

DOI:10.19651/j.cnki.emt.2519427

基于几何-运动误差联合模型和 轨迹预测的动态 VSLAM 算法*

梁若雨 黄卫华 颜如钰 李睿民

(武汉大学人工智能与自动化学院 武汉 430081)

摘要: 针对高动态环境中目标遮挡导致视觉同时定位和地图构建系统存在定位精度下降甚至失效的问题,本文提出一种基于几何-运动误差联合模型及轨迹预测的动态 VSLAM 算法。不同于依赖语义分割或光流估计方法,本文通过融合相机与 IMU 信息,将对极约束几何误差与 IMU 预积分运动误差联合建模,并结合概率模型实现动态目标检测与遮挡状态判断,在遮挡场景下保持较高鲁棒性。为提升动态目标跟踪的连续性与精度,进一步引入基于扩展卡尔曼滤波的轨迹预测对目标位姿进行估计。同时,设计联合因子图模型优化相机、地图点与动态目标特征点,并设计动态平滑运动因子以抑制目标运动突变,从而减少累计误差。最后在 KITTI 跟踪数据集与真实场景的实验表明,相较于基于几何约束和目标跟踪的动态 SLAM 方法,本文算法在高动态环境中目标遮挡场景下具有良好的位姿估计精度与动态目标跟踪性能。

关键词: 视觉同时定位和地图构建;几何-运动误差联合模型;扩展卡尔曼滤波;轨迹预测;联合因子图

中图分类号: TN98 **文献标识码:** A **国家标准学科分类代码:** 510.99

Dynamic VSLAM algorithm based on joint geometry-motion error model and trajectory prediction

Liang Ruoyu Huang Weihua Yan Ruyi Li Ruimin

(School of Artificial Intelligence and Automation, Wuhan University of Science and Technology, Wuhan 430081, China)

Abstract: Aiming to address the problem of decreased localization accuracy or even failure in visual simultaneous localization and mapping systems caused by object occlusion in highly dynamic environments, this paper proposes a dynamic VSLAM algorithm based on a joint geometric-motion error model and trajectory prediction. Unlike methods that rely on semantic segmentation or optical flow estimation, this approach fuses camera and IMU information to jointly model the epipolar geometric error and IMU pre-integrated motion error, and employs a probabilistic model for dynamic object detection and occlusion state estimation, maintaining high robustness under occluded conditions. To improve the continuity and accuracy of dynamic object tracking, an Extended Kalman Filter-based trajectory prediction is introduced for object pose estimation. Meanwhile, a joint factor graph model is constructed to optimize the camera, map points, and dynamic object feature points, where a dynamic motion-smoothing factor is designed to suppress abrupt object motion and reduce accumulated errors. Finally, experiments on the KITTI tracking dataset and real-world scenarios demonstrate that, compared with geometry-based and object-tracking-based dynamic SLAM methods, the proposed algorithm achieves superior pose estimation accuracy and dynamic object tracking performance in object occlusion scenarios within highly dynamic environments.

Keywords: visual simultaneous localization and mapping; joint geometry-motion error model; extended Kalman filter; trajectory prediction; joint factor graph

0 引言

视觉同时定位与地图构建 (visual simultaneous

localization and mapping, VSLAM) 是通过视觉传感器帮助移动机器人在未知环境下进行实时位姿估计与场景重建,已广泛应用于自动驾驶、无人系统及虚拟现实等场

收稿日期:2025-07-22

* 基金项目:国家自然科学基金(62303359)项目资助

景^[1-2]。目前,大多数主流 VSLAM 系统是在静态假设的前提下进行环境地图构建,虽然能在一定程度上能够克服动态目标干扰、场景遮挡及环境噪声等因素的影响,但是 VSLAM 对光照变化非常敏感、无法准确估计场景的尺度等劣势限制了其在复杂场景中的应用。惯性测量单元(inertial measurement unit, IMU)具备不易受场景变化的特点,能够扩展 VSLAM 在动态场景下的感知能力,将视觉传感器与 IMU 融合可提高 VSLAM 系统在动态环境下的定位精度,目前已成为 VSLAM 领域的一个研究热点^[3]。

