文章编号: 0494-0911(2011)05-0001-04 中图分类号: P237 文献标识码: B

基于切片技术的激光点云粗差探测

罗德安,宋晓华,廖丽琼,吴志群
(北京建筑工程学院 测量工程系,北京 100044)

Point Cloud Outlier Detection Based on Slice Technology

LUO Dean, SONG Xiaohua, LIAO Liqiong, WU Zhiqun

摘要:借鉴数据挖掘领域内的离群点探测技术,提出适合于海量点云数据的、基于切片技术的粗差探测算法,该算法将三维点云经过切片降为二维点,有效解决了三维环境下邻域搜索和粗差指标计算效率低下的问题,使算法整体效率和易用性得以显著提高。

关键词:点云; 离群点; 误差

一、引 言

由于扫描过程中外界环境因素(如树木、移动的行人、车辆及飞鸟等)对扫描目标的遮掩,扫描对象反射特性差异(如完全透射及吸收等)及多路径效应等的影响,原始扫描点云中必然存在大量和所需信息无关的扫描数据点,这些信息的存在,必然影响到后续的三维模型建立[1]。因此,在通过密集点云生成三维模型之前,必须对原始数据进行必要的数据清理,去掉可能存在的粗差或错误(也称之为"脏数据")。一般来说,对于扫描数据中较为明显的粗差和无关信息,可以在交互环境下,通过人机交互将其裁剪掉。但对于和扫描对象相距较近的粗差信息,由于数据量较大且较为分散,很难通过人机交互模式逐一剔除,为此,设计一些简单有效的自动算法来探测和剔除这些粗差信息具有重要的理论价值和实用价值。

基于激光点云的粗差探测和传统的测量数据(如三角网控制数据、高程控制数据)粗差探测有着明显的差异,主要表现在以下几个方面: ① 离散性,所有的观测量都是相对独立的个体,个体之间没有直接的关联关系(如拓扑关系); ② 海量数据,激光扫描获得的数据量极其巨大(几百万甚至几千万个点极其普遍),不可能和传统的测量数据一样,构建平差模型对其进行检验; ③ 要求算法具有简单高效的特性,由于其海量数据特性,太过复杂的算法将占用大量的计算时间,即使其在精度和可靠性上有优势,也无法满足激光点云粗差剔除的要求。

在对国内外已有粗差探测方法进行充分研究的基础上,笔者借鉴数据挖掘领域内的离群点探测技术(outlier detection),提出了适合于海量点云数据的、基于切片技术的粗差探测算法。本文完整描述了该算法的基本思想、实现方法及步骤,并借助于真实数据进行了相应的试验与分析,证明了本算法的有效性及实用性。

二、相关研究

离群点发现是数据挖掘领域一个重要的研究方向,其目的是发现数据库中不具备数据一般特性的数据对象。所发展起来的算法可应用于许多领域,如网络入侵检测、电信和信用卡欺骗及气象预报等。这里提及的离群点和激光点云数据中的粗差有着相同的性质,所以,数据挖掘中的离群点探测算法也可以用来探测激光点云中的粗差信息。

已有的离群点探测方法大致可分为以下几类[2]: 基于统计分布(distribution-based)的算法、基于深度(depth-based)的算法、基于距离(distance-based)的算法、基于聚类(cluster-based)的算法和基于密度(density-based)的算法。基于分布的算法涉及的大多数分布模型只能直接应用于单变量特征空间,难以应用于多维空间,而且这种模型要求预先知道数据的分布,而通常情况下难以满足该要求; 基于深度的算法是一种基于计算几何的方法,这种方法通过计算不同层面的 k-凸面来查找离群点,凸面的外层认为是离群点,但该算法有个明显的缺点,随着维数增加,时间及空间开销巨量增加(即维数灾难); 基于距离的算法认为在数据库中,若 $p\%$ 的对象与某对象的距离超过阈值 d,则这个对

收稿日期: 2010-05-31
基金项目: 国家自然科学基金资助项目(40871196); 北京市优秀人才培养资助项目(20061D0501700244); 北京建筑工程学院博士基金项目(1005002)
作者简介: 罗德安(1968—),男,四川乐至人,博士,副教授,主要从事地面激光雷达数据处理理论及其应用研究。

