中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2025)11-3535-12

论文引用格式:Qian Y J and Ding D D. 2025. Dynamic point cloud compression method based on inter frame adaptive prediction. Journal of Image and Graphics, 30(11):3535-3546(钱虞杰, 丁丹丹. 2025. 基于帧间自适应预测的动态点云压缩. 中国图象图形学报, 30(11):3535-3546)[DOI:10.11834/jig.240551]

基于帧间自适应预测的动态点云压缩

钱虞杰,丁丹丹*

杭州师范大学信息科学与技术学院,杭州　311121

摘　要:**目的**　动态点云由于其在三维空间的复杂空间相关性和时间相关性,在动态压缩上面临着巨大挑战。为此,提出一种结合了稀疏卷积和最近邻聚合的邻域特征提取算法来提升动态点云压缩效率。**方法**　提出两种方式来捕获点云的时空域运动信息:一方面融合相邻帧点云进行稀疏卷积,另一方面使用目标k-近邻(k-nearest neighbor,kNN)算法。然后,根据提取到的动态特征信息预测当前帧每个体素块的占用概率,实现对当前帧的精确预测。**结果**　所提方法遵循国际运动图像专家组(Moving Picture Experts Group,MPEG)推荐的通用测试条件进行训练与测试,与MPEG提出的动态点云标准压缩方法GeS(inter)以及基于视频的点云压缩方法(video-based point cloud compression,V-PCC)相比,在PSNR(peak signal-to-noise ratio)D1(D2)上平均分别获得94.14%(96.67%)和82.93%(72.57%)的BD-Rate增益;与当前两种较先进的基于学习的动态点云压缩方法相比,在PSNR D1(D2)上平均获得22.53%(26.73%)和14.21%(40.80%)的BD-Rate增益。**结论**　综上,本文方法能够有效地提取动态点云帧间特征,提升了点云压缩效率与重建质量,为动态点云数据的应用和传输提供支持。

关键词:动态点云压缩;帧间特征;目标邻域特征提取;k-近邻(kNN);卷积特征提取

Dynamic point cloud compression method based on inter frame adaptive prediction

Qian Yujie, Ding Dandan*

School of Information Science and Technology, Hangzhou Normal University, Hangzhou　311121, China

Abstract: **Objective** Dynamic point clouds refer to collections of 3D data points that represent the surfaces of objects in a scene at various instances in time. These point clouds are increasingly being utilized across a wide array of applications, including virtual reality, augmented reality, autonomous vehicles, and robotics. Each of these applications demands high-fidelity representations of dynamic environments, which makes the efficient handling of dynamic point clouds a critical area of research. However, the compression of dynamic point clouds presents great challenges due to the complex spatial and temporal correlations inherent in these 3D structures. As the demand for real-time and high-quality 3D data increases, addressing these challenges becomes essential for the advancement of technologies in various fields. **Method** In recognition of the intricate challenges associated with dynamic point cloud compression, this study proposes an innovative method that effectively combines two distinct yet complementary techniques: a convolutional feature extraction algorithm and a target neighborhood feature extraction algorithm. The first component of the proposed method, that is, the convolutional feature

收稿日期:2024-09-10;**修回日期**:2025-03-06;**预印本日期**:2025-03-13

*通信作者:丁丹丹　dandanding@hznu.edu.cn

基金项目:国家自然科学基金项目(62171174)

Supported by:National Natural Science Foundation of China(62171174)

extraction algorithm, plays a crucial role in distilling local spatiotemporal motion information from aligned point clouds. By applying convolutional operations, this method can extract hierarchical feature representations that encapsulate essential spatial patterns and motion dynamics present in the point cloud data. This hierarchical representation is vital for understanding how objects within the scene move and interact over time. The second key aspect of the proposed compression method involves the use of an adaptive *k*-nearest neighbor (*k*NN) algorithm. Given the varying motion amplitudes of dynamic point clouds, capturing long-distance spatiotemporal information effectively is important. The adaptive *k*NN algorithm helps achieve this capability by identifying relevant neighboring points across frames. By utilizing the *k*NN algorithm, the proposed method can identify and focus on critical data points that provide context and continuity over time. Therefore, even points that are not immediately adjacent in space but are relevant to the current data frame can still be considered, which maintains a broader view of the spatial dynamics at play. One of the fundamental innovations of this method is its capability to predict the occupancy probability of each voxel block in the current frame based on the extracted dynamic feature information. This prediction mechanism enables the compression algorithm to accurately determine which areas of the voxel grid are likely to contain meaningful data points. By anticipating these areas, the method can be more selective regarding the data it retains, which allows for efficient compression without sacrificing important details. This predictive capability is especially useful in scenarios where computational resources are limited, as it enables effective data management. **Result** The effectiveness of the proposed method is evaluated against established standards and cutting-edge techniques in the field. The methodology adheres to general testing conditions recommended by the International Moving Picture Experts Group (MPEG) for training and evaluation. In comparative analyses, the proposed method demonstrates impressive performance against existing dynamic point cloud compression standards such as GeS (inter) and MPEG's video-based point cloud compression method (V-PCC). Specifically, the proposed method achieves average Bjontegaard Delta (BD) rate gains of 94.14% (96.67%) and 82.93% (72.57%) on peak signal-to-noise ratio (PSNR) for datasets D1 and D2, respectively. When compared with the two most advanced learning-based dynamic point cloud compression methods currently available, the results remain equally compelling. The proposed method achieves average BD rate gains of 22.53% (26.73%) and 14.21% (40.80%) on PSNR D1 (D2). These metrics further illustrate the superiority of the proposed approach, which showcases its capability to achieve high-quality reconstructions of dynamic point clouds while significantly optimizing data size. The implications of this research extend beyond mere numerical improvements. The proposed compression method represents a meaningful advancement in the field of dynamic point cloud data handling, which provides critical support for applications that rely on the efficient transmission and storage of large volumes of dynamic data. **Conclusion** The dynamic point cloud compression method proposed in this study effectively addresses the intricate challenges posed by dynamic data in 3D spaces through the innovative integration of convolutional feature extraction and adaptive *k*NN techniques. By accurately predicting voxel occupancy and leveraging the strengths of each component algorithm, the proposed method demonstrates significant improvements in compression efficiency while maintaining high-quality point cloud reconstructions. The implications of this research extend beyond mere numerical improvements, which represents a meaningful advancement in the field of dynamic point cloud data handling. This advancement provides critical support for applications that rely on the efficient transmission and storage of large volumes of dynamic data. Future research directions could explore additional enhancements, such as integrating diverse machine learning strategies, improving real-time processing capabilities, and validating the approach with a wider range of dynamic datasets. Ultimately, these improvements drive further innovation in dynamic point cloud technologies. The continuous evolution of this field has strong potential for unlocking new possibilities in immersive experiences, autonomous systems, and various other domains that depend on precise and efficient 3D data management.

