

文章编号: 1002-1582(2026)01-0109-06

环视 BEV 多层级语义车位自适应加权配准算法

史兴领^{1,2}, 朱添翼^{2*}, 黄刘生¹, 檀杰², 赵晓东²

(1. 中国科学技术大学 计算机科学与技术学院, 安徽 合肥 230027)

(2. 科大国创合肥智能汽车科技有限公司 感知融合部, 安徽 合肥 230088)

摘 要: 车载 BEV 空间配准技术作为智能驾驶车辆的重要手段, 一直是自动驾驶领域研究的热点。在 BEV 空间, 如何自适应的提高配准精度, 成为自动驾驶 BEV 空间应用的难点。针对视觉数据深度估计误差会随距离累积、激光数据在低强度场景鲁棒性差和多源配准强依赖于时间同步精度, 时序偏差易引发错位等问题, 提出环视 BEV 多层级语义车位加权配准方法, 即基于环视 BEV 拼接图后的语义识别结果, 获取车位等语义特征信息, 对车位进行多层级跟踪, 确定配准初始位姿, 基于车载传感器成像精度与距离相关原理, 配准平差中引入基于欧式距离的自适应权重计算, 从而提高配准精度和鲁棒性。该方法极大提高了配准精度, 鲁棒性高, 能够较好的基于配准结果进行建图与定位。地下车库环视数据集的建图与定位结果验证了利用本算法进行 BEV 空间配准的精度与效果。

关 键 词: 环视 BEV; 多层级跟踪; 自适应加权; 配准

中图分类号: TP751

文献标识码: A

Multi-level semantic parking spot adaptive weighted registration algorithm for Surround-view BEV

SHI Xinling^{1,2}, ZHU Tianyi^{2*}, HUANG Liusheng¹, TAN Jie², ZHAO Xiaodong²

(1. University of Science and Technology of China, School of Computer Science and Technology, Hefei 230027, China)

(2. Guochuang Hefei Intelligent Automobile Technology Co., Ltd, Perception Fusion Department, Hefei 230088, China)

Abstract: In-vehicle BEV (Bird's Eye View) spatial registration technology is an important tool for intelligent driving vehicles and has consistently been a research focus in the field of autonomous driving. In the BEV space, adaptively improving registration accuracy has become a major challenge for BEV applications in autonomous driving. Issues such as the accumulation of depth estimation errors in visual data over distance, the poor robustness of LiDAR data in low-intensity scenarios, and the strong reliance of multi-source registration on time synchronization, where temporal deviations easily lead to misalignment are addressed. A surround-view BEV multi-level semantic parking space weighted registration method is proposed. Specifically, based on the semantic recognition results from the surround-view BEV stitched image, semantic feature information such as parking spaces is obtained, multi-level tracking of these parking spaces is carried out to determine the initial registration pose, and adaptive weights based on Euclidean distance are introduced into the registration adjustment according to the imaging accuracy of in-vehicle sensors and distance-related principles. This enhances registration accuracy and robustness. The method significantly improves registration precision and robustness, enabling effective mapping and localization based on the registration results. The mapping and localization results from an underground parking garage surround-view dataset validate the accuracy and effectiveness of BEV spatial registration using this algorithm.

Keywords: Surround view BEV; multi-level tracking; adaptive weighting; registration

收稿日期: 2025-05-29; 收到修改稿日期: 2025-12-07

基金项目: 安徽省科技攻关计划项目 (202423k09020037)

作者简介: 史兴领 (1971—), 男, 高级工程师, 硕士, 从事人工智能、自动控制、软件工程与算法以及智能网联汽车相关技术和产品的研发与项目管理工作。xlshi@ustcsoft.com

朱添翼 (1992—), 男, 中级工程师, 硕士, 从事摄影测量与遥感和多传感器融合应用研究。

* 通讯作者: zhu.ty@qq.com

0 引 言

近年来,自动驾驶相关技术通过不断的技术创新和算法突破,成为人工智能与计算机视觉领域的重点研究方向之一^[1]。通过摄像头和激光雷达等多传感器的融合感知,自动驾驶系统可实时构建车身周边三维环境特征,并为规划控制提供基础^[2,3]。传统基于前视摄像头的融合感知,受限于成像视角的影响,很难鲁棒的支持后续规划与控制。鸟瞰图(Bird's Eye View, BEV)由此应运而生,可用直观且用户友好的俯视坐标系的表达方式,有效融合周围环境信息^[4]。

