

LiDAR点云数据的分布式组织及其并行获取方法

王宗跃^{1,2} 马洪超¹ 徐宏根¹ 邬建伟¹

(1 武汉大学遥感信息工程学院,武汉市珞喻路129号,43007)
(2 集美大学计算机工程学院,厦门市集美区银江路185号,361021)

摘要:提出了一种分布式的大规模LiDAR点云并行快速数据读取方法。采用按空间位置均匀格网划分的数据组织形式,研究了在该组织方式下格网大小、数据分布策略对数据获取速度的影响。该方法在实验中表现出良好的并行效果,当有9台数据服务器并且待获取的数据范围较小时,加速比达到8以上。

关键词:LiDAR点云;快速获取;并行获取;平截头体
中图法分类号:P237.3

激光雷达LiDAR(light detection and ranging)是近年来国际上发展十分迅速的遥感技术,它可以实现空间三维坐标的同步、快速、精确获取,为快速获取空间信息提供了简单有效的手段^[1]。机载LiDAR系统与其他遥感技术相比较具有自动化程度高、受天气影响小、数据生产周期短、精度高等技术特点,是目前最先进的能实时获取地形表面三维空间信息和影像的航空遥感系统^[2]。当数据量较大时,快速地获取指定范围内的数据就成为一个迫切需要解决的问题。

为了解决该问题,许多国内外学者提出了对数据不同的组织方式,总体可以分为KD树和格网组织方式两大类。文献[3-5]采用LOD模型(level of detail)及基于点的绘制^[5]等数据模型,Carsten等人在Qsplat^[6]基础上提出了顺序点树模型,由粗到细构建顺序树状结构,并使用紧凑的文件布置方式,黄先锋等人借鉴了Qsplat中紧凑组织数据的做法,采用了多精度排序四叉树存储模型^[7]。

为了缩短这部分时间,鉴于双核CPU、多核CPU、网络速度达到千兆的这种硬件条件和采用并行技术在加速遥感图像处理^[8-10]、并行空间数据库^[11]方面已取得的丰富成果,本文提出了一种分布式数据组织及其并行数据获取的方法。

平截头体(frustum)表示一个头部被截取的四棱锥,代表一个可视区域或一个投影,在显示或数据处理时,它是最常用的指定范围方式之一,因

此以下的指定范围都是指平截头体。

1 LiDAR数据的并行获取方法设计

分布式组织的数据并行获取过程如下。

- 1) 用户给主机发送获取数据的指数。
- 2) 主机给各数据服务器转发指令。
- 3) 各数据服务器接收到指令后,开始按照自己的数据组织方式获取对应数据。
- 4) 各数据服务器获取完一块数据后就给用户发送数据,使网络传输和数据获取成为流水线方式运作,利用了通信与计算的时间重叠来提高运行效率。服务器的查询完成,就给主机发送获取数据完成的消息。
- 5) 当主机收到各数据服务器发回获取完成的消息后,转告用户获取完毕。

2 LiDAR数据的并行获取效率及时间复杂度分析

假定原始数据被分割为 $M \times N$ 的格网,每块平均点数为 nP 。

单计算节点时获取总时间 T_a 等于获取每一块所需时间之和:

$$T_a = \sum_{i=1}^{M \times N} T_i \quad (1)$$

式中, T_i 是获取一块数据的时间, 它包括判断外包盒与平截头体是包含、相离、相交的空间关系的时间 T_i^1 和在各种关系下需要处理的时间 T_{i2} , 所以 $T_i = T_{i1} + T_{i2}$ 。

$$T_{i2} = \begin{cases} 0, \text{相离} \\ T_i^1 + T_i^d, \text{平截头体包含外包长方体} \\ T_i^1 + T_i^d + T_i^e, \text{相交} \end{cases} \quad (2)$$

式中, T_i^1 表示读取该块数据的时间; T_i^d 表示解析该块中所有点云数据的时间; T_i^e 表示判断该块中每一点是否在平截头体中的总时间。它们对应的时间为:

$$T_i^d = \sum_j^{nP} T_q, T_i^e = \sum_j^{nP} T_p \quad (3)$$

1) T_q 为解析一点的时间。由于标准的 Las 文件格式中, 要获取点云的地理坐标需要经过伸缩和平移变换 $x = ax + b, y = cy + d, z = ez + f$, 所以解析一点的时间 T_q 为进行 3 次乘法和 3 次加法所需的时间。

