

# 基于加权静态特征和选择性优化的视觉惯性 SLAM 算法

齐咏生<sup>1,2</sup> 崔光通<sup>1</sup> 刘利强<sup>1,2</sup> 苏建强<sup>1,2</sup> 张丽杰<sup>1,2</sup>

(1. 内蒙古工业大学电力学院, 呼和浩特 010080;

2. 内蒙古自治区高等学校智慧能源技术与装备工程研究中心, 呼和浩特 010080)

**摘要:** 针对单一视觉传感器 SLAM 技术在动态环境中存在精度低, 可靠性差, 导致无法准确估计相机位姿的问题, 提出一种基于加权静态特征和选择性优化的视觉 SLAM 算法 (CW-SLAM)。在前端加入动态特征点检测, 使用 GC-RANSAC 算法分离出内/外点并拟合最优的基础矩阵后, 设计一种基于加权静态特征检测方法, 剔除来自动态特征的错误约束, 提升匹配精度, 将稳定的特征用于后端位姿优化; 采用因子图模型, 构建以视觉为主系统, IMU 为辅系统, 通过引入辅系统 IMU 里程计因子约束主系统误差, 并接收 VIO 因子实现运动预测和位姿优化的全新结构; 提出一种选择性优化策略消除暂时静态目标的影响, 对回环关键帧进行聚类假设后, 根据因子图优化模型建立约束组的选择性优化过滤假阳性回环假设。通过与经典的 SLAM 算法进行对比, 并在 TUM 公开数据集和真实场景中验证了算法有效性, 结果表明本文算法能有效抑制动态和暂时静止目标对位姿估计的影响, 提升了精度和可靠性。

**关键词:** 同时定位与建图; 动态环境; 特征点检测; 选择性优化; 位姿估计

**中图分类号:** TP242.6 **文献标识码:** A **文章编号:** 1000-1298(2026)06-0227-11

**OSID:**



## Visual-inertial SLAM Algorithm Based on Weighted Static Features and Selective Optimization

QI Yongsheng<sup>1,2</sup> CUI Guangtong<sup>1</sup> LIU Liqiang<sup>1,2</sup> SU Jianqiang<sup>1,2</sup> ZHANG Lijie<sup>1,2</sup>

(1. College of Electric Power, Inner Mongolia University of Technology, Hohhot 010080, China

2. Inner Mongolia Autonomous Region University Smart Energy Technology and Equipment Engineering Research Center, Hohhot 010080, China)

**Abstract:** Aiming at the problem that single visual sensor SLAM technology has low accuracy and poor reliability in dynamic environment, which leads to the inability to accurately estimate the camera pose, a visual SLAM algorithm based on weighted static feature points and selective optimization (CW-SLAM) was proposed. Firstly, dynamic feature point detection was added to the front end. After using the GC-RANSAC algorithm to separate the inner/outer points and fit the optimal basic matrix, a weighted static feature detection method was designed to eliminate the error constraints from the dynamic features, improve the matching accuracy, and use the stable features for back-end pose optimization. Secondly, the factor graph model was used to construct a new structure with vision as the main system and IMU as the auxiliary system. By introducing the auxiliary system IMU odometer factor to constrain the main system error, and receiving the VIO factor to realize the motion prediction and pose optimization. Finally, a selective optimization strategy was proposed to eliminate the influence of temporary static targets. After clustering the loopback key frames, the selective optimization of the constraint group was established according to the factor graph optimization model to filter out the false positive loopback hypothesis. Compared with the classical SLAM algorithm, the effectiveness of the algorithm was verified on the TUM public dataset and in the real environment. The experimental results showed that the algorithm can

收稿日期: 2024-12-18 修回日期: 2025-03-13

**基金项目:** 国家自然科学基金项目(62363029)、内蒙古科技计划项目(2021GG0256)、内蒙古自然科学基金项目(2022MS06018)、呼和浩特市“政产学研推用银”创新联合体项目(2023RC-联合体-10)和高校院所协同创新项目(XTCX2023-16)

**作者简介:** 齐咏生(1975—),男,教授,博士,主要从事机器人协同控制技术研究,E-mail: qys@imut.edu.cn

**通信作者:** 刘利强(1975—),男,教授,博士,主要从事计算机视觉和新能源发电技术研究,E-mail: llqiang@imut.edu.cn

effectively suppress the influence of dynamic and temporary stationary targets on pose estimation, and improve the accuracy and reliability.

**Key words:** simultaneous localization and mapping; dynamic environment; feature point detection; selectivity optimization; pose estimation

## 0 引言

同时定位与地图构建 (Simultaneous localization and mapping, SLAM) 技术使自助机器人和车辆能够在不熟悉的环境中导航。其原理是在构建地图的同时确定自身在地图中位置<sup>[1]</sup>。视觉同时定位与地图构建 (Visual SLAM, VSLAM) 是一种流行的变体, 其以相机作为主要传感器。VSLAM 发展迅速, 像 MonoSLA<sup>[2]</sup> 和 PTAM<sup>[3]</sup> 等开创性工作实现了实时六自由度相机跟踪。VSLAM 方法主要分为 2 类: ① 基于特征点的方法, 致力于最小化图像间的重投影误差。② 直接法, 致力于最小化光度误差<sup>[4]</sup>。这 2 类方法经过深入研究, 产生了如 ORB-SLAM<sup>[5]</sup>、LSD-SLAM<sup>[6]</sup>、DVO<sup>[7]</sup> 和 SVO<sup>[8]</sup> 等系统。

尽管视觉同步定位与地图构建技术在过去几十年中取得了显著进展, 但大多研究集中在静态环境, 所观测到的场景中通常包含动态物体, 这限制了其在实际动态场景中的适用性<sup>[9]</sup>。动态视觉 SLAM 的关键挑战在于, 动态对象的存在导致数据关联和特性匹配错误从而降低位姿估计精度<sup>[10-12]</sup>。这种动态性阻碍了 VSLAM 在自主导航和自动驾驶<sup>[13]</sup> 等方面的应用。因此, 迫切需要一种在这种动态环境中高效且精确的 VSLAM 方法。文献[14-16]使用深度学习进行语义分割, 然而这些数据无法识别缺乏训练样本的未知或未标记的动态对象。一些半监督方法试图克服这一限制<sup>[18]</sup>, 但大型网络在效率和泛化能力方面仍存在缺陷。DE-SLAM<sup>[19]</sup> 采用单目视觉结合深度学习实现对环境中动态物体实时检测, 但单一视觉传感器在光线差时, 会使得定位精度下降, PDR<sup>[20]</sup> 使用基于权重的因子图模型来提高抗干扰能力, GNSS<sup>[21]</sup> 提出轻量里程计网络校正 INS 误差, MCG-SLAM<sup>[22]</sup> 采用一种基于多因素约束图优化的紧耦合结构优化多因子约束中的因子节点以完成位姿优化, LDG-CSLAM<sup>[23]</sup> 通过基于点云曲率分析的关键帧提取, 构建因子图模型对自身和相对里程计进行真实的实时优化来降低动态误差, 此模型依赖于复杂 3D 环境特征的提取, DM-VI-SLAM<sup>[24]</sup> 设计视觉辅系统观测因子来约束 IMU 主系统偏差, 但在动态场景下的准确度和可靠性表现并不佳。另一个研究方向是利用几何约束, 通过背景建模和帧间比较和来确定动态信