ORB-SLAM3^[4]具有多传感器融合、多地图管理以及闭环检测机制,是当前 VSLAM 中具有最广泛关注与应用的算法框架之一^[5]。但 ORB-SLAM3 使用随机抽样一致(random sample consensus, RANSAC)算法对场景动态特征的错误匹配进行筛选,并不区分真实场景中目标的动态特征,导致一些包含正确匹配的数据点被错误地排除在外。针对动态目标运动状态判断准确性的问题,一些改进 ORB-SLAM3 算法将深度学习与运动一致性检验方法相结合,对动态目标的动态特征点进行检测并剔除^[6-7]。文献^[8]在 ORB-SLAM3 框架下设计了基于 FastSAM^[9]图像分割网络与光流法的动态 VSLAM 算法,通过自适应阈值稠密光流方法识别动态目标。文献^[10]将语义先验信息引入 ORB-SLAM3 算法框架并提出了一种核线几何方法,针对不同的动态目标制定差异化的动态特征点剔除策略。然而,将动态特征点检测并剔除忽略了连续帧之间动态目标的运动关系,会直接导致定位产生误差、构建地图存在漂移等问题。因此,研究者们对动态目标进行检测后,将相机和动态目标统一纳入后端优化实现相机和动态目标的联合估计,从而提高 ORB-SLAM3 算法的位姿估计精度。文献^[11]采用双模块实现相机和动态目标的定位和跟踪,提

出了一种动态目标感知算法,并通过因子图优化相机位姿、静态目标位姿及运动目标位姿等。文献^[12]在动态目标检测基础上引入多目标追踪实现动态目标的运动估计,并利用光束法平差法(bundle adjustment, BA)优化静态场景、相机姿态、运动目标轨迹和动态目标的检测框。但是,在动态目标遇到遮挡时,基于几何一致性的目标检测和位姿估计系统易出现目标检测精度下降、跟踪中断等问题,且难以维持判断动态目标的运动状态。

综上所述,针对高动态场景中存在目标遮挡的问题,如何对动态目标状态的进行有效估计成为提升 VSLAM 系统鲁棒性和精度的关键。基于此,在 ORB-SLAM3 框架下,本文提出一种基于几何-运动误差联合模型及目标轨迹预测的动态 VSLAM 算法。首先,在采用 YOLO11-seg 网络和 VideoFlow 网络对 RGB 图像和 IMU 数据进行预处理的基础上,融合特征点对极约束几何误差和 IMU 信息运动误差,根据特征点的遮挡和运动方向建立几何-运动误差联合模型实现动态目标检测;其次,提出一种基于扩展卡尔曼滤波(extended kalman filter, EKF)轨迹预测的动态目标追踪方法,通过保持动态目标运动估计,提高遮挡情况下的跟踪精度和稳定性;然后,提出一种联合因子图方法对相机位姿、动态目标位姿和动态目标运动轨迹等进行全局优化,并引入动态目标平滑运动因子对动态目标运动作平滑约束;最后,在 KITTI 跟踪数据集和真实场景中验证本文算法性能。

1 算法框架

本文以传统 ORB-SLAM3 算法为基础框架进行扩展,根据动态场景下潜在目标运动特性,增加了预处理模块、改进了跟踪线程和建图线程。改进后 ORB-SLAM3 算法框架如图 1 所示。

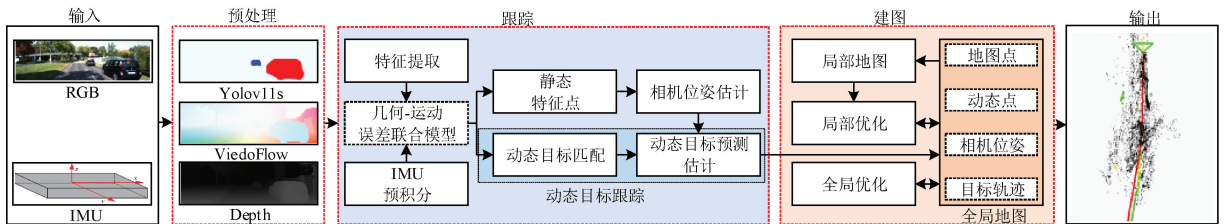


图 1 改进后 ORB-SLAM3 算法框架

Fig. 1 Improved ORB-SLAM3 algorithm framework

在预处理模块中,为了有效检测目标并提高目标识别的精确性和实时性,本文采用 YOLO11-seg 网络对 RGB 图像中目标进行检测生成图像分割掩码;采用 VideoFlow^[13]网络对连续帧之间图像估计光流,并结合 RGB 图像的分割掩码和深度值为跟踪线程中动态目标匹配提供数据。