2) T_p 为判断一个点是否在平截头体中的时间。要判断一个点 (x_0, y_0, z_0) 是否在平截头体中, 需要判断这个点是否都在平截头体 6 个面的前面。设平截头体其中的某个面的方程为 $ax + by + cz = 0$, 若 $ax_0 + by_0 + cz_0 + d \leq 0$, 则该点在该面的后面, 否则在该面的前面。每判断一个面需要进行 3 次乘法、3 次加法和一次判断, 平截头体有 6 面, 所以, T_p 为进行 18 次乘法、18 次加法、6 次判断所需的时间。

当块数据的外包盒与平截头体相离时, 只需判断拓扑关系, 所耗时间很短; 当包含或相交时, 所耗时间较长。因此要提高这部分的并行, 即应让更多的计算节点参与到相交和包含情况下的数据获取中。为了能够实现这一点, 本文提出了以下两个策略: 把格网划分为合适大小; 数据块交叉分配给数据服务器。

如图 1 所示, 灰色的区域表示要获取的指定范围。当数据分块大时如图 1(a) 所示, 只能由一个计算节点来读取该块, 并判断该块中的数据是否落在指定范围之内。而在图 1(b) 中可以由多台数据服务器同时获取数据。划分小块的另一好处是, 当所要获取的指定范围是跨越几块, 如图 2(a) 所示, 若大块的分割将要判断更多的无效区域, 而划分小一些时, 超出指定范围的无效区域将会变小。如图 2(b) 所示, 画斜线部分为无效区域, 灰色部分为待获取区域。

但是也并非分得越小越好, 分得越小, 求各块的外包矩形和平截头体的关系次数就越多, 因此本文为不同数据量和平均获取范围设置不同的分块块数。

设有 4 台数据服务器, 若数据块按顺序分配给各台数据服务器, 如图 3(a) 所示, 图中的数字表示计算节点的序号, 灰色的地方表示获取的指定范围, 只有 2 个计算节点能够参与计算; 若是采用交叉的数据分配方式, 即分块后的数据按照空间位置交叉分配给计算节点, 如图 3(b) 所示, 则有 4 个计算节点参与计算。

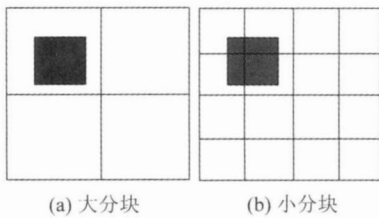


图 1 数据块划分大小
Fig. 1 Data Block Partition Size

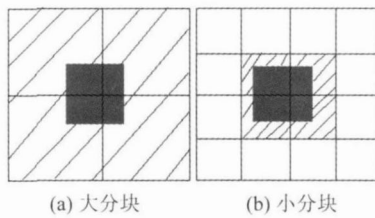


图 2 指定范围跨越多块数据
Fig. 2 Designated Area Spanning Several Data Blocks

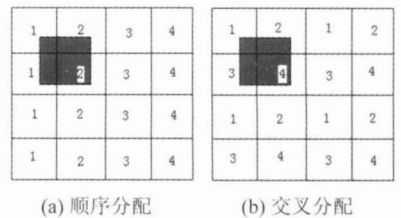


图 3 不同数据分布策略
Fig. 3 Different Data Distribution Strategies

3 应用实例及其分析

在局域网内实现一个测区内多条带的 LiDAR 数据的分布式组织及其对应的并行获取实验。本文使用的数据采集自宜昌长阳山区, 测区范围约为 $5 \text{ km} \times 10 \text{ km}$, 采集系统为 Leica ALS50 激光雷达系统, 飞行高度约为 1300 m , 视

场角 49° , 重叠率 30% ; 航向间隔约为 1.3 m , 旁向间隔约为 1 m , 密度约为 $1.2 \text{ 点}/\text{m}^2$, 数据量约为 2.3 G 。

实验平台的配置为: 软件环境为 Windows2000 操作系统, MPICH2 消息传递并行库; 硬件为 17 台微机通过百兆网卡构成的局域网, 各台机器的配置为 P4, 2 G 的 CPU, 1 G 内存。其中一台为管理节点, 负责接收客户机的请求并通

过 MPI 库管理计算任务,16 个计算节点进行并行数据获取。

实验 1 不同格网划分大小和要获取的指定范围大小对数据获取速度的影响。

在单机上实现 Las 格式点云数据不同大小格

网的数据组织,即将原数据按 X 和 Y 坐标分割为 N 行 M 列。待获取范围为一长方体,该长方体以原数据外包盒的中心为中心,各边长度等于原数据外包盒各边乘系数 $r(0 < r \leq 1)$ 。原数据大小为 64 MB 的实验结果如表 1 所示。

