DOI: 10.3901/JME.2020.12.165

基于信息融合的城市自主车辆实时目标识别*

薛培林¹ 吴愿¹ 殷国栋¹ 刘帅鹏¹ 林乙衡² 黄文涵¹ 张云³

(1. 东南大学机械工程学院 南京 211189;

2. 陆军特种作战学院 桂林 541002;

3. 福特汽车工程研究(南京)有限公司 南京 211100)

摘要: 针对单一传感器感知维度不足、实时性差的问题,提出一种激光雷达与相机融合的城市自主车辆实时目标识别方法。建立两传感器间的坐标变换模型,实现两传感器的像素级匹配。改进 yolov3-tiny 算法,提高目标检测准确率。对激光雷达点进行体素网格滤波,根据点云坡度进行地面分割。建立聚类半径与距离作用模型,对非地面点云进行聚类。引入图像中包络的思想,获取目标三维边界框以及位姿信息;将视觉目标特征与激光雷达目标特征融合。试验结果表明,改进的 yolov3-tiny 算法对于城市密集目标具有更高的识别率,雷达算法能够完整的完成三维目标检测以及位姿估计,融合识别系统在准确率、实时性方面达到实际行驶要求。

关键词: yolo; 聚类; 位姿估计; 激光雷达; 视觉; 信息融合

中图分类号: TG156

Real-time Target Recognition for Urban Autonomous Vehicles Based on Information Fusion

XUE Peilin¹ WU Yuan¹ YIN Guodong¹ LIU Shuaipeng¹ LIN Yiheng²
HUANG Wenhan¹ ZHANG Yun³

(1. School of Mechanical Engineering, Southeast University, Nanjing 211189;

2. The Army Special Operations Academy, Guilin 541002;

3. Ford Motor Research & Engineering (Nanjing) Co., Ltd., Nanjing 211100)

Abstract: Aiming at the problem that the single sensor has insufficient sensing dimensions and poor real-time performance, a real-time target recognition method for urban autonomous vehicles based on the fusion of lidar and camera is proposed. To achieve pixel-level matching of the two sensors, a coordinate transformation model between the two sensors is established; the yolov3-tiny algorithm is improved to increase the accuracy of target detection. Voxel grid filtering was performed on the lidar points, the ground is filtered according to the lidar point slope; the model of clustering radius and distance is established, and non-ground point clouds are clustered; the idea of envelope in images is introduced to obtain the 3D bounding box and pose information of the target; the visual target features are fused with the lidar target features. The experimental results show that the improved yolov3-tiny algorithm has a higher recognition rate for dense urban targets. The lidar algorithm can complete three-dimensional target detection and pose estimation. The fusion recognition system meets the actual driving requirements in terms of accuracy and real-time performance.

Key words: yolo; cluster; pose estimation; lidar; vision; information fusion

0 前言

智能汽车在减少交通事故、缓解交通拥堵、提

高道路及车辆利用率等方面具有巨大潜能,其主要技术可分为感知、决策、控制三大部分。目标识别作为智能汽车感知中的一项基础性工作,对于保障自主车辆安全,提高智能汽车环境理解力具有重要意义,近年来已逐渐成为研究热点。

图像的目标识别技术主要可分为传统机器视觉、深度学习两种^[1]。传统机器视觉主要是

* 国家重点研发计划(2016YFD0700905)、国家自然科学基金(51975118, U1664258)、江苏省重点研发计划(BE2019004-2)和江苏省成果转化项目(BA2018023)资助项目。20191105 收到初稿,20200305 收到修改稿

HOG(Histogram of oriented gradients)+ SVM(Support vector machine)以及上述方法的诸多改进、优化,但存在泛化能力差,识别率低^[2-3]的问题。随着深度学习的兴起,使用深度学习的目标检测方法在精度和速度上都超过了许多传统方法。使用深度学习进行图像目标检测有两种方法。一种方法是基于区域建议,如 Faster R-CNN^[4-6]。该方法首先通过一些卷积层运行整个输入图像以获得特征映射。然后,有一个单独的区域建议网络,使用这些卷积特征来提出可能的检测区域。最后,网络的其余部分给出了这些建议区域的分类。由于网络中有两个部分,一个用于预测边界框,另一个用于分类,这种架构会显著降低处理速度。另一种方法是基于深度学习的回归方法如 yolo(You only look once)、SSD(Single shot multibox detector)等^[7-10]。给定一个输入图像,yolo 首先将图像划分为粗网格。对于每个网格,都有一组基本边界框。对于每个基本边界框,如果 yolo 认为该网格位置中存在对象,则预测偏离真实位置的偏移,置信度得分和分类得分^[7-9]。该方法实时性好,但牺牲了一定的边界框精度。

