

文章编号:1671-6833(2021)02-0080-07

# 基于多源传感器数据融合的三维场景重建

陈义飞<sup>1</sup>, 郭 胜<sup>2</sup>, 潘文安<sup>2</sup>, 陆彦辉<sup>1,3</sup>

(1. 郑州大学 信息工程学院, 河南 郑州 450001; 2. 香港中文大学(深圳) 理工学院, 广东 深圳 518172; 3. 深圳市大数据研究院, 广东 深圳 518172)

**摘要:** 针对特定校园场景重建结果中存在的目标冗余等情况, 提出了一种相机 RGB 位图和激光雷达数据融合的方法。在三维重建领域, 通过数据融合的方法剔除特定场景中无关目标以实现三维场景重建。首先使用轻量级的 LeGO-LOAM 算法, 将不同类型的特征点进行特征提取与匹配, 融合不同时刻的点云完成点云地图的重建; 然后对构建的点云地图中可能存在的无关目标, 借助多源传感器数据和深度学习在计算机视觉领域中的应用技术, 在三维空间中进行目标检测与剔除。对于点云地图建模与目标检测 2 个不同过程, 采用点云配准的方法对其进行算法融合, 最终完成校园环境下的场景重建。实验结果表明, 基于多源数据融合的方法可有效地将三维建模与目标检测 2 个过程进行结合, 完成校园环境下的无冗余目标的点云地图构建。该方法可应用于智慧城市、无人驾驶等领域, 具有实际应用价值。

**关键词:** 数据融合; 三维建模; 深度学习; 目标检测; 特征匹配; 场景重建

**中图分类号:** TP391.9

**文献标志码:** A

**doi:**10.13705/j.issn.1671-6833.2021.02.008

## 0 引言

近年来, 随着激光测量技术的迅速发展和激光测距设备成本的逐渐降低, 利用激光点云数据来重建三维场景模型得到越来越广泛的应用。城市的高速发展对城市三维场景重建提出了新的要求, 其中包括城市三维基础空间信息数据要求保持良好的现势性, 数据更新周期要短, 几何精度要高。随着智慧城市建设的全面推广, 针对目前智慧城市三维建模中存在的成本高、周期长等问题, 探索并采用新技术来解决这些现实问题就显得尤为重要。

基于激光扫描数据的三维场景重建<sup>[1-3]</sup>主要是通过通过对三维激光扫描仪获取的深度数据进行预处理、配准和网格化等处理来实现的。相对于传统的基于图片的三维场景重建方法, 基于激光数据的三维场景重建技术能够恢复出几何信息准确性更高、真实感更强的三维模型, 更有利于在计算机中再现客观的真实场景。但是, 在使用激光雷达进行数据扫描的过程中, 通常会采集到与场景无关的数据(例如行人和车辆), 严重影响了三维

场景重建的精度和效果, 如何剔除三维点云模型中的无关数据, 构建高精度的三维模型是三维场景重建的难点。

检测和剔除点云中的无关目标是一项极具挑战的工作, 直接在点云中检测目标计算量较大, 且不易实现。针对这一问题, 本文使用了点云和图像融合的方法来实现特定目标的检测与剔除。以采集的三维点云数据和真实场景中的二维纹理数据为基础, 采用多数据融合的深度学习方法训练网络模型, 进而使用训练后的网络进行点云目标检测与剔除。该方法能够有效提升三维空间中点云目标的处理速度和精度, 使三维场景重建过程更加快速高效。

## 1 整体硬件方案及预处理标定

整体硬件方案设计如图 1 所示, 主要由移动底座、支架、激光雷达及摄像头组成。移动底座和支架控制不同传感器在采集数据过程中的稳定性, 激光雷达和摄像头用来采集不同传感器的数据信息。硬件设计原理如图 2 所示, 硬件传感器有 GPS、IMU、相机和激光雷达。其中 GPS 和 IMU

收稿日期: 2020-09-09; 修订日期: 2020-11-13

基金项目: 国家自然科学基金资助项目(61771431)河南省技术创新引导专项(182106000027); 深圳市科创委面上项目(JCYJ20190813170803617)

通信作者: 陆彦辉(1972—), 女, 河南许昌人, 郑州大学教授, 博士, 主要从事宽带无线通信及大数据相关的研究和应用, E-mail: ieyhlu@zzu.edu.cn.

