

基于熵模型的激光雷达点云帧间编码方法

石元龙¹ 郑明魁¹ 丁志洋¹ 宋广胜¹

SHI Yuanlong ZHENG Mingkui DING Zhiyang SONG Guangsheng

摘要

存储和传输激光雷达点云数据对于许多自动驾驶应用来说是必不可少的。由于激光雷达点云数据的稀疏性和无序性,很难将激光雷达点云数据压缩到较小的体积。因此,文章提出了一种基于熵模型的激光雷达点云帧间编码方法。为应对激光雷达点云序列的时间冗余问题,利用参考点云与待编码点云的位姿信息,提出一种有效消除点云序列中时域冗余的帧间编码方法。为去除点云的空间冗余问题,将原始点云数据转换成适合小波变换的密集二维矩阵数据,通过小波变换能够有效地利用二维矩阵的空间相关性。通过CDF5/3小波变换对二维矩阵进行小波变换得到小波系数,通过对熵模型训练后的熵参数进行算术编码从而得到更加紧凑的比特流。实验结果表明,提出的设计方法与G-PCC、PCL编码方法相比具有较高的编码性能。

关键词

熵模型;点云压缩;小波变换;位姿信息获取

doi: 10.3969/j.issn.1672-9528.2025.04.041

0 引言

随着自动驾驶技术的发展,3D激光雷达已经展现出其对真实世界环境感知的能力以及有效性^[1]。激光雷达即使在环境恶劣的光线条件下也能够提供准确的距离测量,从而探知物体的存在。同时,其应用涵盖了各种重要的自动驾驶任务,如SLAM、物体检测、点云语义分割和车道线检测等。共享和存储激光雷达点云数据对于许多应用都是十分必要的,例如训练数据集、协同驾驶、云服务和远程控制^[2]。在自动驾驶场景中得到的点云通常需要在V2X(车对一切)/V2V(车对车)网络中进行传输,由于带宽和存储容量是有限的,这将对现有通信系统带来巨大的挑战。为了解决这一难题,除了点云简化、重采样,删除部分点的操作来减少数据量外,旨在减少点云空间和时间冗余对高效点云压缩对自动驾驶领域的应用具有重要意义。因此,点云压缩已成学术界和工业界的研究热点。

国家运动图像专家组(MPEG)在八叉树的基础上发布了基于几何的点云压缩框架(G-PCC)^[3]。通过探索相邻和不同层级的八叉树子立方体的相关性以及八叉树空间分布情况,G-PCC还进一步引入了帧内预测并且和直接编码相结合去除孤立点。由于激光雷达点云呈圆形均匀向远处分布,

Sridhara等人^[4]选择在柱坐标系中进行八叉树编码,性能要优于传统基于笛卡尔坐标系的八叉树编码。Schnabel等人^[5]则基于八叉树空间分解,设计了一种特别适用于密集采样点云的无损压缩方法。通过预测哪些单元被占用,所有量化的叶中心坐标都能被有效地编码。

激光雷达技术在点云数据采集方面具有显著优势,但其采集到的点云数据通常稀疏且无序,缺乏显著的相关性。因此部分基于二维投影的点云编码方法通过将点云投影为距离图,再利用现有的图像和视频编码器进行压缩。Houshiar等人^[6]提出了一种混合距离图像编码算法,该算法使用先前编码的相邻像素值来预测像素,从而提高了距离图像的编码效率。Sun等人^[7]对从点云转换的距离图像执行分割,为减少图像中的冗余,对分割的聚类应用预测压缩方法。然而,传统图像编码算法主要针对整数像素值进行设计,难以有效处理点云所具有的复杂高维特征。

针对以上问题,提出了一种基于熵模型的激光雷达点云帧间编码方法。将激光雷达数据转换为适合小波变换的二维数据格式,使用CDF5/3小波变换进行提升变换得到小波系数,并利用深度熵模型对量化后的小波系数训练获得一个精准的熵参数模型用于算术编码,从而有效地减少了连续点云序列的时域冗余。