激光雷达能够获取驾驶环境的三维信息,在障碍物检测、目标测距、测速等方面具有无可比拟的优势。动态障碍物检测与跟踪是激光雷达领域的研究热点,许多学者对此进行了大量的研究^[11-13]。AZIM 等^[12]提出了比值特征法来区分移动障碍物,然而仅使用了数值来判断对象的类型,当区域点云数据稀疏或检测区域被阻塞时,可能会导致较高的漏报率。周俊静等^[13]在确定特征信息后,采用基于距离的车辆聚类算法,基于多特征信息融合对车辆进行识别,并采用确定性方法进行目标关联。然而,多特征信息融合过程繁琐,规则不明确,相关方法无法处理目标的出现和消失。

上述研究均采用单传感器进行目标检测,但是彩色相机的图像信息会受到环境光的影响,而激光雷达又难以获取环境目标色彩语义信息。因此,在复杂的城市交通环境中,单传感器的性能和识别精度较低,不能满足自动驾驶车辆的安全需求。

为了适应交通环境的复杂性和多变性,一些研究使用彩色相机和激光雷达同时检测自动驾驶车辆上的目标,然后通过融合方法为车辆提供足够的环境信息。ASVADI 等^[14]设计了组合 3D Lidar 输出的密集深度图和密集反射图以及从相机输出的彩色图像的三个检测器,基于三个检测器使用卷积神经网络方法提取障碍物信息。XUE 等^[15]提出以视觉为中

心的多传感器融合框架,该框架通过几何约束和语义约束,将相机、激光雷达等信息融合在一起,实现高效的自主定位和障碍物感知,但多传感器融合的过程和算法过于复杂,难以满足实时的要求。

综上所述,目前已有的解决方案仍存在目标检测率低、实时性差的问题,因此提出一种基于激光雷达与视觉融合的实时目标检测方法。主要工作如下:① 在原有 yolov3-tiny 网络基础上,引入 focal loss,提高城市密集目标的检测率;② 改进 DBSCAN(Density-based spatial clustering of applications with noise)聚类算法,提出聚类半径与距离作用模型,提高聚类准确性;③ 在激光雷达目标检测中,引入图像包络原理,完成目标的实时位姿估计;④ 设计的激光雷达与视觉融合实时感知系统,在轻量化自动驾驶平台上的运行频率 10 Hz 左右,满足城市自主车辆实时性要求。

本文的结构安排如下:第 1 节为系统概述,第 2 节为图像与雷达的配准、同步,第 3 节为基于视觉的目标识别,第 4 节为基于雷达的目标检测,第 5 节为基于特征的数据融合,第 6 节为试验部分,第 7 节为结论。

1 系统概述

1.1 硬件系统

课题组采用 40 线激光雷达,该激光雷达中间部分线束更加密集,对于远方目标识别效果更优秀,水平角分辨率可根据需求,选择 0.2° 或 0.4° 。图 1 为雷达线束分布图。

图 1 激光雷达线束分布

雷达垂直视野主要由安装高度和倾斜角度决定。结合车辆物理空间的限制以及城市目标识别的特征,确定安装高度 $H=1.8$ m。根据场景的需要,相机选用普通 USB 相机,像素为 640×480 ,更小的分辨率提高了图像的处理速度。彩色相机安装在顶部激光雷达下,相机的位置尽量水平,以便相机水平角度和旋转近似为 0° 。

最终的传感器布置方案如图 2 所示。

图2 传感器布置

1.2 软件系统

软件系统以机器人操作系统(Robot operation system, ROS)为基础设计^[16]。Tx2 嵌入式开发板负责实时处理相机数据,完成视觉目标识别。工控机负责接收原始点云数据,对原始点云数据完成滤波、降采样等预处理任务,然后基于改进DBSCAN 聚类,以及基于图像包络的位姿估计获取环境目标的三维信息。最后,基于特征的融合策略完成激光雷达与视觉实时融合的感知。系统软件框架如图3所示。