表 1 不同分割大小和不同获取范围所需的时间/s
Tab. 1 Time for Different Partition Size and Area

$M \times N$	$r=1$	$r=1/2$	$r=1/4$	$r=1/8$	$r=1/10$	$r=1/100$	$r=1/1\,000$	$r=1/10\,000$
1 ×1	6.391	3.328	2.438	2.187	2.172	2.094	2.078	2.094
2 ×2	8.484	3.922	2.625	2.281	2.218	2.125	2.141	2.109
3 ×3	8.562	3.797	0.687	0.375	0.328	0.265	0.282	0.265
4 ×4	9.391	2.844	1.063	0.735	0.688	0.593	0.594	0.594
5 ×5	8.281	2.750	1.156	0.219	0.172	0.109	0.109	0.110
6 ×6	8.953	2.437	0.735	0.406	0.361	0.281	0.266	0.265
9 ×9	9.015	2.359	0.781	0.406	0.110	0.031	0.031	0.047
12 ×12	8.360	1.954	0.703	0.219	0.156	0.078	0.063	0.078
18 ×18	8.172	1.875	0.625	0.266	0.109	0.031	0.031	0.031

由表 1 的数据分析发现一个整体的趋势,即当获取的范围小时,分割得越细,时间越短;但当获取的范围大时,分割得越细,时间反而越长。

实验 2 不同格网划分大小及不同数据分布策略对数据获取速度的影响。

影响数据的获取速度不仅与获取数据的范围大小有关,还与范围所在的位置有关。如图 4 所示,其中图 4(a)只需判断第 1 个分块中的数据是否落在获取的范围内,而图 4(b)则需要判断 1~4 分块中的数据是否落在获取的范围内。

因此,在实验之前,先连续随机产生 1 000 个不同大小和位置的平截头体作为该实验获取数据的指定范围,每个平截头体前面 4 个顶点两两之间的空间距离不超过 1 km,后面 4 个顶点两两之间的空间距离都不超过 0.5 km。用 2.3 G 的数据进行测试,实验结果如表 2 所示。

表 2 不同分布策略获取 1 000 次的总时间
Tab. 2 Total Time of Different Distribution Strategies for 1 000 Times

划分块数 ($M \times N$)	顺序分配获取 的时间/s	交叉分配获取 的时间/s	交叉分配时间/ 顺序分配时间/s
1 ×1	33 243	33 293	1.00
3 ×3	3 409	3 205	0.94
9 ×9	1 295	1 090	0.84
18 ×18	576	432	0.75
27 ×27	465	298	0.64
54 ×54	416	287	0.67
162 ×162	316	310	0.78

从表 2 的数据分析可知,只有在不分块时两种策略都只能把数据分布到一台数据服务器上,因此获取的时间相当;而当分块后,交叉分配策略比顺序分配策略的获取速度快;交叉分配与顺序分配的获取时间比随着分块数增多先变小、后变大,其原因是块数少时,顺序分配所能利用的计算节点少,而当分块数很多时,顺序分配时也能利用到较多的计算节点。

实验 3 串行与并行比较。

评价并行计算性能加速比 $S_p = T_p / T_s$ 是评价并行计算性能的主要指标。式中, p 为节点个数; T_p 为在 p 个节点并行计算耗时; T_s 为单机运行耗时。为了测试加速比,分别在单机和 10 台机器的并行环境下进行实验。实验数据从 64 M 到 2.3 G,划分块数为 54 ×54,如图 5 所示,原数据越大,加速比越高。

为了测试该数据组织方式下的并行率,实验数据大小为 2.3 G,划分块数为 54 ×54。计算节点从 1 台变化到 16 台,分别查询 1 000 次,图 6 显示的是不同计算节点数量所对应的加速比。由图中可以看出,当计算节点达到 11 个时,其加速比达到 9.2。由于网络速度及网络开销原因,超过 11 个节点后再增加计算节点,提高速度缓慢。

