

DOI: 10.19650/j.cnki.cjsi.J2107701

基于立体视觉与激光雷达的车辆目标外形 位置融合算法研究*

冯明驰¹, 高小倩², 汪静姝³, 冯辉宗²

(1. 重庆邮电大学先进制造工程学院 重庆 400065; 2. 重庆邮电大学自动化学院 重庆 400065;
3. 重庆理工大学机械工程学院 重庆 400054)

摘要:环境感知技术是智能汽车的关键技术之一,单独使用视觉传感器或激光雷达在目标检测和目标定位方面存在局限性。本文在图像和激光雷达目标检测的基础上,提出了一种基于立体视觉与激光雷达的车辆目标外形位置融合算法。首先,采用深度学习对图像和点云数据进行目标检测,再通过基于目标三维点和目标种类的目标外形位置估计方法确定目标的外形和位置,最后在数据关联后对同一时刻的图像目标及点云目标进行融合处理,获取目标的外形和位置。在KITTI数据集以及实际道路场景上验证了本文算法,实验结果表明,本文方法在检测精度上分别比YOLOv3网络、Point-GNN网络提高了5.72%和1.8%。另外,在20 m内目标外形及位置平均误差分别为4.34%和4.52%。

关键词: 立体视觉;激光雷达;目标检测;传感器融合

中图分类号: TH86 TP391 **文献标识码:** A **国家标准学科分类代码:** 510.4

Research on the fusion algorithm of vehicle object shape-position based on stereo vision and lidar

Feng Mingchi¹, Gao Xiaoqian², Wang Jingshu³, Feng Huizong²

(1. School of Advanced Manufacturing Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; 2. School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, China;
3. College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: Environment perception is one of the key technologies of the intelligent vehicle. However, there are limitations in object detection and object positioning by visual sensor or lidar only. Based on the image and lidar object detection, a fusion algorithm of vehicle object shape-position using stereo vision and lidar is proposed in this paper. Firstly, the deep learning methods are used for object detection on image and point cloud. Then, the shape-position of object is determined by the object shape-position estimation method based on 3D points and object types. Finally, the image object and point cloud object are fused simultaneously after data association and the shape-position of the object are acquired. The proposed algorithm is evaluated on the KITTI data set and actual road scenarios. Experimental results show that the detection accuracy of the proposed method is 5.72% and 1.8% higher than those of the YOLOv3 network and the Point-GNN network, respectively. In addition, the average error of object shape and position within 20 m is 4.34% and 4.52%, respectively.

Keywords: stereo vision; lidar; object detection; sensor fusion

0 引 言

智能汽车^[1]可以提高驾驶员的安全性和舒适性,同

时提高交通系统的效率,而环境感知技术是保证智能汽车安全行驶的关键,它使智能汽车获取周围环境信息,包括障碍物的类别、位置、速度、大小等,为后续智能汽车的路径规划以及决策控制提供可靠的信息^[2]。智能汽车通

收稿日期:2021-04-01 Received Date: 2021-04-01

* 基金项目:重庆市科技局(cstc2019jsex-zdztzxX0050, cstc2019jsex-mbdxX0004)、重庆市教育委员会(KJZD-M201801101, KJQN201801147)项目资助

常配备不同的传感器感知环境信息,如视觉传感器、激光雷达、毫米波雷达等^[3]。在上述传感器中,视觉传感器和激光雷达以其优越的性能成为目标检测以及目标外形位置估计领域中最常用的传感器。

视觉传感器可以提供高分辨率的图像,并赋予丰富的纹理、色彩等信息,能够进行精确地检测^[4]。随着深度学习的兴起,基于深度学习的目标检测方法在精度和速度上都超过许多传统方法。它主要有两种方法,一种是基于候选框提取的目标检测方法,先对图像提取候选框,然后基于候选区域做二次修正得到检测结果。经典模型包括具有卷积神经网络(convolutional neural networks, CNN)^[5]特征的网络,它已从区域卷积神经网络(region-based CNN, R-CNN)^[6]发展到 Faster R-CNN^[7],以及空间金字塔网络(spatial pyramid pooling network, SPP-Net)^[8]。虽然上述网络算法检测精度高,但检测速度较慢,不能满足实时性要求。为了提高检测速度,研究者们提出了另一种方法:基于回归的目标检测方法,直接对输入图像通过网络回归出目标大小、位置及类别,运用比较广泛的检测网络包括单次检测器(single shot multibox detector, SSD)^[9]、统一的实时目标检测(you only look once, YOLO)^[10]一系列算法。同时,基于视觉的目标外形及位置定位算法主要利用单个相机或者多个相机提取目标特征点,根据前后帧特征匹配,估算目标的外形及位置。CHEN 等^[11]根据道路结构的三垂线和点特征提出一种单目视觉定位算法。刘奕博^[12]根据立体视觉原理,对双目图像进行特征点提取及匹配,提出了一种基于快速特征点提取和描述(oriented FAST and rotated BRIEF, ORB)特征的场景流计算方法,并将三维点的平均值作为目标的最终位置。然而,这些方法由于只利用视觉单一传感器,经过图像检测得出的结果缺乏目标的三维位置、方向和物体几何形状等信息。

相比视觉技术,激光雷达具有探测距离远、获取信息准确、夜视能力强等优点^[13]。根据激光雷达点云数据的不同特征表达方式,其目标检测方法可以分为 3 种:1)原始点云数据不进行处理,直接通过神经网络进行目标检测;2)原始点云数据转换为体素格式后进行目标检测;3)原始点云数据投影成二维鸟瞰图后进行目标检测。点云网络(point network, PointNet)^[14]对原始点云数据不做任何改变,直接输入原始点云数据,利用神经网络模型来处理点云数据,学习点与点之间的关系特征后输出目标三维框。体素网络(voxel network, voxelNet)^[15]首先对原始点云进行体素化,通过对体素特征进行学习后输出三维目标框。点云图形神经网络(point graph neural network, Point-GNN)^[16]是一种图神经网络,将点云编码到一个固定半径的近邻图中,从点云数据中检测目标,预测目标的类别及三维框。激光雷达在识别出目标后,通

