction and ME in inter prediction. Furthermore, the proposed method is only applied to geometry videos. Since the generation of attribute videos is based on geometry reconstruction, our method has slight impact on attribute reconstruction.

4. Experiment Results

4.1 Test conditions

The proposed method is implemented in the V-PCC reference software TMC2-v13.0 [23] and the HEVC reference software HM-16.20+SCM-8.8 [24]. The experiments are tested on a PC with an Intel (R) 2.40GHz processor, 128 Gb RAM. We test lossy geometry and lossy attribute defined in V-PCC common test condition(CTC) in both All Intra setting and Random Access setting. We test five DPCs defined in CTC, as shown in **Fig. 1(a)-(e)**. Each DPC is tested with first 32 frames as a good representation of whole frames. CTC also gives five pairs of quantization parameters(QPs) for testing, which cover a wide range of bitrate. The details of QPs are shown in **Table 2**. The proposed method is compared with V-PCC anchor, the occupancy-map-based RDO and the efficient geometry surface coding(EGSC) method in

terms of Bjontegaard Delta rate(BD-rate) [25] in All Intra setting and Random Access setting. The quality of the reconstructed point cloud is evaluated in both geometry reconstruction and attribute reconstruction. For geometry reconstruction, the point-to-point error(D1) and the point-to-plane error(D2) is used as the geometry metrics in CTC. For attribute reconstruction, Luma-PSNR, Cb-PSNR and Cr-PSNR are regarded as the attribute metrics in CTC.

Table 2. QP settings of CTC.

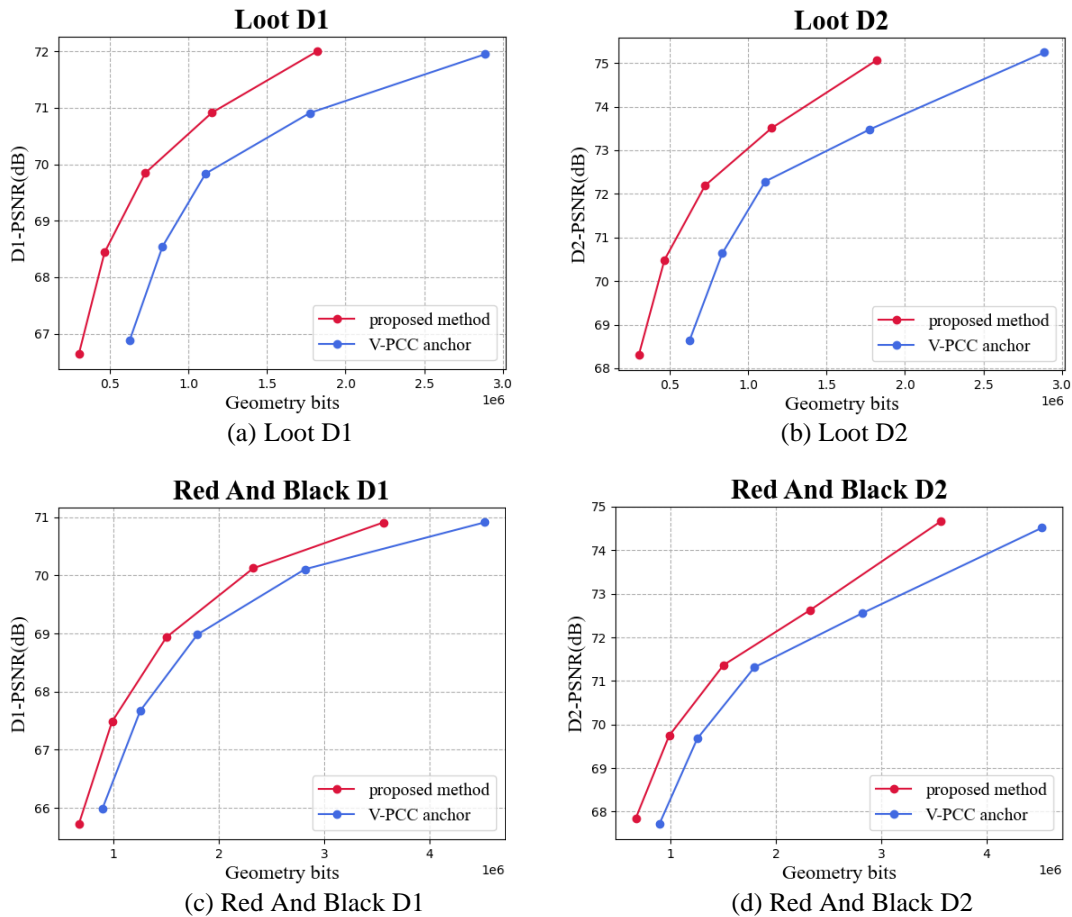| QP setting	Geometry QP	Attribute QP
r1	32	42
r2	28	37
r3	24	32
r4	20	27
r5	16	22