1 基于熵模型的激光雷达点云帧间编码方法

1.1 整体编码框架描述

在获取激光雷达点云时,激光雷达的工作频率是在

1. 福州大学物理与信息工程学院 福建福州 350116
[基金项目] 中国福建光电信息科学与技术创新实验室(闽都创新实验室)资助项目(2021ZR151);福州市科技重大项目(2022-ZD-002)

10~30 Hz, 从而相邻的两帧点云之间的时间间隔较小, 因此在相邻的两帧点云之间会具有较大的重复部分, 从而存在时域的冗余。因此本文设计了基于熵模型的帧间编码方法, 如图 1 所示。给定一个激光雷达点云序列, 将点云序列划分成参考帧点云 (I 帧), 代编码帧点云 (P 帧)。

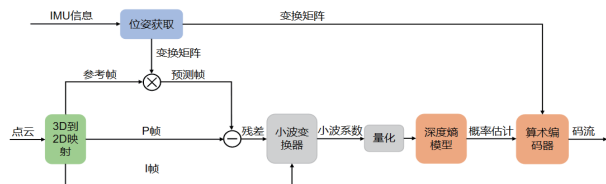


图 1 基于熵模型的激光雷达点云帧间编码框架

激光雷达点云由于其稀疏性的特点不适合基于八叉树的三维编码方式。因此, 本文将每一个三维激光雷达点云帧通过投影过程无损转换为二维矩阵格式。其次, 利用自动驾驶场景中使用的惯性测量单元采集到的位姿信息结合参考帧得到预测帧。再利用 CDF5/3 小波变换对残差数据和参考帧进行处理获取小波系数, 并对小波系数进行量化以控制比特率。由于小波系数数值分布集中, 且各子带之间存在明显的相关性, 利用神经网络获取的精准的熵模型, 最后, 使用算术编码器进行无损编码。

1.2 位姿信息获取

自动驾驶场景中车顶装有 3 种传感器, 分别是激光雷达扫描仪、Global Positioning System/ Inertial Measurement Unit (GPS/IMU) 和深度摄像机。惯性导航系统包含有惯性测量单元 (IMU), 能够检测车辆状态, 加速度等信息。为更加准确地预测 P 帧, 降低更多的时域冗余, 需要借助位姿信息进行点云预测。位姿信息就是计算相邻两帧点云之间最大重叠部分的旋转矩阵和平移矩阵, 从而得到这两帧点云之间的变换矩阵。

旋转矩阵和平移矩阵的计算是利用 IMU 信息中的平移加速度 a 和旋转速率 ω 计算得到。接着通过平移矩阵 T 和旋转矩阵 R , 得到变换矩阵 M , 其公式为:

$$M = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix}_{4 \times 4} \quad (1)$$

设定 i 是 I 帧点云中的任意一点, p' 是 P' 帧点云中的任意一点。 I 帧点云通过变换矩阵预测得到 P' 帧点云, 其计算公式为:

$$p'_{4 \times 1} = M \times i_{4 \times 1} = \begin{bmatrix} R_{3 \times 3} & T_{3 \times 1} \\ 0_{1 \times 3} & 1 \end{bmatrix}_{4 \times 4} \times i_{4 \times 1} \quad (2)$$

通过上述步骤实现了对 P 帧点云的预测。 I 帧点云是已编码的点云, 其与预测点云之间重合的部分不需要再进行编码, 只需编码原始点云 P 与预测点云 P' 之间的差异值即残差。

1.3 小波变换

变换编码是一种去除空间冗余的有效编码方式, 通过将空间域信号转换成频域信号, 将信号的能量进行重新分布, 从而实现信号在变换后的空间压缩编码^[8]。研究发现激光雷达的二维矩阵数据也非常适合使用小波变换来提升编码效率。因此, 选择 JPEG-2000 标准中的 CDF5/3 小波变换。CDF5/3 小波变换可以看成行和列两个方向的一维小波变换。对于一个待变换的激光雷达二维矩阵数据, 先对其行作小波变换, 行变换之后, 再对其进行列变换, 经此变换后会得到 1/4 大小的子系数: LL、LH、HL、HH。接着对子系数 LL 继续分解, 基本过程可以理解为将二维矩阵数据从高尺度向低尺度进行分解的一个过程, 图 2 为对二维矩阵进行三阶小波分解的示意图, 假定从尺度 $J+1$ 开始, 经三级小波变换逐层分别分解到尺度 J 、 $J-1$ 、 $J-2$ 。

