文章编号：1003－1251(2026)01－0008－09

动态场景下基于 YOLO11n 的视觉 SLAM 算法

冯迎宾[1]，雒 艺[1]，王天龙[2]

(1. 沈阳理工大学 自动化与电气工程学院，沈阳 110159；2. 中国科学院机器人与智能制造创新研究院，沈阳 110016)

摘 要：针对动态场景导致视觉定位与建图(simultaneous localization and mapping，SLAM)算法位姿估计精度低和地图质量差等问题，提出一种结合深度学习的动态视觉 SLAM 算法。该算法在 ORB-SLAM3 前端引入轻量化且目标识别率高的 YOLO11n 目标检测网络，检测潜在动态区域，并结合 Lucas-Kanade(LK)光流法识别其中的动态特征点，从而在剔除动态特征点的同时保留静态特征点，提高特征点利用率和位姿估计精度。此外，新增语义地图构建线程，通过去除 YOLO11n 识别到的动态物体点云，并融合前端提取的语义信息，实现静态语义地图的构建。在 TUM 数据集上的实验结果表明，相较于 ORB-SLAM3，该算法在高动态序列数据集中的定位精度提升了 95.02%，验证了该算法在动态环境下的有效性，能显著提升视觉 SLAM 系统的定位精度和地图构建质量。

关 键 词：深度学习；动态视觉定位与建图；YOLO11n；静态语义地图；光流法

中图分类号：TP391.4 **文献标志码**：A **DOI**：10.3969/j.issn.1003－1251.2026.01.002

Visual SLAM Algorithm Based on YOLO11n in Dynamic Scene

FENG Yingbin[1]，LUO Yi[1]，WANG Tianlong[2]

(1. Shenyang Ligong University，Shenyang 110159，China；
2. Institute for Robotics and Intelligent Manufacturing，Chinese Academy of Sciences，Shenyang 110016，China)

Abstract：To address the problems of low pose estimation accuracy and poor map quality in visual simultaneous localization and mapping (SLAM) algorithms in dynamic scenes，a dynamic visual SLAM algorithm incorporating deep learning is proposed. This algorithm integrates a lightweight and high-accuracy object detection network，YOLO11n，into the frontend of ORB-SLAM3 (Oriented FAST and Rotated BRIEF SLAM3) to detect potential dynamic regions. The Lucas-Kanade (LK) optical flow method is then employed to identify dynamic feature points within these regions. By removing dynamic feature points while retaining static ones，the algorithm improves the utilization of feature points and the accuracy of pose estimation. Additionally，a semantic map construction thread is introduced to build a static semantic map by removing point clouds of dynamic objects identified by YOLO11n and integrating semantic information extracted in the frontend. Experimental results of the TUM dataset demonstrate that，compared to ORB-SLAM3，the proposed algorithm improves localization accuracy by 95.02% on highly dynamic sequence datasets，verifying its effectiveness in dynamic environments and significantly enhancing the localization accuracy and map quality of visual SLAM systems.

Key words：deep learning；dynamic SLAM；YOLO11n；static semantic map；optical flow method

收稿日期：2025－04－17

基金项目：辽宁省科学技术计划项目(2023JH2/10700006)

作者简介：冯迎宾(1986—)，男，教授，博士；王天龙(1986—)，通信作者，男，副研究员，博士。

同时定位与建图(simultaneous localization and mapping,SLAM)是机器人在未知环境中进行自主导航的关键技术,其核心目标是实时估计自身位姿并构建环境地图。视觉 SLAM[1]在动态场景下面临严峻挑战,传统基于静态假设的算法易受运动物体干扰,导致定位精度下降与地图漂移[2]。为提升系统在动态环境下的定位精度与鲁棒性,当前主流 V-SLAM[3]方法主要针对图像中的动态特征点进行处理,分为基于几何的方法和基于深度学习的方法。Zeller 等[4]提出了一种直接全光学里程计(DPO)算法,该方法通过多视图几何跟踪与映射,直接对微透镜阵列生成的微图像进行处理,并基于连续光场间的对应关系分析创建半稠密深度图,但存在地图质量欠佳、系统鲁棒性较弱的问题。Zou 等[5]提出了基于特征点的重投影误差实现静态与动态特征区分,虽然计算量轻但仅适用于低动态干扰场景。Yu 等[6]在 ORB-SLAM2 的基础上融合语义分割网络 SegNet 与对极几何约束提出了 DS-SLAM 算法,该算法提升了动态环境下的定位精度,并创建了语义地图,但受限于语义分割的高计算代价,算法难以满足实时性要求。刘建军等[7]针对特殊环境下算法对特征点提取不足的问题,从特征点提取角度进行了优化。Bescos 等[8]提出了 DynaSLAM,利用 Mask R-CNN 分割场景物体,并结合多视图几何剔除动态特征点,但存在特征点利用率低的问题。Zhao 等[9]利用 Mask R-CNN 检测潜在运动目标,并结合光流法剔除动态特征点,确保点云地图仅包含静态部分,但系统的实时性差。Su 等[10]提出了 RTD-SLAM,采用 YOLOv5s[11]进行目标检测,并结合语义信息与光流提取动态特征点,该算法在实时性与鲁棒性方面表现较优,能够高效处理动态环境中的特征点剔除任务,确保系统在复杂场景下仍保持稳定运行与快速响应能力。Gao 等[12]提出了动态特征点剔除算法,通过改进 YOLOv5 目标检测网络增强动态目标识别能力,并结合多视图几何约束构建动态特征判别模型,有效剔除场景中的动态干扰特征,提升了动态环境下 SLAM 系统的定位鲁棒性。李嘉铭等[13]提出一种融合目标检测与多视图几何约束的动态 SLAM 系统,通过运动概率模型实现动态特征点的判别与剔除。但该方法平均每帧图像处理耗时过长,难以满足实时性需求,且在大规模视野场景下,系统对动态环境的鲁棒性不足,稳定性表现欠佳。

