



上海交通大学学位论文

一种基于多源传感器信息融合的 路侧导航增强单元

姓 名：杨嘉业

学 号：519021910359

导 师：张欣

学 院：航空航天学院

学科/专业名称：航空航天工程

申请学位层次：学士学位

2023 年 5 月

**A Dissertation Submitted to
Shanghai Jiao Tong University for Bachelor Degree**

**A ROADSIDE UNIT FOR NAVIGATION
ENHANCEMENT BASED ON
MULTI-SENSORS INFORMATION FUSION**

Author: Yang Jiaye
Supervisor: Zhang Xin

School of Aeronautics and Astronautics
Shanghai Jiao Tong University
Shanghai, P.R.China
May 14th, 2023

上海交通大学

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全知晓本声明的法律后果由本人承担。

学位论文作者签名：

日期： 年 月 日

上海交通大学

学位论文使用授权书

本人同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。

本学位论文属于：

公开论文

内部论文，保密 1 年/ 2 年/ 3 年，过保密期后适用本授权书。

秘密论文，保密 ____ 年（不超过 10 年），过保密期后适用本授权书。

机密论文，保密 ____ 年（不超过 20 年），过保密期后适用本授权书。

（请在以上方框内选择打“√”）

学位论文作者签名：

日期： 年 月 日

指导教师签名：

日期： 年 月 日

摘 要

在自动驾驶的落地过程中，车路一体化系统是尤为重要的一环。车路一体化是指利用无线网络，将车端与路端紧密相连，实现车端与路端的信息交换、信息共享。目前，车辆终端导航定位主要依赖于全球卫星导航系统（GNSS），但该系统受限于卫星相关误差、传播途径相关误差、接收机相关误差等限制，对车辆的定位、测速精度有限。

针对这一问题，我们希望在路侧单元装备激光雷达与相机，通过基于激光雷达与视觉融合的目标跟踪方法，完成车辆的路侧定位。然后将定位结果发回车端，结合车端的 GNSS 定位结果，通过基于优化的方法得到车辆位置的预测值，补偿 GNSS 在脆弱场景下的定位误差，达到导航增强的目的。

我们希望采用 EagerMOT 作为目标跟踪方法，并在 KITTI^[1]数据集上验证了其效果。我们在由清华大学与百度公司发布的 DAIR-V2X 公开数据集上训练了二维与三维目标检测器。其中，二维目标检测器采用了 YOLOv4 框架，三维目标检测器采用了 PointRCNN 框架。最后我们提出了基于因子图优化的车辆定位补偿方法。

关键词：GNSS，路侧单元，目标跟踪，目标检测，因子图，导航增强

ABSTRACT

In the process of autonomous driving, Vehicle-Road Integration System is a particularly important part. Vehicle-Road Integration System refers to the use of wireless networks to closely connect the vehicle end and the road end to realize information exchange and information sharing. At present, the navigation and positioning of vehicle mainly depends on the global satellite navigation system (GNSS). But the system is limited by satellite-related errors, propagation path related errors and receiver relevant errors, thus limiting the accuracy of vehicle positioning and speed measurement.

In response to this problem, we hope to equip roadside units with LIDARs and cameras. And positioning the vehicles on the road end by object tracking method based on the fusion of LIDAR and vision. Then the positioning results will be sent back to the vehicle end, combined with the GNSS positioning results on the road end, predicting the vehicle position based on optimization method. This approach will compensate the positioning errors of GNSS in fragile scenes, and achieve the purpose of navigation enhancement.

We hope to adopt EagerMOT as the object tracking method. Its performance is evaluated on KITTI dataset. We trained 2D and 3D object detectors on the DAIR-V2X public dataset released by Tsinghua University and Baidu Inc. The 2D object detector adopts YOLOv4 framework and the 3D object detector adopts PointRCNN framework. Finally, we propose a vehicle positioning compensation method based on factor graph.

Key words: GNSS, roadside unit, object tracking, object detection, factor graph, navigation enhancement

目 录

摘 要	I
ABSTRACT	II
第一章 绪论	1
1.1 研究背景与研究意义	1
1.2 国内外研究现状	2
1.2.1 车路协同研究现状	2
1.2.2 激光雷达目标跟踪研究现状	3
1.2.3 视觉目标跟踪研究现状	4
1.2.4 激光雷达与视觉融合研究现状	5
1.3 研究内容与研究路线	6
1.3.1 研究路线	6
1.3.2 研究内容	6
1.4 本章小结	7
第二章 公开数据集框架	8
2.1 车路协同数据集综述	8
2.2 数据集标注与标定	9
2.2.1 KITTI 数据集介绍	9
2.2.2 KITTI 数据集标注格式	10
2.2.3 KITTI 数据集标定格式及坐标转换	11
2.2.4 DAIR-V2X 标注与标定	13
2.3 本章小结	14
第三章 目标跟踪器搭建	15
3.1 EagerMOT 原理	15
3.1.1 检测器融合	15
第四章 目标检测器训练	16
4.1 2D 目标检测器	16
4.2 3D 目标检测器	16
4.3 本章小结	16