在BEV感知技术中,BEV配准技术作为其重要处理方式之一,将车载多传感器的数据映射到鸟瞰图中进行配准和融合,有效克服多源和多时序数据一致性受限的问题。其中,在环视BEV配准(Surround-view BEV Registration)过程中,需要将多个摄像头的图像实时转换到BEV空间下进行处理,这对算法的鲁棒性提出了更高的要求^[5,6]。

针对BEV配准技术,国内外目前主要有三类解决方法^[7-12]。(1)是以纯视觉为主的技术进行BEV配准,该方法可视化效果较好,传感器成本可控,但深度估计误差会随距离累积,对配准算法要求较高;(2)是以激光雷达为主的技术进行BEV配准,该方法传感器精度较高,配准精度效果较好,但受限于激光反射强度,低强度场景鲁棒性较差;(3)是以多传感器融合进行BEV配准,该方法鲁棒性较好,适用于多种场景,但配准强依赖于传感器的时间同步等精度,时序偏差易引发错位等问题。

本文针对上述问题,提出一种环视BEV多层级语义车位自适应加权配准算法:该方法基于车载环视BEV拼接数据,基于语义分割^[13,14]结果进行配准点的选取,创新的在配准位姿初始化时引入多层级^[15-17]跟踪和配准过程中引入空间距离权重^[18,19],从而提高配准效果和定位精度。文中利用地下车库采集环视和自车状态数据,对提出的算法进行验证,证明了算法的有效性。

1 BEV感知与配准技术相关研究

Samani E U 和刘俊杰等^[7,8]结合传统透视变换或 transformer 等网络架构,将视觉鱼眼图像转换到车辆BEV空间下生成BEV图,并对BEV空间下的视觉图进行语义特征分割和深度估计,通过估计后的特征点位置和语义信息进行配准,该方法语义分割和可视化效果较好,传感器成本可

控,但方法依赖准确的深度信息,而深度估计误差会随距离累积,在距车速距离较远的场景下配准效果较差;刘天弋和Park G等^[9,10]基于激光雷达传感器获取到的三维点云,可直接获取车辆BEV空间下的二维坐标点,基于强度对关键点进行提取并做后续的坐标配准,该方法传感器精度较高,成功配准后的精度效果较好,但受限于激光反射强度,低强度场景鲁棒性较差;Kim M 和 Cai H 等^[11,12]结合上游视觉和激光数据的特点进行融合,结合视觉丰富的语义特征和激光雷达的高精度三维点云特性,取长补短,提升感知精度,基于融合后的结果进行配准,该方法鲁棒性较好,适用于多种场景,但配准强依赖于传感器的时间同步等精度,时序偏差易引发错位等问题。

2 研究方法

2.1 技术流程

对于视觉BEV图像,由于视觉图像精度在BEV空间下随距离的不同,其精度也不同,综合来说物方坐标距起始点越远,其在图像上的精度越低^[20]。同时在地下车库场景中依赖车辆行驶数据存在较大定位误差,在配准过程中初始化位姿误差较大将对配准误差产生较大影响^[21]。本文针对这些困难,通过对输入的BEV拼接图像进行语义特征点的选取,通过多层级语义车位跟踪初始化配准位姿,在车身坐标系下进行特征点配准权重的计算提高配准效果,在首帧全局坐标系进行BEV配准和位姿输出,优化行驶数据误差,提高建图和定位精度。具体流程如图1所示。

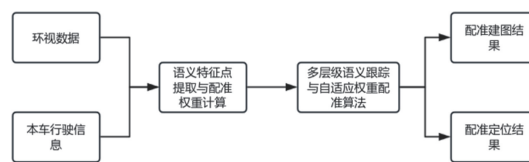


图1 配准技术概念流程

2.2 配准权重计算

配准权重计算是对加权欧式距离模型(Weighted Euclidean Distance Model, WEDM)^[22]的扩展,引入车辆中心点从而计算语义特征点在本车坐标系下的权重。