综上所述,现有算法普遍存在定位精度低、特征点利用率低以及地图构建质量差等问题。为此,本文提出一种结合轻量化 YOLO11n 的动态视觉 SLAM 算法。该算法基于 ORB-SLAM3 框架[14],在跟踪线程中引入动态特征点剔除模块,利用 YOLO11n 目标检测网络识别潜在动态区域。为解决在动态区域内将特征点全部剔除导致的资源浪费问题,本文采用 Lucas-Kanade(LK)光流法[15]的筛选策略,结合光流信息与时间一致性约束,保留可能属于静态背景的特征点,在减少动态目标干扰的同时,提高特征点利用率,增强 SLAM 系统的定位精度与地图构建质量。本文增加了语义地图构建线程,利用红绿蓝 - 深度(RGB-D)传感器获取的深度信息生成点云,并剔除 YOLO11n 检测到的动态物体点云,确保构建的点云地图真实反映静态环境。此外,前端提取的语义信息被用于后端地图构建,以生成静态语义地图,从而进一步提升 SLAM 系统在复杂环境下的稳健性,并促进人机交互,为智能导航和服务机器人等应用提供更精准的信息支撑。

1　动态特征点剔除

1.1　YOLO11n 算法

YOLO11n 是 YOLO11 系列的轻量级目标检测网络,专为嵌入式设备和实时应用场景设计。与标准的 YOLO11 相比,YOLO11n 具有更高的计算效率和更小的模型体积,同时具有良好的检测性能。为满足动态环境下对 SLAM 系统高实时性的要求,本文选取 YOLO11n 作为目标检测的算法模型。YOLO11n 架构如图 1 所示,主要由输入层、主干(Backbone)网络、颈部(Neck)网络和头部(Head)网络四部分组成。输入层接收图像数据,为后续特征提取提供原始信息。Backbone 部分依次堆叠 C2PAS、SPPF、多个交替的 C3K2 和批归一化激活(CBS)模块,通过 C2PAS 完成初始特征提取、SPPF 增强尺度不变性、C3K2 通过实现多路径条件判断的高效特征融合以及利用 CBS 提升特征表达,从而提取丰富且具代表性的基础特征。Neck 部分包含上采样(Upsample)层与特征拼接(Concat)层,前者将低分辨率特征图恢复为高分辨率,后者融合不同尺度特征图,使模型能综合利用多层次特征,增强对不同尺度目标的检测能力。Head 部分由 CBS、深度可分离卷积(DSC)与标准二维卷积(Conv2d)构成,先由 CBS 模块和 DSC 模块对融合特征进一步处理,再经 Conv2d

层预测目标类别与位置,完成目标检测任务。

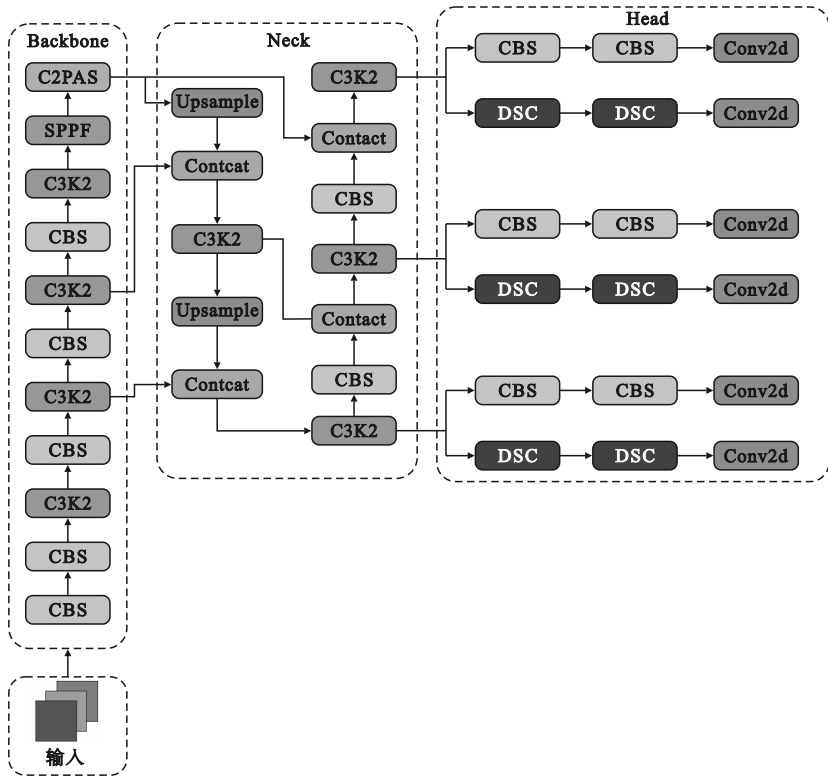


图 1 YOLO11n 算法网络结构

Fig. 1 Network structure of YOLO11n algorithm

1.2 LK 光流法

光流描述了像素在连续图像间的运动,稀疏光流仅计算特征点的运动,在特征跟踪和匹配中,采用稀疏的 LK 光流法计算视觉里程计提取的特征点光流场。首先对移动物体的特征点做如下假设。