图3 系统软件框架

2 配准与同步

2.1 坐标系定义

建立传感器的自身坐标系,确立不同传感器之间的转换关系,各传感器的建立如图4所示。

图4中, $o_c x_c y_c z_c$ 为建立的相机坐标系,相机的光心作为相机坐标系的原点 o_c ,光轴的前方定义为 z 轴的正方向,下方定义为 y 轴的正方向,车体右侧定义为 x 轴正方向。为了简化计算,在这里将世界坐标系和雷达坐标系设为一致,即 $o_1 x_1 y_1 z_1$ 。激光雷达坐标系定义为,激光雷达底部为坐标原点,雷达上方为 z 轴正方向,前方为 x 轴正方向,左侧为 y 轴正方向。所有坐标系均符合右手准则。

2.2 空间同步

激光雷达和相机输出的数据必须在时间和空间上同步,才能实现两者的信息融合。

传感器的标定即空间同步,主要为获取相机内

图4 传感器坐标系

部参数和外部参数。相机内部参数主要包括相机焦距、像素中心点、畸变系数等。外部参数包括相机坐标系相对于雷达坐标系的旋转和平移。

设点在雷达坐标系中坐标为 $(x_1, y_1, z_1)^T$ ，对应的相机坐标系中的点坐标为 $(x_c, y_c, z_c)^T$

$$P_c = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = R \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + T = RP_1 + T \quad (1)$$

式中， R 和 T 分别为相机坐标系相对于雷达坐标系的旋转与偏移。该点最后成像到像素中坐标为 $(u, v)^T$ 。

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = KP_c = K \left(R \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + T \right) = K(RP_1 + T) \quad (2)$$

式中， K 称为相机内参。

相机内参可通过张正友标定法求出。在求解相机外参方法中，LEPETIT 等^[17]提出了一种非迭代 EPnP 算法，该算法将三维点表示为四个虚拟控制点的加权和，问题简化为估计相机坐标系中这些控制点的坐标，实现在不需要额外时间的情况下提高精度。

基于此算法，通过提取激光雷达和图像中相对应的特征点作为 EPnP 算法输入，共提取 36 组特征数据，完成外参获取。图 5 为特征点提取，图 5a 为图像特征点提取，图 5b 为激光雷达特征点提取。

三维点首先从雷达坐标系变换到相机坐标系，最后成像到像素平面，此过程即可完成像素级融合。

2.3 时间同步

自动驾驶车辆在实际探测周围环境时，不同时间环境信息可能千差万别。因此要确保不同传感器采集到的信息为同一时刻的，才能进行融合。

图5 特征点提取

为激光雷达和相机分别创建一个数据采集线程，同时以采样频率低的雷达为准，设置相同的采集帧速率为 10 帧/s，在数据处理前进行时间戳匹配，实现了时间上的数据匹配。

3 基于视觉的目标识别

3.1 Yolov3-tiny 网络结构

Yolov3 算法是由 REDMON 等^[7-9]提出的通用目标检测模型。在基础的图像特征提取方面，yolov3 借鉴残差网络(Residual network)的做法，在一些层之间设置了快捷链路连接，在网络层数加深的情况下也能够提升算法分类的 mAP(Mean average precision) 值以及对小目标检测的准确率。yolov3-tiny 是 yolov3 模型的简化版，该算法融合了最新的特征金字塔网络(Feature pyramid networks, FPN)，模型结构简单，计算量小，检测精度高，并且可以在移动端或者设备端运行^[18]。网络结构如下图 6 所示。

图6 Yolov3-tiny 网络结构

Yolov3-tiny 主要由卷积层和池化层组成。图片像素的输入尺寸为 416×416 ，其主干网络采用 7 层卷积加池化网络提取特征，分类网络采用的是 13×13 、 26×26 分辨率的探测器。

3.2 Focal loss

Yolov3-tiny 使用基于单级回归的策略，运行速度极快，同时也导致算法检测到的边界框精度低、对目标的分类准确度不高、对密集目标检测的效果差^[18]等结果。城市目标的特点正是行人车辆密集，

