

# 动态场景下基于稀疏光流的语义视觉 SLAM 算法<sup>\*</sup>

侯余鑫<sup>1,2</sup> 介婧<sup>1,2</sup> 侯北平<sup>1,2</sup> 郑慧<sup>1,2</sup> 于爱华<sup>1,2</sup>

(1. 浙江科技大学自动化与电气工程学院 杭州 310023; 2. 浙江省智能机器人感知与控制国际科技合作基地 杭州 310023)

**摘要:** 针对复杂动态场景中视觉 SLAM 因动态干扰而导致的特征点匹配精度下降及地图构建误差增大等问题, 提出一种结合语义分割与稀疏光流的动态视觉 SLAM 算法。首先, 引入自适应阈值策略, 有效提升算法在复杂环境中获取特征点的能力; 其次, 将 DY-Conv 模块嵌入 U-Net 语义分割网络中, 并结合 LK 稀疏光流场, 实现对动态物体的精准检测与分割, 有效提高视觉 SLAM 在动态场景中的特征匹配精度和鲁棒性; 最后, 基于 TUM 数据集和实际场景进行算法有效性验证。实验结果表明, 改进 U-Net 算法的平均分割精度由原算法的 92.1% 提高到 94.5%, 而本文所提出的语义视觉 SLAM 算法相比 ORB-SLAM3, 图像处理速度提升 60.13%, 并且在三组高动态序列公开数据集上的位姿估计精度分别提升 43.75%、77.33% 和 64.00%; 另外基于 TUM 数据集和实际场景所生成的稠密三维点云地图, 进一步表明本文算法可有效抑制动态因素的干扰, 从而提升地图构建的准确性。

**关键词:** 复杂动态环境; 动态阈值; 稀疏光流法; 语义分割; 视觉 SLAM

中图分类号: TP391.41; TN98 文献标识码: A 国家标准学科分类代码: 510.1050

## Semantic visual SLAM algorithm with sparse optical flow in dynamic scenarios

Hou Yuxin<sup>1,2</sup> Jie Jing<sup>1,2</sup> Hou Beiping<sup>1,2</sup> Zheng Hui<sup>1,2</sup> Yu Aihua<sup>1,2</sup>

(1. College of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China;

2. Zhejiang International Sci-Tech Cooperation Base for Intelligent Robot Perception, Hangzhou 310023, China)

**Abstract:** Aiming at the problems of feature point matching accuracy degradation and map construction error increase caused by dynamic interference in visual SLAM in complex dynamic scenes, a dynamic visual SLAM algorithm combining semantic segmentation and sparse optical flow is proposed. Firstly, an adaptive thresholding strategy is introduced to effectively improve the algorithm's ability to acquire feature points in complex environments. Secondly, the DY-Conv module is embedded into the U-Net semantic segmentation network and combined with the LK sparse optical flow field to achieve accurate detection and segmentation of dynamic objects, which effectively improves the feature matching accuracy and robustness of visual SLAM in dynamic scenes. Finally, the validity of the algorithm is verified based on the TUM dataset and real scenes. Experimental results show that the improved U-Net algorithm increases the average segmentation accuracy from 92.1% of the original algorithm to 94.5%. Meanwhile, the proposed semantic visual SLAM algorithm improves image processing speed by 60.13% compared to ORB-SLAM3, and enhances pose estimation accuracy by 43.75%, 77.33% and 64.00% on three high-dynamic sequence public datasets, respectively. Additionally, the dense 3D point cloud maps generated based on the TUM dataset and real-world scenarios further demonstrate that the proposed algorithm can effectively suppress the interference of dynamic factors, thereby improving the accuracy of map construction.

**Keywords:** complex dynamic scenes; dynamic thresholding; sparse optical flow method; semantic segmentation; visual SLAM

## 0 引言

自主导航是机器人面对复杂动态环境中实现高效运行

的核心技术, 而同时定位与地图构建技术(simultaneous localization and mapping, SLAM)是机器人需要首要解决的关键问题, 具有重要的研究价值和应用意义<sup>[1]</sup>。其中视

觉SLAM(visual simultaneous localization and mapping,VSLAM)作为SLAM技术的重要分支,主要依赖外部摄像头作为感知设备,能够以较低的成本在未知环境当中实现机器人的自主定位与地图构建<sup>[2]</sup>,为机器人的路径规划和动态避障提供可靠支持。

现有的VSLAM算法根据视觉里程计对图像的处理方式不同,可分为特征点法和直接法两大类<sup>[3]</sup>。特征点法通过提取图像中的显著特征点并进行特征匹配,进而估计相机的运动轨迹。其代表性算法包括ORB-SLAM2<sup>[4]</sup>和ORB-SLAM3<sup>[5]</sup>等。与之相比,直接法则是基于图像的像素灰度信息直接估计相机运动,其代表性算法包括DTAM<sup>[6]</sup>和DSO<sup>[7]</sup>。然而,这些VSLAM算法面对复杂动态场景存在明显的局限性,动态物体的干扰会使算法的响应速度降低,从而影响算法的整体性能。

为降低动态物体对VSLAM算法的干扰,当前主流的方法可分为两类,其中一类是基于几何的方法,Ni等<sup>[8]</sup>在ORB-SLAM3的基础上,设计了新型地图点筛选机制,显著提升了动态地图点的识别能力。Zhou等<sup>[9]</sup>采用汉明距离度量结合仿射变换约束,构建了融合几何与语义信息的动态特征点剔除模型。另一类是基于深度学习的方法,符强等<sup>[10]</sup>提出一种可区分室内高低动态环境的VSLAM,根据所处的环境结合YOLOv8-Seg实例分割网络进行动态特征点的剔除,保证跟踪精度。王爽等<sup>[11]</sup>提出一种基于全景分割与多视图几何的VSLAM方法,将全景分割网络与多视图几何约束结合,并且通过光流法有效处理动态物体。栾添添等<sup>[12]</sup>提出了一种基于傅里叶梅林变换的VSALM算法,克服了传统SLAM难以检测非先验运动物体的弊端。Chang等<sup>[13]</sup>提出将Deeplabv3+和光流结合的VSLAM算法,成功消除动态特征点对系统的影响。程强<sup>[14]</sup>等提出一种基于ORB-SLAM3主体框架和深度学习技术的SLAM算法,首先利用YOLOX目标检测网络获得潜在的动态特征点,然后将几何法与光流法相结合,不断调整阈值,剔除动态特征点,保留静态特征点。此外,Li等<sup>[15]</sup>提出了一种基于改进的YOLOv8与ORB-SLAM3融合的方法,在YOLOv8框架中加入了平衡卷积方法,提高了动态环境中姿态估计的准确性。