过点云对目标进行定位,梁林勋等^[17]提出一种基于相位估计方法对处理后的激光雷达信号进行目标方位估计和测距,实现目标的准确定位。虽然,激光雷达在获取目标的姿态和形状方面比视觉传感器具有优势,但是激光雷达易受雨、雪和雾等恶劣天气影响,同时由于激光雷达的安装位置,近距离会缺失点云、纹理等信息。

虽然视觉传感器和激光雷达都可以单独检测目标并对目标外形位置进行估计,但每个传感器只能获取局部信息。因此,两种传感器需要协同工作,才能在复杂多变的交通环境中完成环境感知技术任务。基于视觉传感器和激光雷达的融合算法可以分为前融合和后融合^[18]:1)前融合包含了数据级融合和特征级融合。其主要方法先对原始图像数据和点云数据的特征进行融合,再检测及定位;2)后融合也叫决策级融合,先分别对两传感器获取的数据进行目标检测,将目标检测结果进行融合及定位。对于前融合,Rovid 等^[19]提出了一种神经网络方法来融合图像和激光雷达点云,从而构成稠密点云,并基于点云对目标的位置进行估算。对于后融合,Sualeh 等^[20]先分别利用网络对图像和点云进行检测,根据三维点云找到目标检测框对角线的两个角点以及第三个角点,从而对目标的外形及位置进行更新。

以上方法虽然在结果方面取得了不错的成绩,但还是存在一些不足:1)为了兼顾激光雷达测距较远的优点,激光雷达的安装位置较高会造成它对近处场景缺失点云。2)视觉传感器和激光雷达只能获取目标的局部信息,如果直接利用局部信息对目标进行外形位置估计,与实际目标的外形位置相差较大。

针对以上问题,同时为了能够准确地、快速地检测出车辆前方目标,并精确获取目标外形位置。本文利用视觉和激光雷达两种传感器和后融合技术,提出了一种基于立体视觉与激光雷达的车辆目标外形位置融合算法。

本文的组织结构如下:第 1 节,算法描述,包括基于视觉传感器的目标检测及目标外形位置估计,基于激光雷达的目标检测及目标外形位置估计,图像与点云目标融合方法;第 2 节,实验验证和结果讨论;第 3 节,结论。

1 算法描述

1.1 基于视觉传感器的目标检测及目标外形位置估计方法

1) 基于 YOLOv3 的图像目标检测方法

YOLOv3^[21]是一种基于深度卷积神经网络进行的目标检测算法,它在 YOLOv1^[22]、YOLOv2^[23]基础上进行网络结构的改进。YOLOv3 主要由 Darknet-53 特征提取网络和多尺度融合特征网络组成,其中 Darknet-53 由 53 个卷积层组成,包括了 1×1 和 3×3 两种卷积层,并结合了

残差网络,输出3种不同尺度的特征图,最后输出目标二维框大小及类别。

YOLOv3 主要检测原理是当原始图像输入到网络输入层时,输入层首先将原始图像大小转变成预先设定的图像分辨率大小,并将图像分成 $S \times S$ 大小的网格。若图像中某个目标的中心落入其中一个网格中,那么该网格就负责检测这个目标。

2) 图像目标获取三维点

为了进一步获取目标的信息,本文首先对左右图像检测到的目标进行特征提取,主要采用 ORB 特征点提取方法,之后利用汉明距离原理完成特征点之间的匹配。

在完成特征点匹配后,利用立体视觉原理^[24] 计算特征点的三维坐标,如式(1)所示。

$$\begin{cases} x_c = \frac{b \cdot x_{left}}{x_{left} - x_{right}} \\ y_c = \frac{b \cdot y_{left}}{x_{left} - x_{right}} \\ z_c = \frac{b \cdot f}{x_{left} - x_{right}} \end{cases} \quad (1)$$

式中: (x_{left}, y_{left}) 、 (x_{right}, y_{right}) 分别表示某一匹配特征点在左图像和右图像的图像坐标。 b 表示两相机之间的基线, f 表示相机之间的焦距。

由于利用目标检测框对目标提取的特征点会包含背景特征信息,为了将背景信息剔除掉及显示效果,本文将三维点投影在二维 XOZ 平面上进行聚类,建立如图1所示的相机、激光雷达、车辆坐标系。

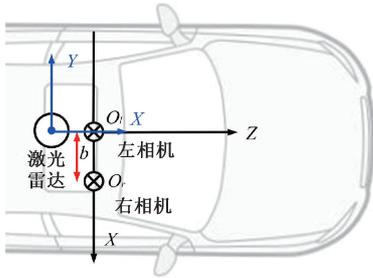


图1 相机、激光雷达、车辆坐标系

Fig. 1 The coordinate system of camera, lidar and vehicle

其中以左相机为主建立车辆坐标系, X 轴与两相机之间的基线 b 重合向右, Z 轴沿着左相机的光轴向前, Y 轴朝下。针对目标的三维点利用重心法计算出 k 时刻第 t 个目标的重心,如式(2)所示。

$$Object3D_k^t = \bar{U}_{i=0}^n(x_k^t, y_k^t, z_k^t) \quad (2)$$

式中: $Object3D_k^t$ 表示 k 时刻第 t 个目标三维点的集合。

$$G(x, y, z)_k^t = \frac{Object3D_k^t}{n} \quad (3)$$

式中: $G(x, y, z)_k^t$ 表示 k 时刻第 t 个目标所有三维点的重心。

本文在获取目标三维点重心后,根据 YOLOv3 检测算法检测出的目标以及所属类别,对目标三维点进行聚类,其目的是为了能够获取只属于目标的三维点信息。步骤如下:

(1) 根据目标检测算法对二维目标检测框进行 ORB 特征提取并进行匹配,利用立体视觉原理计算出匹配点的三维坐标;

(2) 根据式(2)、(3)计算出 k 时刻第 t 个目标的重心,如图2所示为图像目标三维点的聚类过程。首次计算三维点的重心为 C_1 ,检测类别为 Car,根据 Car 外形大小选取范围;