象可视为离群点，该算法的效率受限于邻域搜索机制，此外，阈值的确定也存在一定的问题；基于聚类的算法按照一定的方式对样本进行聚类操作，当操作完成时，剩下的未聚类数据即为离群点；基于密度的算法通过数据空间的所有维度来计算对象的距离，进而计算对象的可达密度，最后通过局部的偏离度来判断离群点，和基于距离的算法类似，算法存在邻域检索效率和阈值选择问题。

上述算法在相关领域（如入侵检查、电子欺骗等）已经得到广泛的应用，并证明其卓有成效，但是这些算法对于激光点云数据处理来说，是不能直接加以利用的。如前所述，这里存在缺乏先验信息、拓扑信息缺失、数据量巨大、算法涉及的阈值难以确定，以及计算效率低等问题，通过综合分析及比较发现，基于距离、基于密度和基于聚类 3 种类型的算法较适合激光点云数据的粗差探测，但是这些算法都需要加以改进，以提高其效率，降低其计算时的空间及时间代价。

地面激光雷达技术是近年来出现的新技术，对其理论及应用的研究还不完善，从已有的文献看，对激光点云粗差探测方法的研究就更为有限。Sotoodeh在 2006 年提出了基于局部离群因子（local outlier factor，LOF）的粗差探测方法[3]，尽管在算法中提出了利用构建 kd-tree（一种空间索引结构）来加速三维邻域点的查找，但该算法的时间代价还是过大，此外还存在阈值选定对探测结果带来影响的问题；2007 年 Sotoodeh 又提出了基于层次聚类的粗差探测方法[4]，该算法分两步实现，即粗聚类和精聚类，首先利用 EMST（the Euclidean minimum spanning tree）数据结构消除全局性具有较大尺度的粗差，然后利用 GG（Gabriel graph）数据结构进一步剔除局部性粗差信息，但该算法必须事先构建相应的 EMST 结构树或 GG 图（类似 delaunay triangulation），其时间开销也是较为巨大的，所以算法还有待改进；白志辉等对激光点云的粗差剔除也作过相应的研究[5]，但所提出解决方法的适用范围受到极大限制，仅限于局部地形扫描数据和平面扫描数据的粗差探测。

三、基于切片技术的点云数据粗差探测算法

1. 定 义

离群点判定的关键在于对其离群程度的度量，必须根据一定的标准找到一个合适的度量指标来描述离群点与其邻域的偏离程度，而后根据某设定的阈值来划分离群点和非离群点。对于点云数据中粗差的离群程度，本文选用了文献[6]提出的局部距离离群因子（local distance outlier factor，LDOF）来描述及度量，其相关的定义如下[6]：

定义 1 设 N_p 为 x_p 的 k 邻域集（不包括 x_p），则 x_p 的 k 邻域距离等于 x_p 与 N_p 中所有对象间距离的平均值。更为规范的描述是，设 $dist(x,x')\geqslant 0$ 为任意两观测量间的距离，则 x_p 的 k 邻域距离可定义为

$$\bar{d}_{x_p}=\frac{1}{k}\sum_{x_i\in N_p}dist(x_i,x_p) \tag{1}$$

定义 2 设 N_p 为 x_p 的 k 邻域集（不包括 x_p），则 x_p 的 k 邻域内部距离等于 N_p 中各对象间距离的平均值，定义如下

$$\bar{D}_{x_p}=\frac{1}{k(k-1)}\sum_{x_i,x_i'\in N_p,i\neq i'}dist(x_i,x_i') \tag{2}$$

定义 3 x_p 的 $LDOF$ 定义为

$$LDOF_{x_p}=\frac{\bar{d}_{x_p}}{\bar{D}_{x_p}} \tag{3}$$

$LDOF_{x_p}$ 可以用来表征 x_p 偏离其邻域 N_p 的程度，值越大，偏离程度越高，反之越低。

2. 算法的基本思想

实际的建筑物扫描过程中，其扫描坐标系大部分情况下并不是随意的，对于扫描仪整平后所进行的扫描而言，其扫描坐标系的 Z 轴与建筑物的立面呈平行关系（如图 1 所示），在完成多站配准后，其 Z 轴将仍然平行于建筑物立面。对于其他扫描对象，也可以找到类似的关系，即扫描对象的轴向与扫描坐标系某轴向呈平行关系。事实上，如果不存在上述关系，也可以通过简单的坐标变换来建立上述关系，所以，这里约定本文算法所有的扫描点云数据都满足上述轴向平行条件（如图 2 所示）。