尽管现有的语义VSLAM算法在一定程度上缓解了动态目标的不利影响,但是在室外环境中仍然面临定位精度不足和实时性较差的问题。因此,探索适用于室外环境的VSLAM方法具有重要的现实意义。基于以上分析,本文提出了一种语义VSLAM算法,该算法将改进的U-Net语义分割网络和稀疏光流(lucas-kanade,LK)相结合,实时剔除动态环境下的干扰特征点,以解决VSLAM在室外复杂动态环境当中,特征点匹配精度下降和地图构建误差增大等问题。

本文的研究工作主要包括以下几个方面:

1)将针U-Net网络中的普通卷积更换为动态卷积,提

高语义分割的精度。

2)将阈值策略融入到特征提取算法当中,提高特征点的提取质量和数量。

3)将语义分割与稀疏光流法相结合,对动态区域的特征点进行有效筛选与剔除。

4)新增稠密建图线程,实现高精度的三维稠密地图构建,提升了场景表达的完整性。

## 1 研究基础

### 1.1 相机运动模型描述

为了提升机器人在室外动态环境中的定位精度,本文算法需要依赖丰富的图像信息以确保系统的准确性。综合考虑成本控制与实际应用需求,本文选用intel realsense d435i深度相机作为数据采集设备。该相机不仅具备较高的性价比,还拥有出色的深度感知能力,能够为VSLAM系统提供高质量的图像数据支持,深度相机在室外环境中的运动状态示意图如图1所示。

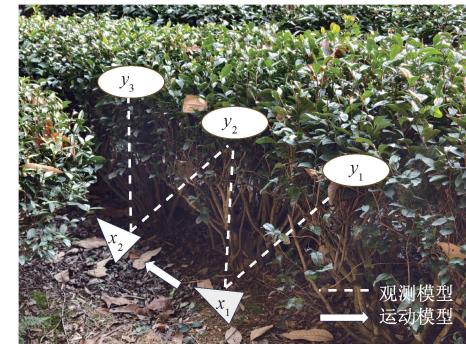


图1 茶园环境下相机运动状态示意图

Fig. 1 Camera movement in a tea garden environment

在VSLAM系统中,机器人的运动状态需要通过运动方程与观测方程进行建模与描述,以实现对机器人动态行为和环境信息的有效表征。从时间  $k-1$  时刻到  $k$  时刻,机器人的相机运动状态可由运动方程表示,具体描述相机在连续时间步中的位姿变化过程,而在  $k$  时刻相机对茶园目标点的观测则可通过观测方程描述,观测方程描述了相机在当前时刻对环境中特征点的深度与二维投影的几何关系。具体的关系如式(1)所示。

$$\begin{cases} x_k = f(x_{k-1}, u_k) + w_k \\ z_{k,j} = h(y_j, x_k) + v_{k,j} \end{cases} \quad (1)$$

式中: $x_k$  表示相机在  $k$  时刻的位置; $u_k$  表示相机在  $k$  时刻拍摄的数据; $w_k$  表示在  $k$  时刻的噪声; $v_{k,j}$  表示当前观测到  $k$  时刻路标点的噪声; $y_j$  表示当前相机观测到的路标点; $z_{k,j}$  表示相机在  $x_k$  处对路标点  $y_j$  的观测数据。

### 1.2 VSLAM算法理论基础

本文所提出的语义VSLAM算法是以ORB-SLAM3为主要框架,该算法对ORB-SLAM2进行了多项优化,显著提升了系统的鲁棒性和定位精度,因此在学术界和工业

界受到了广泛关注<sup>[16-17]</sup>，其核心框架由 tracking、localmapping、loopclosing 和 atlas 四个主要线程构成<sup>[18]</sup>，其功能特点如下：

1) tracking 线程负责处理输入的图像，通过最小化重投影误差实时估计当前帧与地图之间的相对位姿，并完成关键帧的筛选与插入。

2) localmapping 线程负责向活跃地图中添加关键帧和地图点，剔除异常值，并在局部窗口内通过视觉或视觉惯性束调整(bundle adjustment, BA)对地图进行优化。

3) loopclosing 线程通过比较活跃地图与 atlas 中的关键帧进行相似性度量，判断是否执行回环校正或地图合并操作，以消除累积误差。

4) atlas 是由多个未合并地图组成的集合，分为活跃地图和非活跃地图两部分。它通过利用关键帧的数据

地图和非活跃地图两部分。它通过利用关键帧的数据

## 2 融合稀疏光流语义 VSLAM 算法

### 2.1 语义 VSLAM 算法框架

在室外环境中，ORB-SLAM3 存在一些局限性。一方面，室外场景中行人、车辆等动态物体频繁出现，对特征点匹配和定位精度造成明显的干扰；另一方面，复杂多变的光照条件破坏了光度一致性假设，导致定位精度下降。针对上述问题，本文对 ORB-SLAM3 进行了优化与改进，以更好地适应室外复杂场景的实际需求，改进后的 VSLAM 算法框架如图 2 所示。