为此，需要一种针对密集目标检测的改进策略。

现有研究表明，单级识别精度低于二级识别主要是由样本不均衡导致^[19]。在一张输入图片中，目标占的像素比例一般都远小于背景所占的比例，所以在二分类样本中是以负样本为主，由此造成负样本的损失太大，不利于目标的收敛，同时过多的负样本可能会导致模型优化方向并不是我们想要的。

2016 年，SHRIVASTAVA 等^[20]提出 OHEM (Online hard example mining) 算法，通过对损失排序，选出损失

最大的样本进行训练，保证训练的区域都是困难样本。但是该方法是将所有的容易样本都去除了，造成容易的正样本无法进一步提升训练网络的精度。2018 年，LIN 等^[19]针对此问题提出 focal loss 损失函数，通过减少易分类样本的权重，使得模型在训练时更专注于难分类的样本。借鉴此思路，主干网络保持 yolov3-tiny 不变，在分类中引入 focal loss，保持其运算速度前提下，提高对城市密集目标的识别精度。以二分类为例，标准的交叉熵(Cross entropy, CE)损失函数为

$$CE(p, y) = \begin{cases} -\log(p) & y=1 \\ -\log(1-p) & \text{其他} \end{cases} \quad (3)$$

式中， $y \in \{\pm 1\}$ 为真实标签； $p \in [0, 1]$ 是模型对于类别 $y=1$ 所得的预测概率。定义

$$p_t = \begin{cases} p & y=1 \\ 1-p & \text{其他} \end{cases} \quad (4)$$

则， $CE(p, y) = CE(p_t) = -\log(p_t)$ 。

Focal loss 在标准交叉熵损失函数的基础上增加了调制因子，其公式为

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t) \quad (5)$$

式中， $\gamma \geq 0$ ； α_t 为权重因子，其定义为

$$\alpha_t = \begin{cases} \alpha & y=1 \\ 1-\alpha & \text{其他} \end{cases} \quad (6)$$

3.3 改进措施

将 focal loss 引入到 yolov3-tiny 的分类层中，网络算法前向传播的结果均需要经过 Softmax 函数激活以后，再经过交叉熵损失函数来计算预测标签的概率分布与真实标签的概率分布之间的距离即损失函数值，并且在网络训练过程中，通过算法的反向传播不断降低这个损失函数值，从而得到与真实值最接近的预测值。

原交叉熵损失函数对于数据集正样本来说，输出概率越大(易检测样本)，损失函数值越小，对于负样本来说，输出概率越小(易检测样本)，损失函数值越小。由于数据集中正负样本比例极不协调，负样本数量远超正样本数量，在训练过程中，易检测样本的梯度主导网络整个损失函数的梯度下降过程即网络参数优化过程，难检测的包含目标的样本集不能得到很好的训练，为此，在网络的分类层损失函数中引入调制因子，降低易检测样本的权重，使得网络更加关注于难检测样本。

4 基于雷达的目标检测

4.1 基于体素网格滤波

原始雷达点云在应用之前，需要处理两个主

要问题：过多的噪声和过大的密度。在进行环境探测时，受到环境变化、传感器精度、测量误差等影响，激光雷达不可避免会产生噪声。过多的噪声会导致算法错误的解释数据，过大的点云密度会导致点云处理速度缓慢，无法满足城市自主车辆感知实时性要求。

为此，采用体素网格划分的方式进行滤波和降采样。体素对应于图像中的最小单元像素，是将所有输入的点云划分到设定大小的立方体网格中。同时，用立方体中的重心代表整个立方体内的激光点，达到降低点云密度的目的，一定程度上可以消除点云的噪声。

在选择体素大小时，过大的网格参数会导致三维特征信息丢失，过小的参数则仍达不到实时性要求。经过多轮实际测试，选取表现较好的 0.2 m 的立方体进行降采样，体素网格滤波效果如图 7 所示，可以看出点云数量得到缩减。

图 7 体素网格滤波

4.2 基于坡度阈值的地面分割

在进行雷达聚类前，首先进行地面和非地面分割，去除对于目标识别无价值的地面点。在此，采用一种基于坡度阈值的分割算法进行地面点滤除。算法核心思想为设定一个全局坡度阈值(Global slope threshold, GST)和局部坡度阈值(Local slope threshold, LST)。通过计算当前高度是否在合理范围内来滤除地面。具体步骤如下。