语义特征点权重计算分以下三个步骤:(1)提取语义特征点;(2)计算特征点WEDM值;(3)特征点权重赋值。

前端感知采用国产化芯片,采集车库数据进行车位标注进行训练和识别,提取语义特征点^[23,24]。

实验BEV图像中车位特征点感知结果如图2所示。



图 2 特征点配准权重结果

如图 2 所示,车位的四个角点分别是 $p_0-p_3(x, y)$,以中心黑色车辆中心为原点,横坐标为 x ,朝右为正,纵坐标为 y ,朝上为正。

采用一次方程计算其权重系数,设定 20cm 误差为最大限定值,超出 20cm 误差后权重设置为 0;

通过先验计算得出临界距离平均值为 d_l ;

每组坐标 (x, y) 其权重系数 α 计算公式为

$$\alpha = \begin{cases} 1 - \frac{\sqrt{x^2 + y^2}}{d_l}, & (\sqrt{x^2 + y^2} < d_l) \\ 0, & (\sqrt{x^2 + y^2} \geq d_l) \end{cases} \quad (1)$$

2.3 多层级语义跟踪与自适应权重配准算法

多层级语义跟踪与自适应权重配准算法详细流程如图 3 所示。

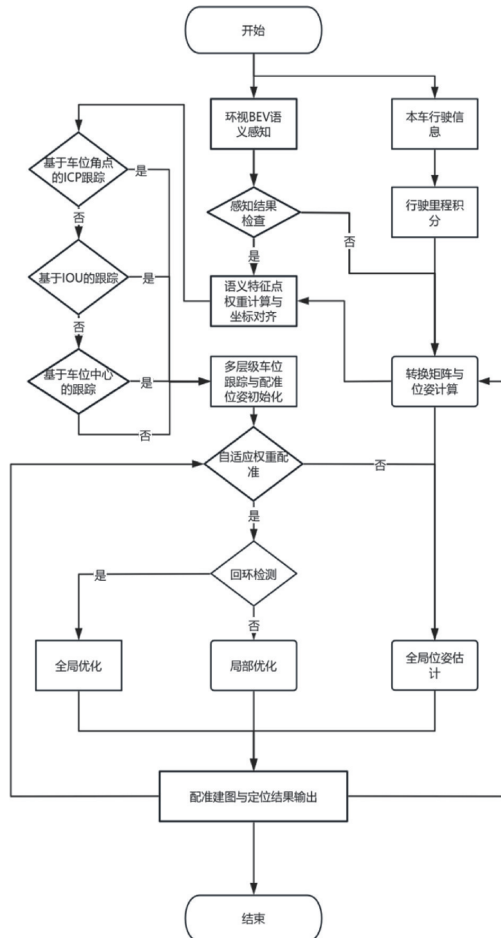


图 3 配准技术详细流程

环视车位语义主要定义为车辆当前状态下环视鸟瞰图 BEV (Bird's Eye View) 拼接空间下感知到的车位角点结果,车位角点如图 4 所示。



图 4 车位角点图

2.3.1 多层级语义跟踪改进算法

本文改进的加入三个层级语义跟踪,优化环视 BEV 车位跟踪效果较差的问题。语义特征点权重计算分以下三个步骤:(1) 最上层为厘米级高精度车位角点 ICP 跟踪;(2) 中层设定为分米级中精度车位 IOU 跟踪;(3) 最下层为米级低精度车位中心点跟踪。

对车语义输入结果,优先进行车位角点的临近点迭代 ICP (Iterative Closest Point) 跟踪,公式表示为

$$\min_{R, T} J = \frac{1}{2} \sum_{i=1}^n \| (p_i - (Rp'_i + T)) \|^2 \quad (2)$$

式中,原始车位角点为 $p'_i(x'_i, y'_i, z'_i)$ 为历史跟踪车位角点坐标,目标点云为 $p_i(x_i, y_i, z_i)$ 的坐标点集(即当前时刻环视感知车位角点坐标), $i \in (1, n)$, n 为点集的数量, R 和 T 为基于起始车辆位姿的旋转矩阵和位移向量。残差值 J 越小,证明匹配越好。

其次对车位进行交并比 IOU (Intersection Over Union) 跟踪

$$IOU = \frac{A \cap B}{A \cup B} \quad (3)$$

式中, A 为当前时刻待匹配感知车位对应四个角点围成的平面; B 为历史跟踪待匹配车位对应四个角点围成的平面。

最后进行的进行低精度车位中心点跟踪:

