

文章编号: 2095-2163(2023)03-0117-07

中图分类号: TP391

文献标志码: A

# 基于16线激光雷达和相机融合的三维目标检测

陈德江<sup>1</sup>, 江灏<sup>2</sup>, 高永彬<sup>1</sup>, 张家豪<sup>1</sup>

(1 上海工程技术大学 电子电气工程学院, 上海 201620; 2 上海振华重工电气集团有限公司, 上海 200125)

**摘要:** 针对16线激光雷达点云过于稀疏而无法进行三维目标检测的问题,设计了一种基于激光雷达和相机融合的三维目标检测方法。首先,对激光雷达和相机进行联合标定,统一两者坐标系,并利用ROS进行时间同步,确保雷达和相机在相同的时间感知到相同的环境。然后,将相机采集的图像作为输入,送入YOLOv5算法中进行二维的目标检测,得到目标在图像中的二维检测框和类别。最后,利用联合标定过的激光雷达获取目标二维像素点对应的三维点云,通过这些雷达点云算出目标的三维坐标和三维边界框,从而得到三维目标检测结果。实验验证:通过搭建硬件平台并在实验场地测试后证明了该方法的可行性。

**关键词:** 三维目标检测; 传感器融合; 联合标定; 16线激光雷达; 相机

## 3D target detection based on fusion of 16-line lidar and camera

CHEN Dejiang<sup>1</sup>, JIANG Hao<sup>2</sup>, GAO Yongbin<sup>1</sup>, ZHANG Jiahao<sup>1</sup>

(1 School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China;

2 Shanghai Zhenhua Heavy Industries Co., Ltd., Shanghai 200125, China)

**[Abstract]** Aiming at the problem that the 16-line lidar point cloud is too sparse for 3D target detection, a 3D target detection method based on the fusion of lidar and camera is designed. First, the lidar and camera are jointly calibrated, the coordinate systems of the two are unified, and ROS is used for time synchronization to ensure that the lidar and the camera perceive the same environment at the same time. Then, the image collected by the camera is used as input and sent to the YOLOv5 algorithm for 2D target detection, and the 2D detection frame and category of the target in the image are obtained. Finally, the jointly calibrated lidar is used to obtain the 3D point cloud corresponding to the 2D pixel points of the target, and the 3D coordinates and 3D bounding box of the target are calculated through these lidar points, so as to obtain the 3D target detection result. Experimental verification is: The feasibility of the method is proved by building a hardware platform and testing it in the experimental site.

**[Key words]** 3D target detection; sensor fusion; joint calibration; 16-line lidar; camera

## 0 引言

近年来,辅助驾驶和自动驾驶正日趋普及,而在其中起着关键作用的就是感知模块的发展。常用的传感器有激光雷达、单目相机、毫米波雷达等,但是单一传感器带来的信息量也是匮乏的,比如普通的单目相机得不出物体的深度信息,仅仅使用激光雷达又无法利用物体丰富的视觉信息。但是通过多传感器融合<sup>[1]</sup>不仅可以获取更丰富的感知信息,还可以通过传感器互补来降低硬件成本,比如利用激光雷达来做三维目标检测,需要至少64线的激光雷

达,因为如果雷达线数太少,则会因为目标点云过于稀疏而导致检测精度变差,但是如果能够将相机和激光雷达结合起来,就能在大幅降低成本的同时获取超过单一传感器的更高精度。

国内外针对激光雷达和相机融合的目标检测方法已经有了一定的进展,大多是基于深度学习方法。Chen等学者<sup>[2]</sup>提出的MV3D-Net<sup>[2]</sup>利用了点云的俯视图、前视图和相机的RGB图片,通过VGG16<sup>[3]</sup>做特征提取后融合在一起进行目标检测。Ku等学者<sup>[4]</sup>提出的AVOD<sup>[4]</sup>利用雷达点云的俯视图和相机的RGB图作为输入,然后利用FPN(Feature Pyramid