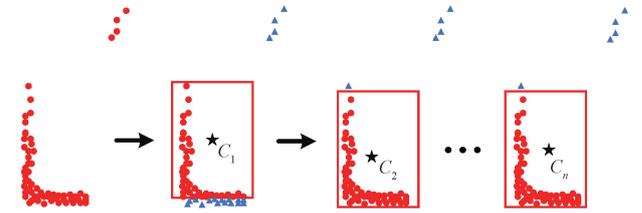


图2 图像目标三维点聚类过程

Fig. 2 The clustering process of 3D points of image object

(3) 再次计算外形框内点的重心为 C_2 ,以同样大小的外形框选取范围。如果框内点有变化,重复以上步骤,计算框内点的重心为 C_n ,根据外形框大小选取范围,直到框内点没有变化为止;如果框内点没有变化,就剔除掉框外的点,保留框内的点作为目标的三维点,完成后续研究工作。图2中五角星表示目标三维点的重心,圆点表示当前阶段所属的目标三维点,三角形代表欲将剔除的背景信息三维点。

3) 基于目标方向及检测类别的目标外形位置估计方法

由于相机视角的缘故,只能获取目标局部信息,利用 ORB 提取的特征点只属于目标的部分特征点。如果直接利用重心法求出的重心点作为物体的中心,与实际目标中心差距较远,对后续的局部路径规划、汽车超车、变道造成影响。如图3所示,其中图3(a)表示根据相机获取的目标局部信息利用重心法求取出的目标中心位置 C' ;图3(b)表示根据目标实际大小获取的目标中心位置 C 。

于是,本文根据检测出的目标类别以及目标方向,对目标的外形及位置进行估计。具体算法流程如下:

(1) 利用最小正矩形框包围目标特征点,并判断目标方向。如果目标在正前方,取底边的中点;如果在左前方,取最小正矩形框的右下点;如果在左侧,取最小正矩形框的右上点;如果在右前方,取最小正矩形框的左下

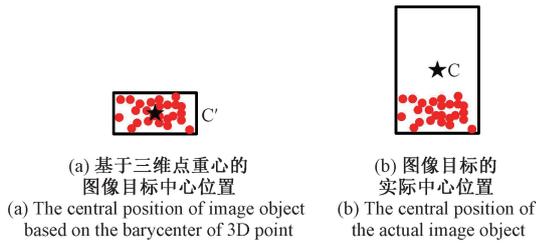


图 3 基于三维点重心的图像目标中心位置及实际图像目标中心位置

Fig. 3 The central position of the image object based on the barycenter of 3D point and the actual image object

点;如果在右侧,取最小正矩形框的左上点,如式(4)所示。如图 4 所示,已知最小正矩形框的左上点坐标为 (x_l, z_l) ,右下点坐标为 (x_b, z_b) ,最小正矩形框的长宽分别为 L_m, W_m ,实际目标外形大小为 W, L, H 。

$$\begin{cases} P_1 = (x_b, z_b) & x_b < -S \text{ and } z_b > L \\ P_2 = (x_b, z_l) & x_b < -S \text{ and } z_l < L \\ P_3 = ((x_b - x_l)/2, z_b) & x_l > -S \text{ and } x_b < S \\ P_4 = (x_l, z_b) & x_l > S \text{ and } z_b > L \\ P_5 = (x_l, z_l) & x_l > S \text{ and } z_l < L \end{cases} \quad (4)$$

式中: P_1, P_2, P_3, P_4, P_5 分别代表目标在左前方时右下点,左侧方时右上点,正前方时底边中点,右前方时左下点,右侧方时左上点, S 表示车道宽。

(2) 根据检测类别,对目标的形状大小进行修正,根据目标大小确定目标的中心位置,如式(5)所示。

$$\begin{cases} C_1 = (x_b - W/2, H/2, z_b + L/2) & x_b < -S \text{ and } z_b > L \\ C_2 = (x_b - W/2, H/2, z_l - L/2) & x_b < -S \text{ and } z_l < L \\ C_3 = ((x_b - x_l)/2, H/2, z_b + L/2) & x_l > -S \text{ and } x_b < S \\ C_4 = (x_l + W/2, H/2, z_b + L/2) & x_l > S \text{ and } z_b > L \\ C_5 = (x_l + W/2, H/2, z_l - L/2) & x_l > S \text{ and } z_l < L \end{cases} \quad (5)$$

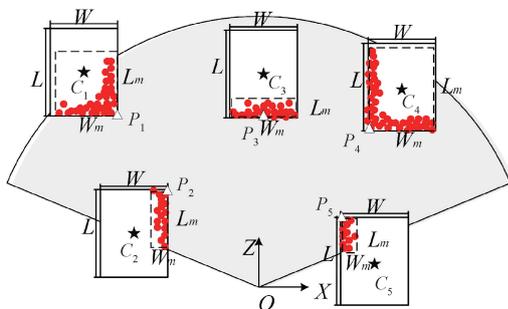


图 4 基于目标不同方向及类别的图像目标外形位置
Fig. 4 The shape-position of image objects based on different directions and categories

1.2 基于激光雷达的目标检测及目标外形位置估计方法

激光雷达是一种以发射激光束测量目标距离、具体位置等特征量的雷达系统。与图像不同,激光雷达获取的点云通常是稀疏的且在规则网格上间隔不均匀。卷积神经网络检测目标,它需要一个规则的网络输入,如果将点云放入规则网络进行卷积操作,会导致点云数据信息的丢失,同时也会造成网络计算量大。于是,本文对于点云检测,采用了 Point-GNN 网络,它是一种图神经网络从激光雷达获取的点云中检测目标。

Point-GNN 主要包含了 3 部分:1) 图形构建;2) T 次迭代的 GNN;3) 边界框合并和打分。其中图形构建将 N 个点云构成集合 P ,接着将集合中的某个点作为顶点,并将该点与在起始设置的固定半径 r 内的相邻点连接构造一个图。之后,图神经网络通过沿边缘聚集特征来提取顶点特征。在 T 次迭代后,不断更新图形中每个顶点的特征,并使用顶点反射强度来预测对象的类别和顶点所属目标的边界框。由于多个顶点可能会位于同一目标对象上,因此神经网络可以输出同一目标对象的多个边界框。为了保证检测目标的精确度,在多个边界框里选择分类分数最高的框,并抑制其他重叠的框,将这些边界框合并到一个,并打出其置信度分数。

激光雷达获取的点云能够直接通过 Point-GNN 网络输出目标三维框,包括目标三维框的大小和目标的三维中心 (x, y, z) 。如图 5 所示为点云目标检测结果。