Key words: dynamic point cloud compression; inter frame features; target neighborhood feature extraction; *k*-nearest neighbor (*k*NN); convolutional feature extraction

0 引 言

动态点云是一种描述三维空间中物体运动的数据表示形式，由一系列点云构成，每个点云表示物体在不同时间点的位置信息。它能够精确地表达目标物体的运动和变形等动态细节，因此广泛应用于虚拟现实（龚靖渝 等，2023）、机器人技术和自动驾驶（龙霄潇 等，2021）等领域。然而，动态点云数据随时间变化，且每一帧点云都包含大量的点，给动态点云压缩带来了巨大挑战。

点云压缩按照点云是否随时间变化分为静态点云压缩和动态点云压缩，目前相关研究主要集中在如何通过传统的方法或基于学习的方法高效存储和传输这些三维时空数据。

传统的动态点云压缩方法主要分为两种：基于二维视频压缩的方法和基于三维模型的方法。

在基于二维的方法上，当前高效的视频编码技术被用于处理从三维点云映射得到的二维图像投影。例如Zhu等人（2021）通过结合全局和区域投影的方法，提高投影的准确性。2019年，国际运动图像专家组（Moving Picture Experts Group，MPEG）提出基于视频的点云压缩方法（video-based point cloud compression，V-PCC）（Schwarz 等，2019）。V-PCC技术利用立方投影和补丁投影等方式生成几何和纹理图像，然后利用高效视频编码器（high efficiency video coding，HEVC）（Sullivan 等，2012）等成熟的视频编码标准来压缩所生成的二维图像序列。尽管将三维数据转化为二维投影难免会引起一些失真和帧间匹配问题，但由于视频编码器的高效性，V-PCC仍然在所有传统动态点云压缩方法中表现出卓越的性能。

在基于三维的方法上，点云通常以八叉树进行构建，通过分析八叉树各节点之间的联系来预测帧间的动态关系。大多数基于三维的方法利用手动设计的特征提取模块来捕获动态点云序列中的帧间运动特征（Santos 等，2021）。de Queiroz 和 Chou（2017）以及Liu等人（2022）利用前后帧点云块的匹配信息处理三维空间中的动态点云，根据点云划分和匹配算法确定点云。MPEG提出基于几何的点云压缩标准（geometry-based point cloud compression，G-PCC）（Graziosi 等，2020）。G-PCC以八叉树结构递归细分体素化点云立方体，利用邻域相关性提高压缩效率。GeS（inter）作为G-PCC的动态解决方案在G-PCC的基础上进一步通过分块来利用块之间的运动信息，增加了对于帧间点云时间相关性的利用，减少重复编码的冗余信息（Lasserre 和 Taquet，2023）。还有其他相关研究基于八叉树结构提出二级概率估计方法，分两步利用上下文信息分别进行帧内和帧间节点的概率估计（王哲诚 等，2023），但这些方法由于八叉树的构建和遍历较为复杂，计算开销较大。

为了克服传统方法的这一挑战，相关学者探索了大量基于深度学习的点云压缩方法，并且首先在静态点云压缩领域得到了广泛的应用。基于学习的静态点云压缩方法主要分为3类：基于点的方法（Gao 等，2021）、基于体素的方法（Guarda 等，2021；Wang 等，2021b）和基于八叉树的方法（Huang 等，2020）。基于点的方法利用点之间的关系提取特征来进行编码和解码，通常通过多层感知机实现（Huang 和 Liu，2019）。然而，对于大规模的点云数据，这类方法需要较多的计算资源和时间来进行训练和推理，限制了其在实时应用中的效率（Pang 等，2022）。基于体素的方法将点云数据转换为三维体素网格，通过学习体素网格的表示和特征来实现压缩和重建。例如，Wang等人（2021a）基于三维稀疏卷积实现体素化点云压缩和分层重建，提高压缩效率，且在大规模点云数据集上取得优秀的性能。朱威等人（2023）通过密集残差结构确保信息在网络中得到充分的传播，然后利用多尺度剪枝针对不同尺度的特征进行优化，提升点云压缩效果。基于八叉树的方法通过将点云数据分割成不同的八叉树节点，然后借助神经网络模型预测每个八叉树节点的占用概率（Que 等，2021），这类方法多用于LiDAR点云这种点分布较稀疏的点云（Fu 等，2022）。

受到静态点云压缩领域基于深度学习方法的启发，基于学习的动态点云压缩方法近年来得到了越来越多的关注。这些方法利用卷积操作从点云序列的连续帧中提取运动特征，通过帧间预测减少了传输比特消耗。相比拥有最好性能的传统动态点云压缩方法V-PCC进一步提升了性能。一种典型的方法是基于运动估计来捕捉动态场景中物体的运动轨迹，利用帧间冗余信息进行压缩。其中，Akhtar等人（2024）利用稀疏卷积将参考帧特征提取到当前帧，并通过多尺度特征融合提取帧间运动信息。而Fan

等人(2022)提出一种基于稀疏卷积的运动估计和补偿方案,首先将参考帧的坐标与当前帧对齐,通过卷积提取帧间特征,然后对这些特征进行残差编码压缩。在解码端,提出用于运动补偿的三维自适应加权插值算法对k个最近邻居进行插值得到特征预测。Xia等人(2023)在Fan等人(2022)的基础上,使用球面k-近邻(k-nearest neighbor,kNN)算法代替稀疏卷积层提取帧间运动信息。但这些基于稀疏卷积或球面kNN的方法只能在固定的感受野内捕获前后帧之间运动信息。最新的基于学习的方法利用场景流网络进行运动向量估计,捕捉细微的运动变化进行运动补偿和压缩(江照意 等,2024),但场景流估计的计算复杂度较高。

受上述工作启发,本文提出一种新的动态点云压缩方法。该方法基于稀疏卷积分析了当前帧和先前重建的帧之间的相关性,通过帧间预测减少了传输比特消耗。在帧间预测方面,所提方法结合了卷积特征提取算法和目标邻域特征提取算法,通过融合相邻帧点云进行稀疏卷积和使用目标kNN算法分别捕捉点云中较小运动幅度区域和较大运动幅度区域的帧间运动信息,自适应增强动态点云压缩效果。实验结果表明,所提方法在MPEG推荐的通用测试条件下,与MPEG提出的动态点云标准压缩方法GeS(inter)以及基于视频的点云压缩方法V-PCC相比,在PSNR(peak signal-to-noise ratio) D1(D2)上平均分别获得94.14%(96.67%)和82.93%(72.57%)的BD-Rate(Bjontegaard Delta rate)增益;与当前两种最先进的基于学习的动态点云压缩方法相比,在PSNR D1(D2)上平均获得22.53%(26.73%)和14.21%(40.80%)的BD-Rate增益。由此可见,所提方法相比以往的动态点云压缩方法实现了更好的性能,特别是在捕捉兼具较大范围和较小范围的运动信息方面表现优异。