第五章 基于优化的路侧导航增强原理..... 17

 5.1 因子图及其增量非线性最优化方法 17

 5.2 基于路侧导航增强的车辆可靠位置推理 17

 5.3 本章小结 17

第六章 全文总结 18

参 考 文 献..... 19

附录 22

攻读学位期间学术论文和科研成果目录..... 23

致 谢 24

第一章 绪论

1.1 研究背景与研究意义

“十三五”期间我国综合交通运输发展取得了显著成效，但与经济社会高质量发展的总体要求相比，仍存在智慧交通发展水平不高、交通基础设施数字化建设亟待加快、交通运输与新技术的融合尚不充分等问题，距离交通强国建设尚有一定差距。随着北斗三号全球卫星导航系统等核心空间基础设施的开通服务、新一代信息技术的发展，构建自主可信、国际领先的时空综合服务体系具备基本条件。2022 国家重点研发计划《广域交通可信导航信号与时空服务系统关键技术》应运而生。本课题主要的研究内容响应了该项目的课题 3 “高精泛源时空感知网络及车路一体化信息融合技术”。

在自动驾驶的落地过程中，车路一体化系统是尤为重要的一环。车路一体化是指利用无线网络，将车端与路端紧密相连，实现车端与路端的信息交换、信息共享。目前，车辆终端导航定位主要依赖于全球卫星导航系统（GNSS），但该系统受限于卫星相关误差、传播途径相关误差、接收机相关误差等限制，对车辆的定位、测速精度有限。以 GPS 为例，近年来，其定位精度（水平，圆概率精度，CEP）达到了 2.3 米，其测速精度达到了 0.2 m/sec（95% 置信度）。路侧终端由于可以预先铺设，其位置，硬件设备都是预先获得精确信息的。且路侧终端由于视野开阔，硬件资源丰富，计算能力强大，可以很好的解决车端对自身定位能力不足的问题。

我们希望在路边设置一种新型路侧单元，利用激光雷达与相机作为传感器对路过车辆进行定位。将定位结果发送回车端，结合车端 GNSS 定位结果，基于优化方法，预测车辆位置，最终实现导航增强的功能。本课题对卫星导航脆弱场景下的车辆安全具有重要意义。

1.2 国内外研究现状

1.2.1 车路协同研究现状

车路协同技术是通过无线通讯技术将车端、路端有机结合起来,实现交通环境数据信息的交换共享,信息处理,从而可以为车辆提供更精确的感知环境信息。

1950 年代末,通用汽车在新泽西州打造了一条埋入大量通信设备的高速公路。这在当时引起轰动,也是车路协同产业发展的雏形。加州 PATH 计划 (Partner for Advanced Transit and Highways) 成立于 1986 年,由加州交通部和加州大学伯克利分校合作建立,是北美第一个专注于现在称为智能交通系统 (Intelligent Transportation Systems, ITS) 主题的组织。其目标是应用电子、通信与自动化等新兴先进科技,增加高速公路的容量与安全性,减少交通堵塞、空气污染和能源消耗。该计划一直参与自动化公路与自动驾驶车辆的研究、发展与测试之中。被连接的车辆可以与其他车辆或者交通设施,如交通信号灯,进行通信。^[2]

车用无线通信技术 (Vehicle to Everything, V2X) 是将车辆与一切事物相连接的新一代信息通信技术,其中 V 代表车辆,X 代表任何与车交互信息的对象,当前 X 主要包含车、人、交通路侧基础设施和网络。借助人,车,路,云平台的全方位连接与信息交互,V2X 可以提升行驶安全,提高交通效率,提供出行信息服务,支持实现自动驾驶等等。大约在 2016 年前后,美国基于 DSRC (Dedicated Short Range Communications) 的 V2X 协议栈基本制定完毕,并有丰田、通用先后量产支持 DSRC 的汽车。2022 年 12 月 13 日,代表整个美国智能交通行业的十大组织联合发出声明,重申了对快速部署 V2X 的支持,并认为 2023 年是 V2X 部署的关键年。

我国工信部早在 2018 年就明确将 5920MHz-5925MHz 划分给 C-V2X (Cellular Vehicle-to-Everything) 并明确表明 C-V2X 是我国唯一使用的技术路线。2020 年发改委联合工信部等其他 10 个单位发布了《智能汽车创新发展战略》,提到“结合 5G 商用部署,推动 5G 与车联网协同建设;开展特定区域智能汽车测试运行及示范应用,验证车辆“人-车-路-云”系统协同性等,支持优势地区创建国家车联网先导区”。

1.2.2 激光雷达目标跟踪研究现状

三维激光雷达是一种主动探测式传感器，它向外发出激光束，返回探测到物体的点云数据信息，从而精确地获得探测到物体的距离信息。激光雷达目标跟踪往往是采用基于检测的跟踪范式进行的，其步骤可以大致划分为目标检测，状态预测，数据关联，状态更新共四个部分。其中，目标检测往往用现有的 SOTA (state of the art) 业界最前沿检测器。状态预测往往用平滑和与滤波方法，根据当前帧的目标的运动状态与位姿，预测下一帧中目标的运动状态与位姿。数据关联将预测的状态值与检测的目标状态匹配在一起，是最重要的步骤。状态更新则根据数据关联的结果，对当前帧目标的运动状态与位姿进行更新。

在目标检测阶段，国内外基于神经网络贡献了不少优秀的算法与思路，按照思路可以分为基于点云的方法，基于体素的方法，基于截面图的方法。^[3]