该并行思想适用于不同的数据组织方式,但还要通过实验比较和理论分析各种数据组织方式的并行效果,找出一种最适合分布式和并行获取的组织方式。

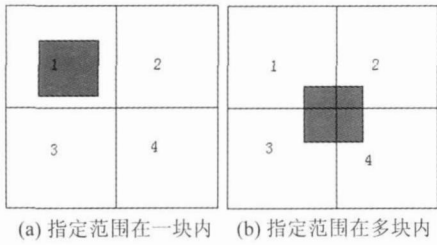


图 4 范围所在位置不同
对获取速度的影响

Fig. 4 Impact of Asseccing Speed for
Different Position of Required Data Area

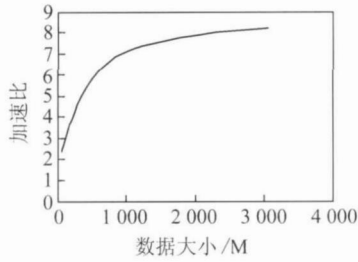


图 5 数据大小对加速
比的影响

Fig. 5 Impact of Accelerating
Ratio for Different Data Size

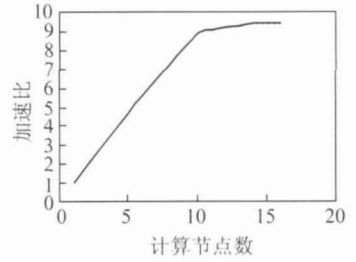


图 6 不同计算节点数所
对应的加速比

Fig. 6 Accelerating Ratio for Diff-
erent Number of Computing Nodes

参 考 文 献

- [1] 朱士才. LiDAR 的技术原理以及在测绘中的应用[J]. 现代测绘, 2006, 4(7): 12-13
- [2] 张玉方, 程新文, 欧阳平, 等. 机载 LiDAR 数据处理及其应用综述[J]. 工程地球物理学报, 2008, 5(1): 119-124
- [3] 杨崇源, 张继贤, 林宗坚. 虚拟地形场景绘制中的实时 LOD 算法[J]. 测绘学报, 2001, 30(2): 133-139
- [4] Hu Y. Automated Extraction of Digital Terrain Models, Roads and Buildings Using Airborne Lidar Data [D]. Calgary: Univ of Calgary, 2003
- [5] Fabio R. From Point Cloud to Surface: the Modelin-gand Visualization Problem[C]. ISPRS Tarasp-Vulp-era, Switzerland, 2003
- [6] Rusinkiewicz S, Marc L. Qsplat: A Multi Resolution Point Rendering System for Large Meshes[C]. Pro-cession of SIGGRAPH, New Orleans, 2000
- [7] 黄先锋, 陶闯, 江万寿, 等. 机载激光雷达点云数据的实时渲染[J]. 武汉大学学报·信息科学版, 2005, 30(11): 975-978
- [8] Meisl P G, Ito M R, Cumming I G. Parallel Synthet-ic Aperture Radar Processing on Workstation Net-works[C]. The 10th International Parallel Processing Symposium, Honolulu, HI, 1996
- [9] Yang Chaotung, Chang Chihli, Hung Chichu, et al. Using a BEOWULF Cluster for a Remote Sensing Application[C]. The 22nd Asian Conference on Re-mote Sensing, Singapore, 2001
- [10] 蒋艳凰, 杨学军, 易会战. 卫星遥感图像并行几何校正算法研究[J]. 计算机学报, 2004, 27(7): 944-945
- [11] 赵春宇, 孟令奎, 林志勇. 一种面向并行空间数据库的数据划分算法研究[J]. 武汉大学学报·信息科学版, 2006, 31(11): 962-965

第一作者简介: 王宗跃, 博士生, 主要从事 LiDAR 数据分析处理、模式识别和计算机视觉。

E-mail: wangzongyue1979@163.com

An Approach for Distributed Organization and Paralleled Access of LiDAR Point Clouds Data

WANG Zongyue^{1,2} MA Hongchao¹ XU Honggen¹ WU Jianwei¹

(1 School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China)

(2 Computer Engineering College Jimei University, 185 Ynjiang Road, Jimei, Xiamen 360021, China)

Abstract: A fast paralleled data access approach for large scale distributed point clouds is proposed in this paper. First, large scale point cloud data is uniformly distributed in the servers, where data is saved according to some data organization. When data in a definite range is required, the servers will immediately access the data by its own organization. This approach is fit for different data organizations. In the experiment, the accelerating ratio reaches 8 when the number of severs is 9 and the required data area is much less than the total data.

Key words: LiDAR point clouds; fast access; paralleled obtaining; frustum

About the first author: WANG Zongyue, Ph.D cadidate, majors in proccession of LIDAR data, pattern recognition and computer vision.
E-mail: wangzongyue1979@163.com