在跟踪线程中,本文根据对极约束获取特征点的几何误差和利用 IMU 预积分获取特征点的运动误差,结合特征点遮挡情况和运动方向构建几何-运动误差联合模型。

针对图像遮挡对动态目标跟踪精度和稳定性的影响,引入 EKF 预测机制对动态目标运动轨迹进行预测,进一步确定目标的真实运动状态,由此减少在连续帧之间动态目标对 VSLAM 算法定位精度的影响,并提高目标追踪在受到遮挡时的跟踪能力。

在建图线程中,为了消除前端累计误差和提高相机和动态目标位姿估计精度,在全局优化中设计联合因子图模型对地图点、相机位姿和动态目标跟踪线程中的误差进行全局优化,输出优化后包含地图点、动态点、相机位姿和目

标轨迹的全局地图。

2 几何-运动误差联合模型构建

对极约束通过基础矩阵描述图像特征点对的几何关系,但易受遮挡或极线方向运动影响。本文提出了一种基于几何-运动误差联合模型的动态目标检测方法,利用 IMU 预积分获取的运动信息对几何误差进行补充,并通过概率模型判断特征点的遮挡情况和运动状态。

2.1 问题描述

特征点几何误差示意图如图 2 所示。图 2 中, O^k, O^{k+1} 表示不同时刻下相机中心;特征点 P_n 在 k 时刻相机坐标系位置为 ${}^c P_n^k$, ${}^{uv} p_n^k$ 和 ${}^{uv} p_n^{k+1}$ 分别为特征点 P_n 在 k 时刻和 $k+1$ 时刻下像素坐标系匹配的特征点对; e_n^k, e_n^{k+1} 为极点,与 ${}^{uv} p_n^k$ 和 ${}^{uv} p_n^{k+1}$ 的连线分别为极线 l_n^k 和 l_n^{k+1} ; ${}^c P_n^{k+1}$ 为 P_n 在 $k+1$ 时刻相机坐标系位置。

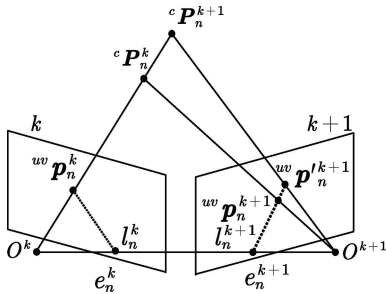


图 2 几何误差示意图

Fig. 2 Geometric error diagram

若 P_n 为静态特征点, ${}^{uv} p_n^k$ 和 ${}^{uv} p_n^{k+1}$ 满足对极约束:

$$({}^{uv} p_n^{k+1})^T \mathbf{F} {}^{uv} p_n^k = 0 \quad (1)$$

式中: \mathbf{F} 为基础矩阵。

设 d_n^v 为 ${}^{uv} p_n^{k+1}$ 到极线 $l_n^{k+1} = (A, B, C)^T$ 的距离为几何误差,则有:

$$d_n^v = \frac{|({}^{uv} p_n^{k+1})^T \mathbf{F} {}^{uv} p_n^k|}{\sqrt{(A^2 + B^2)}} = 0 \quad (2)$$

当 d_n^v 越大时,特征点 P_n 为动态特征点的概率越高。当特征点出现遮挡时,错误特征点匹配会使得 d_n^v 增大,而该特征点可能是静态特征点。

此外,当相机坐标系位置 ${}^c P_n^{k+1}$ 从 k 时刻沿极线方向发生运动,投影点 ${}^{uv} p_n^{k+1}$ 的几何误差 $d_n^v = 0$,但此时 ${}^c P_n^{k+1}$ 为动态特征点,即特征点 P_n 沿极线方向运动时仍满足式(1)对极约束,并且几何误差较小。