基于点云的方法直接输入点云进行目标分类分割任务。由于点云是三维不规则信息，典型的卷积网络不方便对点云直接处理。过去的方法大多是划分为规律的空间网格，或者投影到某一截面，从而利用卷积网络处理。但这样将引入冗余信息，并损害原始数据的自然特征。2017 年，Charles R. Qi 的团队提出了 PointNet 网络^[4]，直接处理点云，考虑到了点云的无序性、空间相关性与旋转不变性。用空间变换网络 (spatial transformer network) 将点云正则化预处理，也即在空间中旋转对齐。用多层感知机 (MLP) 将数据从低维投射到高维，避免池化操作中信息损失过多。再用对称性的 Max 池化函数解决点云输入的无序性，提取出一个高维向量作为全局特征，以此为基础进行后续的分割分类。同年，该团队还提出 PointNet++ 网络^[5]，在 PointNet 的基础上，进一步考虑了点云的局部信息。首先，利用最远点采样法 (FPS) 对整个点云进行局部采样，选出若干个中心点。再为每个中心点在一定半径的局部区域内选择 k 个临近点。最后对每一个局部区域都用 PointNet 提取局部特征，并以此为基础进行后续的分割分类任务。2019 年，香港中文大学的 Shi S 团队提出了 PointRCNN 模型^[6]，该模型是首个两阶段的基于点云的网络。在第一阶段，模型将整个场景分割为前景点与背景点，对前景点特征提取后生成预选框。在第二阶段通过置信度预测与预选框优化获得最终的检测结果。

基于截面图的方法的代表作是 Chen X 等人在 2017 年提出的 MV3DNet

模型^[7]，该模型同时融合了点云的剖视图特征与 RGB 图片特征，分别在点云的俯视图、前视图与 RGB 图像中提取特征。在俯视图特征中计算候选区域后投影到前视图和图像中，经过兴趣区域 (ROI) 整合到同一纬度再输入网络中融合。

基于体素的方法的代表作是 2018 年 Y. Zhou 等人提出的 VoxelNet 模型^[8]。首先将点云划分为等间距的 3D 体素，经过了点的随机采样与归一化处理后，又引入了体素特征编码器 (Voxel Featured Encoding, VFE)，每个非空体素都进行了特征提取。然后，这些特征被输入 3D 卷积神经网络 (Convolutional Middle Layers) 进行特征抽象。最后通过提议区域生成网络 (Region Proposal Network, RPN) 进行目标检测。

在数据关联步骤中，常见的算法有全局最近邻法 (Global Nearest Neighbor, GNN)、多假设跟踪算法 (Multiple Hypothesis Tracking, MHT)、联合概率数据关联 (Joint Probability Data Association, JPDA) 等，该方案的目标匹配精度高，但匹配速度慢、计算成本高。2020 年，Weng X 等人提出了 AB3DMOT 算法^{[9][10]}，该算法将匈牙利匹配算法与卡尔曼滤波估计结合，实现了快速的目标关联。

1.2.3 视觉目标跟踪研究现状

单目视觉跟踪早期主要采用传统的滤波方法，如卡尔曼滤波 (Kalman Filter)、粒子滤波 (Particle Filter)、均值漂移 (Meanshift) 等。近年来流行的研究框架可以分为相关滤波 (Correlation Filter) 框架与孪生网络 (Siamese Network) 框架两个大方向。

2010 年，Bolme 团队首次将相关滤波方法用在了跟踪领域，提出了误差最小平方和滤波器 (Minimum Output Sum of Squared Error filter, MOSSE)^[11]，用最小化均方误差的思路产生滤波器，进而获得跟踪目标的新位置。2012 年，Henriques 等人在 MOSSE 的基础上提出了循环结构检测方法 (Circulant Structure with Kernel)^[12]，一方面修改损失函数为岭回归形式，再引入核函数求解，另一方面引入循环矩阵，达到密集采样与提高运算效率的效果。2014 年，Henriques 等人提出了核相关滤波算法 (Kernel Correlation Filter, KCF)^[13]，该方法在 CSK 算法的基础上，用方向梯度直方图 (Histogram of Oriented Gradient, HOG) 多通道特征替换了原来的单通道灰度特征，并采用了高斯

核函数求解岭回归问题。同一年, Danelljan 等人提出了 DSST(Discriminative Scale Space Tracker)^[14]方法, 在 MOSSE 的基础上, 用两个滤波器分别应对尺度和位置的变化, 同时引入 HOG 特征。

孪生网络框架, 采用两个成对的结构一样的神经网络, 网络之间共享权值、参数等信息。该框架可以接受两个输入, 并对两个输入进行相同的变换, 通过比较输出的两个结果的欧氏距离判断输入之间的相似性。这一思路最早在 1993 年被 J Bromley 应用在美国支票的签名验证场景上^[15]。在目标跟踪问题里, 一个输入可以是初始帧的目标区域, 以此作为模板, 而将后一帧中的候选区域作为第二个输入。孪生网络要做的就是找到两帧间相似度最高的候选区域, 如这一系列的开山之作 SiamFC 框架^[16]。