欧式距离计算公式表示为

$$dis = \sqrt{(PC_i - PC'_j)^2} \quad (4)$$

式中, i 为当前的时刻感知车位的 id; j 为历史跟踪结果对应的 id; PC_i 为第 i 个车位中心的坐标; PC'_j 为第 j 个跟踪结果车位中心转换到当前时刻车身坐标系下的坐标。

2.3.2 自适应权重配准算法

本文重点在基于优化后的车位跟踪结果的自适应权重配准算法，分以下三个步骤：(1) 车位点云生成；(2) 车位点云权重计算；(3) 车位自适应配准；

(1) 车位语义感知结果转换车位点云：

基于原始 BEV 图像长和宽，生成灰度值全部为 0 的掩膜图，在掩膜图像上基于车位角点像素坐标按矩形角点顺序依次连线，并将连线像素灰度值置为 255；

遍历掩膜图中的 255 灰度值结果生成点云，转换如下：

转换公式

$$\begin{cases} x=(u-w/2)\cdot p_x \\ y=(v-h/2)\cdot p_y \end{cases} \tag{5}$$

式中， (x,y) 为当前车身系下的空间坐标， (u,v) 掩膜图像上遍历的灰度值为 255 的横纵像素坐标， w 和 h 分别为图像的长和宽， p_x 和 p_y 分别为掩膜图像素在 BEV 空间下的长和宽的分辨率，车位点云由 (x,y) 组成的点集构成车位点云 P_c 。

具体实施时 BEV 图像 w 和 h 设定为 512 像素，单个像素在 BEV 空间下的长和宽分辨率 p_x 和 p_y 对应物理世界尺寸大小为大小为 0.05m。

(2) 车位点云权重计算：

基于当前车身系下的点云坐标 (x,y) ，结合式 (1) 的先验权重计算公式，计算出车位点云权重值。

(3) 车位自适应配准：

对步骤 (1) 中获得的车位点进行自适应权重临近点迭代 ICP 配准；

设定配准误差阈值 H ，具体实施时设定为小于 0.2m，配准公式详见式 (2)。

自适应过程基于点云权重进行进行三步自适应处理：

第一步，将权重系数 $\alpha>0$ 的 p_i 和 p'_i 点云进行提取，配准出点云数最全面的结果；

第二步，将权重系数 $\alpha>0.4$ 的 p_i 和 p'_i 点云进行提取，配准出点云数次全面的结果；

第三步，将权重系数 $\alpha>0.7$ 的 p_i 和 p'_i 点云进行提取，配准出点云数最少的结果；

对比三次配准误差，优先输出提取点云数量提取较多且误差符合阈值 H 的配准结果。

3 实验结果与分析

实验选取了地下车库影像数据集，车库影像

数据来自于安徽合肥国创智能汽车地下车库无人车采集环视 BEV 拼接图像数据 2000 帧，分辨率 512×512 ，其中一半的偶数帧用于建图实验，另一半的奇数帧用于定位实验。国产化芯片使用黑芝麻 A1000，数据处理环境为：2.3GHz 八核 CPU，16GB 内存，Ubuntu18.04 系统。验证过程中对数据分别进行了传统配准建图和定位、配准优化建图和定位实验，并对配准效果进行了对比，配准精度分别与标准地图和轨迹进行对比，结果如图 5~10；实验量化结果对比如下表所示。