在下一实验阶段使用。由于安装后的相机和激光雷达具有各自不同的坐标系,需要先对传感器进行标定,如图 3 所示为标定示意图,保证在数据收集过程中空间坐标系一致。



图 1 整体硬件结构图

Figure 1 Overall hardware structure

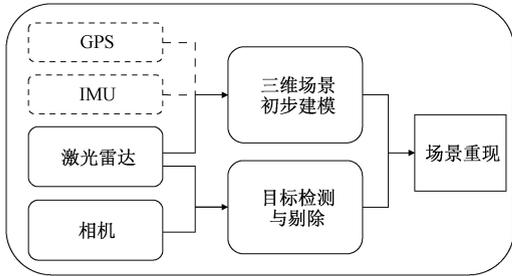


图 2 硬件设计原理

Figure 2 Hardware design schematic

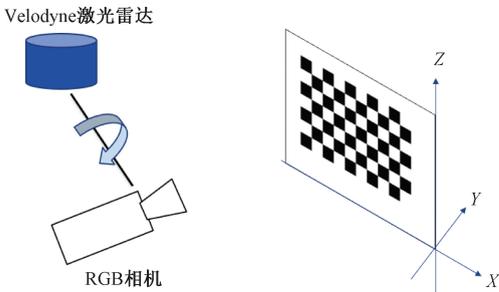


图 3 标定示意图

Figure 3 Calibration diagram

在机器人操作系统 (robot operating system, ROS) 中时间同步通过设定对应的采集帧率、录制图像和点云流的方式来实现,而空间同步只能通过标定的方式来实现。本文使用 ROS 中的多源传感器标定模块实现空间同步的标定,标定方法为张正友标定法,标定装置由激光雷达、相机、黑白棋盘格组成。多源传感器标定分为 2 步。

第 1 步是单个相机的标定<sup>[4]</sup>,主要目的是建立相机成像的几何模型,保证后续算法的稳定性。世界坐标系下的点与成像的相机坐标系中的点的对应关系为:

$$Z_c m = K[R/T]M. \quad (1)$$

式中:  $Z_c$  表示比例因子;  $K$  是相机内参矩阵;  $[R/T]$  是相机的外参矩阵;  $m$  和  $M$  分别代表像素坐标和世界坐标。

整个标定过程需要求解的模型参数就是相机的内参矩阵、外参矩阵和畸变参数,将式(1)展开可得:

$$Z_c \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{bmatrix} f_x & 0 & u_0 & 0 \\ 0 & f_y & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (2)$$

式中:  $Z_c$  表示比例因子;  $u, v$  表示像素坐标;  $f_x, f_y$  分别表示  $x$  轴、 $y$  轴的归一化焦距,  $u_0, v_0$  表示图像原点在相机坐标系的中心;  $[R/T]$  表示相机的外参矩阵,也即世界坐标系到像素坐标系的旋转平移矩阵;  $X_w, Y_w, Z_w$  表示世界坐标。

本文所用摄像头是无畸变的广角摄像头 ( $FOV=120^\circ$ ),以保证视角的广阔性与准确性。安装完硬件平台后需求解参数矩阵,如式(1)所示,本文使用黑白棋盘格标定法求解内参矩阵  $K$  和外参矩阵  $[R/T]$ 。本文借助 ROS 中相机标定工具解算多个点对关系,最终得到相机的内、外参矩阵:

$$K = \begin{bmatrix} 1341.95 & 0 & 359.35 \\ 0 & 1342.67 & 158.30 \\ 0 & 0 & 1 \end{bmatrix}; \quad (3)$$

$$[R/T] = \begin{bmatrix} 1293.19 & 0 & 361.61 & 0 \\ 0 & 1312.08 & 156.66 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