息<sup>[25]</sup>。基于像素的操作无需先验知识即可实现精确的运动分割。然而, 此类技术计算成本依然很高。最近提出一种利用语义信息和几何特征的方法, 例如 PL-SLAM<sup>[26]</sup>。但当物体呈暂时静态而开始运动时, 仍存在局限性, 这些暂时静止的物体可能会由于误报而导致回环出现严重错误, GNC<sup>[27]</sup> 联合最小求解器来计算后端的鲁棒性, 以减少误报回环在优化中的影响。但由于 GNC 专注于瞬时误报闭环, 因此无法处理由暂时静止物体引起的误报回环。

针对上述问题, 基于 ORB-SLAM3 算法, 本文设计一种视觉惯性 SLAM 方法, 以消除由动态特征和暂时静态物体引起的误报回环对位姿估计的影响。在 ORB-SLAM3 线程前端使用基于图割的 GC-RANSCA<sup>[28]</sup> 的内/外点判定所得到的静态/动态特征点的先验估计。在后端通过在视觉里程计因子图中加入 IMU 里程计因子约束偏差。在算法回环检测部分, 设计一种选择性全局优化的策略, 引入 B-R 正则化因子<sup>[29]</sup> 将来自真/假阳性的回环约束组分为多个假设。最后在含有视觉信息和惯性信息的高/低动态环境下的数据集和真实场景中进行仿真实验, 以验证方法有效性。

## 1 原理

### 1.1 极线几何和基础矩阵

在视觉 SLAM 的理论基础中, 通过极线几何和本质矩阵, 能够建立多个相机视图之间的几何关系。考虑 2 个相机视图之间的关系可以通过基础矩阵  $F$  来表示。2 个相机视图图像坐标分别为  $\mathbf{x}$  和  $\mathbf{x}'$ , 与本质矩阵的关系为

$$\mathbf{x}'^T \mathbf{E} \mathbf{x} = 0 \quad (1)$$

式中  $\mathbf{E}$ ——本质矩阵

式(1)表达了极线几何的基本原理, 对于在 1 幅图像中找到的特征点, 在另 1 幅图像上可能位于一条线上。基础矩阵包含相机内参信息以及 2 个相机之间的运动关系, 计算式为

$$\mathbf{F} = \mathbf{K}^{-T} [\mathbf{t}]_{\times} \mathbf{R} \mathbf{K}^{-1} \quad (2)$$

式中  $\mathbf{t}$ ——平移向量

$[\mathbf{t}]_{\times}$ —— $\mathbf{t}$  的反对称矩阵

$\mathbf{K}$ ——相机内参矩阵

$\mathbf{R}$ ——旋转矩阵

通过分解基础矩阵,可以恢复相机运动姿态。

### 1.2 回环检测

词袋模型 (DBoW2) 是一种用于快速检索和匹配场景的技术,广泛应用于回环检测中。

对于某幅图像  $A$ ,其特征点可对应多个单词,组成 DBoW2,即

$$A = \{ (w_1, \eta_1), (w_2, \eta_2), \dots, (w_i, \eta_i) \} \stackrel{def}{=} \mathbf{v}_A \quad (3)$$

- 式中  $w_i$  ——字典树的某个叶子节点
- $\eta_i$  ——  $w_i$  对应权重
- $\mathbf{v}_A$  ——用向量描述 1 副图像  $A$

## 2 动态场景下 SLAM 算法

### 2.1 系统概述

本文算法 (CW - SLAM) 主要根据 ORB - SLAM3 框架进行修改,由前端、局部建图、后端、回环检测和地图构建 5 个线程组成;前端设计动态特征检测线程,后端构建一种以视觉为主系统、IMU 为辅系统的因子图优化结构来代替原始图优化方式,回环检测部分提出一种选择性全局优化的策略。算法总框架如图 1 所示。

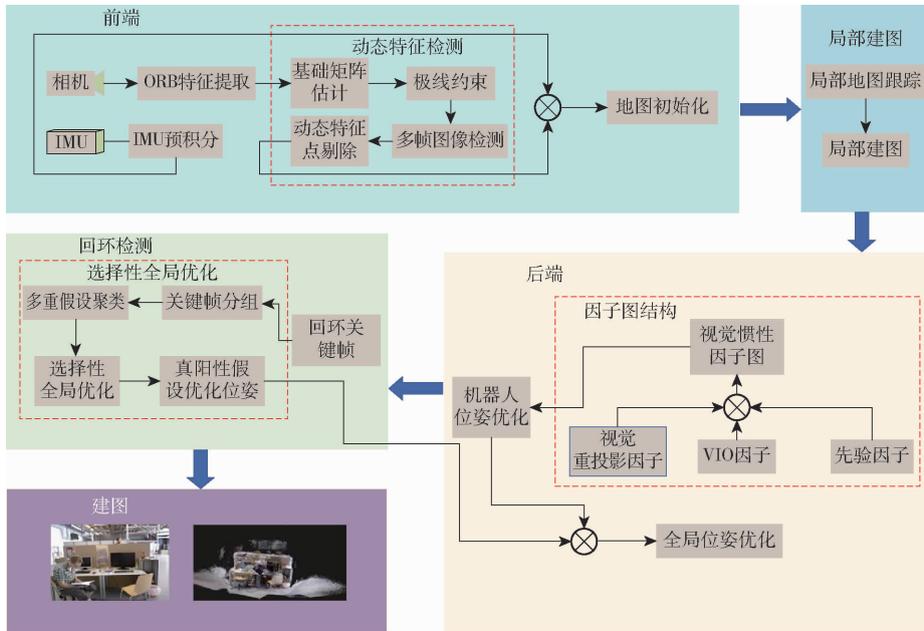


图 1 算法总框架

Fig. 1 General framework of algorithm

在经相机和 IMU 数据对齐后进行 ORB 特征提取,使用动态特征检测机制,将先前匹配的错误约束特征 (例如移动的人) 利用 GC - RANSAC 算法筛选剔除后拟合出最优的基础矩阵,使用估计的基础矩阵赋予动态和静态特征点的先验值,依据多帧数据加权检测的方式来检测动态地标,以提升特征匹配精度,将稳定的特征用于位姿优化,从而进行地图初始化和局部建图。后端优化部分由视觉重投影因子和 VIO 因子等共同构建全局因子图代替原图优化结构,以提高机器人优化位姿的准确度。此后,为消除暂时静态物体对环境建模的影响,在 ORB - SLAM3 回环检测部分加入选择性全局优化的策略,通过特征关键帧分组、多重假设聚类 and 选择性全局优化的方式来消除假阳性闭合回路的关键帧对系统的影响。最后将局部优化和回环检测优化后的位姿联合进行全局位姿优化,完成建图。