因此,特征点被遮挡或动态特征点沿极线方向运动,单纯依靠几何信息难以判断特征点的动态属性。

2.2 几何-运动误差联合模型

特征点 P_n 遮挡示意图如图 3 所示。图 3 中, ${}^{uv} p_n^k$ 和 ${}^{uv} p_n^{k+1}$ 为 P_n 在 k 时刻和 $k+1$ 时刻的像素坐标系投影点, θ_n^k 为 ${}^{uv} p_n^k$ 和 ${}^{uv} p_n^{k+1}$ 之间视角, θ_n^{k+1} 为 ${}^{uv} p_n^{k+1}$ 和 ${}^{uv} p_n^{k+1}$ 之间视角。设 θ_{th} 为遮挡阈值,若 $\theta_n^k > \theta_{th}$ 则 P_n 为遮挡。

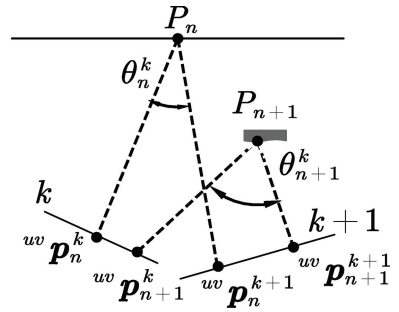


图 3 特征点遮挡示意图

Fig. 3 Feature point occlusion diagram

根据特征点运动在极线方向投影判断特征点是否沿极线方向发生运动。特征点 P_n 对应极线的单位方向向量为:

$$\mathbf{v} = (\cos(\arctan2(-A, B)), \sin(\arctan2(-A, B))) \quad (3)$$

极线方向的投影为:

$$S = |({}^{uv} p_n^{k+1} - {}^{uv} p_n^k) \cdot \mathbf{v}| \quad (4)$$

设运动阈值为 S_{th} ,若 P_n 在极线方向的投影大于其设定运动阈值,即 $S > S_{th}$,则 P_n 可能沿极线 l_n^{k+1} 运动时,此时特征点几何误差仍很小,因此需要结合运动误差进一步判断特征点的运动状况。

特征点运动误差取决于实际观测值和 IMU 提供的相机运动信息,其示意图如图 4 所示。图 4 中, ${}^c P_n^{k+1}$ 为 P_n 在 $k+1$ 时刻相机坐标系下的实际观测值。

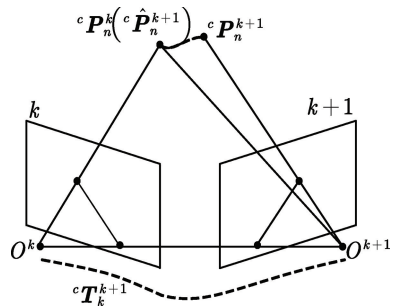


图 4 运动误差示意图

Fig. 4 Motion error diagram

P_n 在 $k+1$ 时刻相机坐标系下的预测位置为:

$$\hat{{}^c P_n^{k+1}} = {}^c P_n^k \mathbf{T}_k^{k+1} \quad (5)$$

式中: \mathbf{T}_k^{k+1} 为 IMU 预积分计算 k 时刻相机坐标系变换到 $k+1$ 时刻相机坐标系的变换矩阵。

P_n 的运动误差为:

$$d_n^h = |{}^c P_n^{k+1} - \hat{{}^c P_n^{k+1}}| \quad (6)$$

对于静态特征点而言,其运动信息与相机运动保持一致,若两者之间的运动误差越大,说明该特征点不满足静态假设,属于动态特征点的概率越大。因此,当 P_n 沿极线运动时运动误差不受影响。

综上,根据特征点在遮挡和沿极线运动方向时运动状

态分析,利用 IMU 预积分获取的运动信息对几何误差进行补充,定义特征点的几何-运动误差联合模型为:

$$f(P_n) = \begin{cases} d_n^b, & \theta_n^k \leq \theta_{th}, S > S_{th} \\ d_n^v, & \theta_n^k \leq \theta_{th}, S \leq S_{th} \\ \infty, & \theta_n^k > \theta_{th} \end{cases} \quad (7)$$