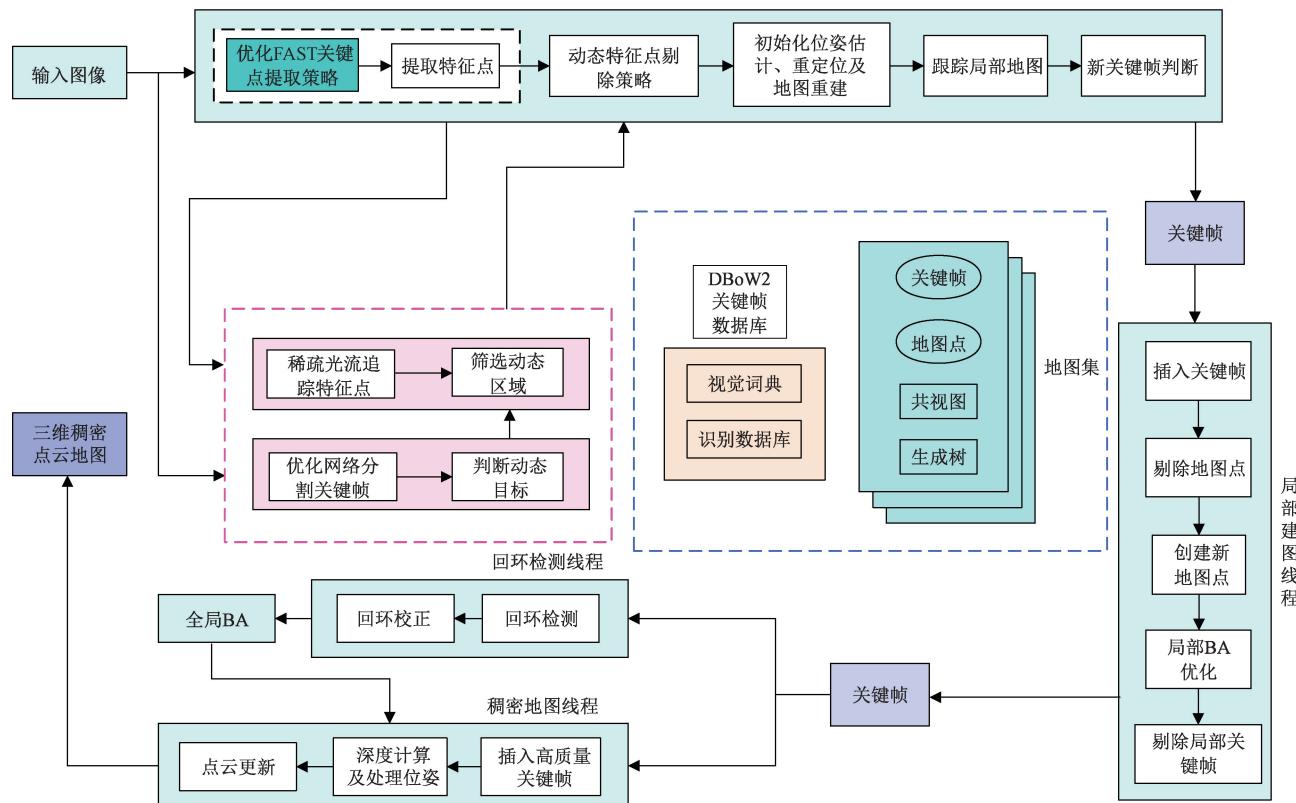


图 2 融合语义分割网络与稀疏光流法融合的 VSLAM 算法框架

Fig. 2 Visual SLAM algorithm based on fusion of semantic segmentation network and sparse optical flow method

本文所提出的语义 VSLAM 算法在跟踪线程中引入自适应阈值调整策略，以优化特征点提取过程；新增语义分割线程，结合深度学习模型对动态区域进行精确分割；为进一步提高动态特征点剔除的准确性，新增稀疏光流线程，利用稀疏光流法对语义分割未完全处理的区域进行补充优化；最后，新增稠密建图线程，为后续的自主导航与路径规划任务提供支持。

### 2.2 基于自适应阈值的特征点提取策略

ORB 特征提取算法(oriented fast and rotated brief，

ORB)主要由 FAST 关键点检测和 BRIEF 描述子生成两部分组成<sup>[19]</sup>。该算法首先通过 FAST 方法检测图像中的关键点，随后对识别为特征点的区域应用 BRIEF 算法生成特征描述子，从而完成特征点的描述与记录。具体提取流程如下：

从图像中选取任意像素块  $P(x_0, y_0)$ ，其像素值记为  $I(x_0, y_0)$ 。以  $P(x_0, y_0)$  为中心，以 3 个像素为半径绘制 bresenham 圆，该圆包含 16 个编号的像素点，各点像素值记为  $I(x_i, y_i)$ 。设置灰度阈值  $M = 0.1I(x_0, y_0)$ ，并判

断圆上是否存在连续 12 个像素点, 其灰度值相对于中心点灰度值的变化满足特定条件, 具体如式(2)所示。

$$-M < I(x_i, y_i) - I(x_0, y_0) < M \quad (2)$$

满足这一条件的像素点将被判定为 FAST 关键点, 并用于后续的特征描述子构建过程。FAST 特征提取示意图如图 3 所示。

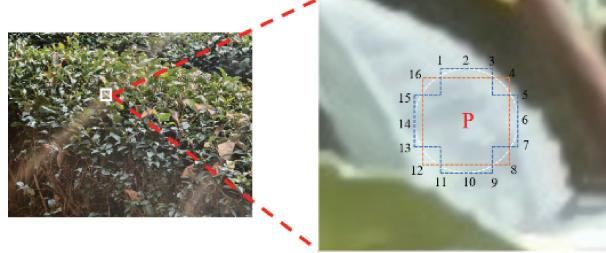


图 3 FAST 特征提取示意图

Fig. 3 Schematic of FAST feature extraction

但是由于室外环境中光照条件复杂多变, 导致特征点提取的数量和质量下降。为了解决这一问题, 本文对特征提取算法进行了改进, 引入自适应阈值策略。该策略的措施如下:

1) 获取图像中像素块周围 16 个像素点的灰度值均值, 具体如式(3)所示。

$$\bar{I} = \frac{1}{HW} \sum_{u,v}^n I(x_u, y_v) \quad (3)$$

式中:  $H, W$  分别为图像的高度和宽度。

2) 计算 16 个像素点的灰度值与像素块灰度均值的平方差  $\epsilon$ , 计算如式(4)所示。

$$\epsilon = \sum_{u,v}^n [I(x_u, y_v) - \bar{I}]^2 \quad (4)$$

3) 计算图像的平均光照强度, 具体如式(5)所示。

$$G_{avg} = \frac{1}{N} \sum_{u,v}^n I(x_u, y_v) \quad (5)$$

式中:  $N$  为图像的总像素数。

4) 设置比例系数  $\beta$ , 取值范围为 0~1, 新阈值  $M'$  的计算公式具体如式(6)所示。

$$M' = \beta \bar{I} \quad (6)$$

5) 最后设置光照强度  $G$ , 根据光照强度的变化自适应的调整比例系数  $\beta$ ,  $\beta$  和  $G$  的关系如式(7)所示。

$$\beta = \beta_0 e^{\frac{G - G_{avg}}{k(G_{max} - G_{min})}} \quad (7)$$

式中:  $\beta_0$  是预设比例系数,  $G_{max} = \max I(u, v)$  为最大光照强度,  $G_{min} = \min I(u, v)$  最小光照强度。

该策略根据光照强度的变化自适应地调整参数, 在特征提取过程中, 设置特征点数量阈值  $N_{min}$ , 当环境光照强度较低时, 图像的整体像素标准差较小, 计算得到的检测阈值也相应的较低, 从而能够提取到足够的特征点。当环境的光照强度较高时, 无法提取到足够的特征点, 这时通过迭代降低检测阈值来保证能够提取到足够的特征

点, 以固定步长  $\Delta\beta = 0.1$  调整比例系数, 直至满足特征点数量  $N \geq N_{min}$  的条件。

### 2.3 基于改进 U-Net 的动态特征点语义分割算法

本文所提出的语义 VSLAM 算法引入了语义分割线程, 以增强其在动态环境中的适应性和处理能力。该线程基于深度学习框架, 将动态卷积模块融入 U-Net 网络进行语义分割, 能够在实时性要求较高的场景中对输入的彩色图像进行高效的处理。

#### 1) DY-conv 模块

动态卷积(dynamic convolution, DY-conv)是一种能够根据输入特征动态生成卷积核权重的方法。该方法的核心思想是通过输入特征自适应调整卷积核权重, 从而增强网络对不同输入特征的适应能力, 其原理如图 4 所示。