### 2.2 基于加权动态特征检测

#### 2.2.1 相机位姿估计

对于每个进入的帧,均需对其相机姿态进行合

理的初始估计。一般通过解决透视点 (PnP) 问题和三维数据关联估计 2 个连续帧之间的自我运动。然而,在动态场景中,由于存在移动物体,三维到二维数据关联会包含不正确匹配。为解决这个问题,必须检测并消除运动物体上的特征点,以便准确估计自我运动。

首先将地标大致标记为静态或动态,然后仅使用静态地标估计自我运动。如图 2 所示,对于具有

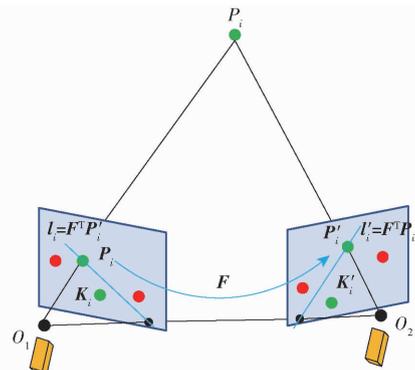


图 2 基础矩阵和对极约束

Fig. 2 Fundamental matrix and epipolar constraint

基础矩阵  $F(K_i, K'_i)$  的图像对  $(K_i, K'_i)$ , 如果  $p' \in K'_i$  位于外极线  $l'_i = F^T p_i$  上, 则三维地标  $P_i$  及其匹配的二维观测值  $(P_i, P'_i)$  可能是静态特征, 否则就是动态特征。因此, 可以利用 GC-RANSAC 算法, 将静态/动态地标识别问题表述为基础矩阵估计过程中的内点/异常点识别问题。具体地, 对于给定的  $n$  对 2D 到 2D 匹配对的集合  $M = \{(P_i, P'_i) | i = 1, 2, \dots, n\}$ , 在 GC-RANSAC 的每次迭代中, 将每个匹配对标记为基础矩阵  $F$  的内点或异常点。这是通过优化能量函数来实现的, 即

$$E(L) = \sum_i B(L_i) + \lambda \sum_{(i,j) \in G} R(L_i, L_j) \quad (4)$$

$(L_i \in \{0, 1\} | i = 1, 2, \dots, n)$

式中  $L$ ——匹配集  $M$  标签分配

$G$ ——相邻的图

$L_i$ ——一对内部匹配点, 取 0

$L_j$ ——对外部离群值, 取 1

能量函数一元项计算式为

$$B(L_i) = \begin{cases} K(\phi(P_i, P'_i, \theta), \varepsilon) & (L_i = 0) \\ 1 - K(\phi(P_i, P'_i, \theta), \varepsilon) & (L_i = 1) \end{cases} \quad (5)$$

其中  $K(\sigma, \varepsilon) = \exp(-\sigma^2 / (2\varepsilon^2))$

式中  $\theta$ ——基础矩阵  $F$  的角度参数

$\phi(P_i, P'_i, \theta)$ ——匹配对  $(P_i, P'_i)$  到基础矩阵  $F$  的距离

$\varepsilon$ ——确定内点和外点的阈值

$\sigma$ ——标准差

成对的能量函数计算式为

$$R(L_i, L_j) = \begin{cases} 1 & (L_i = L_j) \\ (B(L_i) + B(L_j)) / 2 & (L_i = L_j = 0) \\ 1 - (B(L_i) + B(L_j)) / 2 & (L_i = L_j = 1) \end{cases} \quad (6)$$

根据经验, 将  $\lambda$  设为 0.2, 将  $\varepsilon$  设为 0.15。图 3 显示了使用这种基于 GC-RANSAC 的方法选择静



(a) 使用GC-RANSAC前当前帧和帧间的特征匹配



(b) 使用GC-RANSAC后标记为内联的特征点对

图3 GC-RANSAC的静态特征点选择

Fig. 3 Static feature points selection by GC-RANSAC

态特征点的实例, 其中左图为当前帧, 右图为参考帧, 本文将当前帧之后的第 15 帧作为参考帧。

使用估计的基础矩阵推导出动态或者静态的先验值, 以实现精确的动态点检测。具体地, 对于每个二维匹配对  $(P_i, P'_i)$ ,  $P_i \in K_i$ , 其中  $K_i$  和  $K'_i$  分别为当前帧和上一帧, 假设  $P_i$  为对应的三维地标,  $L_i \in K_i$ ,  $L'_i \in K'_i$ , 分别和极线  $l_i = F^T P'_i = (A_i, B_i, C_i)$ ,  $l'_i = F^T P_i = (A'_i, B'_i, C'_i)$  相对应, 计算 2D 特征点和极线之间的距离  $d_i = |l_i p_i| / \sqrt{A_i^2 + B_i^2}$ ,  $d'_i = |l'_i p'_i| / \sqrt{A'^2_i + B'^2_i}$ 。如果  $p_i$  是静态点, 则对极极线距离  $\gamma_i = (d_i + d'_i) / 2$  会很小。因此, 将静态地标的似然定义为  $P_i^\gamma = \exp(-(\gamma_i - \mu_\gamma) / (2\sigma_\gamma^2))$ , 其中  $\mu_\gamma$  为  $\gamma_i$  平均值, 然后继续使用  $P_i^\gamma$  作为每个地标  $P_i$  的静态/动态的识别先验, 以检测动态点。

### 2.2.2 动态特征点检测

通过上述方法可粗略地将三维地标识别为静态或动态。动态点的观测结果相比静态点的观测结果更易出错, 此外, 动态点在重投影点和相应的二维特征点之间存在较大的重投影误差, 这些观测结果促使进行更精确的特征估计。