1.2.4 激光雷达与视觉融合研究现状

传感器信息融合可以分为像素级、特征级和目标级融合^[17]。像素级融合直接融合原始数据, 获取的细节信息最丰富, 因而其准确性和鲁棒性最好。但是像素级融合的前提是信息来自于同类传感器或者传感器拥有同样的量级, 需要传感器之间进行高精度匹配, 因而实时性较低。特征级融合先提取原始数据的特征, 再融合多个特征, 根据目标已有特征对融合特征进行匹配, 获得目标的信息。目标级融合先提取原始数据中的目标信息, 然后融合多个目标信息, 得到最终完整信息。其只对目标信息进行融合, 不受传感器类别的限制, 能够保证实时性。

EagarMOT 是 Aleksandr Kim 等人 2021 年提出的融合框架^[18], 用现成的 2D 与 3D 检测器先得到目标的 2D 与 3D 检测结果, 再用交比 IoU 将同一个目标的两个结果关联在一起得到一个融合实例。在数据关联部分, 该方法采用两阶段进行, 第一阶段对 3D 检测结果和预测结果进行数据关联, 对于没有匹配上的实例和轨迹再进入第二阶段匹配, 利用 2D 检测结果进行数据关联。

1.3 研究内容与研究路线

1.3.1 研究路线

我们希望在路边设置一种新型路侧单元，上面装备了相机、激光雷达等传感器。利用激光雷达的测距测向功能获得深度信息，利用相机的色彩捕捉功能获得图像信息，通过 2d-3d 融合的多目标跟踪方法，获得车辆相对路侧传感器的相对位置姿态。然后，通过路侧与车端的可靠通信，将位姿传回车端，车端利用其作为新增的独立位置约束，结合误差检测与估计理论，提高车辆当前位置估计的精确度，提高车辆导航的可信性。如下图所示：

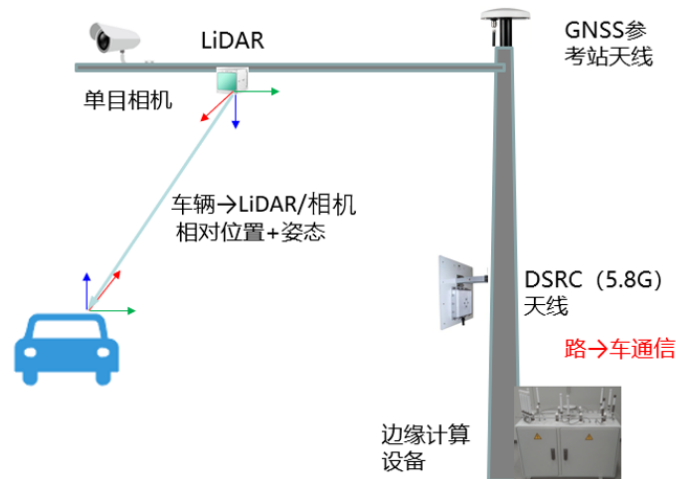


图 1-1 路侧单元架设示意图

但受到实验条件限制，目前硬件平台尚未搭建完全，作者也身处国外，所以只能依托公开数据集进行实验。

1.3.2 研究内容

本文主要依托现有公开数据集，搭建了目标跟踪器，训练了二维与三维检测器，并提出了一种基于因子图的误差补偿方法。本文主要内容如下所示：

第一章为绪论。对本课题的研究背景与意义进行论述，回顾了车路一体化概念的发展历史，对众多国内外基于激光雷达和基于视觉的目标跟踪

方法进行阐述，分析了国内外基于传感器融合的目标跟踪方法，并确定了本文的研究路线。

第二章为现有公开数据集框架。首先总结了近年来适用于自动驾驶与车路协同的公开数据集。再介绍了 KITTI 数据集和 DAIR-V2X 数据集的传感器配置，标签与标定格式，其中包括了激光雷达与摄像头的标定原理及其坐标转换关系。

第三章为目标跟踪器搭建。首先介绍了 EagerMOT 目标跟踪框架的原理，再介绍了目标跟踪常用的评测方法，最后在 KITTI 数据集上复现了该框架，并与论文结果进行了对比。

第四章为目标检测器训练。首先介绍了 2D 检测器 YOLOv4 的原理，训练环境与网络配置，训练结果。再介绍了 3D 检测器 PointRCNN 的原理，训练环境与网络配置，训练结果。最后对检测器训练结果进行了分析。

第五章为基于优化的路侧导航增强原理，首先介绍了基于因子图的优化方法，再针对本课题场景提出了路侧导航增强原理。

第六章为总结与展望，首先总结本文的主要工作，然后展望后续的研究与改进工作。

1.4 本章小结

本章介绍了研究背景与意义，国内外研究现状，研究内容与研究路线。

第二章 公开数据集框架

2.1 车路协同数据集综述

高质量的路侧单元数据集具有重要的工业价值，可以加速路侧的车辆检测模型在车路协同中的迭代优化，在促进创新的学术研究方面也扮演着重要角色。近年来，基于激光雷达与相机的目标检测与跟踪数据集被发布，例如 KITTI^[19]，ApolloScape^[20]，Waymo^[21]，NuScenes^[22]等。

但是，这些数据集往往基于车载传感器，针对自动驾驶场景。而基于路侧单元视角收集的公开数据集则相对稀少，尤其是包含了 3D 点云数据的数据集，这也为我们的研究工作带来了一定的挑战。自去年以来，一些基于路侧激光雷达与相机的数据集陆续发布，这些较新的数据集被总结如下：^[23]