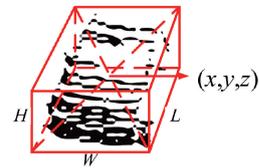


图 5 基于 Point-GNN 的点云检测结果
Fig. 5 The result of point cloud detection based on Point-GNN

1.3 图像与点云目标融合方法

图像目标的外形及位置通过特征点以及检测类别确定,而点云目标的外形及位置直接通过深度学习的方法进行检测后输出目标的三维框大小和中心位置。在分别获取图像目标和点云目标后,根据激光雷达与相机之间的联合标定^[25]获取激光雷达与相机之间的旋转平移矩阵 R, T ,如图 6 所示。

激光雷达与相机标定具体操作步骤如下:

- 1) 首先选取一块不带白边的且小方格大小一样的长方形棋盘格标定板;
- 2) 同时启动左相机和激光雷达,然后两传感器分别捕捉标定板的二维图像和三维点云;
- 3) 利用 MATLAB 完成标定,获取激光雷达与相机之

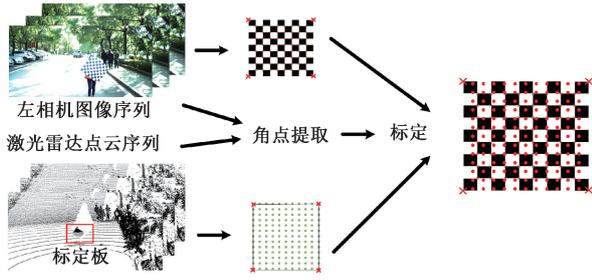


图6 左相机与激光雷达标定流程图

Fig. 6 The calibration flowchart of left camera and lidar

间的旋转平移矩阵。根据标定板的不同姿态获取多组图像和三维点云数据;在图像中检测出标定板,并获取标定板的大小,紧接着在三维点云数据中寻找相同大小的点云区域,将点拟合成线构成矩形;并将多组标定板的顶点二维坐标和三维坐标代入下列公式。

$$\mathbf{Z}_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} [\mathbf{R} | \mathbf{T}] \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} =$$

$$\mathbf{Q} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_l \\ y_l \\ z_l \\ 1 \end{bmatrix} \quad (6)$$

其中, \mathbf{Q} 是一个大小为 3×4 的转换矩阵, f_x, f_y, c_x, c_y 为相机的内参系数,从方程的计算过程可以看出,矩阵中的每个参数都是多个标定内、外参数的复合计算值。只要得到足够数量特征点的对应坐标后,即可计算出变换矩阵 \mathbf{Q} 。

利用获取的标定参数,将点云目标换算到以左相机建立的车辆坐标系下表示,如式(7)所示。

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = (\mathbf{R} | \mathbf{T}) \begin{pmatrix} x_l \\ y_l \\ z_l \end{pmatrix} \quad (7)$$

式中: \mathbf{R}, \mathbf{T} 表示激光雷达和左相机之间的旋转平移矩阵, $(x_l, y_l, z_l), (x_c, y_c, z_c)$ 分别表示激光雷达坐标系、相机坐标系下的点云目标中心位置坐标。

将图像目标和点云目标换算到同一坐标系后,采用数据关联法^[19]进行匹配,并将匹配到的目标利用加权平均法确定匹配目标的最终位置,没有匹配到的目标以原始位置作为最终结果输出。本文采用全局最近邻(global nearest neighbor, GNN)方法匹配同一时刻的图像目标和点云目标。具体算法步骤如下:

1) 设定匹配门限值 G , 图像和点云为同一目标的前提是其中一类目标必须落在另一类目标为中心的

内。如图7所示,为图像目标、点云目标分布情况, C, L 分别代表图像、点云目标,从图中可以看到 L_4 不在任何一个门限内,没有目标与之匹配。在 C_2 关联门中,虽然 L_2, L_3 都满足匹配,但是每一类目标最多只能和另一类其中一个目标匹配。通过计算, L_2 与 C_2 的距离要小于 L_3 与 C_2 的距离。因此,当都满足匹配前提,要求两目标距离最小时的匹配结果为最优匹配结果。

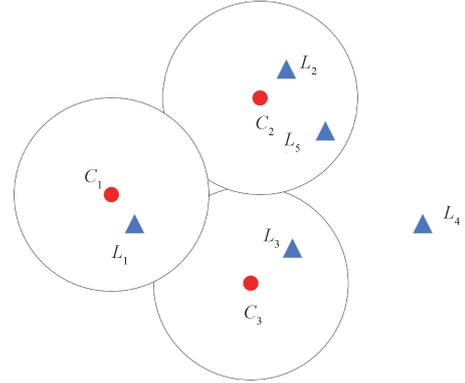


图7 各类目标分布情况示意图

Fig. 7 Distribution diagram of various objects

2) 匹配到的目标表明两种传感器都检测到,并且属于同一目标,且最终目标位置如式(8)所示。若未匹配,表明只有一种传感器检测到或者两种传感器都检测到,但是不满足数据关联法,将此类目标定义为未匹配的图像目标或点云目标,它们的最终位置就是其本身。

$$\begin{cases} x' = \frac{x_l \delta_{cx}^2}{\delta_{lx}^2 + \delta_{cx}^2} + \frac{x_c \delta_{lx}^2}{\delta_{lx}^2 + \delta_{cx}^2} \\ y' = \frac{y_l \delta_{cy}^2}{\delta_{ly}^2 + \delta_{cy}^2} + \frac{y_c \delta_{ly}^2}{\delta_{ly}^2 + \delta_{cy}^2} \\ z' = \frac{z_l \delta_{cz}^2}{\delta_{lz}^2 + \delta_{cz}^2} + \frac{z_c \delta_{lz}^2}{\delta_{lz}^2 + \delta_{cz}^2} \end{cases} \quad (8)$$

式中: $\delta_{lx}, \delta_{ly}, \delta_{lz}$ 分别表示点云目标在 x, y, z 方向距离的平均误差; $\delta_{cx}, \delta_{cy}, \delta_{cz}$ 分别表示图像目标在 x, y, z 方向距离的平均误差; (x_l, y_l, z_l) 表示点云目标在车辆坐标系下的位置; (x_c, y_c, z_c) 表示图像目标在车辆坐标系下的位置; (x, y, z) 表示最终目标位置。