1)像素灰度不变性:同一空间点在不同图像中的灰度值保持不变。

2)小运动假设:相邻帧时间间隔短,像素位置变化小。

3)空间一致性:相邻像素的运动特性相似,局部区域内运动矢量平滑。

LK 光流法通常认为相邻的图像帧随时间变化,图像的灰度可以看作时间的函数,在 t 时刻位于(x,y)处的像素灰度值可以表示为 $I(x,y,t)$,在 $t+\mathrm{d}t$ 时刻,由像素灰度不变性假设可得

$$I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)=I(x,y,t) \tag{1}$$

式中 $I(x,y,t)$和 $I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)$为同一像素点在两帧不同图像中的灰度值。根据小运动假设,对式(1)进行一阶泰勒展开得

$$I(x+\mathrm{d}x,y+\mathrm{d}y,t+\mathrm{d}t)\approx I(x,y,t)+\frac{\partial I}{\partial x}\mathrm{d}x+\frac{\partial I}{\partial y}\mathrm{d}y+\frac{\partial I}{\partial t}\mathrm{d}t \tag{2}$$

由像素灰度不变性假设可知,特征点 t 时刻与 $t+\mathrm{d}t$ 时刻灰度值相等,两边同时除以 $\mathrm{d}t$[16]可得

$$\frac{\partial I}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}+\frac{\partial I}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t}+\frac{\partial I}{\partial t}=0 \tag{3}$$

式中:$\mathrm{d}x/\mathrm{d}t$、$\mathrm{d}y/\mathrm{d}t$ 分别表示特征点沿 x 轴和 y 轴方向的速度,记为 u、v;$\partial I/\partial x$ 和$\partial I/\partial y$ 分别表示图像空间域中 x 轴方向和 y 轴方向的灰度梯度,记为 I_x 和 I_y;$\partial I/\partial t$ 表示时域上的灰度变化率,记为 I_t。由式(3)可得

$$\begin{bmatrix}I_x & I_y\end{bmatrix}\begin{bmatrix}u\\v\end{bmatrix}=-I_t \tag{4}$$

为求解式(4),需要引进额外约束。根据空间一致性假设,选用以特征点为中心的 5×5 像素邻域作为分析单元,可构建关于 u、v 的线性方程组。此时,由 25 个像素点建立的线性方程数量显著超过待求解变量的维度,形成的超定方程组为

$$\begin{bmatrix}I_x & I_y\end{bmatrix}_k\begin{bmatrix}u\\v\end{bmatrix}=-I_{tk},k=1,\cdots,5^2 \tag{5}$$

记 $\boldsymbol{A}=\begin{bmatrix}[I_x,I_y]_1\\ \vdots\\ [I_x,I_y]_k\end{bmatrix}$,$\boldsymbol{b}=\begin{bmatrix}I_{t1}\\ \vdots\\ I_{tk}\end{bmatrix}$,于是方程(5)可

写为

$$\boldsymbol{A}\begin{bmatrix}u\\v\end{bmatrix}=-\boldsymbol{b} \tag{6}$$

利用最小二乘法求解式(6),结果为

$$\begin{bmatrix}u\\v\end{bmatrix}=-(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} \tag{7}$$

在动态场景中,静态和动态特征点的光流矢量存在显著差异。静态特征点的光流矢量主要由相机运动引起,反映了相机的空间位置和姿态变化,而不受物体自身运动的影响。相比之下,动态物体的光流矢量则较为复杂,除了相机运动,还受物体自身运动的影响,表现为相机运动和物体运动的共同作用。图 2 为动态特征点光流示意图,其中 p_1、p_2 和 p_3 为运动状态下的特征点,t_1 与 t_2 为运动状态下的不同时刻。动态场景分析中,其中的虚线标注区域内的静态特征点 p_1 和 p_2 的运动轨迹完全由相机本体运动所主导,而动态特征点 p_3 的轨迹则呈现出相机运动与动态目标运动的复合效应。这种本质性的运动模式差异导致三者在连续图像间的光流场矢量产生显著的非一致性特征:静态特征点位移矢量严格遵循相机运动学模型约束,而动态特征点位移矢量则呈现运动学参数的突变特性。

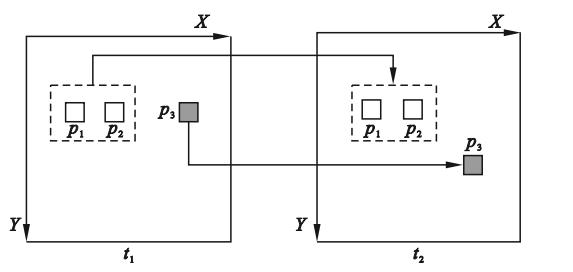


图 2　动态特征点光流示意图

Fig. 2　Schematic diagram of the optical flow of dynamic feature points

在判断动态和静态特征点时,首先计算 YOLO11n 识别出的动态区域外的静态区域特征点光流失量的均值,计算表达式为

$$\begin{bmatrix}\boldsymbol{U}\\\boldsymbol{V}\end{bmatrix}=\frac{1}{N_{\mathrm{static}}}\sum_{k=1}^{N_{\mathrm{static}}}\begin{bmatrix}u_k\\v_k\end{bmatrix} \tag{8}$$

式中:N_{static}表示静态区域的特征点数量;$[\boldsymbol{U},\boldsymbol{V}]^{\mathrm{T}}$为平均光流矢量。然后计算每个特征点光流矢量与静态平均光流矢量之间的二范数,判断是否为静态特征点,计算表达式为

$$\sqrt{(u-U)^2+(v-V)^2}>l \tag{9}$$

式中 l 为静态特征点判定阈值,将其设定为所有静态特征点平均光流矢量的 0.5 倍。若公式(9)计算结果小于预设阈值 l,则该特征点被判定为静态特征点;反之,则判定为动态特征点。

1.3　改进的 ORB-SLAM3 算法框架

ORB-SLAM3 主要包括三个线程:视觉跟踪线程、局部地图构建线程和回环检测线程。这些线程的协同工作使得 ORB-SLAM3 在静态场景中具备高精度的定位与构建地图的能力。然而,在动态环境中,该算法易受到运动物体的干扰,从而影响系统的定位精度与稳定性。