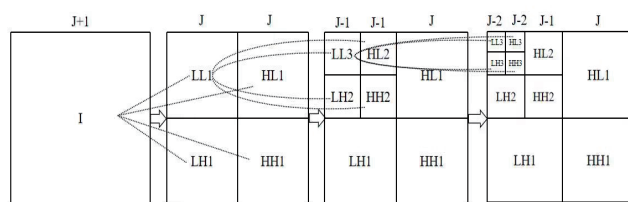


图 2 二维矩阵三阶小波分解的示意图

1.4 深度熵模型压缩

激光雷达二维矩阵数据在同一区域范围分布相对一致, 在经小波变换量化之后相邻数值间具有很强的相关性, 可以设计一个类似 pixelcnn^[9] 的深度熵模型, 在参数中捕获像素之间的依赖关系分布, 依次顺序生成一个像素。同时各层子系数之间也存在着相关性, 为利用这些相关性, 通过训练好的掩膜卷积神经网络为算术编码器建立一个精细有效的熵模型。

在推断时, 输入为待建模子系数 S_j 与前一层级子系数 L_{j-1} , 例如编码 HL_2 时, LL_3 、 HL_3 、 LH_3 、 HH_3 都为上下文信息用于融合。推断是从 LL 开始的, 首先将小波系数缩放至 $[0,1]$ 区间, 从小波系数的左上开始预测系数 $x(1,1)$ 。该过程以光栅扫描顺序继续, 以预测系数 $x_{1,2}, \dots, x_{x,w}, x_{2,1}, \dots, x_{H,W}$ 在处理一个中心像素的信息时, 卷积会考虑其周围的所有系数来计算输出特征图, 而不仅是前面的像素。掩膜分为 A 型掩膜与 B 型掩膜。其从中心开始的权重为 0, 这会阻止网络看到它正在预测的系数以及所有之后的系数, 这称为 A 型掩膜。而由于中心系数在第一卷积层即输入层中被遮挡, 只需要将卷积核的中心设置为 1, 使其能够读取先前卷积层的特征, 而这就是 B 型掩膜。掩膜卷积神经网络体系在整个层中保持输入的空间分辨率, 在每个位置输出概率分布, 从而得到熵参数模型。最后, 使用算术编码器进行无损编码。

2 实验结果

2.1 实验条件和评价指标

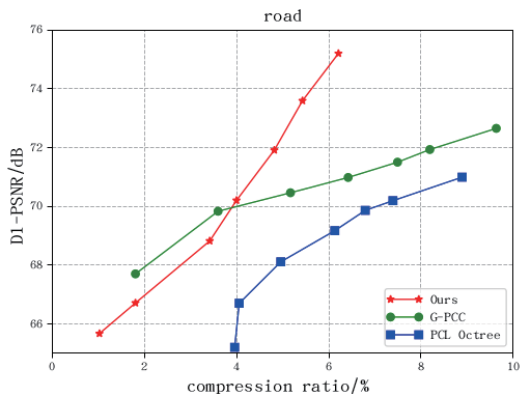
本文所有实验都是在 ubuntu18.04 系统, 电脑配置为英特尔酷睿 i7-7700K 4.20 GHz CPU, NVIDIA GeForce RTX 3090 GPUs, CUDA 版本 10.0 的设备上完成的。实验采用的点云数据集是 KITTI 数据集^[10]中校园、城市、居民区和道路 4 个场景, 每个场景选取 100 帧编码实验。而训练熵模型使用的数据则是随机来源于 KITTI 数据集中的 14 230 帧所生成的小波系数矩阵。实验使用 Adam 优化器对熵模型进行训练, 学习速率为 $1e-5$, 迭代次数为 150 个 epochs。MPEG 为几何压缩提供了两个评估指标, 即 dB 中的 D1-PSNR 和 D2-PSNR, 分别评估点对点误差和点对面误差。本文中只使用 D1-PSNR。因为 D2-PSNR 算法必须拥有地面法线而 KITTI 数据集是不存在法线属性的。