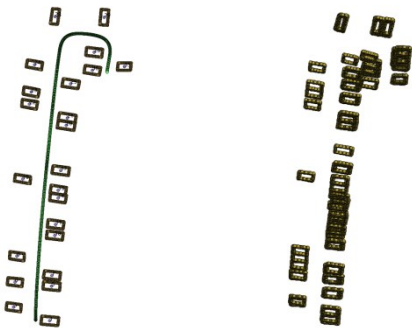


图 5 原始建图 图 6 ICP 配准建图

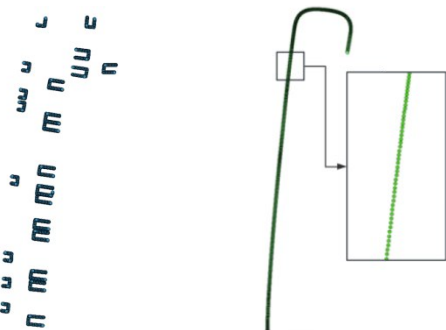


图 7 多层次自适应配准建图 图 8 原始轨迹

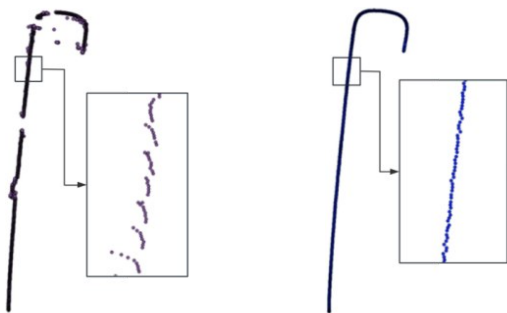


图 9 ICP 配准定 图 10 多层次自适应配准定位

表 1 角点坐标信息

	p_0	p_1	p_2	p_3
感知坐标/m	(-2.93, -1.23)	(-2.95, 1.30)	(-9.09, 1.35)	(-9.01, -1.19)
真实坐标/m	(-2.85, -1.25)	(-2.87, 1.31)	(-8.74, 1.33)	(-8.65, -1.22)

由图 5~7 可以得出，针对环视车位结果，基于传统 ICP 配准的方式，受感知精度的影响会出现大量重影误匹配的情况导致建图误差较大；基于多层级跟踪和自适应加权配准的方式可以有效初始化配准位姿，过滤低精度感知信息的问题，提高建图效果，将原本在米级的建图精度提高到分米级。

由图 8~10，基于传统 ICP 建图后的配准定位过程中，出现定位误差较大，定位失败的情况；基于多层级跟踪和自适应加权配准定位结果有效的控制了定位精度，可将原本米级定位精度，提

升到分米级内，有效解决偶发定位失败的问题。

表 1 给出了图 2 所示的感知坐标与真实坐标结果，结合上诉采用的一次方程计算其权重系数，设定 20cm 误差为最大限定值，通过先验计算得出临界距离平均值为 $d_l = 5.10\text{m}$ 。

由图 11 和表 2 给出了传统配准建图与定位算法和本文优化后的建图与定位算法的精度，可以得出使用本文提出的多层级车位跟踪和自适应加权配准算法在建图精度上提升了 95%，在定位精度上提升了 93%。