将误差联合模型 $f(P_n)$ 映射到 $[0,1]$ 区间进行归一化处理,则概率映射函数为:

$$P_n^D = \frac{1}{1 + e^{-f(P_n)}} \quad (8)$$

当 $P_n^D = 1$ 时,特征点 P_n 为遮挡状态。令动态特征点概率映射函数的阈值为 P_{th} , 根据式(8)概率映射函数 P_n^D 判断特征点运动状态并给出相应决策;若 $P_n^D < P_{th}$, 则 P_n 为静态特征点,用于估计相机位姿;反之,判定 P_n 为动态特征点;若目标掩码区域内动态特征点占比大于动态目标阈值 P_{th}^d , 则该目标为动态目标,对其位姿以及运动进行估计,以降低对静态特征点的影响。

3 基于 EKF 轨迹预测的动态目标跟踪

为了提高 VSLAM 系统在遮挡情况下位姿估计精度和鲁棒性,设计了基于 EKF 轨迹预测的动态目标跟踪方法。该方法通过构造动态目标中心点的状态空间模型,并对其状态进行递推估计,实现对动态目标状态的位姿估计,由此减少动态目标对静态特征点的干扰并提高动态目标跟踪性能。

3.1 动态目标匹配

本文基于光流估计和图像掩码,对动态目标在不同时刻的像素采样点分配唯一标签,构建像素标签集合 L 以区分不同目标。利用光流将前一时刻的特征点及标签传播至当前帧,通过比较相邻两帧的像素标签集合重叠度,判断是否为同一动态目标,若重叠较高,则赋予相同目标标识并进行运动估计。

3.2 动态目标预测估计

设动态目标 i 的特征点 m 在不同时刻的,其运动变换示意图如图 5 所示。图 5 中, $\{w\}$ 为世界坐标系, ${}^w P_m^k = ({}^w P_x^k, {}^w P_y^k, {}^w P_z^k)^T$ 表示 k 时刻动态目标 i 上特征点 m 在世界坐标系下的位置; ${}^w X^k, {}^w X^{k+1}$ 为不同时刻下的相机位姿在世界坐标系的位置; ${}^w H^k, {}^w H^{k+1}$ 表示动态目标 i 在不同时刻的坐标系。

由刚体特性可知, ${}^w P_m^k$ 在 k 时刻和 $k+1$ 时刻之间满足:

$$({}^w H^k)_i^{-1} {}^w P_m^k = ({}^w H^{k+1})_i^{-1} {}^w P_m^{k+1} \quad (9)$$

令 ${}^w T_k^{k+1,m} = ({}^w H^k)_o {}^w T_k^{k+1} ({}^w H^{k+1})_i^{-1}$, 表示动态目标 i 上特征点 m 在两个时刻之间的变换矩阵; 设 ${}^w T_k^{k+1}$ 为 ${}^w H^k, {}^w H^{k+1}$ 之间的变换矩阵。构造 ${}^w P_m^{k+1}$ 的像素坐标 ${}^w p_m^{k+1}$ 的重投影误差:

$$e_m = \left\| {}^w p_m^{k+1} - \pi({}^w T_k^{k+1,m} {}^w P_m^k) \right\|_2 \quad (10)$$

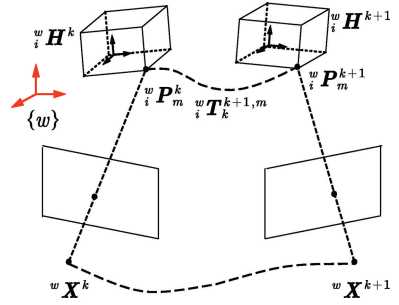


图 5 动态目标运动变换示意图

Fig. 5 Schematic diagram of dynamic object motion transformation

式中: $\pi(\cdot)$ 是投影函数,将相机坐标系投影到像素坐标系。

本文采用列文伯格-马夸尔特(Levenberg-Marquardt)方法对式(10)求解得到 ${}^w T_k^{k+1,m}$, 并以此恢复目标在 k 时刻到 $k+1$ 时刻之间的运动。