第 2 步是相机和雷达的联合标定<sup>[5]</sup>,目的是找到雷达和相机 2 个不同坐标系下的关系。为了求解相机与雷达间的关系,需要定义相机与雷达间的空间位置模型,雷达和相机坐标系下点的坐标关系为:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = [R/T] \times R_0 \times T_{v_2c} \times \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \quad (5)$$

式中:  $u, v$  表示图像坐标;  $X, Y, Z$  表示雷达坐标;  $[R/T]$  是指相机的外参矩阵;  $R_0$  为相机的矫正矩阵;  $T_{v_2c}$  为雷达到相机的投影矩阵。

本文所使用的激光雷达为 VLP-16,它的优点在于体积小,受环境光变化的影响比较小且测距精确。其感知距离为 100 m,质量为 830 g,水平

视角为  $360^\circ$ , 垂直视角为  $30^\circ$ , 误差在 3 cm 左右。工作原理是通过旋转发射红外激光束, 在遇到物体表面时会返回接收, 通过记录接收时间和角度来计算该点距离雷达的相对距离。采集的数据为整个场景中对应时间帧的所有点在雷达坐标系下的三维坐标值以及反射率。

整个联合标定过程使用 ROS 中 Autoware 工具实现。式(5)中,  $[R/T]$  在相机标定过程中已求出, 由于相机无畸变, 此处矫正矩阵  $R_0$  为单位矩阵, 最终所得的雷达到相机的投影矩阵为:

$$T_{v2c} = \begin{bmatrix} 0.039 & 0.105 & 0.994 & -0.058 \\ -0.999 & 0.015 & 0.038 & 0.572 \\ 0.011 & -0.994 & 0.105 & -0.038 \end{bmatrix} \quad (6)$$

## 2 三维场景建图

三维场景建图<sup>[6-8]</sup>是在激光雷达采集数据的过程中同时定位和生成周围环境的点云地图, 如果激光传感器本体静止, 激光只是绕固定轴做选择运动, 点云的注册相对简单。但使用环境中激光通常是运动的, 这就造成建图需要估计运动过程中激光传感器的位姿。

本文三维重建算法 LeGO-LOAM<sup>[8]</sup>是一种轻量版的 SLAM (simultaneous localization and mapping) 算法, 用于实时六自由度姿态估计, 可以在低功耗的嵌入式系统上实现。系统整体框架由 5 个子模块构成, 如图 4 所示, 输入为原始点云数据, 输出为位姿估计(雷达的姿态估计)。下面分别介绍图 4 中每个模块的功能。

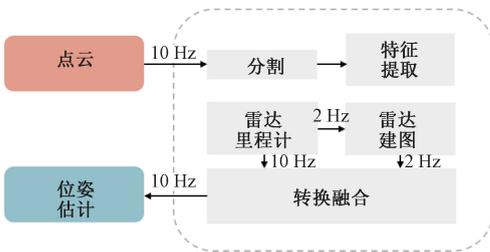


图 4 LeGO-LOAM 整体框架

Figure 4 LeGO-LOAM overall framework

### 2.1 点云的分割

图 4 中的分割模块用于实现点云分割与分簇, 它主要采取基于图像的分割方法将点云分割成许多点簇。将收到的 1 帧点云投影到 1 800 像素  $\times$  16 像素的图像上, 每 1 个收到的点表示 1 个像素, 像素值对应点到传感器的距离。经图像化之后得到 1 个矩阵, 对矩阵的每 1 列进行估计就可以完成对地面的估计, 提取地面点。基于距离