(1) 按照射线形式组织点云。40 线激光雷达点云水平角分辨率为 0.2° ，水平 360° 可分为 1 800 份，然后将每条射线上点按照半径 $r = \sqrt{x^2 + y^2}$ 进行排序。

(2) 按照当前点与前一点高度差是否在局部高度阈值内，当前点与车底高度差是否在全局高度阈值内，进行地面分割。

首先介绍算法所用的变量

$$ds = r_i - r_{i-1} \quad (7)$$

$$general_height = r_i * GST \quad (8)$$

$$local_height = \begin{cases} \min_height & ds < \min_ds \\ ds * LST & ds \geq \min_ds \end{cases} \quad (9)$$

式中， ds 为两点在射线上的距离； $general_height$ 为根据全局坡度阈值求得的当前全局高度阈值； $local_height$ 为求得的局部高度阈值，当两点距离

过近时,则令当前局部高度阈值为设定的最小值。执行时,通过计算当前点与前一点高度差是否在局部高度阈值内,当前点与车底高度差是否在全局高度阈值内以及前一点属性,判断当前点是地面点或非地面点。最终结果如图 8 所示。

(a) 未滤除地面 (b) 滤除地面

图 8 坡度阈值的地面分割

4.3 基于改进 DBSCAN 的聚类

在完成激光点云地面和非地面分割后,即对非地面点云进行聚类。多线激光雷达一次扫描的点数数以万计,同时这些点呈现不规则分布。经典的 K-means 算法及层次聚类算法一般只适用于凸样本集,面临巨大的失效风险^[21]。DBSCAN 是典型的密度聚类算法,其有效性不受数据集的凸以及非凸的统计学特征的限制^[22-24],适用于点云数据。

DBSCAN 算法需要首先确定两个参数:① epsilon: 在一个点周围邻近区域的半径;② minPts: 邻近区域内至少包含点的个数。

然后,任意选择一个既没有指定到一个类也没有指定为外围点的点,计算它的邻域内点的个数,判断是否为核点,通过不断将直接可达点以及间接可达点加入其中。建立簇的过程中,参数 epsilon 和 minPts 对于聚类影响较大。随着目标到雷达距离的增加,点云逐渐变得稀疏,固定参数的聚类方式效果变差。

针对此问题,提出聚类半径与距离作用模型

$$\text{epsilon} = \rho \text{int} \left(\frac{\sqrt{x^2 + y^2}}{L} + 1 \right) \quad (10)$$

式中, ρ 为聚类系数,可根据激光雷达线束进行调整; (x, y) 为激光点水平坐标; L 为步长; int 代表取整。根据此即可将聚类半径根据距离划分为多个嵌套的同心圆,聚类半径随着距离的增大而增大,符合物理规律。图 9 为原始点云及聚类后效果。

(a) 原始点云 (b) 聚类后

图 9 改进 DBSCAN 的聚类

4.4 基于图像包络的目标位姿估计

激光雷达聚类出目标后,即可获得目标所在位置,然而对于物体的姿态信息仍然是缺乏的。如图 9 所示,目标位置信息正确,姿态信息错误。在无人驾驶中,精准的三维估计对于无人车做出不同决策有着至关重要的影响。

因此,提出一种基于图像包络的三维物体位姿估计方法,具体步骤如下:① 将聚类完成的一簇投影至二维平面;② 利用图像包络的方法,寻找这些二维点的最小凸包;③ 提取凸包轮廓的最小外包矩形;④ 该矩形的长宽作为三维边界框的长宽,三维边界框的高,按照该点云簇最高点与最低点差求得;⑤ 求出该矩形框与车体坐标系正前方的夹角作为目标的方位角。

图像包络方面采用 Graham 扫描法^[24],该算法通过极角排序,逆时针遍历,将所有的点包络在一个多边形中。如图 10 所示,点 A-K 为投影到二维的点,凸多边形 ABCDEFG 即为最小包络多边形。

图 10 基于图像包络的目标位姿

然后寻找该多边形的最小包络矩形,图 10 中矩形 SPQR 即为最小包络矩形。图中 oxy 即为像素坐标系,左上角为原点,右侧为 x 正轴,下侧为 y 正轴。将夹角定义为 x 轴旋转到侧边平齐所需角度,顺时针为正,逆时针为负。