$$D1-PSNR = 10 \times \log_{10} \left(\frac{3P^2}{MSE(p2p)} \right) \quad (3)$$

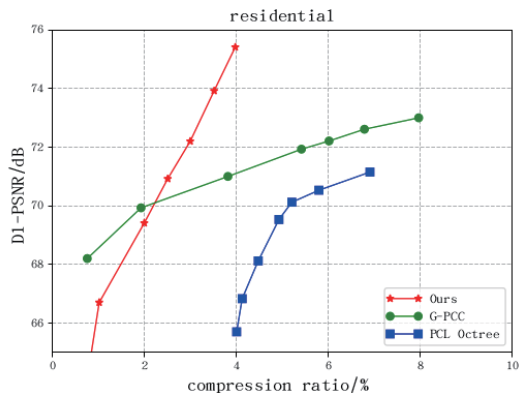
式中: P 为每个序列点云内部点与点之间距离的最大值; $MSE(p2p)$ 为重建点云到原始点云的均方误差。

2.2 本文算法与 G-PCC, PCL 的压缩性能对比

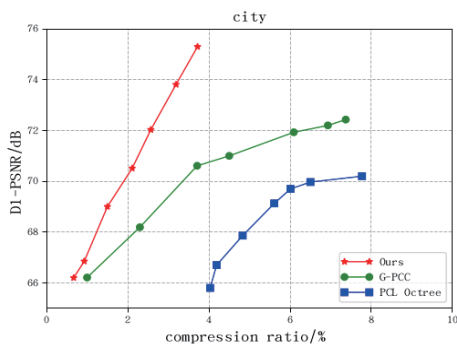
图 3 展示了在 KITTI 4 个不同场景的数据集上, 本文设计的基于熵模型的激光雷达点云帧间编码方法与 G-PCC 以及 PCL Octree 的 R-D 曲线结果对比。



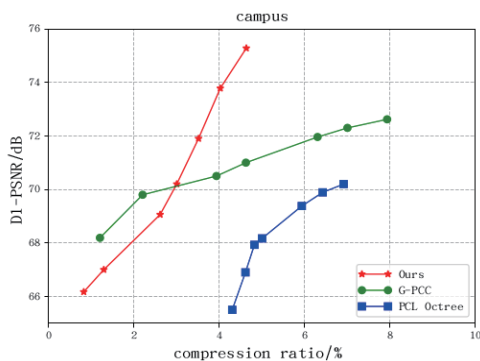
(a) 道路



(b) 居民区



(c) 城市



(d) 校园

图 3 KITTI 数据集的 R-D 曲线图

从图 3 看到, 与 MPEG G-PCC、PCL 基于八叉树的方法相比, 在压缩率大于 4% 的时候, 本文方法拥有最高的 D1-PSNR。同时本方案始终优于 PCL 基于八叉树的方法, 与 MPEG G-PCC 相比也有较大性能提升。

本文算法对城市这个场景编码效果相对较好。因为该场景周围建筑, 围墙等物体在映射成二维格式后特征明显, 有助于提升小波系数的相关性, 同时该场景包含的动态物体较少, 所以预测准确度更高, 降低了更多的时域冗余, 从而实现了整体更优的编码性能。所以本文方法可以在低损失的情况下实现显著高于 MPEG G-PCC 的压缩率, 有效地减少空间和时间冗余, 因此可以实现更好的 R-D 性能。

3 结语

在本文中提出了一种基于熵模型激光雷达点云帧间编码方法用以有效消除三维激光雷达点云的时空冗余, 受到帧间预测编码的启发, 本文结合位姿信息计算参考帧点云与待编码点云之间的变换矩阵, 利用参考帧点云的信息预测待编码点云。该方法在较大程度上消除了点云序列的时域冗余。同时本文将点云映射成二维矩阵, 利用小波变换生得到小波系数, 并通过熵模型神经网络生成熵模型, 有效地编码了 I 帧和预测残差。利用 KITTI 多个场景序列进行测试。实验结果表明, 与 MPEG G-PCC、PCL 相比本文提出的方法在高压压缩率时拥有更优的编码效果。