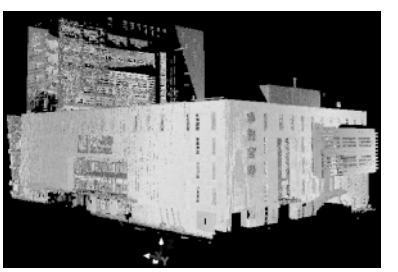

图 1 原始配准点云

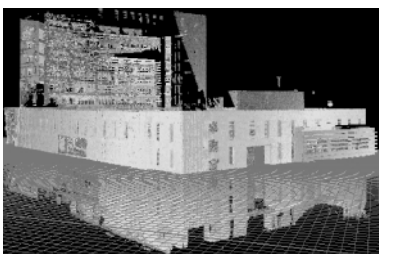

图 2 点云切片

算法的基本思想如下:

1) 对待处理的点云数据，选定某坐标轴向(约定为 Z 轴) 作为参考轴，以一组与该轴向垂直的等间距平面对原点云数据进行切片分割，并将到每一个分割平面的距离小于或等于(分割面上部含等于) 0.5 倍间距的点云数据作为一个独立的数据处理单元(如图 3 所示)。

2) 将独立数据单元的点云数据沿 Z 轴投影到分割平面上(如图 4 所示)。

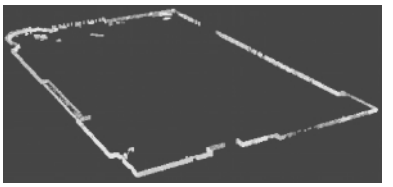

图 3　切片片段(0.5 m)

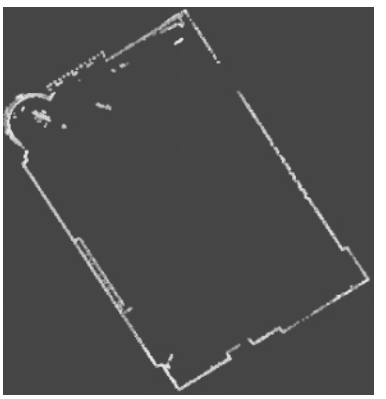

图 4　切片投影

3) 通过二维 kd-tree 建立投影数据的拓扑关系，以加速邻近点查找。

4) 逐点计算 $LDOF$，根据 Top-n 粗差认定原则(即 $LDOF$ 值最大的前 n 个对象视为粗差) 对粗差进行剔除。

5) 逐一完成所有数据处理单元的粗差探测与剔除，并逆向变换回原来的坐标空间，即可获得已完成粗差探测与剔除工作的干净点云数据。

3. 算法实现

1) 算法输入值: 某点云数据集合 D，阈值 d(用以设定切片间距)。

2) 根据切片阈值 d 对点云数据集合 D 进行切片分割，得到一系列新的点云子集 D_i，且满足 $D = \{D_i\}$。

3) 对于每个 D_i，将其投影到对应的切片平面上，形成新的投影点集合 d_i，并对集合 d_i 建立kd-tree。

4) 根据阈值 k(用以设定最近邻域大小) 和式(3)，逐点计算集合 d_i 中的 $LDOF$ 值，并对其进行排序，最后根据阈值 n(自然数，用以确定 $LDOF$ 值最大的前 n 个对象)，去掉前 n 个 $LDOF$ 值相对应的数据对象，得到集合 d'_i。

5) 对每个集合 d'_i 作逆向变换生成新的数据集 D'_i，则最后可得完成粗差探测与剔除的点云集合 $D' = \{D'_i\}$。

四、算法试验及评价

为了对算法进行全面评价，采用了徕卡 Scan-Station2 对某商厦进行了外立面的全面扫描，共获得 8 站扫描数据，在经过配准及在交互环境下去掉无关信息(如地面数据、其他邻近建筑物等) 后得到了该商厦外立面的完整点云数据(如表 1 所示)。该测试数据总的激光点为 3 349 091 个，通过仔细分析及统计获得其外墙立面扫描点为 3 313 449个，离群点数为 35 642 个，大部分离群点为室内目标对穿越玻璃窗到达室内的激光反射形成的光斑点。