4.2 Experimental results

In this section, the experimental results are provided. We compare the proposed method with V-PCC anchor and the occupancy-map-based RDO in both Random Access setting and All Intra setting. The complexity of the proposed method is also presented.
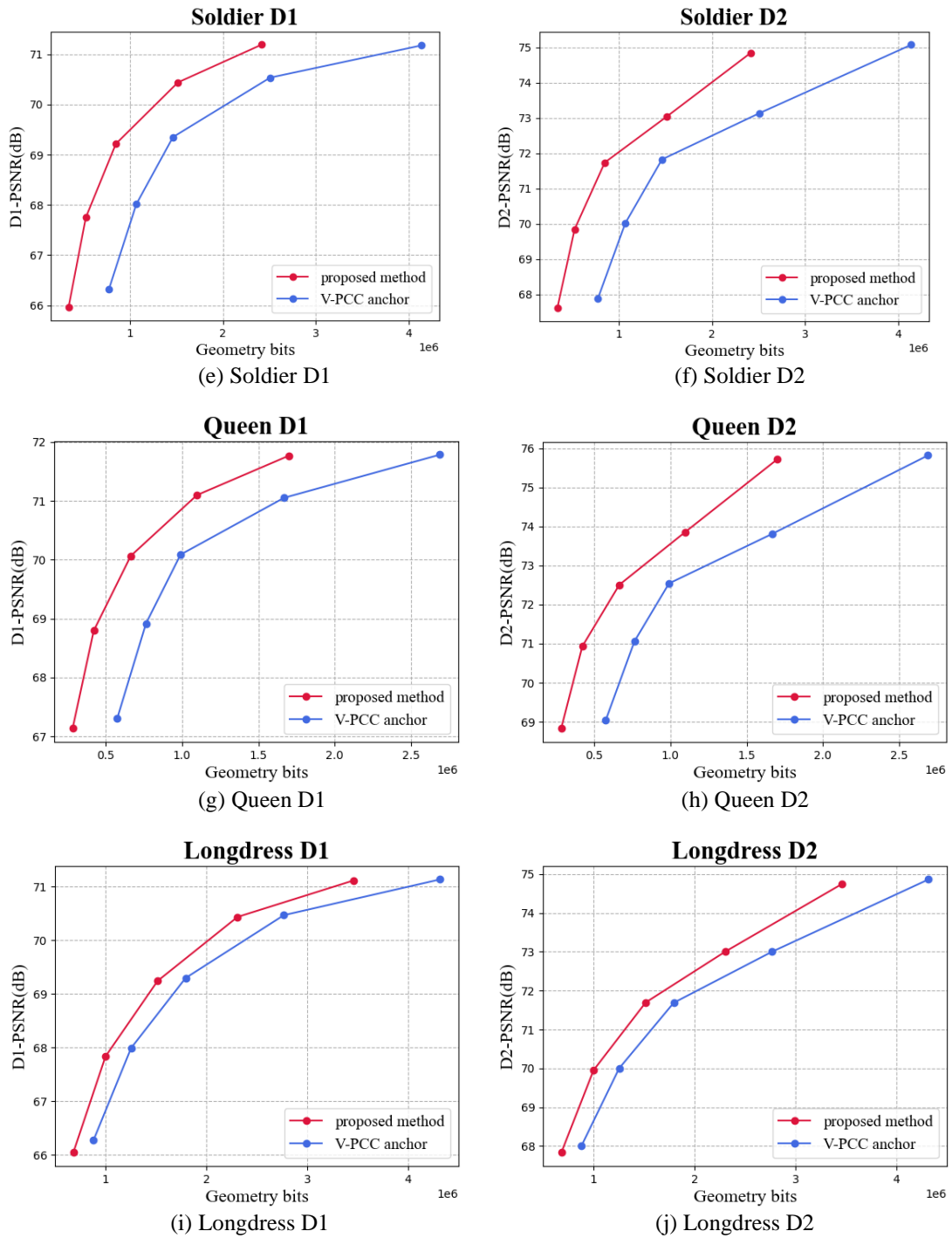
Table 3 shows the performance of the proposed 3D-Distortion based RDO process compared with V-PCC anchor in Random Access setting. It is observed that the proposed method can lead to an average of 31.41% and 31.64% bitrate saving for D1 and D2 metrics. In particular, for the point cloud *Soldier*, the proposed method could save 43.98% and 44.32% bitrate for D1 and D2 metrics respectively. *Longdres*s and *Redandblack* have more complex geometric textures, such as the fold of skirts and curly hair. These textures vary greatly in depth, which inevitably account for a large part of bitrates. Our method mainly aims at improving the RDO process for unoccupied pixels and padded pixels. However, pixels from these complex geometric textures are always occupied pixels, which still use SSE as their distortions. Therefore, these two point cloud sequences could not achieve as much bitrate saving as others. The high bitrate saving in geometry videos inevitably sacrifices little quality of geometry reconstruction, which causes an average of 0.90%, 0.47% and 0.61% bitrate increase in attribute videos for Luma, Cb and Cr. The RD curves for geometry videos in Random Access setting are also provided in **Fig. 6**. It is more obvious that the proposed method could have a significant bitrate saving by observing these RD curves.