在 SLAM 处理过程中, 每个地标都可以在多个关键帧中观察到。记录下相应的二维观测值  $o_j^i$ , 即每个三维地标  $P_i$  在关键帧  $j$  中的位置。计算  $P_i$  和  $o_j^i$  之间的光度再投影误差  $e_j^i$ , 然后通过  $\alpha_i = (\sum_j e_j^i) / \beta_i$  计算平均再投影误差, 其中  $\beta_i$  为  $P_i$  观测总和。根据上文地标  $P_i$  的静态似然先验  $P_i^\gamma$ , 使用所有的观测数据定义第 2 个静态似然:  $P_i^\beta = \exp(-(\beta_i - \mu_\beta) / (2\sigma_\beta^2))$ , 第 3 个是平均重投影误差:  $P_i^\alpha = \exp(-(\alpha_i - \mu_\alpha) / (2\sigma_\alpha^2))$ , 其中  $\mu$  和  $\sigma$  分别表示数量的平均值和标准差。对于每个地标, 有 3 种不同的方式来估计地标  $P_i$  为静态的似然:  $P_i^\alpha$ 、 $P_i^\beta$ 、 $P_i^\gamma$ 。计算这些估计的加权平均值, 以给出  $P_i$  为静态的总体似然:  $P_i^\lambda = \lambda_1 P_i^\alpha + \lambda_2 P_i^\beta + \lambda_3 P_i^\gamma$ , 其中  $\lambda_1 + \lambda_2 + \lambda_3 = 1$ 。如果  $P_i^\lambda$  超过给定阈值  $t$ , 则  $P_i$  最初被标记为静态, 并与静态置信度  $c$  相关联; 否则, 其被标记为动态, 静态置信度为  $1 - c$ 。根据经验设置  $\lambda_1 = \lambda_2 = \lambda_3$ 。根据文献[30], 函数定义为

$$\psi_u(x_i) = \begin{cases} -\lg(c)I & (P_i^\lambda > t) \\ -\lg(1-c)I & (P_i^\lambda < t) \end{cases} \quad (7)$$

式中  $I(\cdot)$ ——指示函数  $x_i$ ——待判断的点

所提方法显著改善了静态/动态点标记的结果。即使对于高度动态的场景, 动态特征点也能被准确分割。



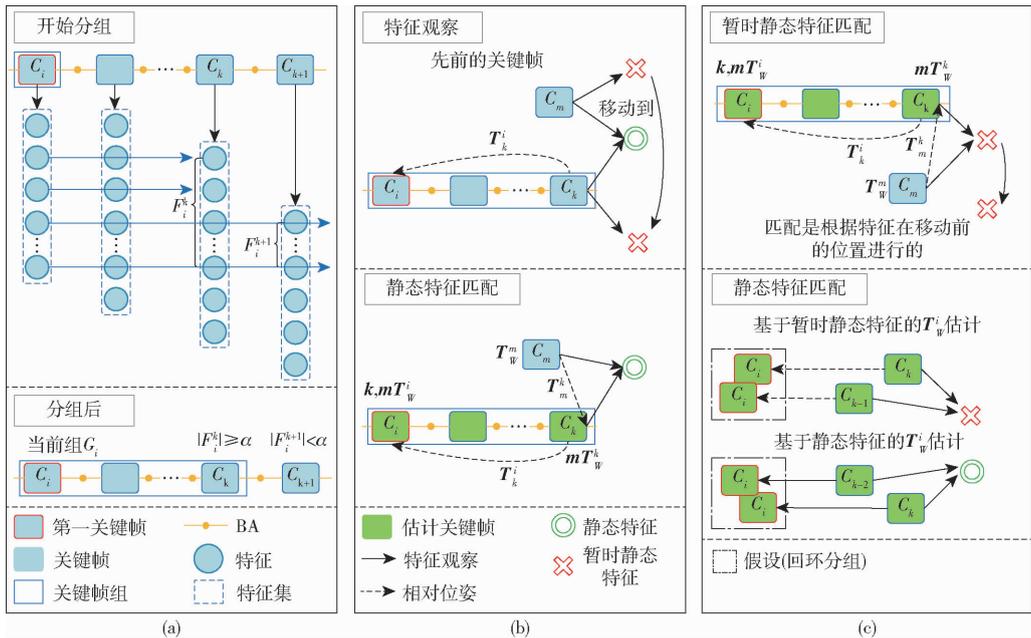


图5 多重假设聚类过程

Fig. 5 Multiple hypothesis clustering process

如图5a所示,在对回环进行分组前,必须对至少共享最少数量的跟踪特征的关键帧进行分组。从第*i*个相机帧 $C_i$ 开始的分组定义为

$$\text{Group}(C_i) = \{C_k \mid |F_i^k| \geq \alpha, k \geq i\} \quad (8)$$

式中  $\alpha$ ——跟踪特征最小数量

$F_i^k$ ——从 $C_i$ 到 $C_k$ 跟踪的特征集,将 $\text{Group}(C_i)$ 表示为 $G_i$

#### 2.4.2 多重假设聚类

按照上述方法对关键帧进行分组后,DBoW2将从 $C_i$ ( $C_k \in G_i$ 且 $m < i$ )开始,识别与当前分组 $G_i$ 中每个关键帧 $C_k$ 相似的关键帧 $C_m$ 。如果没有相似的关键帧,则跳过 $C_k$ 。在为 $k$ 确定最多3个不同的 $m$ 之后, $C_k$ 和这些关键帧之间进行特征匹配,从而得到相对位姿 $T_m^k$ 。利用 $T_m^k$ 可得 $C_k$ 在世界坐标系中的估计位姿 ${}^m T_W^k$ ,即

$${}^m T_W^k = T_m^k T_W^m \quad (9)$$

式中  $T_W^m$ —— $C_m$ 在世界坐标系位姿

然而,很难直接计算当前组中不同关键帧的回环之间的相似性。假设 $C_k$ 和 $C_i$ 之间的相对位姿 $T_i^k$ 足够准确,则 $C_i$ 在世界帧中估计姿态可表示为

$${}^k, m T_W^i = T_i^k m T_W^k \quad (10)$$

式中  $T_W^i$ ——匹配的估计结果

如果用于匹配的特征来自同一对象,即使匹配的 $C_k$ 和 $C_m$ 不同,匹配的估计结果 $T_W^i$ 也会相互接近。因此,在计算回环的估计结果 $T_W^i$ 之间的欧氏距离后,可将欧氏距离较小的回环聚类,剩下的归为另一类。如图5c所示。根据选择的回环群不同,图优化的轨迹结果也不同。因此,每个聚类都可称为

一个假设。为了减小计算成本,通过比较假设中回环的基数,采用前2个假设。当前组 $G_i$ 的这2个假设分别记为 $H_i^0$ 和 $H_i^1$ 。目前还无法区分真阳性和假阳性的假设,因此,下一节将介绍在候选假设中确定真阳性假设的方法。

#### 2.4.3 约束组选择性优化

最近的视觉SLAM算法大多使用图优化。令 $C$ 、 $T$ 、 $L$ 和 $W$ 分别表示关键帧、位姿、回环和所有权重的集合,因子图优化可表示为

$$f = \min_R \left\{ \underbrace{\sum_{i \in D} \|r(T_{i+1}^i, R)\|_{P_{T_i^{i+1}}}^2}_{\text{local edge}} + \underbrace{\sum_{(i,k) \in S} \rho_H \|r(T_k^j, R)\|_{P_S}^2}_{\text{loop closure edge}} \right\} \quad (11)$$