对于目标外形融合,本文根据基于视觉传感器的图像目标外形估计方法确定图像目标的外形大小为 W_1, L_1 , 根据 Point-GNN 网络确定点云目标外形大小为 W_2, L_2 , 同时根据检测算法得出图像目标的置信度为 s_1 , 点云目标的置信度为 s_2 。另外,视觉传感器测距有限,获取目标信息的准确性会随着距离的增加而减小,如式(9)所示。其中, W', L' 为确定最终目标的外形大小, d 代表目标的距离, $d = \sqrt{x^2 + y^2 + z^2}$ 。

$$\begin{cases} W' = \frac{s_1 \times \frac{1}{d}}{s_1 \times \frac{1}{d} + s_2} \times W_1 + \frac{s_2}{s_1 + s_2} \times W_2 \\ L' = \frac{s_1 \times \frac{1}{d}}{s_1 \times \frac{1}{d} + s_2} \times L_1 + \frac{s_2}{s_1 + s_2} \times L_2 \end{cases} \quad (9)$$

目标的外形尺寸由车辆的极限位置之差确定,由误差传递原理可知,外形误差为 $\sqrt{2}$ 倍的位置误差,所以目标的位置精度为 0.7 倍的外形精度。

2 实验验证

为了实现车辆前方目标检测以及目标外形位置估计。本小节主要对描述的算法进行实验验证,分别在公开的 KITTI 数据集和实车上进行测试,从而验证本文算法。

2.1 实验设备

本文的实验平台智能汽车如图 8 所示,它由长安 CS55 改装而成,车顶上为传感器台架,主要放置两个 Basler acA2500-60uc 工业相机和一个 Robosense 128 线激光雷达。实验设备为一台装有 2060 显卡的笔记本电脑,操作系统为 Ubuntu16.04,编程语言为 C++。本文提出的算法在机器人操作系统(robot operating system, ROS)上运行,相机和激光雷达设置相同的采集频率为 10 Hz。算法的实现主要依赖 OpenCV、PCL 完成图像、点云处理的基本操作。



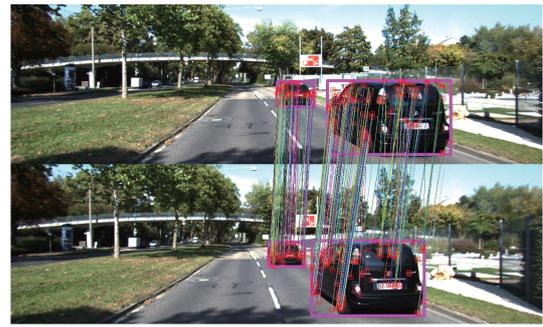
图 8 实验车平台

Fig. 8 Platform of the experimental vehicle

2.2 基于 KITTI 数据集的算法验证

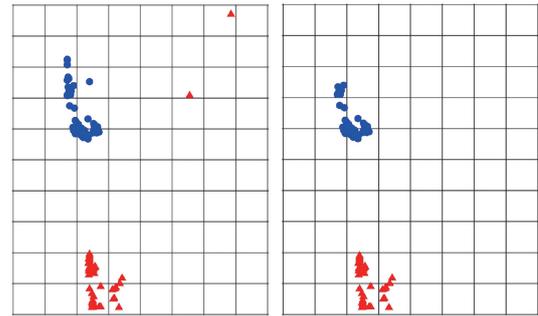
本文为了算法适应各种交通场景,同时由于点云数据集人工标注比较困难,为了保证后期检测的准确性,本文采用公开的 KITTI 图像数据集以及点云数据集分别进行 YOLOv3 和 Point-GNN 网络训练,其中训练集 8 000 张,测试集 500 张。

对于视觉传感器,通过 YOLOv3 网络训练得出的权重文件,进行测试,结果如图 9 所示。



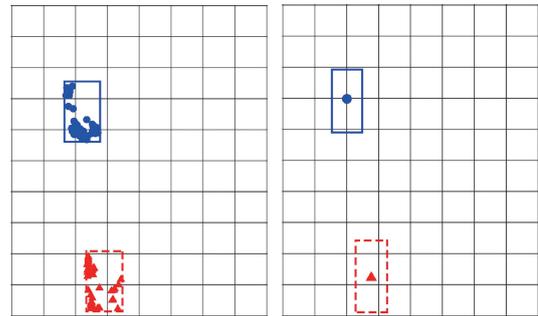
(a) 图像目标检测及特征提取匹配
(a) Image object detection, feature extraction and matching

● 图像目标1 ▲ 图像目标2



(b) 图像目标三维点聚类前后对比
(b) Comparison of pre-clustering results and post-clustering results of image object 3D points

● 图像目标1 ▲ 图像目标2



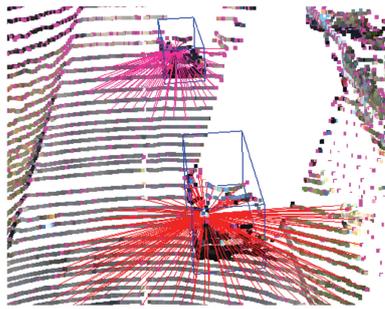
(c) 图像目标外形位置估计
(c) Shape-position estimation of image object

图 9 基于视觉传感器的图像目标外形位置估计结果

Fig. 9 The shape-position estimation result of image object based on vision sensor

对于激光雷达,通过 Point-GNN 网络训练得出的权重文件,进行测试,结果如图 10 所示。

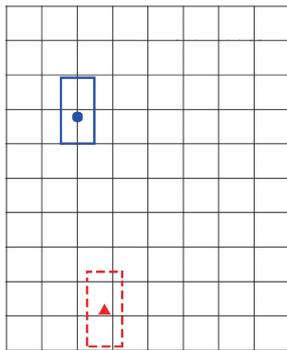
本文对 KITTI 的图像数据集和点云数据集进行测试,并将结果按照本文的方法进行融合,将检测结果与数据集标注结果进行对比,并统计检测结果与真实结果的误差,包括目标检测率和目标误检率。其中,目标检测率表示在不考虑检测类别对与错的情况下,所检测类别数目占测试集所有目标数目的比例,目标误检率是误检数目占所检测类别数目的比例。



(a) 点云目标检测

(a) Object detection of the point cloud

● 点云目标1 ▲ 点云目标2



(b) 点云目标外形位置估计

(b) Shape-position estimation of point cloud object

图 10 基于激光雷达的点云目标外形位置估计结果

Fig. 10 The shape-position estimation result of point cloud object based on lidar