表 2-1 基于路侧传感器收集的公开数据集

数据集	年份	激光雷达	相机	交通场景	天气
IPS300+ ^[24]	2022	2×Robosense Ruby-Lite	2 RGB	城市	白天/夜晚
DAIR-V2X ^[25]	2022	1 300-beam LIDAR	1 RGB	城市高速公路	晴/雨/雾 白天/夜晚
A9-Dataset ^[26]	2022	1 Ouster-OS1 64-beam LiDAR	1 RGB	Autobahn 高速公路	白天

IPS300+: 为推动路侧多模态感知研究在合作车辆基础设施系统，Wang 等人^[24]在 2022 年发布了一个双模态数据集。该数据集配备了路侧激光雷达和摄像头，收集场景是一个城市路口，占地面积 3000 平方米，覆盖半径 300 米。两个感知单元 (IPU) 被安装在交叉路口的对角线上，距离用于数据采集的区域地面 5.5 米。每个感知单元由一个 80 束 RoboSense Ruby-Lite LiDAR 和两个 Sensing-SG5 彩色摄像头组成。数据集包括涵盖不同时间的 14198 帧据。被两个 IPU 收集的每一帧点云数据，都被存储为单个 PCD 文件用于标注。每帧都有平均 319.84 个标签，包括行人、骑车人、三轮车、汽车、公共汽车、卡车和工程车辆等。

DAIR-V2X: 为了加速车路协同自动驾驶的计算机视觉研究和创新，Yu

等人^[25]于 2022 年发布了 DAIR-V2X 数据集。该数据集采集自北京高级别自动驾驶示范区 10 公里的城市道路、10 公里的高速公路和 28 个路口。在 28 个路口各部署了四对 300 束路侧激光雷达和高分辨率摄像头。DAIR-V2X-I 是 DAIR-V2X 的子集，专用于路侧协同感知，包含 10084 帧图像，分别联合标注了图像和路边激光雷达点云数据。注释器详尽地标记了每个图像和点云帧中的 10 个对象类别中的每一个目标，包括不同的车辆、行人和骑车人。

A9-Dataset: 2022 年，Christian 等人^[26]展示了基于德国慕尼黑附近 3 公里长 Providentia++ 试验场路边传感器的 A9 数据集。传感器包括摄像头、雷达和 64 束 Ouster 激光雷达。它们被安装在龙门桥和桅杆上，提供道路景观。数据集提供标记图像和多个路段的激光雷达点云和白天 A9 高速公路上密集交通的不同角度记录。版本 R0 由 1098 个带标签的帧和 14,459 个带标签的 3D 对象组成，包括汽车、拖车、卡车、货车、行人、公共汽车、摩托车、自行车等九类对象。

以上数据集中，IPS300+ 数据集需要申请授权使用，作者较晚才从数据集作者处取得授权，因此没有来得及用于实验。DAIR-V2X 是一个目标跟踪数据集，其帧与帧之间没有时序联系，只能作为检测器的训练数据，不能验证跟踪器的性能。A9-Dataset 虽然为连续采集的数据集，但是激光雷达与相机采集的数据在时间上是分开的，并非对同一场景在同一时间段内采集，无法进行数据融合步骤，且作者所在地区无法注册数据集官网账号。

因此，我们暂时在 DAIR-V2X 数据集上训练检测器，在配套生态健全的 KITTI 数据集上验证跟踪器性能。

2.2 数据集标注与标定

2.2.1 KITTI 数据集介绍

KITTI^[1] (Karlsruhe Institute of Technology and Toyota Technological Institute) 是用于移动机器人和自动驾驶的最受欢迎的数据集之一。该数据集的主要目的是推动以自动驾驶为目标的计算机视觉和机器人算法的发展。它包含多种传感器模式记录的数小时交通场景，包括 2 个高分辨率 RGB、2 个灰度立体相机和激光雷达，高精度 IMU/GPS 导航系统。其示意图如 2-2 所

示，由于我们仅关注激光雷达与相机传感器，所以仅画出这两个传感器的坐标系。

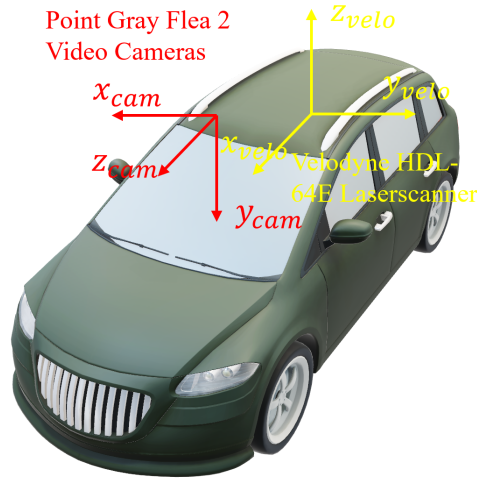


图 2-2 KITTI 数据集采集平台示意图

KITTI 数据集包括了目标检测数据集与目标跟踪数据集，前者为采集时间离散的目标数据集，后者包含了连续的目标信息。我们选择了后者作为跟踪框架的验证集。KITTI 数据集分为训练集与测试集。区别在于前者除了传感器数据与标定文件外还提供标注文件，即真值，用于检测器，跟踪器的训练，后者仅提供传感器数据与标定文件，其真值储存在官方服务器中。目的是防止模型训练者直接在测试集上训练。训练集包含了 20 个时间序列，测试集包含了 28 个时间序列。每个序列将每一帧的图片与点云信息分别储存在 image_01, image_02 与 velodyne 文件夹中。其中 image_01 为左侧彩色相机的采集数据，image_02 为右侧彩色相机的采集数据，二者结合可以进行双目视觉的研究。