1 本文方法

1.1 概述

本文所提方法的动态点云压缩框架如图1所示。其中,$\boldsymbol{X}_t=\{\boldsymbol{C}_{X_t},\boldsymbol{F}_{X_t}\}$和$\boldsymbol{X}_{t-1}=\{\boldsymbol{C}_{X_{t-1}},\boldsymbol{F}_{X_{t-1}}\}$是相邻的两个点云帧,$\boldsymbol{C}_{X_t}$和$\boldsymbol{C}_{X_{t-1}}$是坐标矩阵,$\boldsymbol{F}_{X_t}$和$\boldsymbol{F}_{X_{t-1}}$是指示体素占用的全1特征矩阵。该网络分析了当前帧点云$\boldsymbol{X}_t$和先前的重建帧点云$\boldsymbol{X}_{t-1}$之间的相关性,旨在通过帧间预测来减少传输比特消耗。具体地,$\boldsymbol{X}_t$和$\boldsymbol{X}_{t-1}$首先分别经过两次体素下采样得到$\boldsymbol{y}_t=\{\boldsymbol{C}_{y_t},\boldsymbol{F}_{y_t}\}$和$\boldsymbol{y}_{t-1}=\{\boldsymbol{C}_{y_{t-1}},\boldsymbol{F}_{y_{t-1}}\}$。将$\boldsymbol{y}_t$的特征矩阵置为全1并重新提取特征得到$\boldsymbol{y}_t'$,然后$\boldsymbol{y}_t'$和$\boldsymbol{y}_{t-1}$被送入动态特征提取模块提取帧间特征以及运动信息,动态特征提取模块由基于kNN的目标特征聚合模块和基于卷积的前后帧特征提取模块共同构成。残差特征压缩对$\boldsymbol{F}_{y_t}$和动态特征提取模块提取到的$\boldsymbol{F}$之间的残差进行压缩。同时对坐标信息$\boldsymbol{C}_{y_t}$无损压缩。最后在解码端将通过参数共享得到的$\boldsymbol{F}$和解码得到的残差求和得到$\boldsymbol{y}_t''$,通过点云重构模块得到当前帧的重构$\boldsymbol{X}_t'$。本文提出的所有利用卷积提取特征的操作都基于稀疏卷积实现,它在进行卷积操作时只考虑输入数据中非空体素的位置,忽略空体素的计算,从而减少了计算量和存储空间消耗。

1.2 体素下采样模块

本文体素下采样模块由两个连续的下采样块组成,其主要作用是减少空间冗余并进行特征分析。具体结构如图2(a)所示,所提出的方法采用步长为

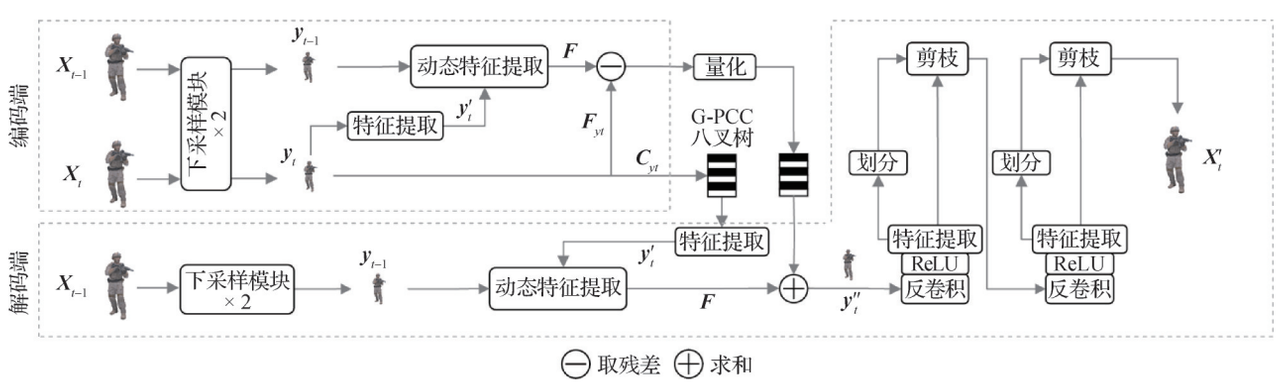


图1 整体框架

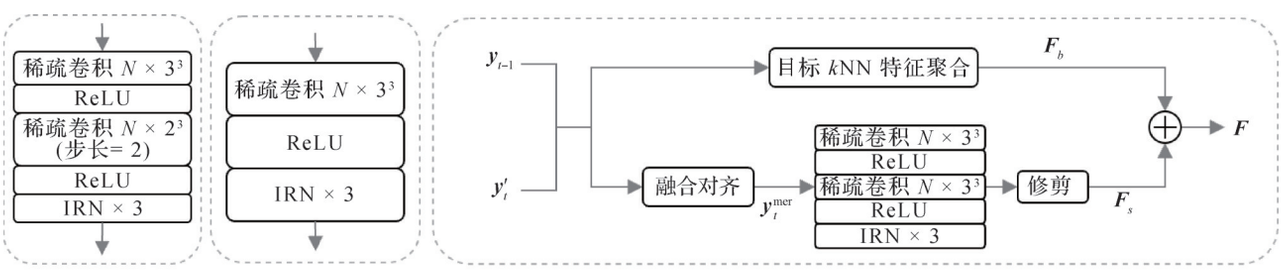
Fig. 1 Overall framework

1和2的两层稀疏卷积以及3个堆叠的初始残差网络(inception residual network，IRN)构成下采样块，当前帧$\boldsymbol{X}_t$和先前重建的帧$\boldsymbol{X}_{t-1}$经过由这样的两个下采样块组成的体素下采样模块得到$\boldsymbol{y}_t$和$\boldsymbol{y}_{t-1}$。其中，IRN的具体结构如图3(b)所示。它除了堆叠了稀疏卷积层组成的两个平行分支外还应用了残差链路，实现了高效的特征提取。

1.3 动态特征提取模块

动态特征提取模块由两部分前后帧运动信息提取模块共同构成，其主要作用是提取参考帧特征信息到当前帧点云上。具体地，所提方法采用特征提取模块得到的$\boldsymbol{y}_t=\{\boldsymbol{C}_{y_t},\boldsymbol{F}_{y_t}\}$和$\boldsymbol{y}_{t-1}=\{\boldsymbol{C}_{y_{t-1}},\boldsymbol{F}_{y_{t-1}}\}$作为输入，分为两个部分来分析时间空间相关性。具体结构如图2(c)所示。