目标检测率=检测类别数目/所有目标数目×100%
 目标误检率=误检类别数目/检测类别数目×100%
 检测结果如图 11 所示。

对于 YOLOv3 网络检测,图像能够较好地检测近处目标,对远处较小目标检测效果比较差,而激光雷达由于近处缺失点云,对近处目标无法检测,如图 11(c) 所示,而且误检率要高于图像。本文将两个检测网络融合,总体上讲,目标检测率有明显地提升,误检率明显下降。如表 1 所示,融合算法目标检测率相比 YOLOv3 提高了 5.72%,相比 Point-GNN 提高了 1.8%。

表 1 测试集检测结果

Table 1 The detection results of test dataset %

	目标检测率	目标误检率
YOLOv3	75.48	5.6
Point-GNN	79.4	6.4
本文融合算法	81.2	3.3

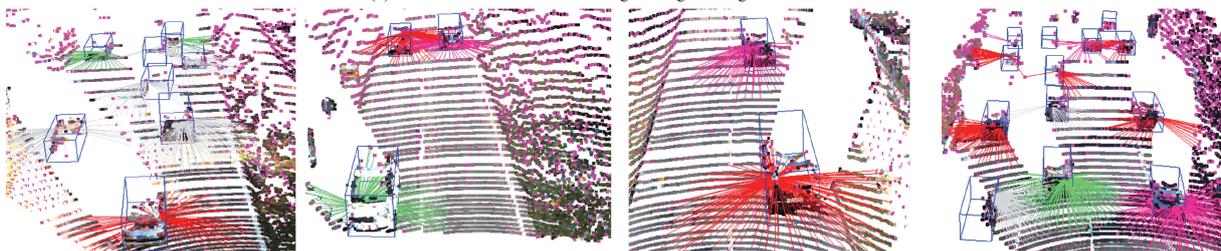
2.3 实车算法验证

首先完成了双目相机的标定以及左相机与激光雷达的标定,对于双目相机,采取小格边长为 100 mm,大小为 18×13 的棋盘格标定板,激光雷达和左相机标定,采取每个小格边长为 70 mm,大小为 10×9 不带白边的棋盘格标定板。各自采集 15 组数据,之后利用 MATLAB 工具箱计算出旋转平移矩阵,结果如表 2 所示。



(a) YOLOv3左右图像检测结果

(a) Detection results of left and right images using YOLOv3



(b) Point-GNN点云检测结果

(b) Detection results of point cloud using Point-GNN



(c) 融合结果

(c) The fusion results

图 11 KITTI 数据集测试结果

Fig. 11 Test results using the KITTI dataset

表 2 双目相机、激光雷达与左相机标定结果

Table 2 Calibration results of binocular camera, lidar and left camera

	双目相机	激光雷达与左相机
R	$\begin{bmatrix} 0.9999 & 0.0059 & 0.0283 \\ -0.0057 & 0.9999 & -0.0078 \\ -0.0283 & 0.0077 & 0.9999 \end{bmatrix}$	$\begin{bmatrix} -0.0111 & -0.9999 & -0.0047 \\ -0.0103 & 0.0048 & -0.9999 \\ 0.9999 & -0.0110 & -0.0104 \end{bmatrix}$
T	$[-0.4489 \quad 0.0008 \quad 0.0141]$	$[0.1125 \quad -0.2198 \quad -0.5908]$

本文经过测试发现由于激光雷达安装位置在车顶处,在智能汽车近距离范围内很难生成点云,而近距离目标信息的获取对智能汽车的防碰撞、控制决策等方面的研究十分重要。如图 12 所示,分别表示激光雷达和视觉传感器近距离观测范围,从图中可以发现激光雷达在智能汽车 5 m 内无法获取周围环境信息,而从图像中能够直接观察到车辆前方有一个障碍物。于是,为了弥补激光雷达的缺陷,本文在近距离范围内采用视觉传感器感知车辆周围环境信息。



图 12 激光雷达和相机的近距离观测范围
Fig. 12 Close observation range of lidar and camera

同时,为了验证本文提出的图像目标外形位置估计方法的准确性,在重庆某高校实际道路上进行了测试,结果如图 13 所示,其中图 13(a)~(d)表示场景 1 下的测试结果,图 13(e)~(h)表示场景 2 下的测试结果。

本文手动测量 20 组目标的实际外形大小和位置距离,并将融合后的目标外形位置大小与测量出的实际外形位置大小进行比较。其中误差值 $error_1$ 、 $error_2$ 分别表示目标的实际外形大小与融合目标外形大小之间的误