2.2.2 KITTI 数据集标注格式

KITTI 数据集将每一帧的标注以 txt 形式储存在了 label 文件夹下，每一行为一个目标信息。每一行以空格为分隔符，共 15 列，表示信息如表 2-2 所示：

表 2-2 KITTI 数据集标注格式

列数	意义	备注
1	类型	共有 Car, Van, Truck 等 8 种类型
1	截断程度	从 0 至 1, 代表了目标被图片边框的截断程度
1	遮挡程度	取整数集 (0,1,2,3), 标志了目标被其他物体遮挡的程度
1	alpha	目标在激光雷达坐标系下位置向量与 x 轴的夹角
4	2d 包围框	目标在图像坐标系下的包围框对角坐标 (x_1, y_1, x_2, y_2)
3	3d 尺寸	目标的三维包围框的大小, 即高度, 宽度, 长度
3	3d 位置信息	目标在激光雷达坐标系下的坐标 (x, y, z)
1	rotation_y	目标自身朝向与激光雷达坐标系 x 轴夹角

2.2.3 KITTI 数据集标定格式及坐标转换

数据集的标定包括了相机内参标定, 外参标定, 雷达-相机标定等部分。

如图2-3所示, 在 KITTI 数据集中一共定义了 4 个坐标系, 激光雷达坐标系 (velodyne coordinate), 相机坐标系 (camera coordinate), 修正的相机坐标系 (rectified camera coordinate), 图像坐标系 (image coordinate)。

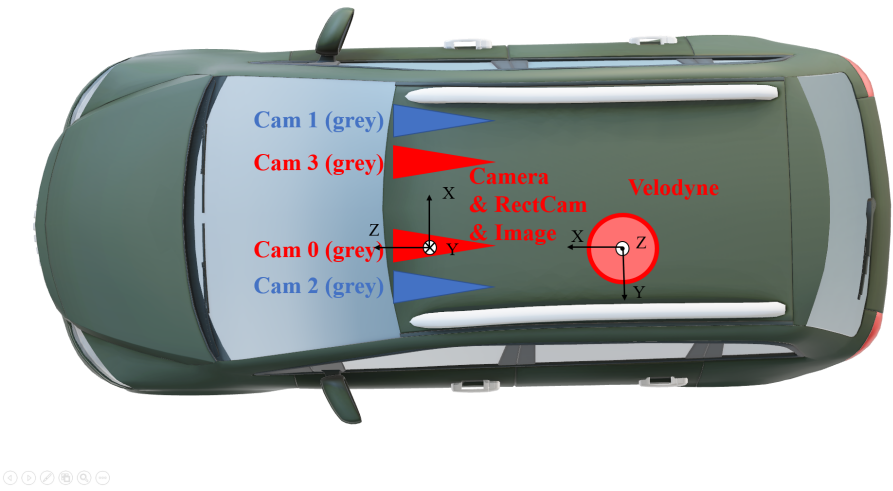


图 2-3 KITTI 数据集坐标系示意图

假设空间中一个点在激光雷达坐标系中的坐标为

$$X_{\text{velo}} = (x_{\text{velo}}, y_{\text{velo}}, z_{\text{velo}}, 1)^T \tag{2.1}$$

在相机坐标系坐标下的坐标为

$$X_{\text{cam}} = Tr_{\text{velo_to_cam}} X_{\text{velo}} \quad (2.2)$$

其中, $Tr_{\text{velo_to_cam}} = [R|t]$, R 为坐标系旋转矩阵, t 为原点间的平移向量。

KITTI 数据集中, 四个相机的图像平面原点的排列与激光雷达坐标系的 y 轴是平行的, 但平面之间并非共面, 为了让四个相机的图像平面共面, 需要将每个相机坐标系进行旋转, 旋转矩阵为 $R_{\text{rect}}^{(i)}$, 其中 i 为四个相机的编号。旋转后, 我们得到了修正的相机坐标系, 在此坐标系中, 该点坐标标记为 $X_{\text{rect_cam}}$, 满足

$$X_{\text{rect_cam}} = R_{\text{rect}}^{(i)} X_{\text{cam}} \quad (2.3)$$

将 0 号修正相机坐标系中的三维坐标系投影至图像坐标系 Y 关系为

$$Y = P_{\text{rect}}^{(i)} X_{\text{rect_cam}} \quad (2.4)$$

其中 $P_{\text{rect}}^{(i)}$ 为第 i 个相机的投影矩阵

$$P_{\text{rect}}^{(i)} = \begin{pmatrix} f_u^{(i)} & 0 & c_u^{(i)} & -f_u^{(i)} b_x^{(i)} \\ 0 & f_v^{(i)} & c_v^{(i)} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (2.5)$$