最后将二维矩形框结果返回到三维边界框中,以目标三维边界框的中心点以及长宽高、姿态的形式输出。最终识别的目标三维边界框如图 11 所示。

图 11 目标位姿效果图

整个算法流程如图 12 所示。

图 12 检测算法流程

5 基于特征的数据融合

基于以上策略，视觉部分已完成图像目标的检测，得到城市目标的二维像素坐标、置信度的信息，激光雷达感知部分已完成城市目标的三维信息获取。同时，前文已完成激光雷达与视觉的时间、空间同步。因此，需要对两者检测到的城市目标特征信息进行融合，以得到更加精确的特征信息。融合的特征及方法如表 1 所示，融合方法根据两感知传感器的优势而定，以激光雷达为主体，确保自动驾驶车辆行驶的安全性，结合相机的二维语义信息，完成融合感知。

表 1 融合策略

激光雷达	相机	最终结果
未识别	未识别	未识别
成功识别	未识别	成功识别，标签为未知，确保不会碰撞
未识别	成功识别	再聚类，确定是否为相机误检，如聚类成功，则识别成功，否则为图像误检。
成功识别	成功识别	成功识别，输出结果

算法执行过程伪代码，如表 2 所示。

表 2 激光雷达与视觉融合算法

算法	激光雷达与视觉融合
输入	1, 激光雷达检测结果(三维 BoundingBox、位姿) 2, 图像检测结果(二维 BoundingBox、置信度)
输出	融合检测结果(三维 BoundingBox、位姿、置信度)
算法流程	<pre> for each 雷达检测的目标 新建融合检测到的目标，三维信息等于雷达检测目标 将雷达目标三维框投影到图像中二维框 for each 图像检测的目标 if(雷达目标对应二维框与图像目标框重合度大于 阈值) a)将对应的融合检测目标标签改为图像目标标签 b)将图像目标标记为已使用 end if end for end for for each 未使用过的图像检测的目标 通过内外参，对应点云再聚类 if 存在聚类成功目标 a)新增融合目标等于聚类密度最大目标，标签为 图像目标标签 end if end for end for return 融合目标 </pre>

6 试验结果

为了测试本文算法的准确性和有效性，在东南大学校园不同地点采集数据，进行了两组试验。

6.1 试验一

第一组试验地点为东南大学九龙湖校区北门停车场，将自身车辆摆放不同位置和角度，通过人工测量前方车辆位置和姿态，记录 10 组数据进行量化评价。试验场景如图 13 所示。

图 13 试验场景 1

当车辆水平放置时，环境目标垂直方向的高度对于无人车的路径规划不产生决定性作用，对于车辆避障、紧急制动等决策不产生干扰。因此，只记录下测试车辆中心点的实际 x 、 y 坐标以及测试车辆与无人车相对角度，然后通过算法检测出目标的坐标位置，将实际数据与检测数据进行对比，计算误差。车辆与无人车相对角度定义为无人车顺时针旋转到测试车辆则角度为正，否则为负。真实数据和检测出的坐标位置如表 3 所示。 x 、 y 坐标单位均为 m，角度单位为 $^{\circ}$ 。

表 3 车辆到激光雷达的测量数据与检测数据

编号	真实坐标 x_i	真实坐标 y_i	真实角度 θ_i	检测坐标 x'_i	检测坐标 y'_i	检测角度 θ'_i
1	10.12	-5.11	-44.3	10.18	-5.04	-42.9
2	10.71	0.12	-2.4	10.64	-0.23	-1.1
3	11.34	5.53	46.1	11.39	5.47	47.9
4	20.17	-5.46	-45.6	20.05	-5.39	-44.7
5	22.32	-5.37	-22.4	22.22	-5.43	-23.8
6	21.45	4.89	18.2	21.35	4.81	16.7
7	22.73	4.97	40.3	22.61	4.98	39.1
8	30.85	-5.39	-50.2	30.68	-5.43	-52.8
9	29.86	0.43	3.2	30.04	0.35	1.4
10	31.27	5.12	44.6	31.11	5.13	43.1

由表 3 可知， x 方向上的平均误差为

$$\epsilon_x = \frac{1}{n} \sum_{i=1}^n |x_i - x'_i| = 0.1130 \text{ m}$$

y 方向上的平均误差为

$$\varepsilon_y = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| = 0.0830 \text{ m}$$