Table 3. Performance of 3D-Distortion based RDO compared with V-PCC anchor (Random Access setting).

Point Cloud Sequences	Geom.BD-GeomRate		Attr.BD-AttrRate		
	D1	D2	Luma	Cb	Cr
Loot	-39.86%	-38.56%	0.60%	0.47%	-2.00%
Redandblack	-17.79%	-19.00%	0.36%	0.09%	0.43%
Soldier	-43.98%	-44.32%	-0.07%	-0.57%	0.50%
Queen	-39.56%	-38.45%	3.14%	2.30%	3.60%
Longdress	-16.45%	-17.88%	0.46%	0.06%	0.55%
Average	-31.41%	-31.64%	0.90%	0.47%	0.61%

Table 4 shows the performance of proposed method compared with V-PCC anchor in All Intra setting. It is shown that the proposed method achieves an average of 6.14% and 6.92% bitrate saving for D1 and D2 metrics. It can be seen that the bitrate saving is more in Random Access setting than in All Intra setting. The reason is that the proposed method is not applied to SATD, the main RDO process in intra prediction because SATD transforms the block as a whole while the proposed method calculates the 3D-Distortion in pixels. Therefore, it is not suitable to apply 3D-Distortions to SATD. However, the proposed method could be applied to most RDO process in inter prediction. It is certain that the proposed method could achieve more bitrate saving in Random Access setting because it is used more frequently in inter prediction than in intra prediction.

(a) Loot D1

(b) Loot D2

(c) Red And Black D1

(d) Red And Black D2

(e) Soldier D1

(f) Soldier D2

(g) Queen D1

(h) Queen D2

(i) Longdress D1

(j) Longdress D2

Fig. 6. RD curves for D1 in Random Access setting.

Table 4. Performance of 3D-Distortion based RDO compared with V-PCC anchor(All Intra setting).

Point Cloud Sequences	Geom.BD-GeomRate		Attr.BD-AttrRate		
	D1	D2	Luma	Cb	Cr
Loot	-6.43%	-7.03%	0.05%	0.76%	-0.36%
Redandblack	-6.48%	-7.71%	-0.08%	0.02%	0.12%
Soldier	-5.78%	-6.62%	-0.02%	0.99%	0.33%
Queen	-6.20%	-6.43%	0.63%	-0.30%	-0.43%
Longdress	-5.82%	-6.81%	-0.04%	-0.15%	-0.05%
Average	-6.14%	-6.92%	0.11%	0.26%	-0.08%

Table 5 shows the performance of proposed method compared with the occupancy-map-based RDO in [18]. It can be seen that the proposed method has an average of 4.90% and 5.84% bitrate saving in Random Access setting while 0.53% and 1.56% bitrate saving in All Intra setting for geometry metrics D1 and D2 respectively. The performance also indicates that the proposed method is very effective for padded pixels. More bitrate saving is achieved when 3D-Distortions are applied to padded pixels.

Table 5. BD-rate of 3D-Distortion based RDO compared with the occupancy-map-based RDO.

Point Cloud Sequences	Random Access		All Intra	
	D1	D2	D1	D2
Loot	-5.07%	-5.21%	-0.20%	-0.91%
Redandblack	-2.47%	-3.97%	-0.69%	-2.00%
Soldier	-8.65%	-9.81%	-0.42%	-1.69%
Queen	-5.96%	-6.20%	-0.96%	-1.39%
Longdress	-2.37%	-4.03%	-0.38%	-1.82%
Average	-4.90%	-5.84%	-0.53%	-1.56%

Table 6 shows the performance of proposed method compared with the EGSC method in [16]. The proposed method has an average of 24.42% and 24.25% bitrate saving in Random Access setting while 2.03% and 2.10% bitrate saving in All Intra setting for geometry metrics D1 and D2. The reason of more bitrate saving is that the proposed method takes pixels from all layers into consideration while the EGSC method only processes pixels on far layers during their RDO process.