式中  $T_{i+1}^i$ ——2个相邻关键帧 $C_i$ 、 $C_{i+1}$ 之间的局部位姿

$T_k^j$ —— $C_j$ 、 $C_k$ 之间的相对位姿

$P_{T_i^{i+1}}$ ——局部位姿协方差矩阵

$P_S$ ——回环协方差矩阵

$r(T_{i+1}^i, R)$ ——关键帧*i*和*i+1*之间的位姿残差

$r(T_k^j, R)$ ——关键帧和回环之间的相对位姿残差,用于闭环约束

式(11)中第1项表示相邻关键帧之间的相对位姿进行优化,确保估计的位姿与相机移动时的观测值保持一致。第2项表示通过回环检测来优化轨迹,利用之前已经观察到的关键帧来修正整个轨迹,减少累计漂移。

对于分组 $G_i$ 的2个假设,引入一个受B-R对

偶启发的正则化因子,则因子图优化可表示为

$$f' = \min_{\mathbf{T}, \mathbf{w}} \left\{ \sum_{i \in C} \|\mathbf{r}(\mathbf{T}_{i+1}, \mathbf{T})\|_{P_{T_{i+1}}}^2 + \sum_{H_i \in H} \left( \sum_{(j,k) \in H_0^i} \frac{w_0^i}{|H_0^i|} \|\mathbf{r}(\mathbf{T}_k, \mathbf{T})\|_{P_L}^2 + \sum_{(j,k) \in H_1^i} \frac{w_1^i}{|H_1^i|} \|\mathbf{r}(\mathbf{T}_k, \mathbf{T})\|_{P_L}^2 + \lambda_i \Phi_l^2(w_i) \right) \right\} \quad (12)$$

式中  $w_0^i, w_1^i$ ——假设的权重

$w_i$ ——权重之和

$H$ ——假设集合

其中第1项和式(11)一致,第2项和第3项引入不同的假设集合  $H_0^i$  和  $H_1^i$ 。每个假设对应一组回环检测,这些回环可能来自静止物体或动态物体。 $\lambda_i \in \mathbf{R}^+$  常数参数,回环正则化因子  $\Phi_l$  定义为

$$\Phi_l(w_i) = 1 - w_i = 1 - (w_0^i + w_1^i) \quad (13)$$

式中  $w_0^i, w_1^i \in [0, 1]$ , 将  $w_i$  除以每个假设的基数来确保权重不受假设中的回环数影响。通过假设的权重  $w_0^i, w_1^i$ , 算法可选择信任度较高的假设,并降低错误假设(例如由动态或暂时静态物体引起的回环)的影响。

式(12)使用交替优化求解。因为当前状态  $\mathbf{T}$  可以从 IMU 预积分和先前状态中估计出来,首先在固定位姿  $\mathbf{T}$  的基础上,优化闭环假设权重  $w_0^i$  和  $w_1^i$ 。权重越大,表示特征或回环假设的可信度越高。更新权重后,重新优化位姿  $\mathbf{T}$ , 以最小化相邻关键帧间的位姿和回环约束的残差,使得轨迹和约束更一致。因此,通过交替优化位姿和权重,可以逐步过滤掉错误的闭环约束,同时保持系统对正确回环的敏感性。所以,当由于多个暂时静止的对象而导致所有假设都是假阳性时,所有权重可接近0,因此,可防止由假阳性假设引起的故障。

### 3 实验验证

为了评估所提出算法准确性、鲁棒性和定位精度,进行数据集和真实场景下的实验对比。公开的 SLAM 评估数据集 TUM-RGBD<sup>[31]</sup> 为评估复杂动态环境中的 SLAM 系统提供了传感器数据和地面实况,该数据集提供了多个包含室内动态物体的序列。选择数据集中的高动态的 fr3\_walking 序列进行评估,在该序列中,2个人围着1张桌子走动,并在移动相机的同时改变桌子位置。由于算法是在 ORB-SLAM3 的基础上构建的,因此将 ORB-SLAM3 作为基准来和所提算法进行比较。ORB-SLAM3 通过融合 IMU 与积分和跟踪线程,提供了稳健而精确的视觉惯性里程测量。

数据集运行测试平台为台式计算机,其硬件配置为 Intel Corei7-10700F,主频为 2.9 GHz,内存为 32 GB,显卡为 GTX1060,搭载系统为 Ubuntu 20.04, cmake 版本为 3.17.3,未使用 GPU 加速。

#### 3.1 选择性全局优化有效性

为了验证本文提出的消除假阳性回环的方法对相机姿态估计的性能和有效性,将这种系统(不包括选择性全局优化)与完整的 SLAM 系统进行对比。通过评估 TUM-RGB-D 数据集中 5 个动态序列的绝对轨迹误差(Absolute trajectory error, ATE)均方根误差(RMSE)和标准差(SD)来进行验证。ATE 衡量估计轨迹与真实轨迹之间的全局误差;而相对位姿误差(Relative pose error, RPE)主要描述相隔固定时间间隔内的局部精度,2个指标有助于研究 SLAM 系统在不同环境条件下的性能。表1为使用和未使用选择性全局优化的实验结果。由表1可以看出,对于高动态序列(名称以“行走”开头的序列,即快速移动的人或相机),所提算法获得的 ATE 的 RMSE 和 SD 明显低于 ORB-SLAM3,在最后一类动态较少情况下,也取得了较好结果。

表1 选择性全局优化有效性验证

Tab.1 Validation of effectiveness of selective global optimization

序列	m			
	不含选择性全局优化		本文算法	
	RMSE	SD	RMSE	SD
walking-xyz	0.032 5	0.015 5	0.012 9	0.006 6
walking-halfsphere	0.045 8	0.020 5	0.017 8	0.008 3
walking-static	0.034 6	0.009 6	0.013 7	0.004 0
walking-rpy	0.095 5	0.059 8	0.036 0	0.023 8
sitting-halfsphere	0.021 1	0.010 0	0.016 2	0.007 9

#### 3.2 在数据集中判断目标真实动态性测试

图6为 TUM-RGBD 数据集下 walking 序列中的 4 帧图像和 2 幅最终建图结果,其中图 6c、6d 为当前帧后的 15 帧,图 6a~6d 为运行结果。从图 6b、6d 可明显看出所提算法很好地去除了该序列中人和椅子上潜在的动态特征点。图 6e、6f 为 ORB-SLAM3 和本文算法在 walking-xyz 序列下的稠密建图结果。由图 6e、6f 可知,当视频序列中存在移动的人作为动态干扰时,ORB-SLAM3 会造成数据关联错误,且稠密地图纹理较多,同时移动的人身上的点云存在于稠密地图中,建图结果不准确,点云数量可用性较差,而本文算法建图效果纹理少,移动的人身上点云被剔除,点云数量可用性高。