方向角的平均误差为

$$\varepsilon_\theta = \frac{1}{n} \sum_{i=1}^n |\theta_i - \theta'_i| = 1.5400^\circ$$

从表 3 以及误差统计可以看出, 所提出的方法 x 、 y 两个方向上的误差以及方向角的误差, 相对于车身来说极小, 在误差允许范围内, 满足智能车的需求。

6.2 试验二

第二组试验地点在机械楼与纪忠楼之间主路之间, 如图 14 所示。该场景车辆行人、车辆密集, 符合一般城市街区场景。为了对算法进行定量评价, 随机截取了数据中的 50 帧, 对速度与检出率进行评估, 测试算法的可靠性。试验中, 车辆最高车速 40 km/h, 检测范围为道路范围内前方 40 m 内。

图 14 试验场景 2

实际检出如图 15 所示, 从图 15a 中可以看出激光雷达检测出路面上的三维目标, 在联合视觉之后, 图 15b 中最终的目标具有了二维和三维的多维度信息。

图 15 检出结果

6.2.1 运行速度

在自动驾驶平台车上进行实际测试, 系统运行频率稳定在 10 Hz 左右, 如图 16 所示。

```
xue@xue:~$ rostopic hz /fused_bounding_boxes
subscribed to [/fused_bounding_boxes]
average rate: 10.038
min: 0.097s max: 0.102s std dev: 0.00184s window: 9
average rate: 9.976
min: 0.092s max: 0.108s std dev: 0.00418s window: 19
average rate: 10.003
min: 0.092s max: 0.108s std dev: 0.00370s window: 29
average rate: 9.997
min: 0.092s max: 0.108s std dev: 0.00329s window: 39
average rate: 10.012
min: 0.092s max: 0.108s std dev: 0.00332s window: 49
average rate: 10.007
min: 0.092s max: 0.108s std dev: 0.00317s window: 59
```

图 16 系统运行速度

6.2.2 目标检测准确率

为了定量分析感知系统表现, 在此给出准确率与检出率定义, 表 4 为准确率统计结果。

$$\text{检出率} = \frac{\text{识别到目标数}}{\text{当前帧目标数}} \quad (11)$$

$$\text{准确率} = \frac{\text{识别到且标签正确目标数}}{\text{当前帧目标数}} \quad (12)$$

表 4 为目标检测准确率 %

方案	行人 (准确率)	车辆 (准确率)	目标检出率
视觉(未改进)	84.3	88.1	88.5
视觉(改进)	86.5	91.8	92.1
信息融合	86.5	91.8	95.7

从表 4 可以看出, 行人目标相对较小, 因此准确率略低于车辆。同时, 激光雷达与视觉融合的策略保障了目标较高检出。即使在目标信息不明朗的前提下, 也可以保障车辆驾驶安全。

7 结论

(1) 验证了激光雷达与视觉融合的感知策略的正确性。提出的激光雷达与视觉融合方法, 既满足感知的容错要求, 增加目标识别的信息维度, 又在实时性上面达到实际需求, 为两种传感器融合提供了一种解决方案。

(2) 在原有 yolov3-tiny 网络基础上, 引入 focal loss, 可提高城市密集目标的检测率。

(3) 在激光雷达感知部分, 提出了兼顾准确性与实时性的雷达感知策略。其中, 利用聚类半径与距离作用模型, 可提高 DBSCAN 聚类准确性; 在激光检测中引入图像包络原理, 完成目标的实时姿态估计。

参 考 文 献

- [1] WU H, LIU Q, LIU X. A review on deep learning approaches to image classification and object segmentation[J]. TSP, 2018, 1(1): 1-5.
- [2] CHERIFI D, KADDARI R, HAMZA Z, et al. Infrared face recognition using neural networks and HOG-SVM[C] // 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART). IEEE, 2019: 1-5.
- [3] LUO J H, LIN C H. Pure FPGA implementation of an HOG based real-time pedestrian detection system[J]. Sensors, 2018, 18(4): 1174-1189.
- [4] CHENG B, WEI Y, SHI H, et al. Revisiting rcnn: On