为解决上述问题,本文在 ORB-SLAM3 的基础上引入目标检测算法,并结合动态特征点剔除策略优化系统性能,算法框架如图 3 所示。首先,基于 COCO 数据集训练 YOLO11n 模型提取图像中的语义信息;然后,识别人、动物、车辆等运动概率较大的目标,并将其所在区域标记为动态区域;接着,在跟踪线程中剔除动态区域内的特征点,且利用 LK 光流法的筛选策略,保留动态区域中的静态特征点,从而在减少动态目标对特征匹配干扰的同时提高特征点的利用率;最后,在语义地图构建线程中结合语义信息,生成静态语义地图。该方法有效抑制了动态干扰,在复杂场景中能够实现更精准的定位与高质量的地图构建。

传统方法通过目标检测算法识别动态物体,并直接剔除所有动态物体特征点。然而这种方法会误删除位于动态区域中的静态特征点,降低特征点的利用率,图 4 展示了 LK 光流剔除动态特征点过程。如图 4(a)、4(b)所示,动态检框中同时包含人体和显示器,显示器显然是静态物体,其特征点被误剔除,影响了后续位姿估计的精度。为解决该问题,本文通过构建运动语义初始化与光流动态校验的双层机制,不再简单去除动态目标框及潜在动态目标框内的所有特征点,而是通过光流分析筛选出静态背景点予以保留,仅剔除真正的动态特征点,从而提高特征点的利用率。首先,利用运动目标检测算法获取动态目标边界框,将人体运动实体特征点划归动态集合 $\boldsymbol{D}$,将显示器等具有表观运动但物理静止的特征点归类为伪动态集合 $\boldsymbol{F}$,其余特征点纳入静态集合 $\boldsymbol{S}$。随后引入光流动态校验机制:采用 LK 光流法计算集合中的特征点光流矢量。针对静态集合 $\boldsymbol{S}$,通过公式(8)计算其平均光流矢量,建立静态特征点运动补偿基准。将伪动态集合 $\boldsymbol{F}$ 中的特征点光流矢量代入式(9)进行动态性校验。若大于预设阈值

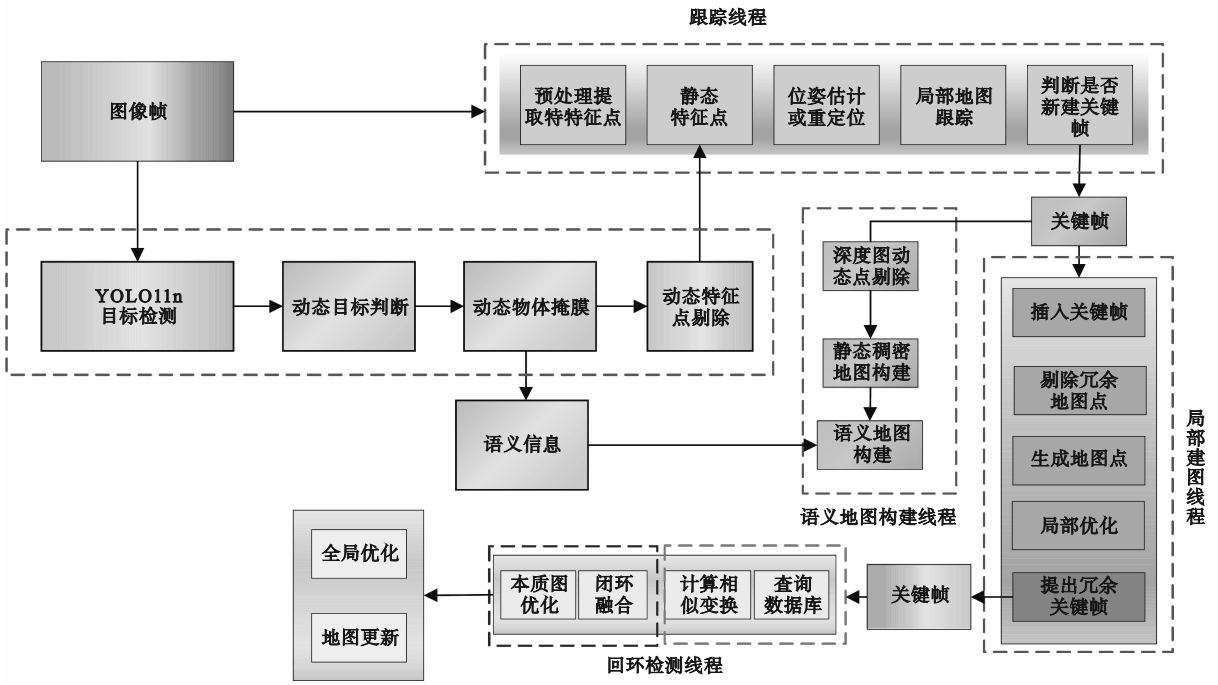


图 3 改进的 ORB-SLAM3 算法框架

Fig. 3 The improved algorithm framework of ORB-SLAM3

l,判定该特征点具有真实运动属性,将其迁移至动态集合 $\boldsymbol{D}$ 并执行伪动态特征剔除;若小于预设阈值 l,则重新归类至静态集合 $\boldsymbol{S}$。图 4(c)的结果表明,该方法有效剔除了动态检测框内人体上的动态特征点,同时保留了检测框中显示器的静态特征点,从而降低了动态特征点对系统的干扰,进一步提升了 V-SLAM 的定位精度。