**基金项目:** 国家自然科学基金(61802253)。

**作者简介:** 陈德江(1997-),男,硕士研究生,主要研究方向:三维目标检测、深度学习;江灏(1981-),男,电气工程师,主要研究方向:港口自动控制;高永彬(1988-),男,博士,副教授,主要研究方向:人工智能、三维视觉;张家豪(1996-),男,硕士研究生,主要研究方向:图像处理、机器视觉。

**通讯作者:** 江灏 Email:jianghao@zpmc.com

**收稿日期:** 2022-05-04

Network)进行特征提取,同样获得了精确的检测结果。Qi等学者<sup>[5]</sup>提出的Frustum-PointNet先利用RGB图像中目标的二维边界框进行点云预处理,将对应的雷达点云分离出来,然后将这些属于对应目标的点云输入到PointNet<sup>[6]</sup>中进行特征提取,从而回归出三维目标检测结果。但是这些方法都是依赖于64线的激光雷达,当前64线的激光雷达价格昂贵,对于一些大型车辆,如轮胎吊大车、重型卡车等,如果想要激光雷达全覆盖,就需要至少2~4个激光雷达,这将带来昂贵的安装成本。

本文考虑到64线激光雷达的局限性,基于16线激光雷达和相机两种传感器,设计了一种三维目标检测方法。首先,将雷达和相机进行空间和时间上的联合标定,获取相互间的刚体变换矩阵,通过刚体变换统一坐标系后再将雷达点云按照相机的内参矩阵投影到图片上,获取二维像素点和三维雷达点云之间的对应关系。然后,将相机图片送入YOLOv5算法进行二维目标检测<sup>[7]</sup>,得出目标的二维像素范围,利用联合标定过的雷达获取这些二维像素对应的三维点云,最后利用这些三维点云算出目标的三维坐标和边界框。通过实际场景测试证明了该方法的可行性。

## 1 硬件平台的搭建和标定

硬件的合理安装和精准标定是算法有效运行的基础,本文设计的实验平台如图1所示。图1中,左侧为16激光雷达,右侧为球形相机,中间通过钢架进行固定。因为雷达内包含转子,在雷达运行的过程中会产生较强的晃动,所以必须做好固定措施。为了能够适应室外天气,采用带球罩的相机,方便遮风挡雨,而且能够自由调整方向,但是球形相机畸变较为严重,所以需要做好相机的标定工作<sup>[8]</sup>。在进行相机标定时,需要采集各个角度的标定板图片再进行标定,这样标定出的相机内参更精确。

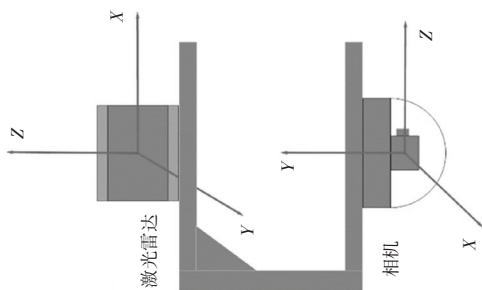


图1 硬件平台设计图

Fig. 1 Hardware platform design diagram

多传感器融合除了进行空间上的同步,还需要

进行时间上的同步。空间上的同步即多传感器联合标定统一坐标系,时间上的同步是指同步传感器获取数据的时间戳,保证2个传感器是在相同的时间感知到同样的物体。

### 1.1 激光雷达和相机的联合标定

激光雷达和相机有着各自的坐标系<sup>[9]</sup>,参见图1中的坐标轴。其中,激光雷达的X轴朝前,Y轴朝左,Z轴朝上。相机的X轴朝右,Y轴朝下,Z轴朝前,需要通过联合标定的方式算出2个坐标系之间的刚体变换矩阵。刚体变换的公式为:

$$RP_i + t = P_c \quad (1)$$

其中, $R$ 为 $3 \times 3$ 的旋转矩阵; $t$ 为 $3 \times 1$ 的平移向量; $P_i$ 为激光雷达下物体的坐标; $P_c$ 为相机坐标系下物体的坐标。