将点分组为多个聚类, 同一类的点具有相同的标签, 而由于室外环境中会有很多噪声点, 在连续 2 帧里都在同一位置出现的可能性不大, 因此直接剔除掉, 不参与帧间匹配。本文采用的方法是计算每个点簇中点的个数, 将所有个数小于 30 的点簇都剔除掉, 这里设置为 30 是将采集数据过程中移动的行人目标剔除掉。最终得到点的类别(地面点或者分割点)和点的深度值。

### 2.2 特征提取

图 4 中的特征提取子模块为点云特征提取, 先在水平方向上将深度图分为多个子图像, 计算每个点的曲率:

$$c = \frac{1}{|S| \cdot \|r_i\|} \left\| \sum_{j \in S, j \neq i} (r_j - r_i) \right\| \quad (7)$$

式中:  $c$  表示曲率, 本文实验中设定为 0.1;  $S$  表示对应的点云集;  $r$  表示点到雷达的欧式距离,  $r_i$  与  $r_j$  表示点云集中的不同点。在子图像中比较曲率与设定的阈值(0.1)大小将点分为 2 大类, 分别为平面点或者角点。

### 2.3 雷达里程计

雷达里程计子模块为雷达的运动估计, 通过相邻帧点对线、点对面的匹配, 边缘点到线的距离为:

$$d_1 = \frac{|(P_{t,i}^M - P_{t-1,j}^M) \times (P_{t,i}^M - P_{t-1,l}^M)|}{|P_{t-1,i}^M - P_{t-1,l}^M|} \quad (8)$$

式中:  $P_{t,i}^M$ 、 $P_{t-1,j}^M$ 、 $P_{t-1,l}^M$  分别表示  $t$  时刻边缘特征点集  $M$  中第  $i$  个点、 $(t-1)$  时刻距离  $i$  最近的  $j$  点和  $l$  点。

而平面点到面的距离为:

$$d_2 = \frac{|(P_{t,i}^M - P_{t-1,j}^M) \times (P_{t-1,j}^M - P_{t-1,l}^M) \times (P_{t-1,j}^M - P_{t-1,m}^M)|}{|(P_{t-1,j}^M - P_{t-1,l}^M) \times (P_{t-1,j}^M - P_{t-1,m}^M)|} \quad (9)$$

式中:  $P_{t,i}^M$ 、 $P_{t-1,j}^M$ 、 $P_{t-1,l}^M$ 、 $P_{t-1,m}^M$  分别表示  $t$  时刻平面特征点集  $M$  中第  $i$  个点、 $(t-1)$  时刻距离  $i$  最近的  $j$  点、 $l$  点和  $m$  点,  $j$ 、 $l$  和  $m$  3 个点可以确定一个平面。

本文使用 Levenberg-Marquardt<sup>[6]</sup>方法(L-M)优化点到线和点到面的最小距离来完成前后帧点云的匹配, 估计雷达的相对位姿。L-M 优化可以较快地找到最优值, 从而获得姿态变换来进行空间约束, 执行闭环检测, 消除模块的误差。每帧输出的相对位姿主要由旋转平移值确定, 旋转值为  $[\theta_{pitch}, \theta_{yaw}, \theta_{roll}]$ , 平移值为  $[t_x, t_y, t_z]$ , 分别代表雷达在 XYZ 轴上的各个旋转量和偏移量。

### 2.4 雷达建图以及转换融合

雷达建图和转换融合子模块主要用于雷达运

动前后帧间位姿的转换和点云地图的构建。其中前后帧点云间的关系为:

$$P_k^I = R_{k-1}^I P_{k-1}^I + T_{k-1}^I \quad (10)$$

式中:  $P_k^I$  与  $P_{k-1}^I$  分别表示  $k$  时刻和  $(k-1)$  时刻下的点云特征点集;  $R_{k-1}^I$  与  $T_{k-1}^I$  表示前后帧特征点云间的旋转矩阵和平移向量。