上半部分即基于kNN的目标特征聚合模块如图3(a)所示，首先输入当前帧经过下采样处理后的点云坐标$\boldsymbol{C}_{y_t}$以及对应的特征$\boldsymbol{F}_{y_t}$，还有前一帧参考点云体素下采样获得的坐标$\boldsymbol{C}_{y_{t-1}}$以及对应的多尺度特征$\boldsymbol{F}_{y_{t-1}}$。对于当前帧y_t中的每个点坐标$\boldsymbol{C}_{y_t}$，所提方法在参考帧$\boldsymbol{y}_{t-1}$的点云坐标$\boldsymbol{C}_{y_{t-1}}$中动态地寻找其最近的k个邻居点，并进一步计算两者之间的局部注意力权重。具体地，邻居点与中心点关系越密切，其权重就越大；反之，权重就越小。通过这种邻域构建方式，所提方法能够更好地捕捉局部的运动信息。最后，将参考帧中这k个邻居点的特征根据计算得到的注意力权重进行加权融合，得到当前帧点云对应的融合特征$\boldsymbol{F}_b$，此处的k设置为32。这种融合方式能够有效地将参考帧的运动信息传递到当前帧，增强了帧间的相关性。相比现有的Akhtar和Fan提出的方法，所提出的基于kNN的目标特征聚合模块能够进一步提取到运动幅度较大区域以及分布较稀疏点云的局部特征。

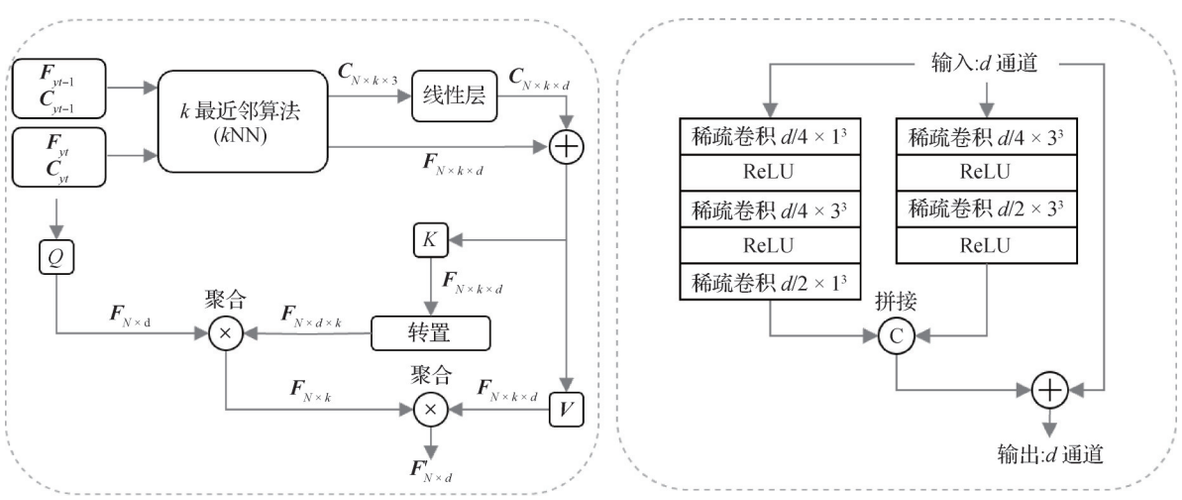


(a) 下采样块　(b) 特征提取模块　(c) 动态特征提取模块

图2　关键模块的网络结构

Fig. 2　Network architecture of key modules

((a) downsampling block; (b) feature extraction block; (c) dynamic feature extraction block)



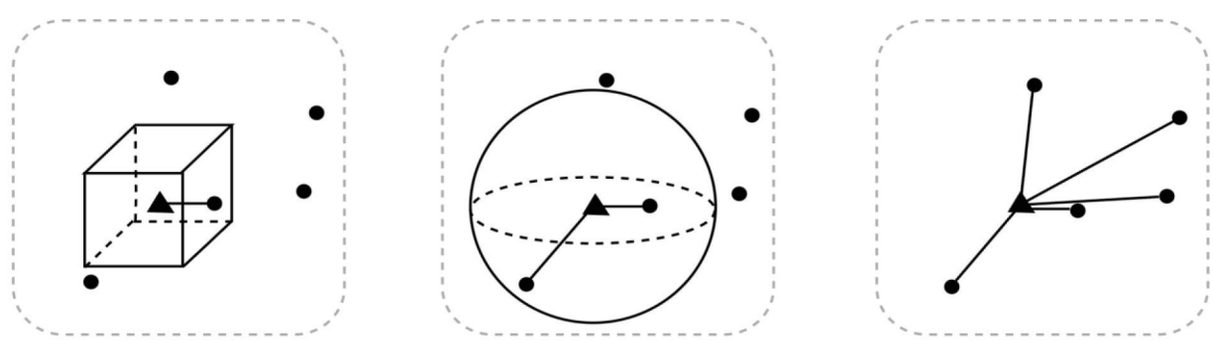
(a) 基于kNN的目标特征聚合模块　(b) IRN特征提取结构框架

图3　kNN和IRN的网络结构

Fig. 3　Network architecture of kNN and IRN

((a) target feature aggregation block based on kNN; (b) IRN feature extraction structural framework)

图4可视化了所提方法与Akhtar等人(2024)和Fan等人(2022)提出的基于卷积的方法以及Xia等人(2023)提出的基于球面kNN的方法的区别,以8iVFB数据集中Soldier序列两个连续帧点云的特定位置为例,其中,参考帧用圆形的点表示,当前帧用三角形的点表示。如图4(a)(b)所示,Akhtar等人(2024)、Fan等人(2022)以及Xia等人(2023)的方法分别使用基于目标卷积、对齐前后帧后卷积、球面kNN算法等方式进行前后帧特征提取,且都以当前帧(三角形)点云坐标为中心,在固定大小的卷积感受野内或限定在球面区域内的邻域提取参考帧(圆形)点云的特征信息。这类方法在处理局部运动较大的区域时,会出现无法准确找到前后帧对应位置的问题,从而降低了特征聚合的效率。而所提出的基于kNN目标特征聚合模块如图4(c)所示,以当前帧(三角形)的每个点为中心,在参考帧(圆形)点云中找到其k个最近邻点。这种方式相比于固定感受野或球面区域的特征提取,能够更好地处理局部运动较大的情况,准确地将参考帧中的特征信息聚合到当前帧上。与之前的方法(图4(a)(b))当前帧(三角形)点云只能提取到1个和2个参考帧(圆形)点云的特征信息相比,所提方法能够更全面地捕捉前后帧之间的对应关系和运动信息,为后续的点云压缩编码提供更有价值的特征表示。以$k=5$为例,所提方法(图4(c))在当前帧(三角形)点云成功提取到5个参考帧(圆形)点云的点的特征信息。



(a) Akhtar等人(2024)和Fan等人(2022)　(b) Xia等人(2023)　(c) 本文

图4　目标特征聚合模块可视化

Fig. 4　Visualization rendering of target feature aggregation block

((a) Akhtar et al. (2024)and Fan et al. (2022); (b) Xia et al. (2023); (c) ours)