#### 3.3 ORB-SLAM3 和其他先进 SLAM 算法对比

为了评估所提算法准确性,图 7a~7d 分别表示 ORB-SLAM3 和本文算法在 TUM-RGBD 数据集中



图6 动态点剔除和建图结果

Fig. 6 Dynamic point cutting and mapping results

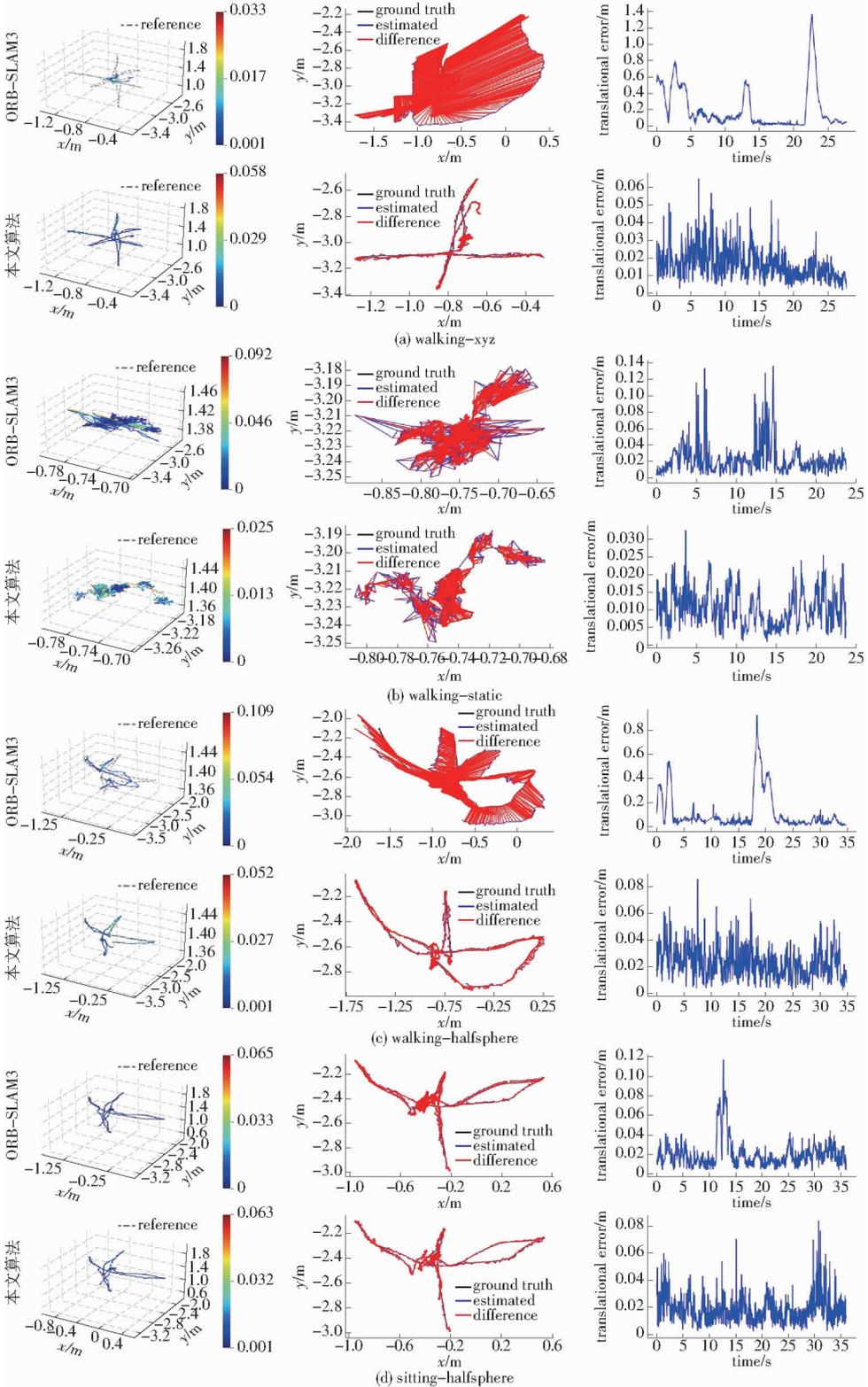


图7 三维轨迹误差热力图、绝对轨迹误差和相对轨迹误差

Fig. 7 3D trajectory error heat map, absolute trajectory error map and relative trajectory error map

的 4 个高/低动态序列 (walking - xyz、walking - static、walking - halfsphere、sitting - halfsphere) 的三维轨迹误差热力图、绝对轨迹误差 (ATE) 和相对位姿误差 (RPE) 图。如图 7 所示, 在高动态环境下, ORB - SLAM3 系统估计的运动轨迹与真实轨迹存在较大的差异, 甚至在某些区域产生错误的轨迹。相反, 所提算法估计的运动轨迹和真实轨迹高度重叠。

表 2、3 为本文算法与 ORB - SLAM3 对比结果。算法精确度通过绝对轨迹误差 (ATE) 的均方根误差 (RMSE)、误差中值 (Median) 和标准差 (SD), 平移相对位姿误差 (t. RPE) 的均方根误差 (RMSE) 和标准差 (SD) 进行评估。与 ORB - SLAM3 相比, 所提算法在 ATE 和 t. RPE 方面有显著提高; ATE 方面, 在典型的高动态序列 (walking - xyz) 中, ORB -

SLAM3 的 RMSE 从 0.286 4 m 下降到本文算法的 0.012 9 m, 降低 95.50%, 对于其他的高动态序列, 所提算法的 RMSE 与 ORB - SLAM3 相比降低 36.50% ~ 93.40%; 但在低动态序列下, 改进程度不显著。RMSE 仅降低 28.95%。这主要是因为 ORB - SLAM3 本身是为低动态环境而设计, 能够很好地处理低动态场景, 能取得良好结果。此外, 为了证明所提 SLAM 系统先进性, 将所提算法与其他最先进的动态 SLAM 系统 (即用于动态环境的在线可视 SLAM (OVD - SLAM<sup>[32]</sup>)、通过加权特征实现动态场景的鲁棒 VSLAM (WF - SLAM<sup>[33]</sup>) 进行对比 (表 4)。由表 4 可知, 本文算法在高动态序列中明显优于 OVD - SLAM 和 WF - SLAM, 除“sitting - halfsphere”序列外, 本文所提算法误差最小。

表 2 绝对轨迹误差对比

Tab. 2 Comparison of absolute trajectory error

m

序列	ORB - SLAM3			本文算法		
	RMSE	Median	SD	RMSE	Median	SD
walking - xyz	0.286 4	0.225 4	0.133 8	0.012 9	0.011 1	0.006 6
walking - halfsphere	0.226 8	0.189 2	0.092 4	0.017 8	0.012 5	0.008 3
walking - static	0.013 7	0.010 5	0.007 4	0.008 7	0.005 2	0.004 0
walking - rpy	0.154 2	0.111 9	0.072 9	0.036 0	0.021 2	0.023 8
sitting - halfsphere	0.022 8	0.018 1	0.011 8	0.016 2	0.012 3	0.007 9