本文实验场景为某大学校园实验楼下,如图 5 所示为 LeGO-LOAM 建模算法在实际测试过程中的效果。

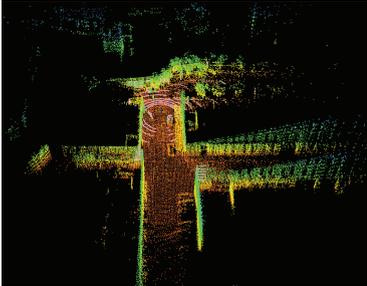


图 5 LeGO-LOAM 测试效果

Figure 5 Test result of LeGO-LOAM

### 3 多源数据融合

在激光雷达建图的过程中总会产生一些和地图信息无关的障碍物、车辆以及行人。在制作三维地图时需要对这些目标进行检测和剔除。目的就是保留关键信息,提高三维场景建模的速度和准确度。本文借助深度学习技术在目标检测方面的算法来解决三维重建中遇到的问题。三维点云的数据是稀疏且无序的。目前成熟的二维目标检测方法并不能直接应用于三维点云中。

2017 年, Charles 等<sup>[9]</sup> 和 Qi 等<sup>[10]</sup> 提出用 PointNet++ 网络对点云中的目标进行分类与分割,解决了长期以来原始点云特征提取学习的瓶颈,但 PointNet++ 所需的计算复杂度太高,且适用于室内场景,不能满足室外场景的需求。国内学者对三维目标的特征提取提出了不同的三维处理模型<sup>[11-12]</sup>。2018 年, Qi 等<sup>[13]</sup> 又提出了轻量化、高精度的 Frustum PointNets 实现室外场景的三维多目标检测模型。

基于应用场景的相似性和任务同一性,本文沿用了与 Frustum PointNets 思路一致的网络结构。但校园场景中存在着与地图信息无关的目标,比如行人、车辆,在最终场景地图中需要对其进行检测以及剔除<sup>[14-15]</sup>。因此本文提出了一种在校园场景下检测和剔除多类目标的多源数据融合的目标检测与剔除网络 D-FPN (delete-frustum

PointNet)。

D-FPN 整体网络设计流程如图 6 所示,包括目标检测、目标剔除、单帧三维场景重建。下面简单说明图 6 中各模块的作用。

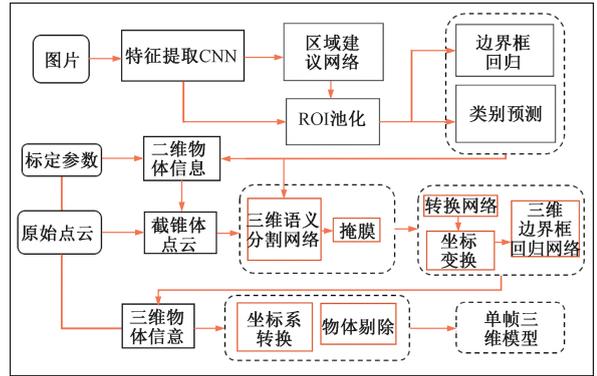


图 6 D-FPN 网络整体流程

Figure 6 D-FPN overall network flowchart

### 3.1 多源数据实现目标检测

图 6 中特征提取、区域建议网络、ROI 池化是一个类似于 Faster-RCNN 的二维图像目标检测网络,先使用特征提取卷积层来提取图像的特征,接着用区域建议网络获取感兴趣区域,ROI 池化用于修正位置,最终由边界框回归模块和类别预测模块来输出目标的位置和类别信息。本文所使用的训练数据为 KITTI 数据集<sup>[16]</sup>中的行人和车辆数据,二维图像目标检测模块经神经网络训练后,在测试集上对车辆识别的平均精准率为 94.3%,对行人识别的平均精准率为 86.7%,两类目标的平均精准率为 91%。图 7 所示为本文算法在 KITTI 验证集上对于行人和车辆的检测效果。