动态特征提取模块的下半部分设计是由于大部分位置的点云都在较小范围内运动,因此所提方法在目标邻域特征提取算法的基础上设计了基于卷积的前后帧特征提取模块,通过对齐点云的方法进一步提取到更多的帧间特征信息,以适应不同的运动情况。具体地,所提出的方法将前后帧的下采样$\boldsymbol{y}_{t-1}$和$\boldsymbol{y}_t$合并到一起得到$\boldsymbol{y}_t^{\mathrm{mer}}$,具体对应到每一个点的合并操作定义为

$$\boldsymbol{y}_u^{\mathrm{mer}}=\begin{cases}\boldsymbol{y}_{t-1,u}\oplus\boldsymbol{y}_{t,u} & u\in\boldsymbol{C}_{t-1}\cap C_t\\ \boldsymbol{y}_{t-1,u}\oplus\boldsymbol{0} & u\in\boldsymbol{C}_{t-1}\wedge u\notin\boldsymbol{C}_t\\ \boldsymbol{0}\oplus\boldsymbol{y}_{t,u} & u\notin\boldsymbol{C}_{t-1}\wedge u\in\boldsymbol{C}_t\end{cases}\tag{1}$$

式中,$\boldsymbol{y}_u^{\mathrm{mer}}$是$\boldsymbol{y}_{t-1,u}$和$\boldsymbol{y}_{t,u}$的合并。$\boldsymbol{y}_{t-1,u}$和$\boldsymbol{y}_{t,u}$是在点云任意位置$u$处定义的输入前后帧特征向量,$\oplus$是前后帧之间的合并操作。所有合并得到的$\boldsymbol{y}_u^{\mathrm{mer}}$共同构成$\boldsymbol{y}_t^{\mathrm{mer}}$,再通过两层卷积网络和堆叠的初始残差网络将参考帧的特征信息聚合到当前帧点云上,并通过修剪操作保留与当前帧坐标信息一致点云的特征作为$\boldsymbol{F}_s$。

最后,所提方法将两部分前后帧运动信息提取模块提取得到的特征信息$\boldsymbol{F}_b$和$\boldsymbol{F}_s$相加得到动态特征信息$\boldsymbol{F}$,同时覆盖了不同大小幅度的运动,实现自适应帧间特征提取的目的。

1.4　残差压缩模块

在动态点云压缩方案中,由于帧间点云关系密切,具有局部几何相似性,因而采用残差编码的方式传输特征来进一步节省比特流消耗。具体地,当前帧点云两次体素下采样得到的特征信息$\boldsymbol{F}_{y_1}$减去动态特征提取模块提取到的动态特征信息$\boldsymbol{F}$得到对应的残差。编码端特征提取完成后使用G-PCC无损压缩八叉树结构(Graziosi等,2020)的下采样坐标信息

$\boldsymbol{C}_{y_t}$，同时使用基于熵模型的算术编码器压缩量化后的残差。

1.5 点云重建模块

在解码端，得到量化后的残差信息以及坐标$\boldsymbol{C}_{y_t}$，与编码端相同，分别使用参数共享的下采样和特征提取模块得到$\boldsymbol{y}_{t-1}$和$\boldsymbol{y}_t'$，然后一起送入动态特征提取模块得到$\boldsymbol{F}$，将$\boldsymbol{F}$与残差信息相加得到的特征$\boldsymbol{F}_{y_t}'$用于点云重建。随后通过两个连续的上采样块进行分层重构，上采样块使用步长和卷积核均为2的稀疏转置卷积进行点云上采样，随后使用特征提取模块生成每个体素的占用概率，最后根据占用概率通过自适应剪枝的方式（Wang等，2021a）修剪假体素。

1.6 损失函数

点云有损压缩以提高压缩率、提升点云重建的质量为目的，所以在训练时使用比特率—失真（rate-distortion，R-D）权衡损失函数进行端到端优化。该损失函数由二进制交叉熵（binary cross entropy，BCE）和压缩比特率组成，R-D总损失计算为

$$J_{\text{loss}} = R + \lambda D \tag{2}$$

式中，J_{loss}由R和D两部分组成，R为点云压缩比特率，D为衡量预测值与真实值之间差异的损失函数L_{BCE}，参数λ用来控制码率，训练得到不同码率的多个模型。编码前的特征量化过程通过添加均匀噪声$\mu \sim u(-0.5, 0.5)$来近似。点云压缩比特率R计算为

$$R_{\hat{\boldsymbol{F}}_y} = \frac{1}{N}\sum_{i}^{N} -\log_2(p_{\hat{\boldsymbol{F}}_{yi}}) \tag{3}$$

式中，$\hat{\boldsymbol{F}}_y$为编码前点云特征图，$p_{\hat{\boldsymbol{F}}_{yi}}$为每个体素对应的预测概率，$R_{\hat{\boldsymbol{F}}_y}$利用特征图计算码率的大小。

L_{BCE}损失函数计算为

$$L_{\text{BCE}} = \frac{1}{N}\sum_{i} -(x_i \log(p_i) + (1-x)\log(1-p_i)) \tag{4}$$

式中，x_i表示当前体素的真实占用情况（1表示占用，0表示未占用），p_i为网络预测的体素占用概率，L_{BCE}的值越小，说明每个体素预测的概率p_i越接近真实占用情况x_i，从而不断优化L_{BCE}，得到更高质量的重建点云。分层重建可表示为

$$D = \frac{1}{N}\sum_{1}^{K} L_{\text{BCE}}^{(k)} \tag{5}$$

式（5）对每个尺度的失真进行平均，k是尺度指数。

2 实 验

2.1 实验设置

所有实验使用的数据集均依照MPEG点云压缩委员会推荐的通用测试条件（common test conditions，CTC）（Alexandre等，2022）。在训练集的选取上，所提出的方法采用10 bit精度8iVFB数据集（d'Eon等，2017），包括Longdress、Loot、Redandblack和Soldier 4个序列，每个序列包括300帧动态点云。具体地，每帧分割为100 000～200 000个点组成的小块，最终得到7 432个小块作为所提出的动态点云编解码模型的训练数据集。在测试集的选取上，为了在精度上保持和训练集的一致，所提方法采用量化为10 bit精度的Owlii动态序列点云（Yi等，2017），包括Basketball Player、Dancer、Exercise和Model 4个序列。将每个序列的前100帧，共计400帧用做测试。在每一个动态点云序列的压缩过程中，第1帧使用PCGCv2（Wang等，2021a）进行压缩，随后的99帧使用本文所提出的动态点云压缩方法进行压缩，k设置为32。

本文所用实验平台为Intel Core i7-8700K CPU，32 GB内存，NVIDIA GeForce RTX 3090GPU的计算机，并借助PyTorch和MinkowskiEninge（Choy等，2019）实现模型训练。网络模型优化器选择Adam，参数β_1和β_2分别设置为0.9和0.999，学习率每15轮训练衰减0.7，Batch-size为4。对每个码率点采用二阶段训练策略，对于前5个epoch，将λ设为20，以加速点云重建模块的收敛；在5个epoch后，将λ设置为原始值，对模型进行另外95次epoch的训练。

在模型评估上，所提方法使用bpp（即每点比特）来衡量每个点所需的比特数量。PSNR D1（点对点峰值信噪比）和PSNR D2（点对平面峰值信噪比）基于图像处理中的峰值信噪比指标进行扩展来评估点云质量，PSNR值越大，失真越小。遵循MPEG CTC（Alexandre等，2022），所提出的测试策略在用10比特精度的Owlii数据集进行测试时，设置点云分辨率参数为1 023。