表 3 平移相对位姿误差对比

Tab. 3 Comparison of translational relative positional errors

m

序列	ORB - SLAM3			本文算法		
	RMSE	Median	SD	RMSE	Median	SD
walking - xyz	0.341 2	0.154 8	0.163 4	0.021 2	0.012 2	0.014 2
walking - halfsphere	0.132 2	0.090 0	0.053 2	0.031 2	0.021 2	0.013 4
walking - static	0.023 5	0.016 6	0.007 1	0.012 3	0.009 1	0.003 8
walking - rpy	0.211 1	0.185 5	0.081 2	0.041 2	0.031 1	0.014 9
sitting - halfsphere	0.018 3	0.013 8	0.012 1	0.013 2	0.010 2	0.009 1

表 4 先进 SLAM 算法绝对轨迹误差对比

Tab. 4 Comparison of absolute trajectory errors of some advanced SLAM algorithms

m

序列	WF - SLAM <sup>[33]</sup>			OVD - SLAM <sup>[32]</sup>			本文算法		
	RMSE	Median	SD	RMSE	Median	SD	RMSE	Median	SD
walking - xyz	0.013 0	0.012 4	0.007 1	0.091 2	0.063 2	0.058 1	0.012 9	0.011 1	0.006 6
walking - halfsphere	0.024 4	0.017 4	0.007 9	0.351 3	0.031 3	0.017 3	0.017 8	0.012 5	0.008 3
walking - static	0.006 8	0.006 4	0.004 2	0.006 2	0.009 4	0.004 1	0.008 7	0.005 2	0.004 0
walking - rpy	0.025 2	0.016 4	0.032 1	0.032 1	0.019 7	0.034 5	0.036 0	0.021 2	0.023 8
sitting - halfsphere	0.017 1	0.011 1	0.008 8	0.010 3	0.007 2	0.004 3	0.016 2	0.012 3	0.007 9

### 3.4 真实场景验证

为了验证所提算法的准确性和有效性, 在室内进行了真实动态场景实验。图 8 为自行搭建的手持平台, 相机型号为 Realsense D435i, 嵌入式设备型号为 NVIDIA ORIN NX, 该设备运行 Ubuntu 20.04 的

操作系统。图 9 为 ORB - SLAM3 和 CW - SLAM 算法在真实场景中沿着路径移动的结果, 由图 9a、9c 可知, ORB - SLAM3 原始算法提取了大量动态特征, 而在图 9b、9d 中移动的人身上特征点基本被去除, 几乎所有的特征都是在静态背景下提取的。



图8 实验平台

Fig. 8 Experimental platforms

图9e、9f为2个算法运动轨迹,由于实验室动态特征点的存在,ORB-SLAM3所运行轨迹始终无法形成一个完整回环,与ORB-SLAM3相比,本文算法估计的轨迹很好地形成了一个回环,图9g、9h为ORB-SLAM3和本文算法在工位上的建图结果,由于移动的人和暂时静态物体的影响导致错误的关联和约束,由ORB-SLAM3所运行的图9g出现了点云重叠状况,建图效果不佳,而在本文算法稠密建图结果(图9h),上述情况并未发生,点云数量可用率高。图9i、9j为本文算法最终的局部稠密地图和全局稠密地图结果。经实际场景验证,本文算法优于ORB-SLAM3算法。

#### 4 结束语

为消除动态目标对位姿估计的影响,提出了一种动态场景下的视觉惯性CW-SLAM,在前端使用GC-RANSAC计算初始位姿,依据极线约束为每个地标分配静态/动态先验值,根据先验值采用多帧加权检测特征的方式来判断动态/静态路标点,用于提升系统匹配精度,从而提高定位精度;在后端构建视觉惯性因子等约束进行联合优化;为消除暂时静态目标对系统定位的影响,在回环检测部分加入选择性全局优化,通过构建约束组的因子图优化方程来筛选假阳性的回环假设,从而进一步提升SLAM系统定位精度。与多个经典的SLAM算法平移相对轨迹误差和绝对轨迹误差对比实验结果表明,算法