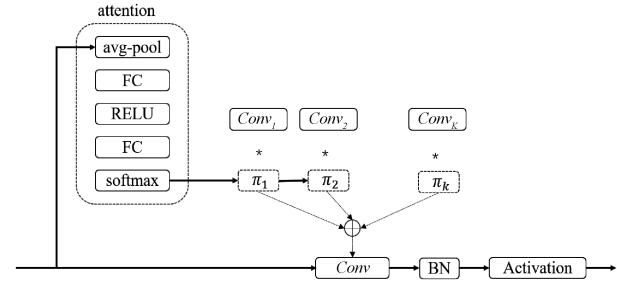


图 4 DY-conv 原理图

Fig. 4 DY-conv schematic diagram

该模块的创新性主要体现在以下两点: 一方面, 通过对注意力机制进行约束, 缩小多个并行卷积叠加的核空间, 降低了模型学习注意力权重的复杂度, 从而简化了模型结构并提高了计算效率; 另一方面, 采用 softmax 归一化函数作为注意力权重系数的输出函数, 并引入热度参数调节 softmax 的分布特性, 进一步加速模型收敛, 同时提升了特征提取的精度和网络的整体性能。

#### 2) 改进 U-Net 网络架构

U-Net 由 ronneberger 等人提出, 在多种计算机视觉任务中表现出卓越的性能。该网络架构由下采样编码器和上采样解码器两部分组成。编码器与解码器之间通过跳跃连接机制横向整合不同层次的特征, 使得下采样过程中提取的细节信息能够与上采样过程中生成的语义特征充分交互。但是随着研究的深入, U-Net 在处理复杂场景时存在特征表达能力不足的问题。

为解决上述问题, 本文使用 DY-conv 模块替换 U-Net 中原有的常规卷积运算, 提高了对图像细节的提取能力, 改进后的 U-Net 模型的网络架构如图 5 所示。

#### 3) 动态特征点语义分割算法

将 DY-conv 模块嵌入 U-Net 语义分割网络之后, 本文采用 libtorch 框架将改进后的 U-Net 网络模型集成到 ORB-SLAM3 的多线程框架当中, 以实现动态环境下特征点的精确剔除。

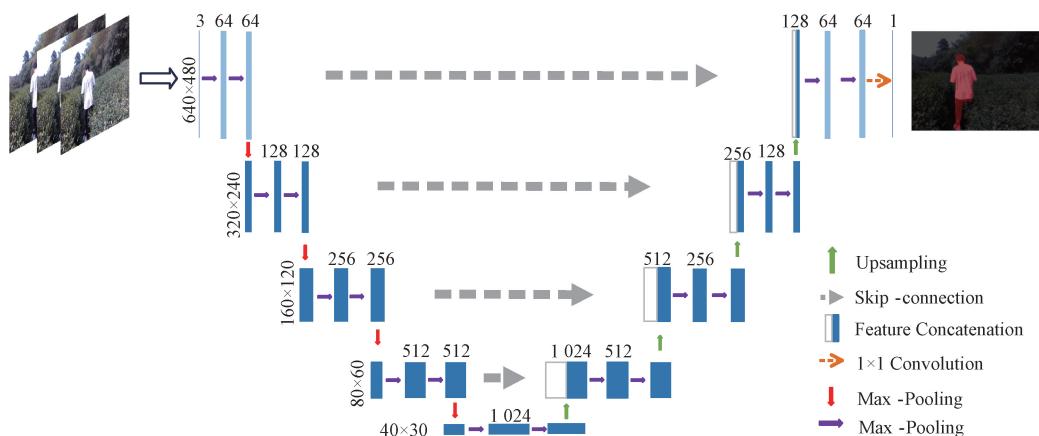


图 5 改进 U-Net 模型的网络架构设计

Fig. 5 Design of network architecture for with improved U-Net model

当新的关键帧图像输入到系统中时,语义分割线程启动 U-Net 网络处理流程,生成一张与输入彩色图相同的语义灰度图,可有效区分场景中的动静态区域。生成动态掩膜的示意图如图 6 所示。



图 6 动态区域的掩膜示意图

Fig. 6 Schematic diagram of the mask of the dynamic region

然后利用语义分割线程生成的动态掩膜数据,将所有落入动态掩膜内的特征点识别为动态特征点,并将其从后续处理流程中剔除,语义分割的效果如图 7 所示。



图 7 实时分割效果示意图

Fig. 7 Schematic diagram of real-time segmentation effect

基于改进 U-Net 的动态特征点语义分割算法的具体流程如算法 1 所示。

#### 算法 1: 基于改进 U-Net 的动态特征点语义分割算法

```

1. 输入: 当前帧(imrgb,imd,mask,tframe), 光流阈值 z,
      prevframe, currframe
2. 输出: 当前帧的静态特征点集合 mdescriptors
3. for (mvkeys.size())//循环处理当前帧中的特征点
4. 获取掩码图像与 mvkeys[i] 特征点对应的像素值
5. if (val != 30)//判断特征点是否在动态区域
6. 将特征点加入 mdescriptors
7. else
8. 将特征点加入 mvkeysdynamic_
9. end

```

#### 2.4 基于稀疏光流的动态特征点剔除算法

尽管在 ORB-SLAM3 系统中融入语义分割网络能够有效剔除大部分动态目标,但在实际场景中的动态目标被遮挡时,会导致特征点被误分配,为解决这一问题,本文将稀疏光流与语义分割相结合,形成一种互补机制,从而更全面地检测动态区域。特征点误分配如图 8 所示。



图 8 特征点误分配示意图

Fig. 8 Schematic diagram of feature point misassignment

稀疏光流是一种估计图像序列中像素运动的技术,其

理论基础建立亮度恒定假设、运动微小假设和空间一致性假设的基础上。首先在相邻帧提取显著特征点,随后,通过计算窗口内每个像素的梯度,构建梯度矩阵和时间梯度向量,并利用最小二乘法求解光流。最终,对提取的所有特征点重复上述步骤,生成稀疏光流场。具体流程如下所示:

根据亮度恒定的假设,对于一个特征点及其邻域内的多个像素点,光流约束方程为:

$$I_x u + I_y v + I_t = 0 \quad (8)$$

式中: $u, v$  分别是特征点在  $x$  轴、 $y$  轴方向上的速度。

然后假设窗口内有  $n$  个像素点,得到  $n$  个线性方程组的方程组:

$$\mathbf{A} \begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{b} \quad (9)$$

其中,  $\mathbf{A} = \begin{bmatrix} I_{x1} \cdots I_{xn} \\ I_{y1} \cdots I_{yn} \end{bmatrix}^T$ ,  $\mathbf{b} = [-I_{t1} \cdots -I_{tn}]^T$ 。