本文采用的多传感器联合标定原理是让2个传感器同时感知到相同的三维点,这些物体在不同的传感器里有各自的坐标,可以通过多组对应点构成方程组,但是这些点在数据采集时不一定是绝对精准的,可能会因为各种原因产生误差,所以不存在绝对解,而是存在一个近似的最优解,让标定后的传感器感知误差最小,误差表达方式为:

$$err = \sum_{i=1}^N \| RP_i^i + t - P_c^i \|^2 \quad (2)$$

故而可以采用类似求解点云配准的方式利用SVD解出2个传感器之间的刚体变换,其中旋转矩阵 $R$ 的求解方法为:

$$\begin{aligned} H &= (A - \text{centroid}_A)(B - \text{centroid}_B)^T \\ [U, S, V] &= \text{SVD}(H) \\ R &= VU^T \end{aligned} \quad (3)$$

其中, $A$ 为激光雷达在激光雷达的坐标系中感受到的一组点; $B$ 为相机在相机坐标系中感受到的同一组点; $\text{centroid}$ 为对应点群的中心。求解出 $R$ 后就可以进一步求出 $t$ ,求解方式见下式:

$$\begin{aligned} R \times A + t &= B \\ R \times \text{centroid}_A + t &= \text{centroid}_B \\ t &= \text{centroid}_B - R \times \text{centroid}_A \end{aligned} \quad (4)$$

求解出旋转矩阵 $R$ 和平移向量 $t$ 后,就可以直接将雷达的坐标系直接转换到相机的坐标系,这时雷达扫描出的点云在经过变换后将处于相机的坐标系中。

### 1.2 激光雷达和相机的时间同步

不同种类的传感器运行频率也并不相同,相机在25 Hz左右,激光雷达在20 Hz左右。传感器除了扫描频率,数据传输速率也不一样,所以相同时间

发生的事情从传感器到达计算机的时间很可能并不一致,数据存在时间错位,如图2所示,所以需要给传感器产生的数据打上时间戳,并且采用合理的同步算法进行数据同步。

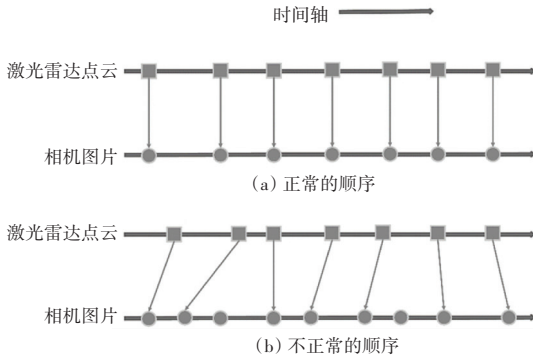


图2 传感器数据同步

Fig. 2 Sensor data synchronization

本文采取的方式是利用 ROS 中的 Message Filters 进行时间同步。ROS 是最常用的机器人操作系统之一,将传感器数据抽象成 Topic 进行数据的发布和订阅,并通过消息队列的方式对数据进行缓存。Message Filters 同时订阅激光雷达驱动和相机驱动发布的 Topic,但只有在两者时间戳满足时间差要求时才会触发回调函数对传感器数据进行处理,而过期或者时间戳相差较大的数据则会舍弃掉。这样就可以保证雷达和相机是在相同的时间感知到相同的环境。

## 2 算法实现

本文主要应用场景是基于激光雷达和相机两种传感器融合起来在大车前进的车道线内进行三维目标检测,将检测的结果反馈给大车的控制中心。通过相机进行车道线检测和二维目标识别,并通过雷达赋予目标三维信息,从而得到目标的三维坐标和边界框,对具体算法实现流程,可给出研究论述如下。

### 2.1 车道线检测

本文针对车道线内的兴趣区域进行目标检测,所以先通过相机进行车道线检测。当前对于车道线检测已经有了很多有效的方法,但是大多基于深度学习的方式,相对复杂。而本文场景中的车道线具有一定的规律,不存在弯曲的部分,所以可以通过霍夫变换算法进行车道线提取<sup>[10]</sup>,算法流程如图3所示。