2.2 实验对比

本文方法与MPEG提出的动态点云标准压缩方法GeS（geometry-based solid content test model）（inter）（Lasserre和Taquet，2023）以及V-PCC（video-

based point cloud compression)(Schwarz等,2019)进行了对比。并且选取了3种最先进的基于学习的动态点云压缩方法：Akhtar等人(2024)、Fan等人(2022)以及Xia等人(2023)提出的方法,使用上述3种方法在MPEG公开提案中的结果进行对比,其中,Akhtar等人(2024)方法只给出了PSNR D1上的结果,因此只对比PSNR D1。如表1所示,所提出的动态点云压缩方法相比于传统或基于学习的方法具有明显优势,具体地,相比两种MPEG提出的编码方法GeS(inter)和V-PCC,所提方法大幅领先,在PSNR D1(D2)上平均分别有94.14%(96.67%)和82.93%(72.57%)的显著BD-Rate增益;相比目前性能较佳的Fan等人(2022)和Xia等人(2023)方法,在PSNR D1(D2)上也平均有22.54%(26.72%)和14.21%(40.80%)的BD-Rate增益;同时,所提方法也优于Akhtar等人(2024)的方法,在PSNR D1上平均有34.80%的BD-Rate增益。

此外,不同方法在多个测试序列上的R-D曲线如图5所示,其中,红色曲线代表所提出的方法,蓝色、绿色、棕色曲线分别为Fan等人(2022)、Akhtar等人(2024)、Xia等人(2023)的方法,粉色和黄色曲线分别代表V-PCC和GeS(inter)。显然,在规范的MPEG CTC测试条件上,所提出的方法在较宽的比特率范围内优于对比方法,以上实验结果进一步证实了所提出的方法在动态点云压缩领域的可行性和有效性。

表1　对比不同动态压缩方法在客观指标PSNR D1和PSNR D2上的平均BD-Rate增益

Table 1　The average BD rate gain of different dynamic compression methods on objective indicators PSNR D1 and PSNR D2

/%

对比方法	评价指标	Basketball vox10	Dancer vox10	Exercise vox10	Model vox10	平均
GeS(Lasserre和Taquet,2023)	PSNR D1	−94.29	−94.04	−94.72	−93.50	−94.14
V-PCC(Schwarz等,2019)		−81.94	−83.57	−81.55	−84.66	−82.93
Fan等人(2022)		−26.25	−26.74	−14.66	−22.49	−22.54
Akhtar等人(2024)		−32.71	−37.61	−32.08	−36.78	−34.80
Xia等人(2023)		−16.33	−18.83	−9.13	−12.55	−14.21
GeS(Lasserre和Taquet,2023)	PSNR D2	−97.36	−96.53	−97.86	−94.91	−96.67
V-PCC(Schwarz等,2019)		−72.81	−73.41	−72.63	−71.43	−72.57
Fan等人(2022)		−30.93	−31.54	−18.85	−25.57	−26.72
Akhtar等人(2024)		–	–	–	–	–
Xia等人(2023)		−42.61	−43.80	−40.21	−36.58	−40.80

注:“–”表示对应文献中未对该项指标提供相关信息。

2.3　消融实验

为了证明所设计模块的有效性,本文进行了多组消融实验,在Owlii测试数据集计算所提方法对比去掉基于*k*NN的目标特征聚合模块和去掉基于卷积的前后帧特征提取模块的BD-Rate增益,具体对比结果如表2所示。其中,w/o TKFA表示禁用基于*k*NN的目标特征聚合模块,w/o concat表示禁用基于卷积的前后帧特征提取模块。结果显示所提方法对比去掉基于*k*NN的目标特征聚合模块在PSNR D1和D2上平均有10.95%和11.63%的BD-Rate增益,所提出的方法对比去掉基于卷积的前后帧特征提取模块在PSNR D1和D2上有5.03%和5.52%的BD-Rate增益。这表明,基于*k*NN的目标特征聚合模块和基于卷积的前后帧特征提取模块都可以在一定程度上提高动态点云帧间预测效率。

2.4　复杂度比较

为了证明所设计模块的低复杂度,本研究比较了上述方法的编解码时间,在相同的实验平台Intel Core i7-8700K CPU,32 GB内存,NVIDIA GeForce RTX 3090GPU的计算机上测试量化为10比特精度的8iVFB数据集400帧,并求平均每帧运行时间,具体对比结果如表3所示。所提出的动态点云压缩方