用最小二乘法对方程组进行求解,以获得光流矢量。

$$\begin{bmatrix} u \\ v \end{bmatrix} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b} \quad (10)$$

基于稀疏光流的动态特征点剔除算法的具体流程如算法 2 所示。

#### 算法 2: 基于稀疏光流的动态特征点剔除算法

```

1. 输入: 当前帧(imrgb, imd, mask, tframe), 光流阈值 z,
   prevframe, currframe
2. 输出: 当前帧的静态特征点集合 mdescriptors
3. for(mdescriptors)//mdescriptors 中是否为动态点
4.   计算 mdescriptors mvkeysdynamic 的光流值
5.   计算平均光流(u,v)
6.   if (status[0] == 1)//光流计算成功
7.     if 当前特征点阈值经平方差>z
8.       将特征点加入 mvkeysdynamic_
9.     else
10.       将特征点加入 _mdescriptors
11.   return 集合 mdescriptors
12. end

```

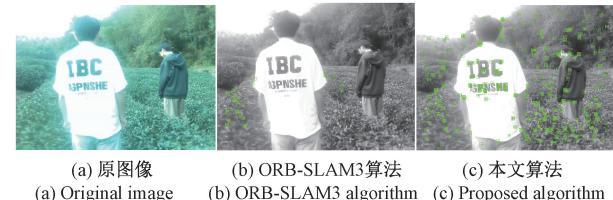
### 3 实验结果与分析

为了验证所语义 VSLAM 算法的性能。本文采用的系统为 Ubuntu18.04, CPU 为 Core TM i7-8700, GPU 为 NVIDIA GeForce RTX 3050Ti, 采用的相机为 intel realsense d435i。采用德国慕尼黑工业大学的 TUM 数据集和自建数据集,其中 TUM 数据集包括低动态数据序列以及高动态数据序列,主要分别模拟静态或缓慢移动的动态场景和快速移动或复杂的动态场景。

#### 3.1 基于自适应阈值的特征点提取实验

为了验证自适应阈值提取特征点策略的有效性,选取

数据集中光照强度较高的一帧与 ORB-SLAM3 算法的特征点提取策略进行对比实验,实验结果如图 9 所示,其中原图像如图 9(a)所示,基于 ORB-SLAM3 算法的特征点提取示意图如图 9(b)所示,基于本文算法的特征点提取示意图如图 9(c)所示。



(a) 原图像 (b) ORB-SLAM3 算法 (c) 本文算法  
(a) Original image (b) ORB-SLAM3 algorithm (c) Proposed algorithm

图 9 特征点提取对比实验

Fig. 9 Feature point extraction comparative experiment

实验结果表明,在光照强度较高的情况下 ORB-SLAM3 算法的特征点提取效果受到明显影响,提取区域特征点稀疏,而本文方法通过自适应阈值调整,有效克服了光照的干扰,可提取到更多分布均匀且稳定的特征点。

#### 3.2 基于语义分割的动态特征点检测实验

为验证改进 U-Net 算法的有效性,本实验选择 PSPNet 模型、SegNet 模型、DeepLabv3+ 模型和 U-Net 模型在 voc2012 数据集进行训练与测试,该数据集包含人、车、猫等 20 种类别,具有较高的多样性和复杂性。为充分评估模型性能,将数据集按照 7:2:1 的比例划分为训练集、验证集和测试集,确保模型在不同数据分布下的泛化能力。具体性能指标如表 1 所示。

表 1 不同模型性能指标对比

Table 1 Comparison of performance indicators of different models

模型	PA/%	MIoU/%
PSPNet	90.7	85.4
SegNet	89.3	80.8
DeepLabv3+	91.6	79.3
U-Net	92.1	83.2
本文算法	94.5	85.3

实验结果表明,与其他模型相比,本文算法的像素准确率达到 94.5%,明显高于 PSPNet、SegNet、DeepLabv3+ 和 U-Net,充分体现了其在全局像素分类精度上的优越性。本文算法的平均交并比达到 85.3%,与 PSPNet 的最高水平 85.4% 接近,表明改进后的算法在边界细节处理方面具备更强的能力。

#### 3.3 基于稀疏光流的动态特征点筛选实验

从 walking\_half 数据集中随机选取相邻的两帧图像,进行动态特征点光流筛选实验,实验结果如图 10 所示,其中前一帧图像如图 10(a)所示,当前帧图像图 10(b)所示,稀疏光流法筛选后的动态特征点结果示意图如

图 10(c)所示。

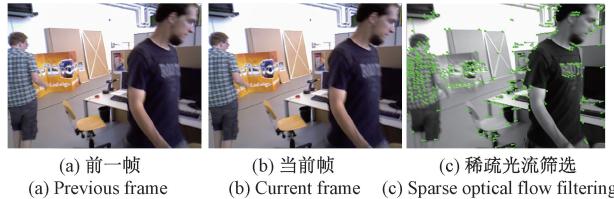


图 10 动态特征点筛选实验

Fig. 10 Dynamic feature point filtering experiment

实验结果表明,稀疏光流法能够有效识别由动态物体引入的干扰特征点,减少了动态因素对系统性能的影响,可以为后续的位姿估计提供了更加稳定和可靠的数据支持。

### 3.4 结合语义与光流技术的动态特征点剔除实验

从 sitting\_half 和 walking\_half 数据集当中分别随机选取一帧图像,对这两帧图像进行动态特征点的剔除实验,实验结果如图 11 所示,其中低动态数据序列 sitting\_half 的原图像如图 11(a)所示,动态特征点剔除的效果如图 11(b)所示,高动态数据序列 walking\_half 的原图像如图 11(c)所示,动态特征点剔除的效果如图 11(d)所示。

实验结果表明,在低动态数据序列中,本文算法能够有效剔除缓慢移动目标周围的特征点;在高动态数据序列中,即使是面对快速移动的目标也可以有效剔除缓慢移动目标周围的特征点,保持较高的剔除精度,有效减少了动态物体对定位和建图的负面影响,验证了本文方法在动