(a) 提取ORB特征点 (b) 动态框中特征点剔除 (c) 保留静态特征点

图 4 LK 光流剔除动态特征点

Fig. 4 LK optical flow rejects dynamic feature points

1.4 语义地图构建

本文将深度相机的深度信息与前端提取的语义信息传至语义地图构建线程,用于构建静态语义地图。首先,通过轻量化 YOLO11n 网络实现动态实体检测,对识别出的运动目标的深度图执行零值掩码处理,深度图剔除动态物体前后对比结果如图 5 所示,以此消除动态物体对静态地图构建的干扰。在剔除动态物体的深度图基础上,根据帧间的一致性准则选取关键帧,并依据针孔成像原理将二维像素阵列映射至三维欧氏空间,其投影关系表达式为

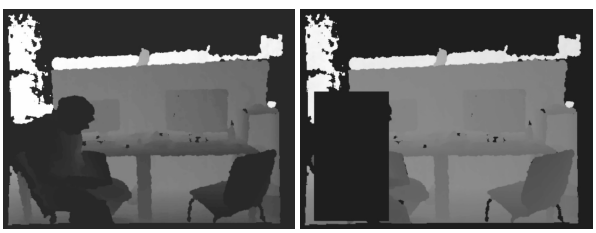
$$\begin{cases} X = \dfrac{\mu' - c_x}{f_x} Z \\ Y = \dfrac{v' - c_y}{f_y} Z \\ Z = d \end{cases} \tag{10}$$

式中:$[\mu', v']^{\mathrm{T}}$ 表示像素点的二维图像坐标;d 为该点的深度值;c_x 和 c_y 为主点坐标,是相机光轴与图像平面的交点在图像坐标系下的坐标值;f_x 和 f_y 分别是相机在 x 轴和 y 轴方向上的焦距;$[X$,

$Y,Z]^{\mathrm{T}}$ 为二维图像坐标对应的三维空间坐标。最后,通过视觉里程计解算的关键帧位姿,将点云从相机坐标系转换至世界坐标系,并基于点云匹配算法实现多帧点云的空间配准,从而生成完整的静态三维点云地图。点云全局坐标转换公式为

$$\begin{bmatrix} x_{\mathrm{w}} \\ y_{\mathrm{w}} \\ z_{\mathrm{w}} \\ 1 \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} x_{\mathrm{c}} \\ y_{\mathrm{c}} \\ z_{\mathrm{c}} \\ 1 \end{bmatrix} + \boldsymbol{T} \tag{11}$$

式中:$[x_{\mathrm{c}},y_{\mathrm{c}},z_{\mathrm{c}},1]^{\mathrm{T}}$ 为三维点云在相机坐标系下的齐次坐标;$[x_{\mathrm{w}},y_{\mathrm{w}},z_{\mathrm{w}},1]^{\mathrm{T}}$ 为点云在世界坐标系下的齐次坐标;$\boldsymbol{R}$ 为旋转矩阵;$\boldsymbol{T}$ 为平移向量。



(a) 原始深度图　　(b) 剔除动态物体后深度图

图 5　深度图剔除动态物体前后对比

Fig. 5　Comparison of depth maps before and after removing dynamic objects

最终,在构建的点云地图上融合前端提取的语义信息,并为其点云赋予对应的颜色标签,图 6 为获得的静态语义地图,采用差异化的颜色机制对电脑和椅子进行可视化表征,提升了地图的可理解性。该语义增强的静态地图不仅能直观反映真实环境的物体分布,还能为机器人导航提供关键信息。

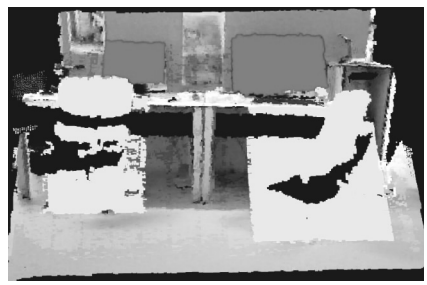


图 6　静态语义地图

Fig. 6　Static semantic map

2　实验结果与分析

2.1　目标检测算法实验

为验证 YOLO11n 目标检测算法在动态场景中的性能,面向 COCO 数据集设计专项训练方案:初始学习率设为 0.001,批量大小为 16,训练 500 轮次并设置早停机制,精度 50 轮后无提升则终止。IoU 阈值提升至 0.5 以强化检测精度。数据集按 9∶1 划分为训练集与验证集,训练前加载预训练权重,通过迭代优化使模型在动态环境中实现目标位置与类别的精准检测,为 SLAM 系统提供可靠输入。表 1 为 YOLO11n 目标检测网络性能测试结果。

表 1　目标检测网络性能测试

Table 1　Object detection network performance testing

网络模型	GFLOPS	参数量/10^6	准确率/%	召回率/%	mAP@0.5/%
YOLO11n	6.5	2.6	89.9	78.4	81.6

由表 1 数据可见,YOLO11n 具备轻量级特性,计算量(GFLOPS)与参数量充分满足动态场景下视觉 SLAM 系统的实时性要求。其 89.9% 的检测准确率与 81.6% 的平均精度均值(mAP@0.5)指标表明,该模型对动态场景中的主流目标(如行人、车辆等)具备较强的识别能力,能够有效辅助 SLAM 算法实现静态背景与动态物体的精准区分,显著降低动态特征误匹配对定位精度的干扰,同时为复杂动态环境下的视觉 SLAM 提供了可靠的语义信息支撑。

2.2　位姿估计误差分析实验

为验证改进算法的有效性,本文选用 RGBD 数据集[17]开展实验。该数据集由 Kinect 深度相机获取 RGB 图像与深度信息。实验选取 3 个图像序列:Sitting_xyz、Walking_halfsphere、Walking_xyz。Sitting_xyz 为小范围运动的低动态场景序列,Walking_halfsphere 与 Walking_xyz 是大范围运动的高动态场景序列[18]。