选取动态目标 i 中心点 ${}^w P_o^k$ 作为速度估计代表点,则动态目标 i 在 k 时刻的速度为:

$${}^w V_o^k = \frac{{}^w P_o^k - {}^w P_o^{k-1}}{\Delta t} \quad (11)$$

式中: ${}^w V_o^k$ 表示动态目标 i 的中心点在 k 时刻的速度; Δt 为 k 时刻到 $k+1$ 时刻的时间间隔。

选取动态目标中心点作为运动轨迹预测代表点,采用 EKF 对动态目标 i 在 k 时刻上中心点 ${}^w P_o^k$ 的进行轨迹预测。设运动目标 i 的状态向量为:

$${}^w D_o^k = [{}^w P_o^k, {}^w V_o^k, {}^w a_o^k]^T \quad (12)$$

式中: ${}^w a_o^k = ({}^w a_x^k, {}^w a_y^k, {}^w a_z^k)^T$ 为动态目标 i 的中心点 ${}^w P_o^k$ 在世界坐标系下的加速度。

设 U^k 是控制输入,根据式(12)定义动态目标的状态转移函数和状态协方差预测模型为:

$$\begin{cases} {}^w D_o^{k+1|k} = f({}^w D_o^k, U^k) + W^k \\ P^{k+1|k} = F^k P^{k|k} (F^k)^T + Q^k \end{cases} \quad (13)$$

式中: W^k 是过程噪声且服从高斯分布。 $f(\cdot)$ 为目标运动模型; $P^{k+1|k}$ 为 $k+1$ 时刻先验协方差; $P^{k|k}$ 为 k 时刻后验协方差; F^k 是状态转移函数对状态 ${}^w D_o^k$ 的雅可比矩阵; Q^k 是过程噪声协方差矩阵。

在 $k+1$ 时刻,动态目标 i 的中心点像素坐标投影为实际位置观测函数和观测残差为:

$$\begin{cases} {}^{uv} \hat{p}_o^k = \pi(({}^w X^{k+1})_i^{-1} {}^w P_o^k) \\ e^{k+1} = {}^{uv} p_o^{k+1} - {}^{uv} \hat{p}_o^{k+1} \end{cases} \quad (14)$$

式中: ${}^{uv} p_o^k$ 为实际位置观测值。

获取观测值后,状态更新如下:

$$\begin{cases} K^{k+1} = P^{k+1|k} (H^{k+1})^T (H^{k+1} P^{k+1|k} (H^{k+1})^T + R^{k+1})^{-1} \\ {}^w \hat{D}_o^{k+1|k+1} = {}^w D_o^{k+1|k} + K^{k+1} e^{k+1} \\ P^{k+1|k+1} = (I - K^{k+1} H^{k+1}) P^{k+1|k} \end{cases} \quad (15)$$

式中: \mathbf{H}^{k+1} 是观测函数关于状态 ${}^w_i \mathbf{D}_o^{k+1}$ 的雅可比矩阵; \mathbf{R}^{k+1} 是观测噪声的协方差矩阵; ${}^w_i \hat{\mathbf{D}}_o^{k+1}$ 为动态目标 i 中心点 ${}^w_i \mathbf{P}_o^k$ 在 $k+1$ 时刻最终状态估计; ${}^w_i \mathbf{D}_o^{k+1}$ 为 k 时刻的先验估计。

构造中心点 ${}^w_i \mathbf{P}_o^k$ 在相机图像中预测值和实际观测值之间的重投影误差为:

$$e_i = \| {}^{uv} \mathbf{p}_o^{k+1} - {}^{uv} \hat{\mathbf{p}}_o^{k+1} \|_2^2 \quad (16)$$

更新状态最小化重投影误差 e_i , 使预测值和观测值的差异逐渐减小。在动态目标被遮挡时, 即动态目标在 $k+1$ 没有匹配到, 可以通过 ${}^w_i \mathbf{D}_o^k$ 的预测状态来恢复 $k+1$ 的动态目标中心点的状态; 在匹配到相同的动态目标时, 通过 ${}^w_i \mathbf{P}_o^k$ 的目标的实际观测对预测过程修正, 实现动态目标的实时位姿估计和运动变换估计。

4 联合因子图全局优化

在传统 VSLAM 的全局优化中, 采用因子图模型对地图点和相机位姿进行全局优化^[14]。然而, 动态目标的识别、位姿估计与轨迹预测易产生累计误差。为此, 本文提出一种联合因子图模型, 将相机、地图点和动态目标的观测误差、运动误差与预测误差统一建模并进行联合优化, 同时引入平滑运动因子, 对动态目标施加平滑约束, 以抑制运动突变。