- awakening the classification power of faster rcnn[C] // Proceedings of the European Conference on Computer Vision (ECCV). 2018: 453-468.
- [5] CHAO Y W, VIJAYANARASIMHAN S, SEYBOLD B, et al. Rethinking the faster r-cnn architecture for temporal action localization[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 1130-1139.
- [6] HE K, GKIOXARI G, DOLLÁR P, et al. Mask r-cnn[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 2961-2969.
- [7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 779-788.
- [8] REDMON J, FARHADI A. Yolo 9000: Better, faster, stronger[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 7263-7271.
- [9] REDMON J, FARHADI A. Yolov3: An incremental improvement[EB/OL]. [2018-04-08]. <https://arxiv.org/pdf/1804.02767.pdf>.
- [10] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]// European Conference on Computer Vision. Springer, Cham, 2016(83): 21-37.
- [11] ASVADI A, PREMEBIDA C, PEIXOTO P, et al. 3D lidar-based static and moving obstacle detection in driving environments: An approach based on voxels and multi-region ground planes[J]. Robotics and Autonomous Systems, 2016, 83(S1): 299-311.
- [12] AZIM A, AYCARD O. Detection, classification and tracking of moving objects in a 3D environment[C]// 2012 IEEE Intelligent Vehicles Symposium. IEEE, 2012: 802-807.
- [13] 周俊静, 段建民, 杨光祖. 基于雷达测距的车辆识别与跟踪方法[J]. 汽车工程, 2014, 36(11): 1415-1420
ZHOU Junjing, DUAN Jianmin, YANG Guangzu. A vehicle identification and tracking method based on radar ranging[J]. Automotive Engineering, 2014, 36(11): 1415-1420.
- [14] ASVADI A, GARROTE L, PREMEBIDA C, et al. Multimodal vehicle detection: fusing 3D-lidar and color camera data[J]. Pattern Recognition Letters, 2018, 115(Suppl. 1): 20-29.
- [15] XUE J, WANG D, DU S, et al. A vision-centered multi-sensor fusing approach to self-localization and obstacle perception for robotic cars[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(1): 122-138.
- [16] WHITNEY D, ROSEN E, ULLMAN D, et al. Ros reality: A virtual reality framework using consumer-grade hardware for ros-enabled robots[C]// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2018: 1-9.
- [17] LEPETIT V, MORENO-NOGUER F, FUA P. Eppn: An accurate o(n) solution to the pnp problem[J]. International Journal of Computer Vision, 2009, 81(2): 155-166.
- [18] 刘军, 后士浩, 张凯, 等. 基于增强 Tiny YOLOV3 算法的车辆实时检测与跟踪[J]. 农业工程学报, 2019, 35(8): 118-125.
LIU Jun, HOU Shihao, ZHANG Kai, et al. Real-time vehicle detection and tracking based on enhanced tiny YOLOV3 algorithm[J]. Journal of Agricultural Engineering, 2019, 35(8): 118-125.
- [19] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980-2988.
- [20] SHRIVASTAVA A, GUPTA A, GIRSHICK R. Training region-based object detectors with online hard example mining[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 761-769.
- [21] ZHANG S, LI S, HU J, et al. An iterative algorithm for optimal variable weighting in K-means clustering[J]. Communications in Statistics-Simulation and Computation, 2019, 48(5): 1346-1365.
- [22] WANG D, LU X, RINALDO A. DBSCAN: Optimal rates for density-based cluster estimation[J]. Journal of Machine Learning Research, 2019, 20(170): 1-50.
- [23] STARCZEWSKI A, CADER A. Determining the Eps parameter of the DBSCAN algorithm[C]// International Conference on Artificial Intelligence and Soft Computing. Springer, Cham, 2019: 420-430.
- [24] WIBOWO A, SANTOSO H B, RACHMAT C A, et al. Mapping and grouping of farm land with graham scan algorithm on convex hull method[C]// 2019 International Conference on Sustainable Engineering and Creative Computing (ICSECC). IEEE, 2019: 121-126.

作者简介: 薛培林, 男, 1995 年出生。主要研究方向为无人驾驶系统感知技术。

E-mail: 220170269@seu.edu.cn

殷国栋(通信作者), 男, 1976 年出生, 博士, 教授, 博士研究生导师。主要研究方向为先进电动汽车、车辆系统动力学与控制、智能无人汽车、智能网联汽车。

E-mail: ygd@seu.edu.cn