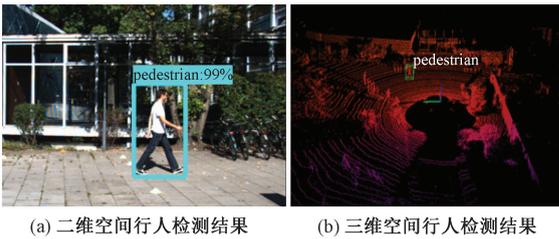
图 7 车辆行人检测

Figure 7 Vehicle and pedestrian detection

二维图像目标检测结果作为后一阶段网络的输入信息。在得到图像中目标所在的位置后,通过投影获取包含目标的截锥体点云,即得到一个小范围的点云场景。再使用 Pointnet 三维语义分割子网络完成对目标的实例分割 (semantic seg-

mentation),得到一个包含目标每个点的语义掩膜。通过一个转换网络和坐标变换,将所得的目标点云转换到坐标系的中心位置,方便接下来对点云进行边界框的回归。Pointnet 三维边界框回归子网络是点云的回归网络,需要对目标点云进行旋转、平移,得到目标点云在相机坐标系下的中心位置以及长、宽、高。

经过在 KITTI 数据集上训练后,本文三维点云目标检测模块在测试集上对车辆和行人识别的平均精准率分别为 79.3% 和 53.1%。网络的测试效果如图 8 所示,图 8(a)为二维空间行人检测结果,图 8(b)为三维空间中的行人在预测的三维包围框中。



(a) 二维空间行人检测结果 (b) 三维空间行人检测结果

图 8 特定场景下行人检测

Figure 8 Pedestrian detection in images

### 3.2 目标剔除

三维点云建模所得到的地图是整个校园内的场景,但其中存在着一些冗余的信息。图 9 中物体剔除目的是剔除校园场景地图中无关的信息,即可能出现的行人与车辆。

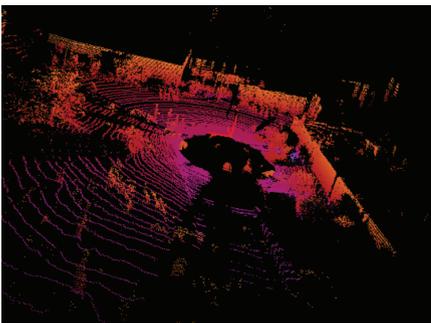


图 9 剔除目标后的单帧场景

Figure 9 Single scene after frame culling target

针对所检测到的与地图信息无关的目标,下一步需要将其在整个场景去除。由于二维图像中存在纹理连续和空间遮挡,无法单独将目标剔除,而三维点云中并不存在严格的遮挡与纹理,目标单独存在于三维空间中,因此可单独将目标分离出来。如图 8 所示,利用目标三维边界框信息,结合标定参数把相机坐标系中的信息经过刚体旋转变换到雷达坐标系中。最终把位于边界框内的目

标点归一化到原点,完成目标的剔除并得到单帧无目标场景的三维模型,再进行可视化。在图 8 的行人检测基础上,经目标剔除后的单帧激光雷达点云场景如图 9 所示。

## 4 场景重现

### 4.1 点云配准与融合

在实际的数据采集过程中,无法匹配的动态噪声点会通过建图过程消除掉,但是对于静态目标,必须结合建图和检测过程来去除。

对应多帧点云建图结果和单帧点云目标剔除结果具体实现思路有 2 种:第 1 种是先对每一帧场景进行处理,将处理后的单帧场景进行动态 SLAM 匹配建图,不过这样无法保证处理的帧率满足建图的实时性,导致存在丢帧问题,所构建地图误差太大无法使用;第 2 种思路是先完成建图过程,对所建的地图和单帧检测的结果进行配准,这样融合处理后的结果,能够保证场景的高效性与准确性。