图3中,通过相机采集到的 RGB 格式的图片为

图3(a),然后将图片从 RGB 域转到 HSV 域得到图片,即图3(b),进行转换主要是为了根据可行驶区域的颜色过滤出可能是车道线的部分,有助于排除干扰信息。在本算法中,过滤条件为  $[H, S, V]$  的数值介于  $[20, 60, 30]$  和  $[40, 200, 255]$  之间,阈值设置范围相对较广能够保证算法在各种光照下的检测鲁棒性。接下来再通过 Gaussian 平滑滤波和 Candy 滤波得到可能是车道线的轮廓图,见图3(c),根据得到的轮廓图使用霍夫变换算法,对轮廓图中的直线进行检测,直到找到所有的直线。由于找到的直线不一定属于车道线,需要根据直线的长短、斜率等信息进行分类。在这里,先使用 K-means 聚类算法将直线分为2类,一类属于左侧车道线,一类属于右侧车道线,最后再根据左、右侧车道线的斜率特征做进一步过滤。将两侧车道线的集合取均值就得到了完整的车道线表达式,处理后的车道线检测效果见图3(d)。

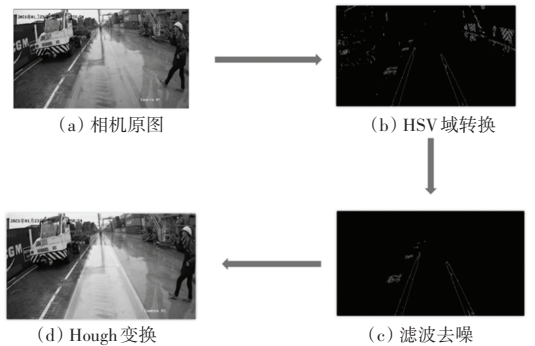


图3 车道线检测算法流程

Fig. 3 Lane detection algorithm flow chart

### 2.2 二维目标检测

确定兴趣区域后,再检测兴趣区域内的目标,本文需要精度高、运行速度快的二维目标检测算法。YOLO 系列是具有高性能的二维目标检测算法,本文采用最新版本的 YOLOv5。YOLOv5 实现了端到端的目标检测,输入图片即可输出检测结果,检测剪度高、且检测速度快。YOLOv5 的算法网络结构<sup>[11]</sup>如图4所示。

图4中,主要包括骨干网络、颈部网络和检测头。骨干网络通过 CSPNet 结构减少 GPU 显存占用,避免重复的梯度信息,加速模型训练和运行,让模型更加轻量 and 高效,其检测效果如图5所示。除了较高的检测精度和极快的运行速度,YOLOv5 本身基于 COCO<sup>[12]</sup>数据集进行训练,有着丰富的检测种类和预训练模型,本文的应用场景主要检测人、汽