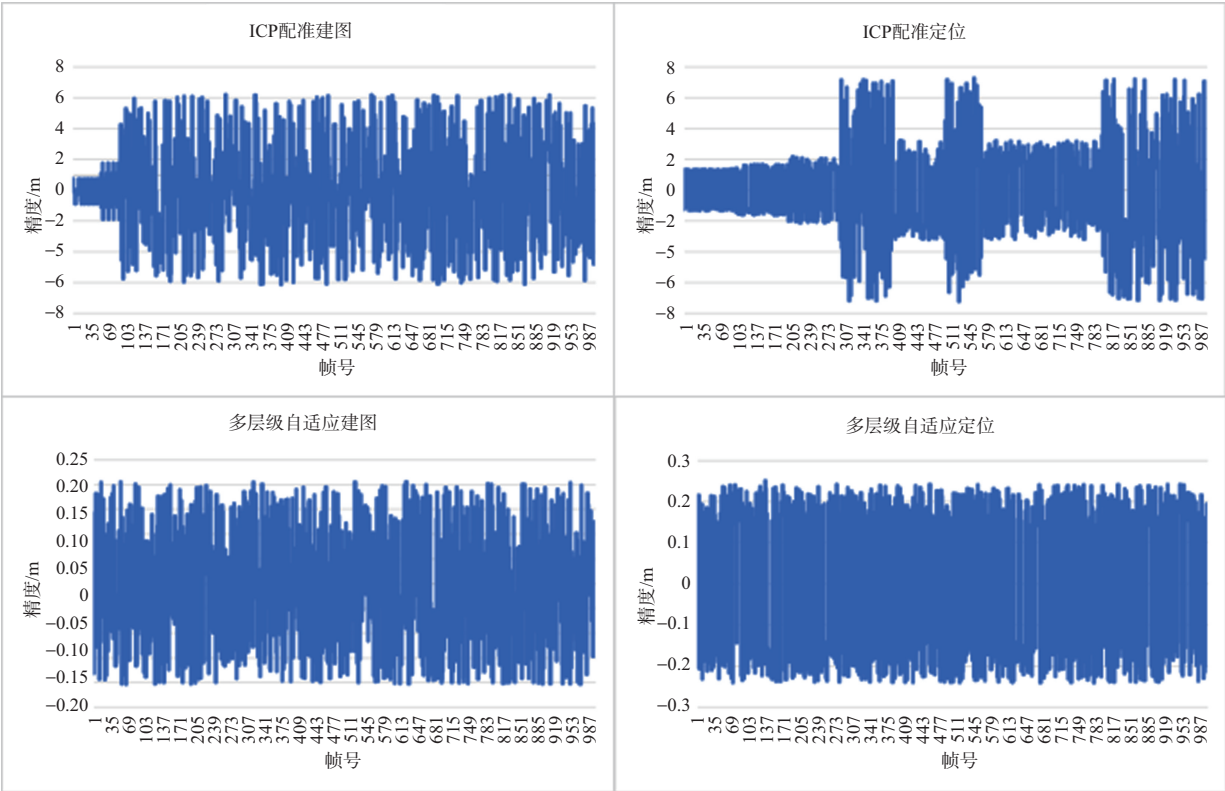


图 11 数据折线图

表 2 数据信息

	ICP 配准 建图	ICP 配准 定位	多层级 自适应建图	多层级 自适应定位
平均误差/m	2.54	2.76	0.11	0.19
最大误差/m	6.15	7.22	0.20	0.25

4 结 论

本文基于地下车库场景获取车辆行驶环视数据，对 BEV 配准算法进行了研究，针对纯视觉 BEV 像素精度随距离变化的特性和地下车库无 GPS 定位等问题，提出了一种车载 BEV 空间多层级车位跟踪和自适应加权配准算法，该方法可有效解决地下车库场景 BEV 配准精度和定位较差的问题。实验证明，本文提出的算法取得了较好的配准效果，极大提高了定位精度，可用于城市地

下车库车载数据的处理。

对于将环视图像转换到 BEV 空间下进行处理的方法仍存在数据处理时间较长的问题，需增加预处理或调整网络来改进。此外在融合方式方面，可尝试多种方法，如加权部分可以与深度信息进行多次融合取最优结果等。

参考文献：

[1] Clavijo M, Jiménez Felipe, et al. The development and prospects of autonomous driving technology[J]. Applied Sciences, 2023,13(9):2076—3417.

[2] Yang B, Li J, Zeng T. A review of environmental perception technology based on multi-sensor information fusion in autonomous driving[J]. World Electric Vehicle Journal, 2025, 16(1):20.

[3] 林俊钦, 胡晓明, 闫达远. 场景轮廓的动态规划立体匹配算法[J]. 光学技术, 2013(1):5.

Lin Junqin, Hu Xiaoming, Yan Dayuan. Dynamic programming stereo matching algorithm for scene profiles[J]. Optical