Table 6. BD-rate of 3D-Distortion based RDO compared with the EGSC method.

Point Cloud Sequences	Random Access		All Intra	
	D1	D2	D1	D2
Loot	-31.82%	-30.62%	-1.29%	-2.81%
Redandblack	-13.25%	-14.51%	-2.44%	-3.55%
Soldier	-35.15%	-34.99%	-2.41%	-1.71%
Queen	-29.10%	-27.92%	-1.21%	-0.06%
Longdress	-12.80%	-13.19%	-2.78%	-2.35%
Average	-24.42%	-24.25%	-2.03%	-2.10%

The complexity of the proposed method is shown in **Table 7**, which is compared with V-PCC anchor in both Random Access setting and All Intra setting. The complexity of the anchor is defined as 100%. It is revealed that the increase encoding time of the proposed method is an average of 6.9% and 4.3% respectively, which is mainly from the search of points in K-Dimensional space.

Table 7. Complexity of 3D-Distortion based RDO compared with V-PCC anchor.

Point Cloud Sequences	Random Access	All Intra
Loot	102.2%	102.1%
Redandblack	105.9%	103.9%
Soldier	109.5%	103.7%
Queen	107.3%	107.2%
Longdress	109.8%	104.9%
Average	106.9%	104.3%

5. Conclusion

In this paper, we propose a 3D-Distortion based RDO process which replaces traditional SSE with 3D-Distortions for padded pixels and unoccupied pixels during full RDO process. For padded pixels, they are reconstructed to 3D space. Their 3D-Distortions are obtained by finding the nearest points in 3D space with the help of K-Dimensional tree. The square of geometric distance between reconstructed points and nearest points is defined as their 3D-Distortions. For unoccupied pixels, their 3D-Distortions are set to "0". Experimental results show that the proposed algorithm achieves an average of 31.41% and 6.14% bitrate saving for D1 metric in Random Access setting and All Intra setting on geometry videos compared with V-PCC anchor. In the future, we will further work on improving RDO process for V-PCC. We will try to figure out deeper relationship between 3D-Distortions and SSE of these pixels. We will also try to apply our method to attribute videos to get more bitrate saving.

References

[1] Z. Zhang, B. Cao, J. Guo, D. Weng, Y. Liu and Y. Wang, "Inverse Virtual Reality: Intelligence-Driven Mutually Mirrored World," in *Proc. of 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 735-736, Aug. 2018. Article (CrossRef Link)

[2] FuchsHenry, StateAndrei, and BazinJean-Charles, "Immersive 3D Telepresence," *Computer*, vol. 47, no. 7, pp. 46–52, Jul. 2014. Article (CrossRef Link)

[3] C. Timmerer, "Immersive Media Delivery: Overview of Ongoing Standardization Activities," *IEEE Communications Standards Magazine*, vol. 1, no. 4, pp. 71-74, Dec. 2017. Article (CrossRef Link)

[4] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *Proc. of 2011 IEEE International Conference on Robotics and Automation*, pp. 1-4, Aug. 2011. Article (CrossRef Link)

[5] W. Zhu, Z. Ma, Y. Xu, L. Li and Z. Li, "View-Dependent Dynamic Point Cloud Compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 765-781, Feb. 2021. Article (CrossRef Link)

[6] D. Wang, W. Zhu, Y. Xu, Y. Xu and LeYang, "Visual Quality Optimization for View-Dependent Point Cloud Compression," in *Proc. of 2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1-5, May. 2021. Article (CrossRef Link)

[7] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *Proc. of 2011 IEEE International Conference on Robotics and Automation*, pp. 1-4, Aug. 2011. Article (CrossRef Link)

[8] E. S. Jang et al., "Video-Based Point-Cloud-Compression Standard in MPEG: From Evidence Collection to Committee Draft [Standards in a Nutshell]," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 118-123, May. 2019. Article (CrossRef Link)

[9] Graziosi D, Nakagami O, Kuma S, "An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, pp. 1-17, Sep. 2020. Article (CrossRef Link)

[10] Mammou, A. M. Tourapis, D. Singer, and Y. Su, "Video-based and Hierarchical Approaches Point Cloud Compression," document m41649, Macau, China, Oct. 2017. Article (CrossRef Link)

[11] L. Li, Z. Li, S. Liu and H. Li, "Efficient Projected Frame Padding for Video-Based Point Cloud Compression," *IEEE Transactions on Multimedia*, vol. 23, pp. 2806-2819, May. 2021. Article (CrossRef Link)