为量化本文算法在动态环境下的性能表现,本文设计了与 ORB-SLAM3 算法的对比实验。采用线条与阴影结合的方式实现轨迹及误差的可视化:以黑色实线表征真实轨迹,黑色虚线表示预测轨迹,通过密集的阴影区域直观反映预测位姿与真实位姿之间的误差分布。在该可视化体系下,阴影区域的覆盖范围与系统定位精度呈负相关,即阴影区域越紧凑、重叠程度越低,意味着预测轨

迹与真实轨迹的偏差越小,系统定位精度越高。图 7 ~ 8 分别为 Walking_xyz 序列真实轨迹与预测轨迹对比结果以及 Walking_halfsphere 序列真实轨迹与预测轨迹对比结果。由图 7、图 8 可见,在高动态场景中,相较于 ORB-SLAM3,改进算法剔除了动态特征点,大幅提升了定位精度,因此轨迹误差更小。反观 ORB-SLAM3,在动态场景尤其是存在动态物体快速运动的环境下,轨迹漂移现象表明系统精度受到严重影响。图 9 为低动态场景的 Sitting_xyz 序列真实轨迹与预测轨迹对比结果,由图 9 可知,改进算法与 ORB-SLAM3 的预测轨迹均与真实轨迹高度接近,二者定位精度相近。

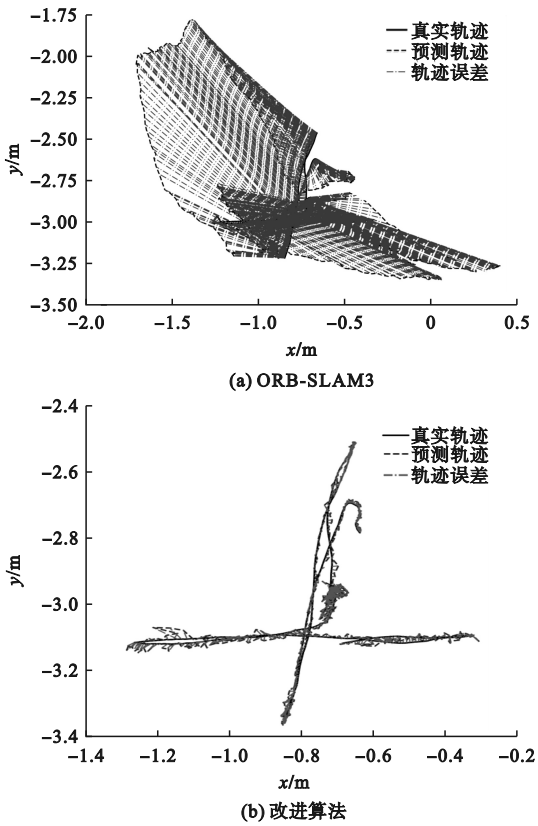


图 7 Walking_xyz 序列真实轨迹与估计轨迹对比

Fig. 7 Comparison of the real trajectories and the estimated trajectories of the Walking_xyz sequence

视觉同步定位与建图(V-SLAM)系统的量化评估体系主要由定位精度、计算效率和算法复杂度等三大核心要素构成[18]。其中,定位精度作为关键性能指标,直接影响系统性能的综合评价。本研究采用绝对轨迹误差(ATE)作为轨迹估计质量的评估基准,构建由均方根误差(RMSE)、标准差(SD)、均值误差(Mean error)及中值误差(Median error)组成的多维量化分析框架。

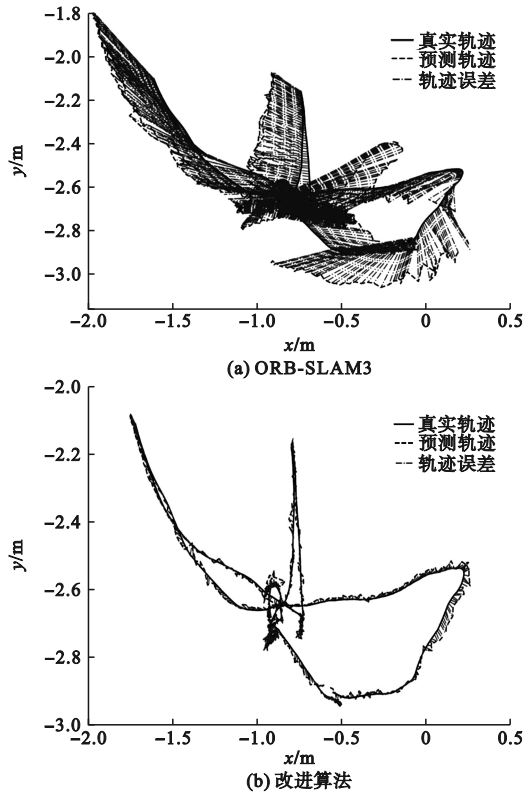


图 8 Walking_halfsphere 序列真实轨迹与估计轨迹对比

Fig. 8 Comparison of the real trajectories and the estimated trajectories of the Walking_halfsphere sequence

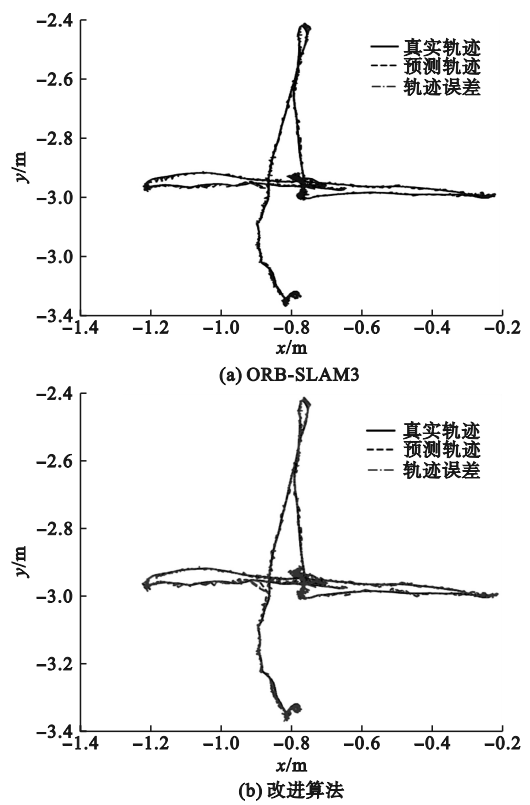


图 9 Sitting_xyz 序列真实轨迹与估计轨迹对比

Fig. 9 Comparison of the real trajectories and the estimated trajectories of the Sitting_xyz sequence