投影矩阵 P_{rect} 是相机内参 K 与相机外参 $[R|t]$ 的乘积。

$$P_{\text{rect}} = K[R|t] \quad (2.6)$$

为了简洁起便, 此处省略了相机编号 i 。

相机内参 K 是相机坐标系下一个点的坐标到相机图像坐标系的变换矩阵, 包含了相机像素点的长宽 u, v , 相机坐标系原点在图像坐标系下的坐标 (u_0, v_0) , 相机焦距 f 等信息。

$$K = \begin{pmatrix} f_u^{(i)} & 0 & c_u^{(i)} & -f_u^{(i)} b_x^{(i)} \\ 0 & f_v^{(i)} & c_v^{(i)} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad (2.7)$$

式中

$$f_u^{(i)} = \frac{f}{u} \quad f_v^{(i)} = \frac{f}{v} \quad c_u^{(i)} = u_0 \quad c_v^{(i)} = v_0 \quad (2.8)$$

相机外参 $[R|t]$ 是世界坐标系到相机坐标系的变换矩阵。在 KITTI 数据集中，由于相机坐标系已作过修正，外参中旋转矩阵 $R = I$ ， I 为单位矩阵。由于 4 个相机沿着相机坐标系 x 轴排列，平移向量 t 仅在 x 轴方向有分量 $-b_x^{(i)}$ ，即

$$t = (b_x^{(i)}, 0, 0)^T \quad (2.9)$$

结合式(2.6)可以发现第一列最右侧为 $-f_u^{(i)} b_x^{(i)}$ 。注意(2.6)中的相机内参 K 由 3×3 矩阵被补齐为了 4×4 矩阵， $K(4,4) = 1$ ，其余为 0。

2.2.4 DAIR-V2X 标注与标定

DAIR-V2X 数据集在路侧单元架设了 1 个 RGB 相机与 1 个激光雷达，如图2-4所示

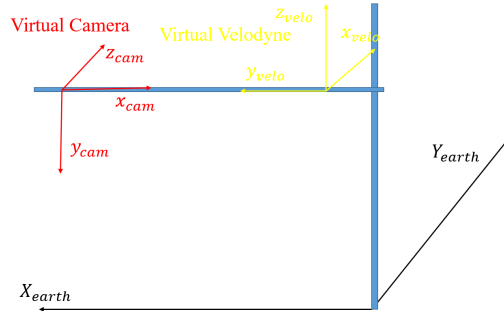


图 2-4 DAIR-V2X 数据集坐标系示意图

现实中激光雷达和相机与地面均有大约 11° 的夹角，这导致目标激光雷达坐标系中的高度与现实中的高度不对应，给研究带来了一定的麻烦。数据集为方便起见，沿着平行于地面的方向建立了虚拟的 x 轴和 y 轴，并以此建立了虚拟激光雷达坐标系，方便研究。

但论文作者并未将相机坐标系 $x-y$ 平面旋转至与地面平行，这使得相机坐标系与激光雷达坐标系之间不仅仅只有 $x-y$ 平面上的旋转。这一点与

KITTI 的基本设置不符合。由于主流的可视化工具，目标检测器，目标跟踪器不少在 KITTI 上开发，这小小的区别会导致可视化结果，检测器训练出现异常。如图2-5所示，直接在 DAIR-V2X 数据集进行可视化或检测器训练时，检测器接受的 3D 包围框是与真值有角度偏差的。

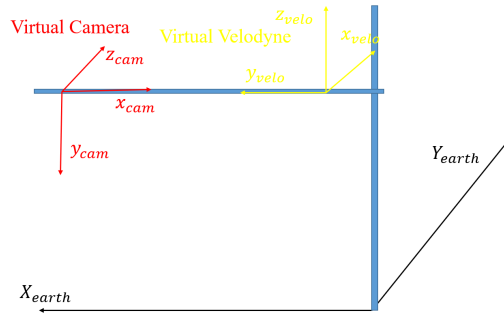


图 2-5 DAIR-V2X 数据集坐标系示意图

因此我们利用标定文件，计算出相机坐标系 $x-y$ 平面与地面的夹角，将其旋转至水平，重新生成标注与标定文件。

2.3 本章小结

本章总结了近年来适用于自动驾驶与车路协同的公开数据集。再介绍了 KITTI 数据集和 DAIR-V2X 数据集的传感器配置，标签与标定格式，其中包括了激光雷达与摄像头的标定原理及其坐标转换关系。

第三章 目标跟踪器搭建

为了实现路侧单元对车辆的对目标跟踪任务，我们采用了 EagerMOT 框架。由于缺少合适的路侧目标跟踪数据集，我们在 KITTI 数据集上对该框架进行了性能测试。

3.1 EagerMOT 原理

EagerMOT 框架结合了互补的 2D 和 3D（例如 LiDAR）目标信息，这些信息是从预训练的物体检测器中获得的。框架的总体概述如图所示。作为每一帧的输入，我们的方法采用一组 3D 边界框检测 3dDt 和一组二维检测 2dDt。然后，观察融合模块 (i) 将来自 2D 和 3D 的检测关联起来相同的对象，(ii) 两阶段数据关联模块跨时间关联检测，并且，基于可用的检测信息（全 2D+3D，或部分）我们更新跟踪状态和 (iv) 我们采用简单的跟踪管理初始化或终止轨道的机制。该公式允许所有检测到的对象与轨迹相关联，即使它们在图像域或 3D 传感器。这样，我们的方法就可以从短遮挡中恢复并保持近似 3D 当其中一个探测器发生故障时的位置，而且重要的是，我们之前可以在图像域中跟踪远处的物体物体进入 3D 感应范围。一旦物体进入感应范围，我们可以顺利地初始化一个 3D 运动模型对于每个轨道。