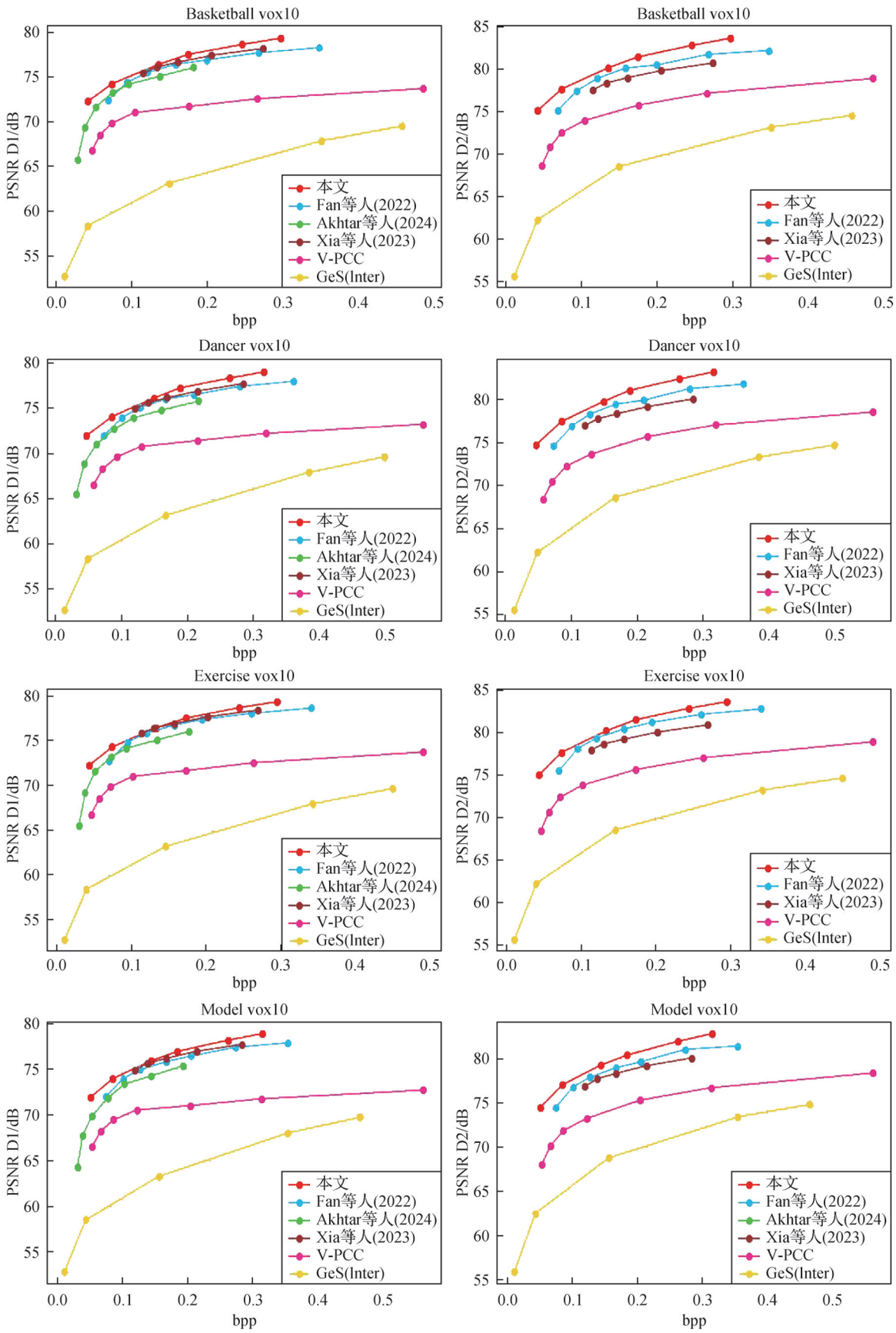


图 5　对比不同方法在多个数据集上的 R-D 曲线图

Fig. 5　Comparing R-D curves of different methods on multiple datasets

表2　结构消融实验结果对比

Table 2　Comparison of structural ablation experiment results

/%

结构	评价指标	Basketball vox10	Dancer vox10	Exercise vox10	Model vox10	平均
w/o TKFA	PSNR D1	−11.67	−7.03	−13.37	−11.73	−10.95
w/o concat		−5.73	−2.14	−7.60	−4.66	−5.03
w/o TKFA	PSNR D2	−12.29	−7.89	−14.15	−12.18	−11.63
w/o concat		−6.12	−2.32	−8.23	−5.39	−5.52

法运行时间远小于V-PCC，这是由于V-PCC在执行三维到二维的投影操作复杂度高；对比Fan节省0.2 s/帧，仅略长于Akhtar和Xia的方法；对比MPEG提出的编码方法GeS(inter)仅略微延长了运行时间，这与重建效率上的大幅提升相比是可以接受的。此外，对比不同基于学习的方法的模型大小，结果如表4所示。所提方法模型大小为2.5 MB，这与目前可获得的同样基于深度学习的方法的模型大小处于同一水平，但所提方法在重建效率上取得了明显的提升。

表3　不同动态压缩方法在测试序列的平均编解码时间

Table 3　Average encoding and decoding time of different dynamic compression methods on test sequences

/(s/帧)

方法	编码	解码	总计
V-PCC(Schwarz等，2019)	81.01	1.25	82.26
GeS(Lasserre和Taquet，2023)	0.77	0.11	0.88
Fan等人(2022)	0.96	1.04	2.00
Xia等人(2023)	0.87	0.87	1.74
Akhtar等人(2024)	0.47	0.68	1.15
本文	0.79	1.01	1.80

表4　不同基于学习的方法的模型大小

Table 4　Model sizes for different learning based methods

方法	模型大小/MB
Fan等人(2022)	2.83
Akhtar等人(2024)	2.03
本文	2.50

3　结　论

本文提出一种基于深度学习的动态点云压缩方法，旨在实现高效的动态点云压缩并提升点云重建质量。所提方法结合了对齐点云后的卷积特征提取算法和目标邻域特征提取算法，自适应提取到不同运动幅度点云的帧间特征信息，从而提高特征提取效率。所提方法以G-PCC无损编码器压缩下采样后的坐标信息，以熵编码器压缩量化后的残差信息，并通过参数共享帮助解码端更准确地预测当前帧点云体素的正占用概率，从而提高点云重构效率。通过在MPEG点云压缩委员会推荐的数据集上进行实验验证，证明了所提方法具有较高的压缩效率和较低的复杂度，对于实际应用具有良好的指导意义。此外，所提方法可以覆盖绝大多数的动态点云数据集。然而，理论上，在点云运动幅度过大时，所提方法可能会出现kNN算法对应关系寻找出错的情况。因此，未来的一个研究方向是针对稀疏程度更显著、运动幅度不可控的LiDAR点云进行动态压缩。

参考文献(References)

Akhtar A, Li Z and van der Auwera G. 2024. Inter-frame compression for dynamic point cloud geometry coding. IEEE Transactions on Image Processing, 33: 584-594 [DOI: 10.1109/TIP.2023.3343096]

Alexandre Z, Graziosi D and Ali T. 2022. Dataset split for AI-PCC. m58754. ISO/IEC JTC 1/SC 29/WG 7.

Choy C, Gwak J and Savarese S. 2019. 4D spatio-temporal ConvNets: Minkowski convolutional neural networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Bench, USA: IEEE: 3070-3079 [DOI: 10.1109/CVPR.2019.00319]

de Queiroz R L and Chou P A. 2017. Motion-compensated compression of dynamic voxelized point clouds. IEEE Transactions on Image Processing, 26(8): 3886-3895 [DOI: 10.1109/TIP.2017.2707807]

d'Eon E, Harrison B, Myers T and Chou P A. 2017. 8i voxelized full bodies — a voxelized point cloud dataset. ISO/IEC JTC1/SC29

Joint WG11/WG1 (MPEG/JPEG) Input Document WG11M40059/WG1M74006, 7(8): #11

Fan T Y, Gao L Y, Xu Y L, Li Z and Wang D. 2022. D-DPCC: deep dynamic point cloud compression via 3D motion prediction [EB/OL]. [2024-09-10]. https://arxiv.org/pdf/2205.01135.pdf

Fu C Y, Li G, Song R, Gao W and Liu S. 2022. OctAttention: octree-based large-scale contexts model for point cloud compression//Proceedings of 2022 AAAI Conference on Artificial Intelligence. [s.l.]: AAAI Press: 625-633 [DOI: 10.1609/aaai.v36i1.19942]

Gao L Y, Fan T Y, Wan J Q, Xu Y L, Sun J and Ma Z. 2021. Point cloud geometry compression via neural graph sampling//Proceedings of 2021 IEEE International Conference on Image Processing. Anchorage, USA: IEEE: 3373-3377 [DOI: 10.1109/ICIP42928.2021.9506631]

Gong J Y, Lou Y J, Liu F Q, Zhang Z W, Chen H M, Zhang Z Z, Tan X, Xie Y and Ma L Z. 2023. Scene point cloud understanding and reconstruction technologies in 3D space. Journal of Image and Graphics, 28(6): 1741-1766 (龚靖渝, 楼雨京, 柳奉奇, 张志伟, 陈豪明, 张志忠, 谭鑫, 谢源, 马利庄. 2023. 三维场景点云理解与重建技术. 中国图象图形学报, 28(6): 1741-1766) [DOI: 10.11834/jig.230004]