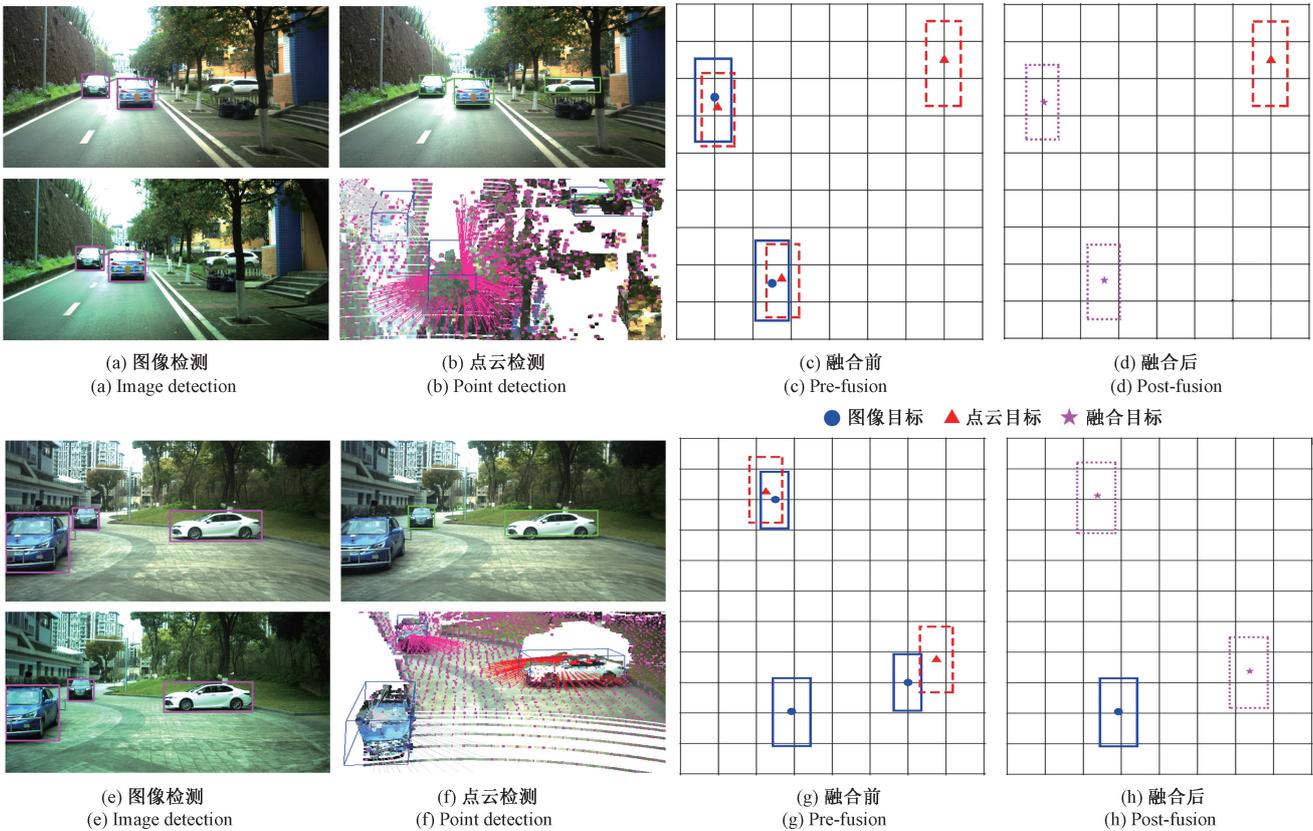


图 13 实车环境下的车辆检测及外形位置估计结果

Fig. 13 The results of vehicle detection and shape-position estimation in real vehicle environment

差,目标的实际位置与融合目标位置之间的误差,如式(10)、(11)所示。

$$error_1 = \frac{\sqrt{(W' - W)^2 + (L' - L)^2}}{\sqrt{W^2 + L^2}} \times 100\% \quad (10)$$

$$error_2 = \frac{\sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2}}{\sqrt{x^2 + y^2 + z^2}} \times 100\% \quad (11)$$

式中: W 、 L 、 W' 、 L' 分别表示实际目标外形大小和融合后的目标外形大小。 (x, y, z) 、 (x', y', z') 分别表示实际目标位置和融合后的目标位置。

在本文研究中,为了与点云数据集类别统一,将图像数据集里的车辆目标统称为 Car,并对车辆目标外形大小赋予相同的值。在实验中,图像目标与实际目标的外形误差为固定值其可以得到控制。因此,本文不对图像目标外形进行误差计算,只对激光雷达和融合后的目标外形与实际测量出的目标外形进行比较,并将误差结果进行统计。如图 14 所示,其中图 14(a)表示实际目标外形与计算出的目标外形误差,图 14(b)表示实际目标位置与计算出的目标位置误差。

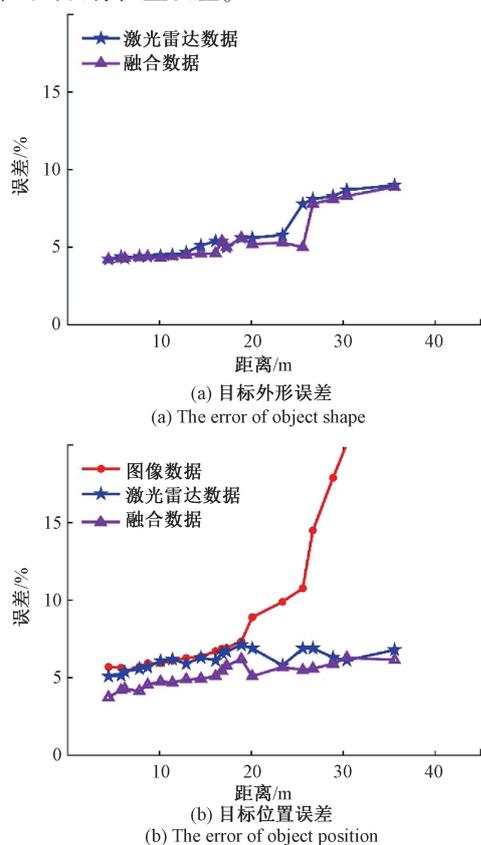


图 14 实际目标与视觉传感器、激光雷达、融合目标的外形位置误差

Fig. 14 The shape-position error between actual object and vision sensor, lidar, fusion object

表 3 实际目标与视觉传感器、激光雷达、融合目标的外形位置平均误差

评价指标	评价指标			
	视觉传感器	激光雷达	融合结果	
外形	0~20 m 平均外形误差	-	4.76	4.34
	20~40 m 平均外形误差	-	7.61	7.23
位置	0~20 m 平均定位误差	6.19	5.25	4.52
	20~40 m 平均定位误差	15.02	6.42	5.75

通过计算发现,在 20 m 以内,融合后的目标外形大小与实际的目标外形大小平均误差为 4.34%,目标位置平均误差为 4.52%,如表 3 所示。在 20 m 以后,由于视觉传感器测距有限,误差会随着距离的增加逐渐变大。同时受到噪声影响,目标位置精度为 0.7 倍的外形精度不是那么明显,但是可以大致反应出该关系。

3 结 论

本文提出了一种基于立体视觉与激光雷达的车辆目标外形位置融合算法,先对图像和点云进行目标检测,然后采用外形位置估计方法确定目标外形和位置,最后利用融合方法对目标进行融合,从而确定目标的最终外形和位置。通过 KITTI 数据集以及实车测试,实验结果表明,本文方法在检测性能上比单个网络目标检测率分别提高了 5.72% 和 1.8%,将融合后的目标外形框大小与实际目标外形框大小进行比较,在 20 m 以内平均误差大小为 4.34%,满足误差要求,在目标定位上与实际的目标位置相差小。总体来说,本方法能够为智能汽车提供可靠的周围环境信息。

在后续的研究中,将继续改进目前的深度神经网络和融合算法。目前在实际实验中考虑的车辆种类有限,以后将进一步对车辆类别进行细分,每种车辆类别赋予一个外形尺寸,从而提高车辆目标的外形检测精度。

参考文献

- [1] SOULTANA A, BENABBOU F, SAEL N. Context-awareness in the smart car: Study and analysis [C]. Proceedings of the 4th International Conference on Smart City Applications, Casablanca, Morocco, 2-4 October, 2019: 1-8.
- [2] 张艳辉,徐坤,郑春花,等. 智能电动汽车信息感知技术研究进展[J]. 仪器仪表学报, 2017, 38(4): 794-805.