本文基于第 2 种思路使用 LeGO-LOAM 进行初步建模,其坐标系与单帧所采数据并不一致,需要将不同坐标系下点云统一到同一坐标系下进行点云的配准<sup>[17-19]</sup>,如图 10 所示。在配准算法中,采用基于最小二乘的最优配准方法 ICP<sup>[18]</sup>算法,计算点云地图与单帧点云间的最优刚体变换矩阵,即 2 个坐标系下的刚体旋转矩阵  $[R/T]$ 。本文实验在 ROS 系统中进行,采用 PCL(point cloud library)点云库中的 ICP 点云配准算法工具,将单帧点云下目标检测与剔除结果和点云建模地图融合,最终实现整个场景的重现,并保存了无冗余信息的地图模型。

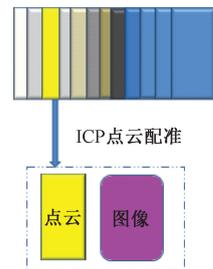


图 10 单帧与多帧点云间的配准

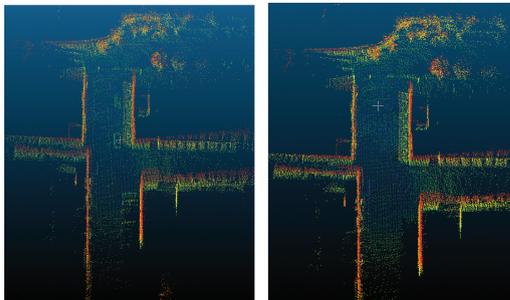
Figure 10 Match between single frame and multiframe point cloud

### 4.2 场景重现

本文实现了基于多源数据的三维场景重建系统,整个系统主要是基于 ROS 系统来完成。首先利用 ROS 将相机和雷达采集的数据传入系统并

保存,然后使用 LeGO-LOAM 点云建模算法完成对采集的离线数据的建图并保存地图模型。同时利用 D-FPN 算法检测每帧 RGB 图像和点云数据,将每一帧图像和点云及其在雷达坐标系下的目标边界框信息进行保存。调用 ROS 系统下 PCL 点云库中的 ICP 配准工具将离线点云地图和含目标的单帧点云进行一一配准得到转换关系并保存,最终在 ROS 中利用保存的点云地图、单帧点云、转换关系将检测到的三维目标统一转换到点云地图中,完成无冗余信息的三维场景重建,保存最终模型,为项目下一步研究做准备。

如图 11 所示为实际实验场景中的测试结果,图 11(a)为 LeGO-LOAM 算法建模结果,图 11(b)为系统输出结果,可以看出图 11(a)中的实验楼下停放的车辆目标被完整剔除了。



(a) LeGO-LOAM算法建模结果 (b) 系统输出结果

图 11 场景重建

Figure 11 Scene reconstruction

## 5 结论

基于搭建的硬件平台与软件框架,以激光雷达和广角相机采集的多源数据为基础,使用一种轻量级的 LeGO-LOAM 算法初步完成点云地图的重现。针对地图中与场景无关的数据,采用计算机视觉中多源传感器数据融合的目标检测方法对无关目标进行剔除,然后使用点云配准方法实现数据的融合,最终在特定校园场景中实现了无关目标剔除的三维场景重现,达到了预期的目标。未来在实际应用中,针对硬件实现的成本以及效率问题,下一步可能会通过加入 GPS、IMU 等不同类型的传感器来完成模型精度的提升。在算法设计层面,可通过多种算法集成的方式来节约成本。

## 参考文献:

[1] 张勤,贾庆轩.基于激光与单目视觉的室内场景三维重建[J].系统仿真学报,2014,26(2):357-362.  
[2] 辛俊伟,罗艳,程朋根.基于 LiDAR 技术的古建筑复杂曲面三维重建[J].测绘与空间地理信息,

2017,40(7):65-67.