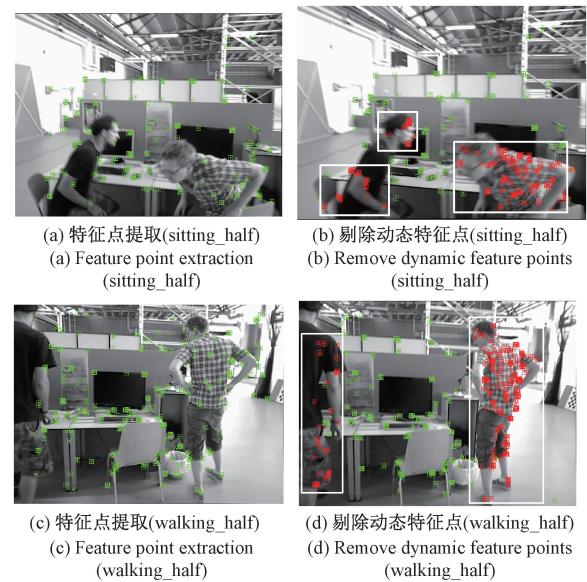


图 11 动态场景下的特征点剔除效果图

Fig. 11 Effect diagram of feature point elimination

态环境中的有效性与适用性。

### 3.5 轨迹精度评估与对比实验

本实验采用绝对轨迹误差 (absolute trajectory error, ATE) 和相对轨迹误差 (relative pose error, RPE) 作为评价指标来验证本文算法的定位精度,对本文算法和 ORB-SLAM3 算法在 TUM 数据集上进行 ATE 的对比,实验结果如图 12 和图 13 所示。

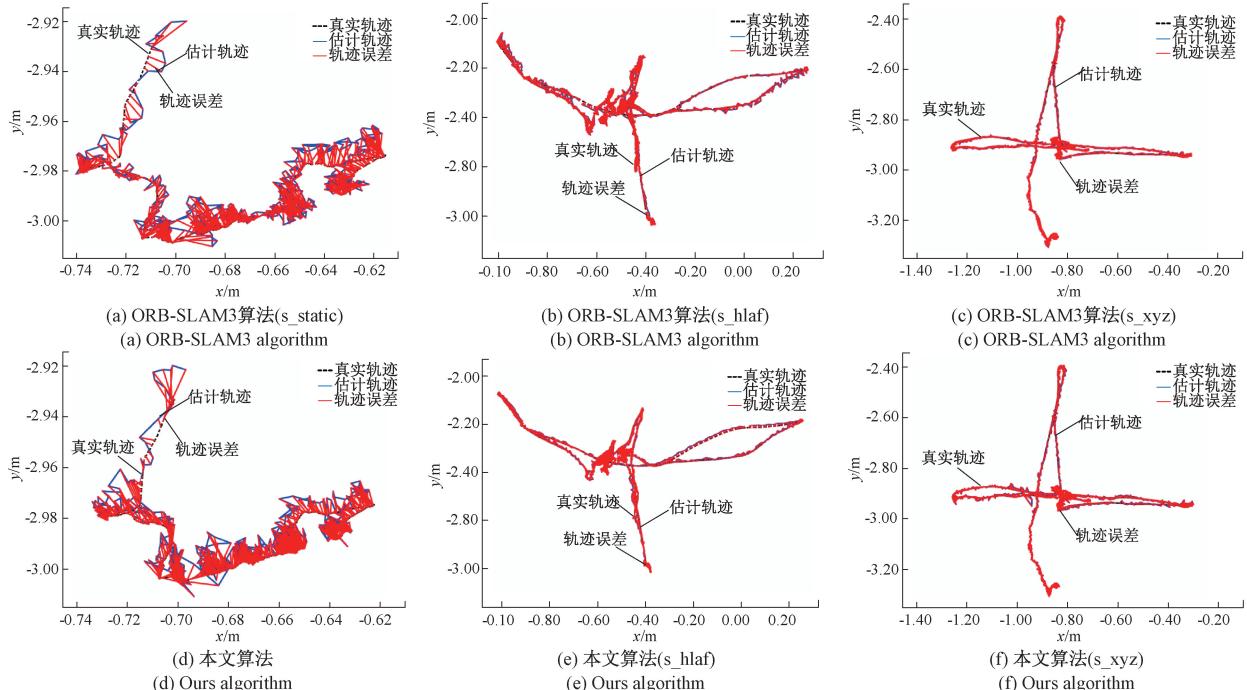


图 12 TUM 低动态数据集下 ORB-SLAM3 和本文方法的绝对轨迹误差对比

Fig. 12 Comparison of absolute camera trajectory errors in low dynamic scenes under TUM dataset

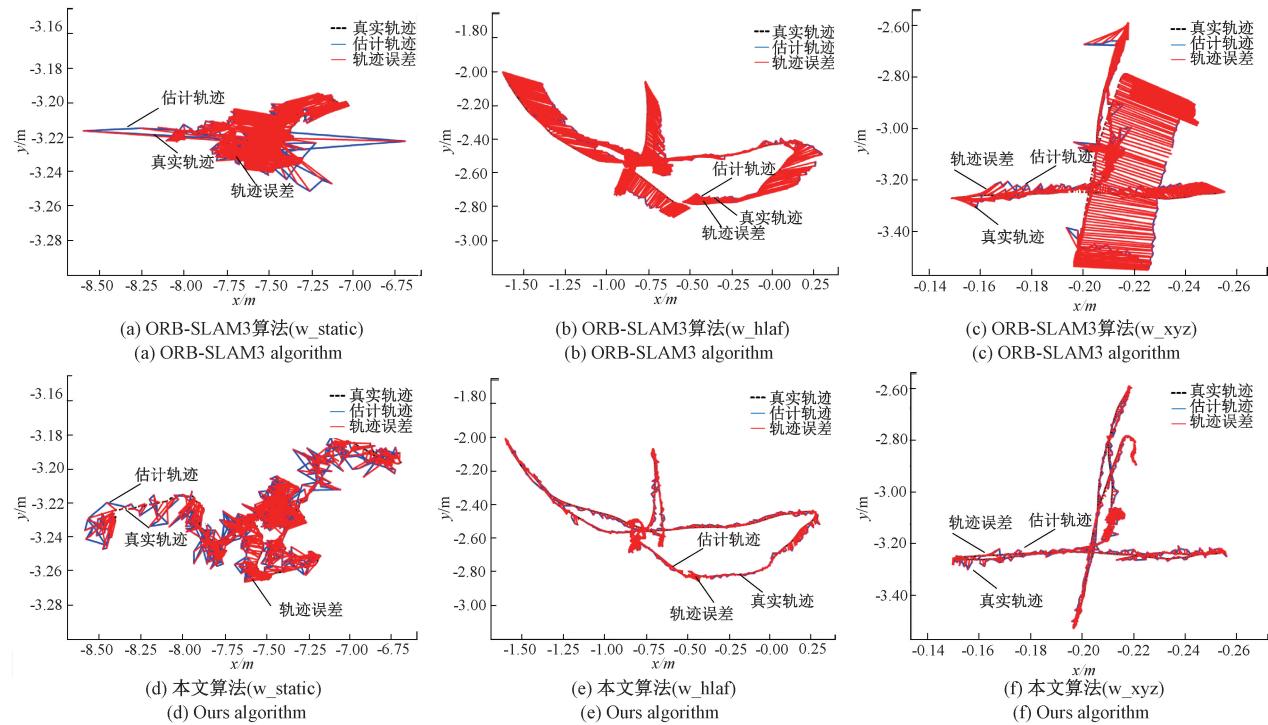


图 13 TUM 高动态数据集下 ORB-SLAM3 和本文方法的绝对轨迹误差对比

Fig. 13 Comparison of absolute camera trajectory errors in high dynamic scenes under TUM dataset