参考文献:

- [1] HECHT J. Lidar for self-driving cars[J]. Optics and photonics news,2018,29(1):26-33.
- [2] WOOPEN T, LAMPE B, BODDEKER T,et al.UNICARagil - disruptive modular architectures for agile, automated vehicle concepts[C]//Aachen Colloquium Automobile and Engine Technology 2018. Berlin:Springer,2018:663-694.
- [3] Point cloud compression category 13 reference software, TMC13-v14.0[EB/OL]. (2021-10-22)[2024-08-06].htx.int-evry.fr/software/MPEG/PCC/TM/mpeg-pcc-tmc13/tree/release-v14.0.
- [4] SRIDHARA S N, PAVEZ E, ORTEGA A. Cylindrical coordinates for lidar point cloud compression[C]//2021 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2021: 3083-3087.
- [5] SCHNABEL R, KLEIN R. Octree-based point-cloud compression[C]//Proceedings of the 3rd Eurographics/IEEE VGTC conference on Point-Based Graphics. NewYork:ACM,2006:111-121.
- [6] HOUSHIAR H, NÜCHTER A. 3D point cloud compression using conventional image compression for efficient data transmission[C]// 2015 XXV International Conference on Information, Communication andAutomation Technologies (ICAT).

Piscataway:IEEE,2015: 1-8.

- [7] SUN X B, MA H, SUN Y X, et al.A novel point cloud compression algorithm based on clustering[J].IEEE robotics and automation letters, 2019, 4(2): 2132-2139.
- [8] SANG L. Wavelet transform based video compression algorithm and its application in video transmission[J]. Computer modeling & new technologies, 2013, 17(5B): 60-63.
- [9] OORD A V D, KALCHBRENNER N, KAVUKCUOGLU K. Pixel recurrent neural networks[C]// Proceedings of the 33rd International Conference on International Conference on Machine Learning. NewYork: ACM, 2016:1747-1756.
- [10] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. International journal of robotics research, 2013, 32(11):1231-1237.

【作者简介】

石元龙（2000—），男，安徽芜湖人，硕士研究生，研究方向：点云编码。

郑明魁（1976—），男，福建福州人，博士，副教授，研究方向：点云编码与视频编码。

（收稿日期：2024-12-06）

（上接第 175 页）

- [20] GUAN Y N, LIAO S J, WAGG W Y. AParC-DETR: accelerate DETR training by introducing Adaptive position-aware circular convolution[J]. The visual computer, 2024, 41: 1319-1333.
- [21] DASGUPTA S S, RAY S N, TALUKDAR P. HYTE: hyperplane-based temporally aware knowledge graph embedding[C]//Proceedings of the 2018 conference on empirical methods in natural language processing. Brussels: ACL, 2018: 2001-2011.
- [22] JIA N N, YAO C Y. A brief survey on deep learning-based temporal knowledge graph completion[J]. Applied sciences, 2024, 14(19): 8871.
- [23] SUCHANEK F M, ALA, M, BONALD T, et al. Yago 4.5: a large and clean knowledge base with a rich taxonomy[C]// Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval. NewYork:ACM, 2024: 131-140.

- [24] NIAZMAND E, VIDAL M E. Discovering relationships among properties in wikidata knowledge graph[J]. Big data analytics and knowledge discovery, 2024(18): 388-394.
- [25] XU C J, NAYYERI M, ALKHOURY F, et al. Tero: a time-aware knowledge graph embedding via temporal rotation[C]// Proceedings of the 28th International Conference on Computational Linguistics.Brussels: ACL, 2020: 1583-1593.

【作者简介】

许运生（1992—），男，广东惠州人，硕士研究生，研究方向：语义 Web、深度学习。

陈美（2000—），女，广西陆川人，硕士研究生，研究方向：语义 Web、机器学习。

朱巧燕（1994—），女，广东惠州人，硕士研究生，研究方向：系统可靠性。

许珂（1994—），男，广东惠州人，硕士研究生，研究方向：系统可靠性。

（收稿日期：2024-12-07）