Graziosi D, Nakagami O, Kuma S, Zaghetto A, Suzuki T and Tabatabai A. 2020. 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, 9(1): #e13 [DOI: 10.1017/ATSIP.2020.12]

Guarda A F R, Rodrigues N M M and Pereira F. 2021. Adaptive deep learning-based point cloud geometry coding. IEEE Journal of Selected Topics in Signal Processing, 15(2): 415-430 [DOI: 10.1109/JSTSP.2020.3047520]

Huang L L, Wang S L, Wong K, Liu J and Urtasun R. 2020. OctSqueeze: octree-structured entropy model for LiDAR compression//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 1310-1320 [DOI: 10.1109/CVPR42600.2020.00139]

Huang T X and Liu Y. 2019. 3D point cloud geometry compression on deep learning//Proceedings of the 27th ACM International Conference on Multimedia. Nice, France: ACM: 890-898 [DOI: 10.1145/3343031.3351061]

Jiang Z Y, Zou W Q, Zheng S H, Song C and Yang B L. 2024. Variable rate compression of point cloud based on scene flow. Journal of Zhejiang University (Engineering Science), 58(2): 279-287, 333 (江照意, 邹文钦, 郑晟豪, 宋超, 杨柏林. 2024. 基于场景流的可变速率动态点云压缩. 浙江大学学报(工学版), 58(2): 279-287, 333) [DOI: 10.3785/j.issn.1008-973X.2024.02.006]

Lasserre S and Taquet J. 2023. Report for EE 13.60 on dynamic dense coding with G-PCC. m61562. ISO/IEC JTC 1/SC 29/WG

Liu J Y, Wang J, Sun L H, Pei J and Zhu Q. 2022. Cluster-based point cloud attribute compression using inter prediction and graph Fourier transform//Proceedings of the 14th International Conference on Digital Image Processing. Wuhan, China: SPIE: 943-949 [DOI: 10.1117/12.2644218]

Long X X, Cheng X J, Zhu H, Zhang P J, Liu H M, Li J, Zheng L T, Hu Q Y, Liu H, Cao X, Yang R G, Wu Y H, Zhang G F, Liu Y B, Xu K, Guo Y L and Chen B Q. 2021. Recent progress in 3D vision. Journal of Image and Graphics, 26(6): 1389-1428 (龙霄潇, 程新景, 朱昊, 张朋举, 刘浩敏, 李俊, 郑林涛, 胡庆拥, 刘浩, 曹汛, 杨睿刚, 吴毅红, 章国锋, 刘烨斌, 徐凯, 郭裕兰, 陈宝权. 2021. 三维视觉前沿进展. 中国图象图形学报, 26(6): 1389-1428) [DOI: 10.11834/jig.210043]

Pang J H, Lodhi M A and Tian D. 2022. GRASP-Net: geometric residual analysis and synthesis for point cloud compression//Proceedings of the 1st International Workshop on Advances in Point Cloud Compression, Processing and Analysis. Lisboa, Portugal: Association for Computing Machinery: 11-19 [DOI: 10.1145/3552457.3555727]

Que Z Z, Lu G and Xu D. 2021. VoxelContext-Net: an octree based framework for point cloud compression//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 6038-6047 [DOI: 10.1109/CVPR46437.2021.00598]

Santos C, Gonçalves M, Corrêa G and Porto M. 2021. Block-based inter-frame prediction for dynamic point cloud compression//Proceedings of 2021 IEEE International Conference on Image Processing (ICIP). Anchorage, USA: IEEE: 3388-3392 [DOI: 10.1109/ICIP42928.2021.9506355]

Schwarz S, Preda M, Baroncini V, Budagavi M, Cesar P, Chou P A, Cohen R A, Krivokuća M, Lasserre S, Li Z, Llach J, Mammou K, Mekuria R, Nakagami O, Siahaan E, Tabatabai A, Tourapis A M and Zakharchenko V. 2019. Emerging MPEG standards for point cloud compression. IEEE Journal on Emerging and Selected Topics in Circuits and Systems, 9(1): 133-148 [DOI: 10.1109/JETCAS.2018.2885981]

Sullivan G J, Ohm J R, Han W J and Wiegand T. 2012. Overview of the high efficiency video coding (HEVC) standard. IEEE Transactions on Circuits and Systems for Video Technology, 22(12): 1649-1668 [DOI: 10.1109/TCSVT.2012.2221191]

Wang J Q, Ding D D, Li Z and Ma Z. 2021a. Multiscale point cloud geometry compression//Proceedings of 2021 Data Compression Conference. Snowbird, USA: IEEE: 73-82 [DOI: 10.1109/DCC50243.2021.00015]

Wang J Q, Zhu H, Liu H J and Ma Z. 2021b. Lossy point cloud geometry compression via end-to-end learning. IEEE Transactions on Circuits and Systems for Video Technology, 31(12): 4909-4923 [DOI: 10.1109/TCSVT.2021.3051377]

Wang Z C, Wan S, Wei L and Yang F Z. 2023. Lossless geometry coding method for dynamic point clouds under octree structure. Journal of Xi'an Jiaotong University, 57(12): 11-19 (王哲诚, 万帅,

魏磊, 杨付正. 2023. 八叉树结构下动态点云的无损几何编码方法. 西安交通大学学报, 57(12): 11-19)[DOI: 10.7652/xjtuxb202312002]

Xia S T, Fan T Y, Xu Y L, Hwang J N and Li Z. 2023. Learning dynamic point cloud compression via hierarchical inter-frame block matching//Proceedings of the 31st ACM International Conference on Multimedia. Ottawa, Canada: ACM: 7993-8003 [DOI: 10.1145/3581783.3613793]

Yi X, Yao L and Wen Z Y. 2017. Owlii dynamic human mesh sequence dataset//ISO/IEC JTC1/SC29/WG11 m41658. Macau, China: MPEG/JPEG

Zhu W, Zhang Y H, Ying Y, Zheng Y Y and He D F. 2023. A dense residual structure and multi-scale pruning-relevant point cloud compression network. Journal of Image and Graphics, 28(7): 2105-2119 (朱威, 张雨航, 应悦, 郑雅羽, 何德峰. 2023. 结合密集残差结构和多尺度剪枝的点云压缩网络. 中国图象图形学报, 28(7): 2105-2119)[DOI: 10.11834/jig.220047]

Zhu W J, Ma Z, Xu Y L, Li L and Li Z. 2021. View-dependent dynamic point cloud compression. IEEE Transactions on Circuits and Systems for Video Technology, 31(2): 765-781 [DOI: 10.1109/TCSVT.2020.2985911]

作者简介

钱虞杰,男,硕士研究生,主要研究方向为点云压缩。
E-mail: 1160206543@qq.com
丁丹丹,通信作者,女,副教授,主要研究方向为视频、点云编码与压缩、多媒体通信。E-mail:dandanding@hznu.edu.cn