表 2 为改进前后算法性能对比结果。由表 2 数据可知,在低动态环境的 Sitting_xyz 序列中,动态物体因动作幅度较小,对位姿估计结果的影响较为有限,但实验结果显示,改进算法仍展现出一定提升效果。这一现象表明,即便在动态干扰较弱的场景下,通过优化动态物体的特征处理机制,仍能进一步降低环境噪声对定位的潜在干扰,体现了改进算法对细微动态变化的鲁棒性。在高动态场景下,本文算法通过剔除动态框中的动态特征点,仅保留静态特征点进行系统轨迹估计,显著提升了定位精度。相较于 ORB-SLAM3 算法,本文算法在 Walking_halfsphere 和 Walking_xyz 序列中的 RMSE 误差平均提升了 95.02%。上述结果表明,本文算法通过动态特征剔除机制有效抑制了高动态场景下的干扰因素,定位精度显著优于传统 ORB-SLAM3 算法,验证了改进策略的有效性和鲁棒性。

表 2　ORB-SLAM3 与改进算法 ATE 的性能对比

Table 2　Comparison of ATE performance between ORB-SLAM3 and the improved algorithm　m

数据集	ORB-SLAM3				改进算法			
	RMSE	SD	Mean	Median	RMSE	SD	Mean	Median
Sitting_xyz	0.008 2	0.004 0	0.007 1	0.006 4	0.008 1	0.004 2	0.006 9	0.006 1
Walking_halfsphere	0.288 2	0.105 2	0.268 3	0.255 3	0.017 5	0.009 2	0.014 9	0.012 9
Walking_xyz	0.283 1	0.127 5	0.252 8	0.242 6	0.013 4	0.007 3	0.011 1	0.009 3

本文算法仍存在以下局限性:其一,光照变化鲁棒性不足,当场景出现室内外切换等剧烈光照突变时,光流法的亮度恒定假设易被打破,静态物体表面因光照变化产生的像素灰度波动可能被误判为伪动态特征,导致静态背景基准计算偏差;其二,遮挡与特征稀疏场景适应性有限,动态物体大面积遮挡静态背景时,特征点数量锐减会引发光流矢量统计失效,特别是在密集人群场景中,若静态特征点占比低于 10%,公式(8)的平均光流计算将无法准确反映背景运动,造成伪动态特征校验失灵。针对上述问题,从以下方向优化:一是设计自适应特征保留机制,对半静态物体(如缓慢移动的家具)构建运动概率模型,通过时序光流变化趋势判断其动态属性,避免误归为完全动态特征;二是强化光照鲁棒性优化,引入光度不变性特征描述子(如归一化 RGB 特征)或融合 IMU 数据,建立光照变化补偿模型,缓解亮度假设失效对光流计算的干扰,提升在复杂环境下进行特征点分类的准确性。表 3 为改进算法与 ORB-SLAM3 算法每帧耗时对比结果。

表 3　改进算法与 ORB-SLAM3 算法每帧耗时对比

Table 3　Comparison of the time per frame between the improved algorithm and the ORB-SLAM3 algorithm ms

数据集	ORB-SLAM3	本文算法
Sitting_xyz	25.3	26.1
Walking_halfsphere	27.8	31.5
Walking_xyz	28.2	32.8

由表 3 可知,静态场景 Sitting_xyz 中,改进算法每帧耗时 26.1 ms,与 ORB-SLAM3 的 25.3 ms 基本持平,满足实时需求。动态场景(Walking_halfsphere、Walking_xyz)中,改进算法耗时增至 31.5 ~ 32.8 ms,较 ORB-SLAM3 增加 12% ~ 15%,主要因为动态特征检测剔除的计算开销。但改进算法在动态场景中定位精度得到显著提升,且实时性仍满足多数应用要求。

通过实验将本文算法与 DynaSLAM 和 DS-SLAM[19]算法在动态数据集上的运行结果进行了对比,采用 ATE 的 RMSE 误差作为评估指标。为消除硬件设备等外部因素的影响,引入相对提升率 ρ,以直观反映改进算法相较于原 ORB-SLAM3 算法在相同实验条件下的性能提升,其计算表达式为

$$\rho = \left(1 - \frac{\alpha}{\beta}\right) \times 100\% \tag{12}$$

式中:α 代表原 ORB-SLAM3 算法所获取的绝对轨迹误差;β 为改进算法所得到的绝对轨迹误差。表 4 为改进算法与其他动态 SLAM 算法的 ATE 对比结果。

表 4　改进算法与其他动态 SLAM 算法的 ATE 对比

Table 4　Comparison of the ATE between the improved algorithm and other dynamic SLAM algorithms　%

数据集	DS-SLAM	DynaSLAM	本文算法
Walking_halfsphere	89.90	88.80	93.93
Walking_xyz	92.60	93.30	95.26