- [3] 刘晋,禹鹏,白隽璋,等.基于三维激光扫描技术的犯罪现场重建[J].刑事技术,2017,42(6):476-481.  
[4] 李云翔.相机标定与三维重建技术研究[D].青岛:青岛大学,2009.  
[5] 曹明玮,钱烨强,王冰,等.基于监督式学习的全景相机与激光雷达的联合标定[J].机电一体化,2018,24(1):3-9,34.  
[6] ZHANG J, SINGH S. Low-drift and real-time lidar odometry and mapping[J]. Autonomous robots, 2017, 41(2):401-416.  
[7] DISSANAYAKE M W M G, NEWMAN P, CLARK S, et al. A solution to the simultaneous localization and map building (SLAM) problem[J]. IEEE transactions on robotics and automation, 2001, 17(3):229-241.  
[8] SHAN T X, ENGLLOT B. LeGO-LOAM: lightweight and ground-optimized lidar odometry and mapping on variable terrain[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). New York: IEEE, 2018:4758-4765.  
[9] CHARLES R Q, HAO S, MO K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2017:77-85.  
[10] QI C R, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space [EB/OL]. (2017-06-07) [2019-12-01]. <https://arxiv.org/abs/1706.02413>.  
[11] 张艺琨,唐雁,陈强.基于多特征融合的三维模型检索[J].郑州大学学报(工学版),2019,40(1):1-6.  
[12] LI Y Y, BU R, SUN M C, et al. PointCNN: convolution on x-transformed points[C]//32nd Conference on Neural Information Processing Systems. New York: ACM, 2018:820-830.  
[13] QI C R, LIU W, WU C X, et al. Frustum PointNets for 3D object detection from RGB-D data [C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018:918-927.  
[14] 庄加福.基于机载激光雷达的复杂场景车辆类目标检测[D].武汉:华中科技大学,2013.  
[15] 叶语同,李必军,付黎明.智能驾驶中点云目标快速检测与跟踪[J].武汉大学学报(信息科学版),2019,44(1):139-144,152.  
[16] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. The international journal of robotics research, 2013, 32(11):1231-1237.  
[17] LIU Y S, HU Z C, GAO S, et al. In situ analysis of

major and trace elements of anhydrous minerals by LA-ICP-MS without applying an internal standard [J]. *Chemical geology*, 2008, 257(1/2): 34-43.

- [18] RUSINKIEWICZ S, LEVOY M. Efficient variants of the ICP algorithm [C]// *Proceedings 3rd International*

*Conference on 3-D Digital Imaging and Modeling*. New York: IEEE, 2001: 145-152.

- [19] 戴静兰, 陈志杨, 叶修梓. ICP 算法在点云配准中的应用 [J]. *中国图象图形学报*, 2007, 12(3): 517-521.

## 3D Scene Reconstruction Based on Multi-source Sensor Data Fusion

CHEN Yifei<sup>1</sup>, GUO Sheng<sup>2</sup>, PUN Man-On<sup>2</sup>, LU Yanhui<sup>1,3</sup>

(1.School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China; 2.School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen 518172, China; 3.Shenzhen Research Institute of Big Data, Shenzhen 518172, China)

**Abstract:** Aiming at the target redundancy in reconstruction of certain scenes, in this paper a data fusion method of the camera RGB bitmap and lidar data was employed to solve the problem. In the field of 3D reconstruction, this method of data fusion could eliminate the irrelevant targets in the specific scene and reproduce the three-dimensional scene. The lightweight LeGO-LOAM algorithm was used to extract and match different types of feature points at first, and point clouds were merged at different times to complete the reproduction of the point cloud map. For the irrelevant targets in the constructed point cloud map, with the help of multi-source sensor data and deep learning application technology in the field of computer vision, object detection and elimination are accomplished in three-dimensional space. For the two different processes of point cloud modeling and target detection, the method of point cloud registration was adopted to fuse the algorithm and finally complete the scene reproduction in the campus environment. Experimental results showed that the method based multi-source data fusion could effectively combine the two processes of 3D modeling and object detection. The method proposed in this paper could be applied to smart cities, unmanned driving and other fields, and should have practical application value.

**Key words:** data fusion; 3D modeling; deep learning; object detection; feature matching; scene recurrence