车和卡车三种目标,这3类目标现已存在于预训练模型的检测范围中。针对部分车辆分类不准确的问题,

通过采集和定制当前场景下的数据集,基于预训练模型还可以做进一步的优化与完善,提升准确率。

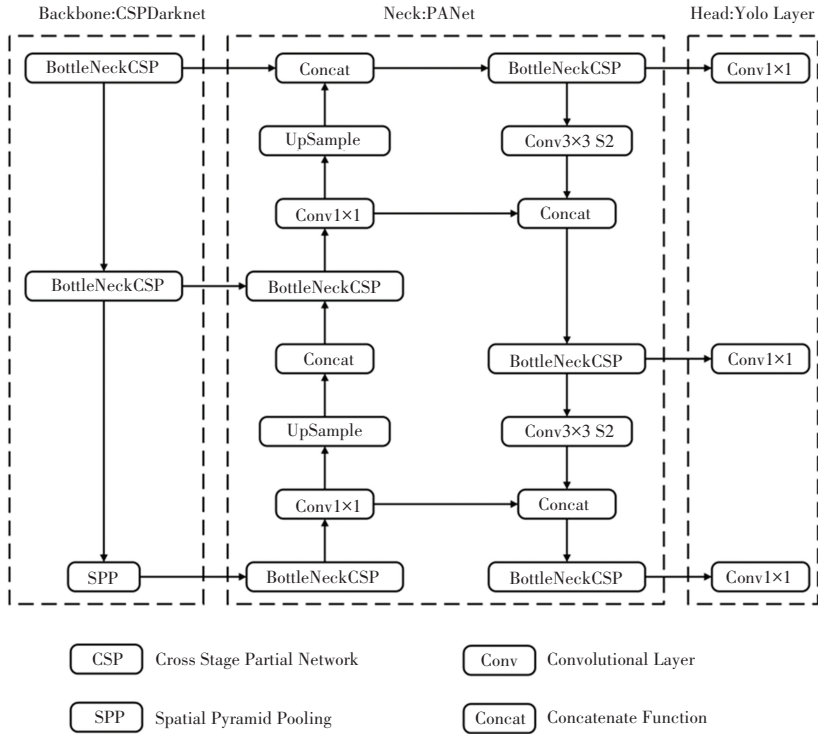


图4 YOLOv5网络结构图

Fig. 4 YOLOv5 network structure diagram



(a) 行人检测

(b) 汽车检测

图5 YOLOv5检测效果

Fig. 5 YOLOv5 detection effect

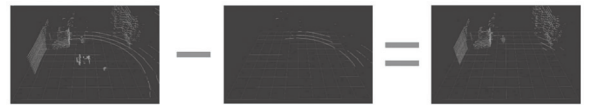
### 2.3 激光雷达地面点云去除

激光雷达在扫描时有很大一部分点云出自地面,这部分点云对于本文的应用场景属于无效信息,并且会拖慢算法运行速度,因此需要过滤掉。常用的地面去除算法是基于RANSAC<sup>[13]</sup>的平面拟合算法,通过多次尝试将地面所处的平面拟合出来,再进行去除,但是计算量较大,而且本文应用场景中有很多平面高台,所以并不适合在该场景中应用。本文采用栅格立柱过滤法<sup>[14]</sup>。方法中,将点云空间按照栅格划分,栅格竖起来形成立柱,立柱中的最高点为 $H_{ij}$ ,最低点为 $h_{ij}$ ,这里的 $ij$ 为立柱的索引值,如果该立柱满足布尔表达式:

$$H_{ij} - h_{ij} < s \quad (5)$$

其中,  $s$  为高度阈值,即立柱中的高度差不超过

阈值,则该立柱中的点云为地面。如果超出阈值,则说明该立柱中的点云有较大的起伏,可能属于路面上的物体。地面点云去除算法的效果如图6所示。



(a) 全部点云

(b) 地面点云

(c) 有效点云

图6 地面点云去除效果

Fig. 6 Ground point cloud removal effect

由图6可知,图6(a)为16线激光雷达扫描环境后输出的全部点云,图6(b)为算法检测出的地面部分,将其剥离之后可以得到非地面部分的有效点云,见图6(c)。

### 2.4 三维目标检测

相机在成像的过程中丢失了深度信息<sup>[15]</sup>,所以通过相机进行目标检测只能获取二维的结果,即只能得出目标在图像中的位置和目标的种类,但是无法知道目标的实际大小和实际的三维坐标。研究可知,通过联合标定的激光雷达可以补足这个缺陷。首先将激光雷达投影到相机成像平面上并保留对应关系,投影算法为:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \stackrel{def}{=} \frac{1}{Z} \mathbf{K} \mathbf{P} \quad (6)$$

其中,  $P$  为雷达中的三维点, 用  $X, Y, Z$  表示三维坐标, 经过相机内参  $\mathbf{K}$  的转换后得到像素坐标  $u, v$ , 即在图像中的坐标, 这样就可以由每个雷达反射点找到在图片中的位置, 同时也可以通过像素点得到对应雷达反射点的三维空间坐标, 即可以获取二维像素点和三维雷达点云的映射关系。通过激光雷达和相机的联合标定可以在将雷达点云变换到相机坐标系后, 接着再利用相机标定得到的内参矩阵  $\mathbf{K}$ , 这样就具备了将雷达点云投影到相机图片上的所有条件, 投影效果如图 7 所示。