[12] V. Zakharchenko, "V-PCC codec description," Document ISO/IEC JTC1/SC29/WG11 N18487, Geneva, CH, March. 2019. Article (CrossRef Link)

[13] G. J. Sullivan, J. -R. Ohm, W. -J. 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. Article (CrossRef Link)

[14] L. Li, Z. Li, V. Zakharchenko, J. Chen and H. Li, "Advanced 3D Motion Prediction for Video-Based Dynamic Point Cloud Compression," *IEEE Transactions on Image Processing*, vol. 29, pp. 289-302, Aug. 2019. Article (CrossRef Link)

[15] Q. Liu, H. Yuan, J. Hou, R. Hamzaoui and H. Su, "Model-Based Joint Bit Allocation Between Geometry and Color for Video-Based 3D Point Cloud Compression," *IEEE Transactions on Multimedia*, vol. 23, pp. 3278-3291, Sep. 2020. Article (CrossRef Link)

[16] J. Xiong, H. Gao, M. Wang, H. Li, K. N. Ngan and W. Lin, "Efficient Geometry Surface Coding in V-PCC," *IEEE Transactions on Multimedia*, 2022. Article (CrossRef Link)

[17] C. Herglotz, N. Genser and A. Kaup, "Rate-Distortion Optimal Transform Coefficient Selection for Unoccupied Regions in Video-Based Point Cloud Compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 11, pp. 7996-8009, 2022. Article (CrossRef Link)

[18] L. Li, Z. Li, S. Liu and H. Li, "Occupancy-Map-Based Rate Distortion Optimization and Partition for Video-Based Point Cloud Compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 1, pp. 326-338, Jan. 2021. Article (CrossRef Link)

[19] C. Qian, W. Lin, J. Chen, Z. Li and H. Gao, "Point cloud trajectory planning based on Octree and K-dimensional tree algorithm," in *Proc. of 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC)*, pp. 213-218, Nov. 2016. Article (CrossRef Link)

[20] J. Vanne, M. Viitanen, T. D. Hamalainen and A. Hallapuro, "Comparative Rate-Distortion-Complexity Analysis of HEVC and AVC Video Codecs," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1885-1898, Dec. 2012. Article (CrossRef Link)

[21] C. -M. Fu, C. -Y. Chen, Y. -W. Huang and S. Lei, "Sample adaptive offset for HEVC," in *Proc. of 2011 IEEE 13th International Workshop on Multimedia Signal Processing*, pp. 1-5, Dec. 2011. Article (CrossRef Link)

[22] Y. Gao, B. Zheng, G. Chen, W. -C. Lee, K. C. K. Lee and Q. Li, "Visible Reverse k-Nearest Neighbor Query Processing in Spatial Databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1314-1327, Sep. 2009. Article (CrossRef Link)

[23] Point Cloud Compression Category 2 Reference Software, TMC2-13.0, 2022. [Online]. Available: https://github.com/MPEGGroup/mpeg-pcc-tmc2

[24] High Efficiency Video Coding test model, HM-16.20+SCM-8.8, 2022, [Online]. Available:https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/tags/

[25] G. Bjontegaard, "Calculation of Average PSNR Differences Between RD-Curves," document VCEG-M33, Austin, Texas, USA, Apr. 2001 [Online]. Available: https://www.researchgate.net/publication/244455155

Yihao Fu received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, Shanghai, China, in 2020, where he is currently pursuing the M.S. degree. His research interests include video-based point cloud compression, rate distortion optimization.
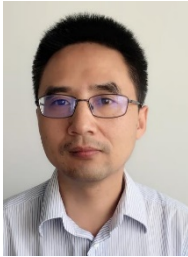
Liquan Shen received the B.S. degree in automation control from Henan Polytechnic University, Jiaozuo, China, in 2001, and the M.E. and Ph.D. degrees in communication and information systems from Shanghai University, Shanghai, China, in 2005 and 2008, respectively. Since 2008, he has been a Faculty Member with the School of Communication and Information Engineering, Shanghai University, where he is currently a Professor. From November 2013 to November 2014, he was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Florida, Gainesville. He has authored or coauthored more than 100 refereed technical papers in international journals and conferences in the field of video coding and image processing. He holds ten patents in the areas of image/video coding and communications. His major research interests include high efficiency video coding, perceptual coding, video codec optimization, 3DTV, and video quality assessment.

Tianyi Chen received the B.S. degree from the School of Communication and Information Engineering, Shanghai University, Shanghai, China, in 2017, where he is currently pursuing the Ph.D. degree. His research interests include video-based point cloud compression, geometry repairing.