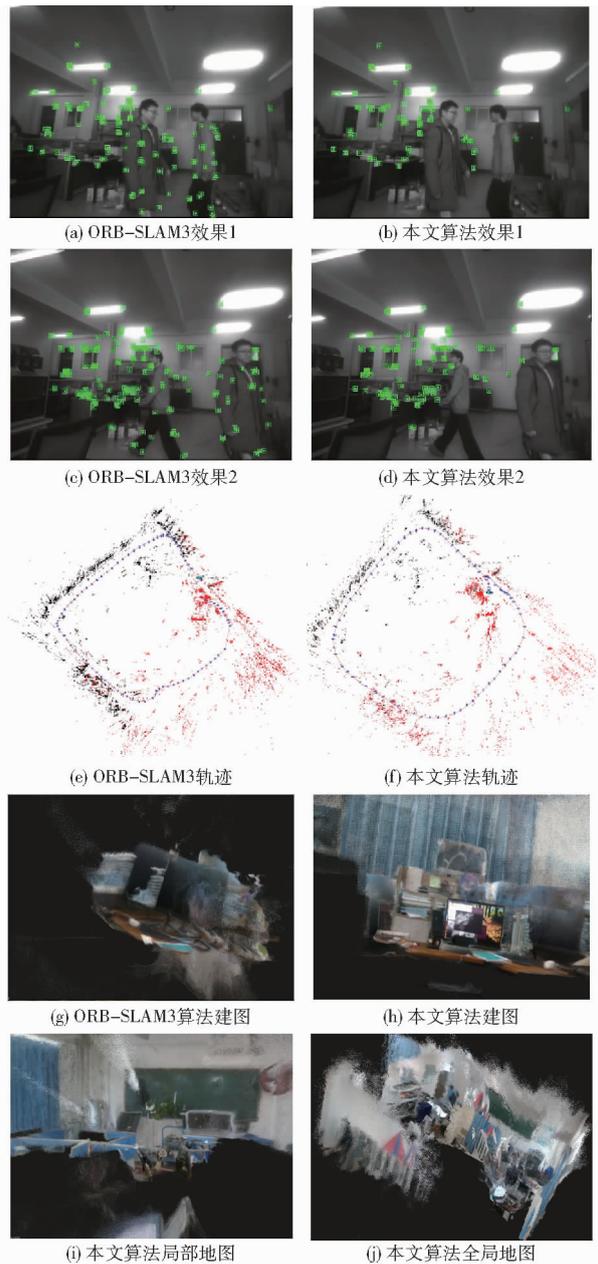


图9 真实场景实验

Fig. 9 Real-world experiments

在高动态环境数据集上的定位精度和准确性有较高提升。

#### 参 考 文 献

- [1] WANG R, WAN W, WANG Y, et al. A new RGB-D SLAM method with moving object detection for dynamic indoor scenes [J]. Remote Sensing, 2019, 11(10): 1143.
- [2] TAN W, LIU H, DONG Z, et al. Robust monocular SLAM in dynamic environments [C] // 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2013: 209-218.
- [3] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving [J]. Journal of Field Robotics, 2020, 37(3): 362-386.
- [4] JINYU L, BANGBANG Y, DANPENG C, et al. Survey and evaluation of monocular visual-inertial SLAM algorithms for augmented reality [J]. Virtual Reality & Intelligent Hardware, 2019, 1(4): 386-410.
- [5] MUR-ARTAL R, TARDÓS J D. ORD-SLAM2: an open-source slam system for monocular, stereo, and RGB-D cameras [J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [6] FU D, XIA H, QIAO Y. Monocular visual-inertial navigation for dynamic environment [J]. Remote Sensing, 2021, 13(9): 1610.
- [7] FAN C, HOU J, YU L. A nonlinear optimization-based monocular dense mapping system of visual-inertial odometry [J].

- Measurement, 2021, 180: 109533.
- [8] SUN Y, LIU M, MENG M Q H. Improving RGB-D SLAM in dynamic environments: a motion removal approach[J]. *Robotics and Autonomous Systems*, 2017, 89: 110-122.
- [9] LI S, LEE D. RGB-D SLAM in dynamic environments using static point weighting[J]. *IEEE Robotics and Automation Letters*, 2017, 2(4): 2263-2270.
- [10] LONG R, RAUCH C, ZHANG T, et al. Rigidfusion: robot localisation and mapping in environments with large dynamic rigid objects[J]. *IEEE Robotics and Automation Letters*, 2021, 6(2): 3703-3710.
- [11] BESCOS B, CAMPOS C, TARDÓS J D, et al. DynaSLAM II: tightly-coupled multi-object tracking and SLAM[J]. *IEEE Robotics and Automation Letters*, 2021, 6(3): 5191-5198.
- [12] ZHAO Y, XIONG Z, ZHOU S, et al. KSF-SLAM: a key segmentation frame based semantic SLAM in dynamic environments [J]. *Journal of Intelligent & Robotic Systems*, 2022, 105(1): 3.
- [13] ISLAM Q U, IBRAHIM H, CHIN P K, et al. FADM-SLAM: a fast and accurate dynamic intelligent motion SLAM for autonomous robot exploration involving movable objects[J]. *Robotic Intelligence and Automation*, 2023, 43(3): 254-266.
- [14] HE X, DING L, LAN Y. DSK-SLAM: a dynamic SLAM system combining semantic information and a novel geometric method based on K-means clustering[J]. *IEEE Sensors Journal*, 2024, 24(14): 23265-23279.
- [15] ISLAM Q U, IBRAHIM H, CHIN P K, et al. MVS-SLAM: enhanced multiview geometry for improved semantic RGBD SLAM in dynamic environment[J]. *Journal of Field Robotics*, 2024, 41(1): 109-130.
- [16] BARNES D, MADDERN W, PASCOE G, et al. Driven to distraction: self-supervised distractor learning for robust monocular visual odometry in urban environments [C] // 2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2018: 1894-1900.
- [17] LI J, ZHANG R, LIU Y, et al. The method of static semantic map construction based on instance segmentation and dynamic point elimination[J]. *Electronics*, 2021, 10(16): 1883.
- [18] KIRILLOV A, GIRSHICK R, HE K, et al. Panoptic feature pyramid networks [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 6399-6408.
- [19] 齐咏生, 陈培亮, 刘利强, 等. 基于单目视觉的动态环境同步定位与多地图构建算法[J]. *农业机械学报*, 2022, 53(4): 280-292.
- QI Yongsheng, CHEN Peiliang, LIU Liqiang, et al. Simultaneous localization and multi-mapping algorithm in dynamic environment based on monocular vision[J]. *Transactions of Chinese Society for Agricultural Machinery*, 2022, 53(4): 280-292. (in Chinese)
- [20] LI Z, JUNNA S, LIAO W. A robust factor graph framework for navigation on PDR/magnetic field integration [J]. *Measurement*, 2025, 245: 116509.
- [21] YU Z, JIANG J, YAN P, et al. A lightweight odometry network for GNSS/INS integration during GNSS outages[J]. *Applied Soft Computing*, 2024, 151: 111143.
- [22] WANG Q, LI W, NIE L, et al. MCG-SLAM: tightly coupled SLAM for multi-factor constraint graph optimisation[J]. *Information Sciences*, 2024, 678: 121075.
- [23] HE K, JIA R, HONG H, et al. LDG-CSLAM: multi-robot collaborative SLAM based on curve analysis, normal distribution, and factor graph optimization[J]. *Journal of Field Robotics*, 2025, 13: 255612.
- [24] 齐咏生, 宋继鹏, 刘利强, 等. 基于点线特征融合的延迟边缘化视觉惯性 SLAM 方法[J]. *农业机械学报*, 2024, 55(12): 373-382.
- QI Yongsheng, SONG Jipeng, LIU Liqiang, et al. Delayed marginalization visual inertia SLAM method based on point and line feature fusion[J]. *Transactions of The Chinese Society for Agricultural Machinery*, 2024, 55(12): 373-382. (in Chinese)
- [25] CAMPOS C, ELVIRA R, RODRÍGUEZ 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.
- [26] MA X, LIANG X. Point-line-based RGB-D SLAM and bundle adjustment uncertainty analysis[J]. *arXiv Preprint, arXiv: 2102.07110*, 2021.
- [27] YANG H, ANTONANTE P, TZOUMAS V, et al. Graduated non-convexity for robust spatial perception: from non-minimal solvers to global outlier rejection[J]. *IEEE Robotics and Automation Letters*, 2020, 5(2): 1127-1134.
- [28] BARATH D, MATAS J. Graph-cut RANSAC [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 6733-6741.
- [29] BLACK M J, RANGARAJAN A. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision[J]. *International Journal of Computer Vision*, 1996, 19(1): 57-91.
- [30] NARITA G, SENO T, ISHIKAWA T, et al. Panopticfusion: online volumetric semantic mapping at the level of stuff and things [C] // 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2019: 4205-4212.
- [31] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems [C] // 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2012: 573-580.
- [32] HE J, LI M, WANG Y, et al. OVD-SLAM: an online visual SLAM for dynamic environments[J]. *IEEE Sensors Journal*, 2023, 23(12): 13210-13219.
- [33] ZHONG Y, HU S, HUANG G, et al. WF-SLAM: a robust VSLAM for dynamic scenarios via weighted features[J]. *IEEE Sensors Journal*, 2022, 22(11): 10818-10827.