3.1.1 检测器融合

a) 获得来自相机的 2D 视频输入与来自激光雷达 3D 信息流的检测结果与. b) 对多模态的检测结果进行数据关联，采用贪心算法，将 3D 检测结果投影到 2D 图像域上，计算二者的交比 (Intersection of Union)。c) 将所有可能的检测配对按照交比进行排序，按照交比从高到低的顺序，当 2D 与 3D 检测结果均尚未配对成功，且交比大于预先设置的阈值 (threshold) 时，产生融合实例. d) 以此得到融合的检测实例集合. 其中，每个实例包含了来自 2D 检测器的 2D 边界框 (bounding box) 与来自 3D 检测器的 3D 位置与 3D 边界框，因此也属于两个单模态检测实例集合与, 即且 e) 将剩下的未匹配的检测结果与作为部分观测结果 (partial observation) 加入检测实例集合中与对应的与中。这种思路虽然简单，但是过去的经验证明了它的鲁棒

性。

第四章 目标检测器训练

4.1 2D 目标检测器

4.2 3D 目标检测器

4.3 本章小结

第五章 基于优化的路侧导航增强原理

5.1 因子图及其增量非线性最优化方法

5.2 基于路侧导航增强的车辆可靠位置推理

5.3 本章小结

本文……

第六章 全文总结

参 考 文 献

- [1] 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.
- [2] 邢亚男. 车路协同感知融合研究[D/OL]. 吉林大学, 2022. DOI: [10.27162/d.cnki.gjlin.2022.006243](https://doi.org/10.27162/d.cnki.gjlin.2022.006243).
- [3] MAO J, SHI S, WANG X, et al. 3d object detection for autonomous driving: a review and new outlooks[A]. 2022.
- [4] 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.
- [5] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in neural information processing systems, 2017, 30.
- [6] SHI S, WANG X, LI H. Pointcnn: 3d object proposal generation and detection from point cloud[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 770-779.
- [7] 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.
- [8] 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.
- [9] WENG X, WANG J, HELD D, et al. Ab3dmot: A baseline for 3d multi-object tracking and new evaluation metrics[A]. 2020.

- [10] WENG X, WANG J, HELD D, et al. 3d multi-object tracking: A baseline and new evaluation metrics[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2020: 10359-10366.
- [11] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]//2010 IEEE computer society conference on computer vision and pattern recognition. IEEE, 2010: 2544-2550.
- [12] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[C]//Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part IV 12. Springer, 2012: 702-715.
- [13] HENRIQUES J F, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE transactions on pattern analysis and machine intelligence, 2014, 37(3): 583-596.
- [14] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Convolutional features for correlation filter based visual tracking[C]//Proceedings of the IEEE international conference on computer vision workshops. 2015: 58-66.
- [15] BROMLEY J, GUYON I, LECUN Y, et al. Signature verification using a” siamese” time delay neural network[J]. Advances in neural information processing systems, 1993, 6.
- [16] BERTINETTO L, VALMADRE J, HENRIQUES J F, et al. Fully-convolutional siamese networks for object tracking[C]//Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part II 14. Springer, 2016: 850-865.
- [17] 王思信. 基于三维激光雷达与视觉融合的车辆跟踪与驾驶行为研究[D]. 武汉理工大学, 2020.
- [18] KIM A, OŠEP A, LEAL-TAIXÉ L. Eagermot: 3d multi-object tracking via sensor fusion[C]//2021 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2021: 11315-11321.

- [19] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 3354-3361.
- [20] HUANG X, CHENG X, GENG Q, et al. The apolloscape dataset for autonomous driving[C]//Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2018: 954-960.
- [21] SUN P, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: Waymo open dataset[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 2446-2454.
- [22] CAESAR H, BANKITI V, LANG A H, et al. nuscenes: A multimodal dataset for autonomous driving[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 11621-11631.
- [23] SUN P, SUN C, WANG R, et al. Object detection based on roadside lidar for cooperative driving automation: a review[J]. Sensors, 2022, 22(23): 9316.
- [24] WANG H, ZHANG X, LI Z, et al. Ips300+: a challenging multi-modal data sets for intersection perception system[C]//2022 International Conference on Robotics and Automation (ICRA). IEEE, 2022: 2539-2545.
- [25] YU H, LUO Y, SHU M, et al. Dair-v2x: A large-scale dataset for vehicle-infrastructure cooperative 3d object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 21361-21370.
- [26] CRESS C, ZIMMER W, STRAND L, et al. A9-dataset: Multi-sensor infrastructure-based dataset for mobility research[C]//2022 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2022: 965-970.

符号与标记（附录 1）

攻读学位期间学术论文和科研成果目录

[1] 张三, 李四. (已录用)

致 谢

致谢主要感谢导师和对论文工作有直接贡献和帮助的人士和单位。致谢言语应谦虚诚恳，实事求是。