实验结果表明,在低动态场景序列下,ORB-SLAM3 算法和本文算法测得的相机轨迹均接近于真实轨迹,都表现出较高的定位精度。但在高动态场景序列下,ORB-SLAM3 算法测得的轨迹随着时间推移逐渐偏离真实轨迹,本文算法在高动态场景中依然能够准确跟随真实轨迹,这说明本文算法有效避免了动态干扰对定位精度的影响。

为了进一步验证本文算法的有效性,分别对 ORB-SLAM3、DS-SLAM、Dyna-SLAM 和本文算法进行 ATE, RPE 的对比实验,实验结果如表 2 和 3 所示。

表 2 绝对轨迹误差 ATE 对比

Table 2 Absolute trajectory error ATE comparison

序列	Dyna-SLAM	ORB-SLAM3	DS-SLAM	Ours
fr3/s/static	—	0.008	0.007	0.008
fr3/s/xyz	0.014	0.013	—	0.010
fr3/s/half	0.018	0.019	—	0.016
fr3/w/static	0.006	0.013	0.008	0.005
fr3/w/xyz	0.016	0.125	0.025	0.009
fr3/w/half	0.031	0.159	0.028	0.024

根据表 2 的数据分析可知,本文提出的算法在三组低动态数据序列上,相较于 DynaSLAM、ORB-SLAM3 和 DS-SLAM 三种算法,绝对轨迹误差均有小幅的提升;而在高动态序列上,相较于 DynaSLAM、ORB-SLAM3 和

表 3 相对轨迹误差 RPE 对比

Table 3 Relative trajectory error RPE comparison

序列	Dyna-SLAM	ORB-SLAM3	DS-SLAM	Ours
fr3/s/static	—	0.007	0.008	0.006
fr3/s/xyz	0.017	0.014	—	0.011
fr3/s/half	0.023	0.022	—	0.022
fr3/w/static	0.010	0.021	0.011	0.009
fr3/w/xyz	0.023	0.073	0.033	0.018
fr3/w/half	0.027	0.091	0.035	0.022

DS-SLAM 三种算法,绝对轨迹误差均有大幅度的提升,特别是在 w\_xyz 数据集上,相比 ORB-SLAM3 算法,提升幅度达到 43.75%、77.33% 和 64.00%。

根据表 3 可知,本文所提出的算法在低动态数据序列上,相对轨迹误差相比 DynaSLAM、ORB-SLAM3 和 DS-SLAM 算法取得了小幅提升;在高动态数据序列上,本文算法在相对轨迹误差的表现同样优于现有算法,尤其在 walking\_xyz 数据集上,提升幅度达到 21.73%、75.34% 和 45.45%。

### 3.6 实时性能评估与对比实验

为验证算法的实时性,在高动态场景下分别对 Dyna-SLAM、DS-SLAM、ORB-SLAM3、PWC-SegNet 和本文算法的单帧平均处理耗时进行对比测试,实验结果如表 4 所示。

表 4 实时性对比结果

Table 4 Real time comparison results

方法	平均耗时 / ms
Dyna-SLAM	>350
DS-SLAM	58.2
ORB-SLAM3	32.1
PWC-SegNet	24
本文算法	23.2

根据表 4 的数据分析可知,相较于 Dyna-SLAM、ORB-SLAM3 和 DS-SLAM 和 PWC-SegNet,实时性分别提升了 93.37%、60.13%、32.1%、3.34%。表明本文算法能够兼顾计算效率和定位精度。

此外,为验证本文算法在实际部署中的可行性,在户外进行 90 min 的多线程实际动态测试,资源占用率如表 5 所示。

表 5 资源占用率

Table 5 Resource utilization rate

运行时长/min	单帧平均耗时 / ms	平均帧率 / fps	内存平均消耗 / MB	CPU 占用率 / %
0~30	22.5	44.4	683	58±4
30~60	23.4	42.7	702	64±7
60~90	23.7	42.2	725	69±3

### 3.7 三维稠密地图重建效果评估

本文所提出的语义 VSLAM 算法新增稠密建图线程,以解决 ORB-SLAM3 在环境细节表达能力不足的局限性。稠密建图线程是基于位姿信息和深度信息得到局部点云图,同时根据新增加的关键帧不断对其进行更新,然后将局部点云地图拼接成三维稠密点云地图。稠密建图线程的详细参数配置表如表 6 所示。

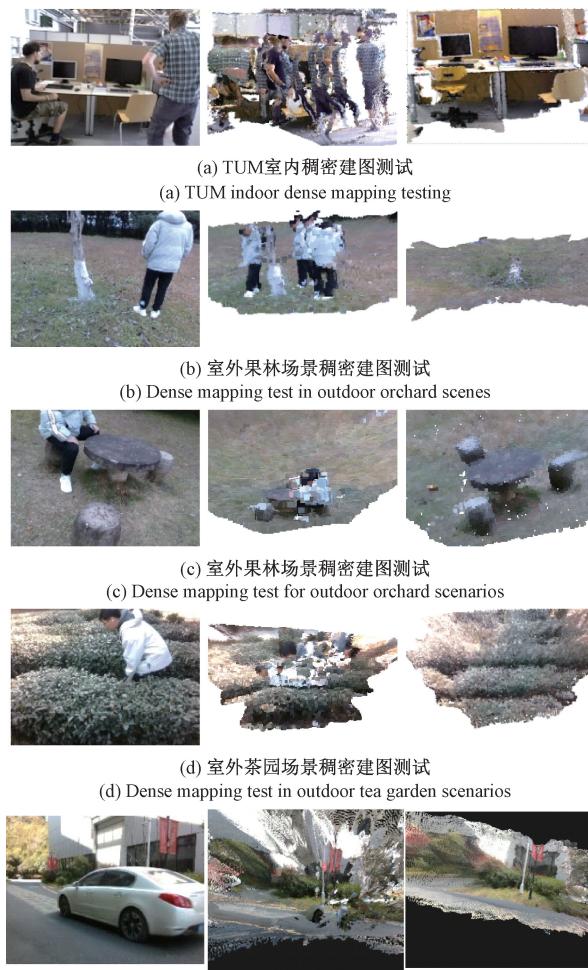
表 6 稠密建图详细参数配置表

Table 6 Dense mapping detailed parameter configuration table

参数名称	取值	单位
关键帧插入间隔	2	s
X 轴体素滤波	0.01	m
Y 轴体素滤波	0.02	m
Z 轴体素滤波	0.15	m
深度有效范围	0.5	m

为验证本文三维稠密建图算法的有效性,分别在 TUM 室内和 4 组室外场景中进行了实验测试,实验结果如图 14 所示。

实验结果表明,本文算法在对动态特征点进行有效剔除后,能够生成精确的三维稠密点云地图,还原真实场景



(e) Dense mapping test for the scenarios of fast-moving vehicles  
Fig. 14 Dense 3D point cloud map construction renderings