1. 切片间距对探测结果的影响分析

通过对不同切片间距进行测试发现(如表 1 所示)，随着切片间距的增加，将外墙立面点误判为离群点(即去真) 的概率并没有显著的变化，而将离群点误判为墙面点的概率却显著增加。这说明随着切片间距加大，离群扫描点投影后使得离群点与其邻域间距离均值变小，从而使其偏离指标计算值(即$LDOF$值) 变小，其结果是使得其在后续的排序中名次向后偏移，从而导致纳伪的概率增大。

表 1　不同切片间距的探测结果

切片间距/m	所属类型	探测结果/个		误判比例/(%)	
		墙面点	离群点	去真	纳伪
0.2	墙面点	3 302 135	11 314	0.34%	
	离群点	273	35 369		0.77%
0.5	墙面点	3 293 121	20 328	0.61%	
	离群点	394	35 248		1.11%
1.0	墙面点	3 267 354	46 095	1.39%	
	离群点	557	35 085		1.56%

2. 算法效率对比分析

为了检验算法的效率，将本文算法(0.5 m 切片间距) 与文献 [3] 和文献 [4] 提出的算法进行了对比分析，其结果如表 2 所示。

表 2 不同算法对比

	本文算法	文献[3]算法	文献[4]算法
运算时间/s	281	482	564
去真概率/(%)	0.61	1.23	0.49
纳伪概率/(%)	0.77	2.52	0.70

从表 2 可发现，本文算法的效率明显优于后两种算法，而其准确度介于两者之间，由于本文算法涉及的切片间距具有可控性，所以，实际应用可以根据需要(准确度优先还是速度优先)来适当选取。

五、结束语

本文提出的算法，从本质上讲是将三维数据通过切片技术及投影变换为二维数据，起到了降维的作用，从而使得原来在三维数据环境下难以使用或效率不高的算法(如基于距离的算法、聚类算法等)能够在低维环境下应用并具有较高效率。试验表明，该算法对海量点云数据的粗差探测具有较高的适应性和效率，并具有一定的自动化水平。和已有类似算法相比，本算法在效率上具有明显优势，在使用上无任何限制，更不需要任何先验信息。

参考文献：

[1] 罗德安，廖丽琼. 地面激光扫描仪的精度影响因素分析[J]. 铁道勘察，2007，33(4)：5-8.

[2] HODGE V J，AUSTIN J. A Survey of Outlier Detection Methodologies[J]. Artificial Intelligence Review，2004，22(2)：85-126.

[3] SOTOODEH S. Outlier Detection in Scanner Point Clouds[J]. ISPRS，2006，36(5)：297-302.

[4] SOTOODEH S. Hierarchical Clustered Outlier Detection in Laser Scanner Point Clouds[J]. ISPRS，2007，36(3)：383-388.

[5] 白志辉，张舒，王响雷，等. 三维激光扫描点云粗差剔除方法研究[J]. 矿山测量，2009(2)：13-15.

[6] BREUNIG M M，KRIEGEL H P，NG R T，et al. LOF：Identifying Density-based Local Outliers[C]// Proc of SIGMOD'00. Dallas：[s. n.]，2000，427-438.

点亮数字中国

——拓普康在京举办移动测量系统在数字城市建设中的应用研讨会

【本刊讯】 近日，拓普康在京举办“点亮数字中国”——移动测量系统在数字城市建设中的应用研讨会。北京拓普康商贸有限公司专业人员介绍了拓普康全方位多领域的测量科技与理念，以及一直以来为中国数字城市建设提供的全面完善的解决方案。

在研讨会中，中国科学院、中国工程院李德仁院士作主题报告《基于 3S 集成的移动测量技术及其应用》，中国测绘科学研究院地图学与地理信息系统研究所所长李成名阐述了数字城市地理空间框架建设技术的思路与展望，美国拓普康定位系统有限公司 Eduardo 高级副总裁为大家展示了拓普康 IP-S2 在全球的应用案例，海南测绘局麦照秋、山西省基础地理信息院吴博义分别发表了 I-S2 移动测量系统在高速公路测量中的应用报告及 IP-S2 在标准化生产作业流程系统中的应用等。

会上，拓普康举办了 GNSS(全球导航卫星系统)国产化最新成果——HiPer IIG 下线仪式，这是为中国客户量身打造的 GNSS 解决方案。

国家测绘局局长徐德明对研讨会给予了大力支持，并高度肯定和赞扬了“点亮数字中国”这一会议主题

参加研讨会的来宾对现场展示的拓普康 IP-S2 移动测量系统表示出了浓厚的兴趣

(本刊编辑部)