- Technology, 2013(1):5.
- [4] Zhao J, Shi J, Zhuo L. BEV perception for autonomous driving: State of the art and future perspectives[J]. Expert Systems with Applications, 2024, 258(000):18.
- [5] Wei M, Li J, Kang H, et al. BEV-CFKT: A LiDAR-camera cross-modality-interaction fusion and knowledge transfer framework with transformer for BEV 3D object detection[J]. Neurocomputing, 2024, 582:127527.
- [6] 赵园园, 曾亚涛. 自动驾驶下的纯视觉BEV融合研究与应用[J]. 计算机应用文摘, 2024, 40(14):107—108.
Zhao Yuanyuan, Zeng Yatao. Research and application of pure visual BEV fusion under autonomous driving[J]. Computer Application Abstracts, 2024, 40(14):107—108.
- [7] Samani E U, Tao F, Dasari H R, et al. F2BEV: bird's eye view generation from surround-view fisheye camera images for automated driving[C]//2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Detroit, MI, USA:IEEE, 2023:9367—9374.
- [8] 刘俊杰. 跨模态引导的纯视觉3D感知算法的研究[D]. 广州: 华南理工大学, 2023.
Liu Junjie. Research on pure visual 3D perception algorithms based on cross-modal guidance[D]. Guangzhou: South China University of Technology, 2023.
- [9] 刘天弋. 面向城市道路环境的车载低线数LiDAR/INS/GNSS融合导航算法研究[D]. 武汉: 武汉大学, 2023.
Liu Tianyi. Research on the fusion navigation algorithm of low-line LiDAR/INS/GNSS in vehicle-mounted low-lines for urban road environments [D]. Wuhan: Wuhan University, 2023.
- [10] Park G, Koh J, Kim J, et al. LiDAR-based 3D temporal object detection via Motion-aware LiDAR feature fusion[J]. Sensors (14248220), 2024, 24(14):4667.
- [11] Kim M, Kim G, Jin K H, et al. BroadBEV: collaborative LiDAR-camera fusion for Broad-sighted bird's eye view map construction[C]//2023 IEEE International Conference on Robotics and Automation (ICRA 2023). London, UK: IEEE, 2023:11125—11132.
- [12] Cai H. BEVFusion4D: learning LiDAR-camera fusion under bird's-eye-view via cross-modality guidance and temporal aggregation[EB/OL]. (2023-03)/[2025-05]. <https://api.semanticscholar.org/CorpusID:257833799>.
- [13] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015. Munich, Germany. Cham: Springer International Publishing, 2015:237—244.
- [14] 杨颖, 岳圣斌, 楚博文, 等. 同尺度和跨尺度增强的U-Net用于视网膜血管分割[J]. 光学技术, 2023, 49(4):487—496.
Yang Ying, Yue Shengbin, Chu Bowen, et al. U-Net with the same scale and cross-scale enhancement is used for retinal vascular segmentation[J]. Optical Technology, 2023, 49(4):487—496.
- [15] 刘利源. 基于多层次语义信息适配的图像语义分割[D]. 天津: 天津大学, 2020.
Liu Liyuan. Image semantic segmentation based on multi-level semantic information adaptation [D]. Tianjin: Tianjin University, 2020.
- [16] 高智, 胡傲涵, 陈泊安, 等. 多层次几何—语义融合的图神经网络地表异常检测框架[J]. 通感学报, 2024, 28(7):2051—2068.
Gao Zhi, Hu Aohan, Chen Bo'an, et al. Multi-level geometric-semantic fusion graph neural network surface anomaly detection framework[J]. Journal of Remote Sensing, 2024, 28(7): 2051—2068.
- [17] 占钟鸣, 李庆武, 余大兵, 等. 基于边缘及多尺度特征融合的显著性目标检测方法[J]. 光学技术, 2024, 50(5):606—612.
Zhan Zhongming, Li Qingwu, Yu Dabing, et al. Senseful target detection method based on edge and multi-scale feature fusion[J]. Optical Technology, 2024, 50(5):606—612.
- [18] Huiyan W, Tengfei Z, Fumin M A. Rough K-means algorithm with self-adaptive weights measurement based on space distance[J]. Computer Science, 2018, 45(7):190—196.
- [19] 周辉奎, 顾牡丹. 自适应加权直方图均衡化的红外图像增强[J]. 光学技术, 2023, 49(6):750—755.
Zhou Huikui, Gu Mudan. Infrared image enhancement of adaptive weighted histogram equalization[J]. Optical Technology, 2023, 49(6):750—755.
- [20] Dereniak E L, Dereniak T D. Geometrical and trigonometric optics: image formation[M]. New York: Cambridge University Press, 2008.
- [21] Jiang S, Qi P, Han L, et al. Navigation system for orchard spraying robot based on 3D LiDAR SLAM with NDT_ICP point cloud registration[J]. Computers and Electronics in Agriculture, 2024:220:108870.
- [22] 王海起, 张腾, 彭佳琦, 等. 空间加权距离的GIS数据Fuzzy C—means聚类方法与应用分析[J]. 地球信息科学学报, 2013, 15(4):536—542.
Wang Haiqi, Zhang Teng, Peng Jiaqi, et al. Fuzzy C-means clustering method and application analysis of GIS data with space-weighted distances[J]. Journal of Earth Information Science, 2013, 15(4):536—542.
- [23] Liu Z, Zhu Y, Zhang C. Change detection method of high resolution remote sensing image based on unet network[C]//2024 5th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). Nanchang, China (Virtual): IEEE, 2024:151—157.
- [24] Zhao X, Yuan Y, Song M, et al. Use of unmanned aerial vehicle imagery and deep learning unet to extract rice lodging[J]. Sensors (14248220), 2019, 19(18):3859.