的几何结构与细节信息。

## 4 结论

本文提出了一种复杂动态场景下融合稀疏光流的语义 VSLAM 算法,有效解决了因动态干扰而导致的特征点匹配精度下降及地图构建误差增大等问题。

首先,使用基于自适应阈值的特征点提取策略,使系统能够在茶园等复杂环境中采集更多数量且质量更高的特征点;其次,将 DY-Conv 模块嵌入 U-Net 网络,并结合 LK 光流法,精确识别并且剔除动态特征点,最后,新增稠密建图模块对高质量静态特征点进行稠密重建,生成了完整的场景地图。本文通过公开数据集和本文构建的室外数据集的实验测试表明,所提出的算法在动态环境下表现出优异的性能。

## 参考文献

- [1] ZHANG F, ZHU K. MG-SLAM: RGB-D SLAM based on semantic segmentation for dynamic

- environment in the internet of vehicles [J]. Computers, Materials & Continua, 2025, 82 (2): 2353-2372.
- [2] 张炳力,米向东,郑达,等.面向自动泊车的可变滑动窗口环视SLAM算法[J].电子测量技术,2024,47(15):109-116.
- ZHANG B L, MI X D, ZHENG D, et al. Surround-view SLAM algorithm with variable sliding window for automatic parking [J]. Electronic Measurement Technology, 2024, 47(15): 109-116.
- [3] 章国锋,黄赣,谢卫健,等.基于单目视觉惯性的同步定位与地图构建方法综述[J].中国图象图形学报,2024,29(10):2839-2858.
- ZHANG G F, HUANG G, XIE W J, et al. A review of monocular visual-inertial simultaneous localization and mapping methods [J]. Journal of Image and Graphics, 2024, 29(10): 2839-2858.
- [4] MUR-ARTAL R, TARDOS J D. ORB-SLAM2: An open-source slam system for monocular, stereo, and rgbd cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- CAMPOS C, ELVIRA R, RODRIGUEZ J J G, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial, and multimap slam[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.
- [6] MANKOTIA V, SUNKARIA R K, GURUNG S. DTAM: A dynamic threshold and monitoring based technique to protect mobile ad-hoc network from black-hole and flooding attacks[J]. Wireless Personal Communications, 2024, 134(3): 1469-1490.
- [7] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(3): 611-625.
- [8] NI J, WANG L, WANG X, et al. An improved visual SLAM based on map point reliability under dynamic environments[J]. Applied Sciences, 2023, 13(4): 271.
- [9] ZHOU Y, TAO F, FU Z, et al. RGB-D SLAM in indoor dynamic environments with two channels based on scenario classification [J]. Measurement Science and Technology, 2023, 34(11): 115007.
- [10] 符强,曾凡治,纪元法,等.针对室内高低动态环境的视觉SLAM算法研究[J].电子测量技术,2024,47(21):159-167.
- FU Q, ZENG F ZH, JI Y F, et al. Research on visual SLAM algorithm for indoor environments with high and low dynamics [J]. Electronic Measurement Technology, 2024, 47(21): 159-167.
- [11] 王爽,刘云平,张柄棋,等.基于全景分割与多视图几何的动态SLAM方法[J].电子测量技术,2024,47(24):149-159.
- WANG SH, LIU Y P, ZHANG B Q, et al. A dynamic SLAM method based on panoptic segmentation and multi-view geometry[J]. Electronic Measurement Technology, 2024, 47(24): 149-159.
- [12] 栾添添,吕奉坤,班喜程,等.高动态环境下的傅里叶梅林变换视觉SLAM算法[J].仪器仪表学报,2023,44(7):242-251.
- LUAN T T, LYU F K, BAN X CH, et al. Visual SLAM algorithm based on Fourier-Mellin transform in high-dynamic environment [J]. Chinese Journal of Scientific Instrument, 2023, 44(7): 242-251.
- [13] CHANG J, DONG N, LI D. A real-time dynamic object segmentation framework for SLAM system in dynamic scenes [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70(5): 1-9.
- [14] 程强,张友兵,周奎.基于改进YOLOX的动态视觉SLAM方法[J].电子测量技术,2024,47(23):123-133.
- CHENG Q, ZHANG Y B, ZHOU K. An improved YOLOX-based dynamic visual SLAM method [J]. Electronic Measurement Technology, 2024, 47 (23): 123-133.
- [15] LI Y, SHEN H, FU Y, et al. A method of dense point cloud SLAM based on improved YOLOV8 and fused with ORB-SLAM3 to cope with dynamic environments[J]. Expert Systems with Applications, 2024, 255(3): 124918.
- [16] 张耀,吴一全,陈慧娴.基于深度学习的视觉同时定位与建图研究进展[J].仪器仪表学报,2023,44(7):214-241.
- ZHANG Y, WU Y Q, CHEN H X. Research progress of visual simultaneous localization and mapping based on deep learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(7): 214-241.
- [17] SALVO D E, BELLUCCI S, CELIDONIO V, et al. Visual localization domain for accurate V-SLAM from stereo cameras[J]. Sensors, 2025, 25(3): 739-739.
- [18] YANG Z, ZHAO K, YANG S, et al. Research on a density-based clustering method for eliminating inter-frame feature mismatches in visual SLAM under dynamic scenes[J]. Sensors, 2025, 25(3): 622-622.
- [19] 杨雪梅,李帅永.移动机器人视觉SLAM回环检测原理,现状及趋势[J].电子测量与仪器学报,2022,36(8):1-12.
- YANG X M, LI SH Y. Principles, current status and trends of loop closure detection in visual SLAM for mobile robots[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(8): 1-12.

### 作者简介

侯余鑫,硕士研究生,主要研究方向为视觉SLAM、路径规划和自主导航。

E-mail:hyxzust@zust.edu.cn

介婧(通信作者),博士,教授,硕士生导师,主要研究方向为智能计算及优化、智能控制、机器学习、基于生物信号的模式识别等。

E-mail:jingjie@zust.edu.cn

侯北平,博士,教授,博士生导师,主要研究方向机器人感知、人工智能、无损检测、AI视频分析、移动机器人。

E-mail:bphou@zust.edu.cn

郑慧,博士,副教授,主要研究方向智能计算及优化、机器学习和目标跟踪研究。

E-mail:huizheng@zust.edu.cn

于爱华,博士,硕士生导师,主要研究方向为信号处理、机器人控制和机器视觉。

E-mail:yuaihua\_seu@163.com