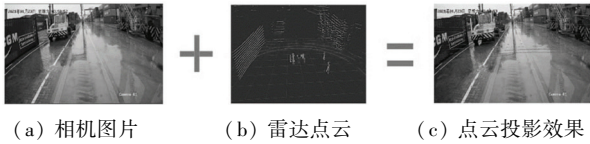


图 7 激光雷达和相机联合标定后的投影效果

Fig. 7 Projection effect after joint calibration of the lidar and the camera

至此, 通过将二维目标检测阶段的结果作为三维目标检测阶段的输入, 将 2D Bounding Box 中的像素点转化成三维的雷达点云, 对这部分雷达点云进行 3D Bounding Box 求解<sup>[16]</sup>, 计算步骤如下:

首先, 计算点云的质心  $P_c$ , 质心是指物体质量的中心, 但是点云是没有质量的, 所以这里假设点的质量为单位 1 即可, 得到点云质心坐标为  $(c_0, c_1, c_2)$ , 计算公式为:

$$P_c = (c_0, c_1, c_2) = \frac{1}{n} \left( \sum_{i=0}^n x_i, \sum_{i=0}^n y_i, \sum_{i=0}^n z_i \right) \quad (7)$$

然后, 计算出点云的协方差, 其中  $P$  已经减去了质心  $P \leftarrow P - P_c$ , 点云归一化后的协方差矩阵为:

$$\mathbf{C} = \begin{pmatrix} cov(x, x) & cov(x, y) & cov(x, z) \\ cov(y, x) & cov(y, y) & cov(y, z) \\ cov(z, x) & cov(z, y) & cov(z, z) \end{pmatrix} \quad (8)$$

最后, 求解出协方差矩阵的特征向量  $(e_0, e_1, e_0 \times e_1)$ , 结合点云信息和特征向量即可解出 3D Bounding Box 的所有信息, 求解公式为:

$$\begin{aligned} \mathbf{R} &= quaternion(e_0, e_1, e_0 \times e_1) \\ D_c &= \frac{1}{2} (P_{max} + P_{min}) \\ \mathbf{t} &= \mathbf{R} \times D_c + (c_0, c_1, c_2) \end{aligned} \quad (9)$$

其中,  $\mathbf{R}$  为 Bounding Box 的旋转;  $P_{max}$  和  $P_{min}$  为点云的极大值和极小值;  $D_c$  为对角线的中点;  $\mathbf{t}$  为

Bounding Box 的平移向量; Bounding Box 的宽、高、深度分别为  $x_{max} - x_{min}, y_{max} - y_{min}$  和  $z_{max} - z_{min}$ 。由此得到了 3D Bounding Box 所需的所有信息, 实现了二维检测到三维检测的升维。

### 3 实验测试

实验平台硬件主要由 16 线激光雷达和球形相机构成, 使用的工控机配置为 Intel(R) Core(TM) i7-8700 CPU @ 3.20 GHz, RTX2060 GPU, 32 G 内存。实验硬件平台与安装位置如图 8 所示。



图 8 硬件平台与安装位置

Fig. 8 Hardware platform and installation location

实验内容为: 将传感器硬件平台安装在大车前方 1.5 m 高处, 大车本身行驶在车道线内, 当传感器在车道线内前方 20 m 内检测到障碍物时, 需要及时将信号传递给大车的控制中心, 控制中心根据障碍物的种类和距离进行告警和制动。

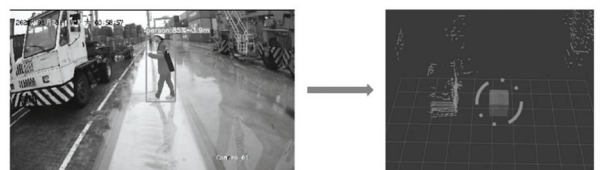
通过设置实验测试场景进行测试, 测试目标包括行人、汽车和卡车三个种类, 让 3 个种类的目标分布在车道线内的不同距离和位置, 测试结果见表 1。