- ZHANG Y H, XU K, ZHENG CH H, et al. Advanced research on information perception technologies of intelligent electric vehicles [J]. Chinese Journal of Scientific Instrument, 2017, 38(4): 794-805.
- [3] JAHROMI B S, TULABANDHULA T, CETIN S. Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles [J]. Sensors, 2019, 19(20): 1-23.
- [4] ZHAO X, SUN P, XU Z, et al. Fusion of 3D LIDAR and camera data for object detection in autonomous vehicle applications [J]. IEEE Sensors Journal, 2020, 20(9): 4901-4913.
- [5] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望[J]. 自动化学报, 2017, 43(8): 1289-1305.
- ZHANG H, WANG K F, WANG F Y. Advances and perspectives on applications of deep learning in visual object detection [J]. Acta Automatica Sinica, 2017, 43(8): 1289-1305.
- [6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014: 580-587.
- [7] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Springer, 2016: 21-37.
- [10] 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, Seattle, 2016: 779-788.
- [11] CHEN X, MA H, WAN J, et al. Multi-view 3d object detection network for autonomous driving [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1907-1915.
- [12] 刘奕博. 基于车载双目相机的目标检测及其运动状态估计[D]. 重庆: 重庆邮电大学, 2020.
- LIU Y B. Object detection and motion state estimation based on vehicle-mounted binocular camera [D]. Chongqing: Chongqing University of Posts and Telecommunications, 2020.
- [13] GENG K, DONG G, YIN G, et al. Deep dual-modal traffic objects instance segmentation method using camera and lidar data for autonomous driving [J]. Remote Sensing, 2020, 12(20): 1-22.
- [14] QI C R, SU H, MO K, 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, 2017: 652-660.
- [15] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3d object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.
- [16] SHI W, RAJKUMAR R. Point-GNN: Graph neural network for 3d object detection in a point cloud [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Electr Network, 2020: 1711-1719.
- [17] 梁林勋, 杨俊杰, 楼志斌. 基于激光雷达的机器人定位信息处理技术研究[J]. 自动化与仪器仪表, 2020(1): 40-43.
- LIANG L X, YANG J J, LOU ZH B. Research on robot positioning information processing technology based on lidar [J]. Automation & Instrumentation, 2020(1): 40-43.
- [18] 郑少武, 李巍华, 胡坚耀. 基于激光点云与图像信息融合的交通环境车辆检测[J]. 仪器仪表学报, 2019, 40(12): 143-151.
- ZHENG SH W, LI W H, HU J Y. Vehicle detection in the traffic environment based on the fusion of laser point cloud and image information. [J]. Chinese Journal of Scientific Instrument, 2019, 40(12): 143-151.
- [19] ROVID A, REMELI V, SZALAY Z. Raw fusion of camera and sparse LiDAR for detecting distant objects[J]. At-Automatisierungstechnik, 2020, 68(5): 337-346.
- [20] SUALEH M, KIM G W. Visual-LiDAR based 3D object detection and tracking for embedded systems[J]. IEEE Access, 2020, 8: 156285-156298.
- [21] REDMON J, FARHADI A. YOLOv3: An incremental

improvement [J]. ArXiv Preprint, 2018, ArXiv: 1804.02767.

- [22] LU J, MA C, LI L, et al. A vehicle detection method for aerial image based on YOLO[J]. Journal of Computer and Communications, 2018, 6(11): 98-107.
- [23] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [24] YANG L, WANG B, ZHANG R, et al. Analysis on location accuracy for the binocular stereo vision system[J]. IEEE Photonics Journal, 2017, 10(1): 1-16.
- [25] PARK Y, YUN S, WON C S, et al. Calibration between color camera and 3D LIDAR instruments with a polygonal planar board[J]. Sensors, 2014, 14(3): 5333-5353.

作者简介



冯明驰,分别2008年和2014年于中国科学技术大学获得学士学位和博士学位,现为重庆邮电大学副教授,主要研究方向为视觉测量、智能汽车环境感知。

E-mail: fengmc@cqupt.edu.cn

Feng Mingchi received his B. Sc. degree and Ph. D. degree both from University of Science and Technology of China in 2008 and 2014, respectively. He is currently an associate professor at Chongqing University of Posts and Telecommunications. His research interests include vision measurement and environmental perception of intelligent vehicles.



高小倩,2018年于重庆邮电大学移通学院获得学士学位,现为重庆邮电大学硕士研究生,主要研究方向为计算机视觉。

E-mail: 429103582@qq.com

Gao Xiaoqian received her B. Sc. degree from College of Mobile Communication at Chongqing University of Posts and Telecommunications in 2018. She is currently pursuing her M. Sc. Degree at Chongqing University of Posts and Telecommunications. Her research interest is computer vision.



汪静姝(通信作者),分别2008年和2013年于中国科学技术大学获得学士学位和博士学位,现为重庆理工大学副教授,主要研究方向为视觉测量、智能制造过程。

E-mail: donotbreeze@163.com

Wang Jingshu (Corresponding author) received her B. Sc. degree and Ph. D. degree both from University of Science and Technology of China in 2008 and 2013, respectively. She is currently an associate professor at Chongqing University of Technology. Her research interests include vision measurement and processing of intelligent manufacturing.



冯辉宗,2004年于重庆大学获得博士学位,现为重庆邮电大学教授,主要研究方向为汽车电子控制系统建模仿真和发动机控制。

E-mail: fenghz@cqupt.edu.cn

Feng Huizong received his Ph. D. degree from Chongqing University in 2004. He is currently a professor at Chongqing University of Posts and Telecommunications. His research interests include modeling and simulation of automobile electronic control system and engine control.