由表 4 可知,相较 DynaSLAM 和 DS-SLAM,本文算法在平均定位精度上分别提升了 3.94% 和 3.68%。

2.3 点云地图对比实验

为验证本文算法在动态场景下构建稠密地图的实际效果,采用 Walking_xyz 作为图像数据集。图 10 展示了由 ORB-SLAM3 构建的点云地图,由于人物在场景中的移动,地图中产生了明显的移动残影。图 11 呈现的是经过本文算法剔除动态点后所得到的静态稠密语义地图。该地图不仅展现出了较高的质量水准,在结构完整性与细节清晰度上表现出色,且蕴含着详细的语义信息,为机器人导航进程提供精准且完备的语义支撑体系。

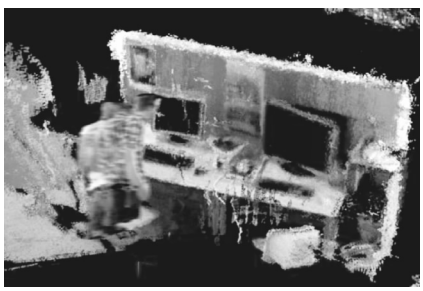


图 10 ORB-SLAM3 构建的稠密点云地图

Fig. 10 Dense point cloud map of ORB-SLAM3

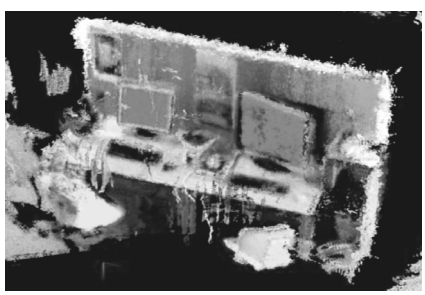


图 11 本文算法构建的静态语义地图

Fig. 11 The static semantic map constructed by the algorithm in this paper

3 结论

本文融合 ORB-SLAM3、轻量化 YOLO11n 目标检测网络与 LK 光流法,提出了适应动态场景的视觉 SLAM 算法。本文算法在 TUM 数据集上完成了验证,实验结果表明,相比 ORB-SLAM3 算法,本文算法在动态场景下位姿估计精度的均方根误差提升了 95.02%,可以完全消除动态特征点的影响,语义地图构建质量也显著提升。相较 DynaSLAM、DS-SLAM 算法,定位精度分别提升了 3.94% 和 3.68%。

参考文献(References):

[1] 田野,陈宏巍,王法胜,等. 室内移动机器人的 SLAM 算法综述[J]. 计算机科学,2021,48(9):223 – 234.
TIAN Y,CHEN H W,WANG F S,et al. Overview of SLAM algorithms for mobile robots[J]. Computer Science,2021,48(9):223 – 234. (in Chinese)

[2] 罗元,沈吉祥,李方宇. 动态环境下基于深度学习的视觉 SLAM 研究综述[J]. 半导体光电,2024,45(1):1 – 10.
LUO Y,SHEN J X,LI F Y. Review of visual SLAM research based on deep learning in dynamic environments[J]. Semiconductor Optoelectronics,2024,45(1):1 – 10. (in Chinese)

[3] 严瀚宇,孙博,马天力,等. 动态场景下融合深度信息的实时语义 SLAM 方法[J]. 传感器与微系统,2025,44(2):139 – 142.
YAN H Y,SUN B,MA T L,et al. Real-time semantic SLAM method fusing depth information in dynamic scenes[J]. Transducer and Microsystem Technologies,2025,44(2):139 – 142. (in Chinese)

[4] ZELLER N,QUINT F,STILLA U. From the calibration of a light-field camera to direct plenoptic odometry[J]. IEEE Journal of Selected Topics in Signal Processing,2017,11(7):1004 – 1019.

[5] ZOU D P,TAN P. CoSLAM:collaborative visual SLAM in dynamic environments[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013,35(2):354 – 366.

[6] YU C,LIU Z X,LIU X J,et al. DS-SLAM:a semantic visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid,Spain:IEEE,2018:1168 – 1174.

[7] 刘建军,卢大威,胡雪花,等. 基于点线特征的快速单目惯性 SLAM 算法[J]. 国外电子测量技术,2022,41(3):14 – 19.
LIU J J,LU D W,HU X H,et al. Fast monocular inertial SLAM algorithm based on point line features[J]. Foreign Electronic Measurement Technology,2022,41(3):14 – 19. (in Chinese)

[8] BESCOS B,FÁCIL J M,CIVERA J,et al. DynaSLAM:tracking,mapping,and inpainting in dynamic scenes[J]. IEEE Robotics and Automation Letters,2018,3(4):4076 – 4083.

[9] ZHAO X,ZUO T,HU X Y. OFM-SLAM:a visual semantic SLAM for dynamic indoor environments[J]. Mathematical Problems in Engineering,2021,2021(1):5538840.

[10] SU P,LUO S Y,HUANG X C. Real-time dynamic SLAM algorithm based on deep learning[J]. IEEE Access,2022,10:87754 – 87766.

[11] 李东宇,王绪娜,高宏伟. 基于驾驶场景的高效多模态融合检测方法[J]. 沈阳理工大学学报,2024,43(3):18 – 25.
LI D Y,WANG X N,GAO H W. Efficient multi-modal fusion detection method based on driving scenes[J]. Journal of Shenyang Ligong University,2024,43(3):18 – 25. (in Chinese)

[12] GAO R Z,LI Z H,LI J F,et al. Real-time SLAM based on dynamic feature point elimination in dynamic environment[J]. IEEE Access,2023,11:113952 – 113964.

[13] 李嘉铭,解明扬,张民,等. 动态环境下基于语义信息与几何约束的视觉 SLAM 系统[J]. 智能科学与技术学报,2023,5(4):477 – 485.
LI J M,XIE M Y,ZHANG M,et al. Visual SLAM based on semantic information and geometric constraints in dynamic environment[J]. Chinese Journal of Intelligent Science and Technology,2023,5(4):477 – 485. (in Chinese)

(下转第 23 页)