表 1 实验检测精度

Tab. 1 Experimental detection accuracy

检测目标	二维目标检测精度 mAP/ %	三维目标检测误差 $F_2$ / m
行人	86.7	0.2
汽车	87.1	0.5
卡车	85.5	0.8

检测效果可视化如图 9 所示。图 9 中, 左侧为车道线检测和二维目标检测的可视化效果, 右侧为雷达点云视角中的三维目标检测效果。



(a) 二维检测结果 (b) 三维检测结果

图 9 目标检测效果可视化

Fig. 9 Visualization of target detection effect

此外,根据自动化港口码头需求,开发了一套后端系统及 UI 界面。其中,后端系统基于 ROS 操作系统开发,将消息通讯抽象为 Topic,满足实时性要

求。UI 界面如图 10 所示,通过相机和雷达的 IP 进行控制,当在车道线上检测到障碍物时进行告警,并且将告警信息记录在日志中,方便后期查询。



图 10 三维目标检测报警系统 UI

Fig. 10 UI of 3D target detection alarm system

## 4 结束语

本文针对 16 线激光雷达输出点云过于稀疏而无法进行三维目标检测的问题,提出了一种基于 16 线激光雷达和相机融合的三维目标检测方法。通过联合标定从空间上统一了激光雷达和相机的坐标系,并利用 ROS 从时间上对齐了激光雷达和相机的数据。接下来,采用相机的三维投影算法在二维像素和三维点云之间建立了映射关系。最后,对目标点云进行包围框的计算,将二维检测结果升级至三维。实验结果表明:该方法可以有效检测车道线内的目标,理论上可以满足大型车辆三维目标检测的需求。

## 参考文献

- [1] GRAVINA R, ALINIA P, GHASEMZADEH H, et al. Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges[J]. Information Fusion, 2017, 35: 68-80.
- [2] CHEN Xiaozhi, MA Huimin, WAN Ji, et al. Multi-view 3d object detection network for autonomous driving [C]//Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA; IEEE, 2017: 1907-1915.
- [3] SENGUPTA A, YE Yuting, WANG R, et al. Going deeper in spiking neural networks: VGG and residual architectures [J]. Frontiers in Neuroscience, 2019, 13: 95.
- [4] KU J, MOZIFIAN M, LEE J, et al. Joint 3d proposal generation and object detection from view aggregation[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, Spain; IEEE, 2018: 1-8.

- [5] QI C R, LIU Wei, WU Chenxia, et al. Frustum pointnets for 3d object detection from rgb-d data[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA; IEEE, 2018: 918-927.
- [6] QI C R, SU Hao, MO Kaichun, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii; IEEE, 2017: 652-660.
- [7] ZHAO Zhongqiu, ZHENG Peng, XU Shoutao, et al. Object detection with deep learning: A review[J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, 30(11): 3212-3232.
- [8] RUSU R B, COUSINS S. 3d is here: Point cloud library (pcl) [C]//2011 IEEE International Conference on Robotics and Automation. Shanghai, China; IEEE, 2011: 1-4.
- [9] PUSZTAI Z, HAJDER L. Accurate calibration of LiDAR-camera systems using ordinary boxes [C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. Venice, Italy; IEEE, 2017: 394-402.
- [10] DU Xinxin, TAN K K. Vision-based approach towards lane line detection and vehicle localization [J]. Machine Vision and Applications, 2016, 27(2): 175-191.
- [11] KUZNETSOVA A, MALEVA T, SOLOVIEV V. Detecting apples in orchards using YOLOv3 and YOLOv5 in general and close-up images [C]//International Symposium on Neural Networks. Cham; Springer, 2020: 233-243.
- [12] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft coco: Common objects in context [C]//European Conference on Computer Vision. Cham; Springer, 2014: 740-755.
- [13] ISA S N M, SHUKOR S A A, RAHIM N A, et al. Point cloud data segmentation using ransac and localization [J]. IOP Conference Series: Materials Science and Engineering, 2019, 705(1): 012004.