本文设计的联合因子图模型如图 6 所示, 图中, 静态地图点 ${}^w \mathbf{P}_s^k$ 、相机位姿 ${}^w \mathbf{X}^k$, 动态目标上的特征点 ${}^w_i \mathbf{P}_m^k$ 、动态目标上的中心点 ${}^w_i \mathbf{P}_o^k$ 和动态目标变换矩阵 ${}^w_i \mathbf{T}_k^{k+1,m}$ 作为待优化变量构成因子图的顶点, 优化变量之间的矩形为优化函数, 构成因子图的边。

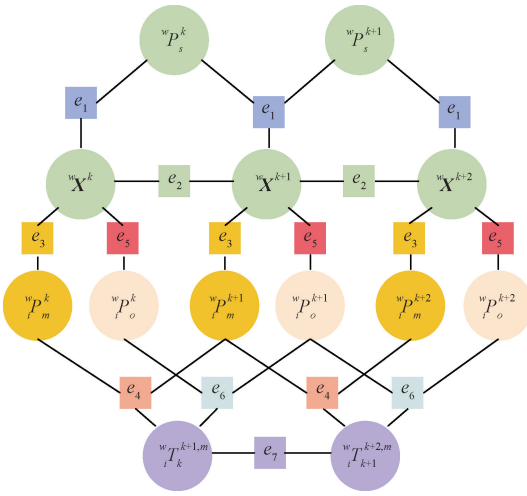


图 6 联合因子图模型

Fig. 6 Joint factor graph model

静态地图点 ${}^w \mathbf{P}_s^k$ 的优化函数时在相机位姿 ${}^w \mathbf{X}^k$ 下的重投影误差为:

$$e_1({}^w \mathbf{X}^k, {}^w \mathbf{P}_s^k) = {}^{uv} \mathbf{p}_s^k - \pi(({}^w \mathbf{X}^k)^{-1} {}^w \mathbf{P}_s^k) \quad (17)$$

式中: ${}^{uv} \mathbf{p}_s^k$ 是静态地图点在像素坐标系下实际观测值。

不同时刻下惯性残差 $e_2(\mathbf{r}_{k,k+1})$ 为:

$$e_2(\mathbf{r}_{k,k+1}) = [\mathbf{r}_{\Delta \mathbf{R}_{k,k+1}}, \mathbf{r}_{\Delta \mathbf{v}_{k,k+1}}, \mathbf{r}_{\Delta \mathbf{p}_{k,k+1}}] \quad (18)$$

式中: $\mathbf{r}_{\Delta \mathbf{R}_{k,k+1}}$, $\mathbf{r}_{\Delta \mathbf{v}_{k,k+1}}$, $\mathbf{r}_{\Delta \mathbf{p}_{k,k+1}}$ 分别为 IMU 从 k 时刻到 $k+1$ 时刻的旋转残差、速度残差和位置残差。

动态目标上的特征点投影误差 $e_3({}^w \mathbf{X}^k, {}^w_i \mathbf{P}_m^k)$ 和运动误差 $e_4({}^w_i \mathbf{P}_m^{k+1}, {}^w_i \mathbf{P}_m^k, {}^w_i \mathbf{T}_k^{k+1,m})$ 为:

$$\begin{cases} e_3({}^w \mathbf{X}^k, {}^w_i \mathbf{P}_m^k) = {}^w_i \mathbf{p}_m^k - \pi(({}^w \mathbf{X}^k)^{-1} {}^w_i \mathbf{P}_m^k) \\ e_4({}^w_i \mathbf{P}_m^{k+1}, {}^w_i \mathbf{P}_m^k, {}^w_i \mathbf{T}_k^{k+1,m}) = {}^w_i \mathbf{P}_m^{k+1} - {}^w_i \mathbf{T}_k^{k+1,m} {}^w_i \mathbf{P}_m^k \end{cases} \quad (19)$$

同样, 动态目标上的中心点的预测误差 $e_5({}^w \mathbf{X}^k, {}^w_i \mathbf{P}_o^k)$ 和运动误差 $e_6({}^w_i \mathbf{P}_o^{k+1}, {}^w_i \mathbf{P}_o^k, {}^w_i \mathbf{T}_k^{k+1,m})$ 为:

$$\begin{cases} e_5({}^w \mathbf{X}^k, {}^w_i \mathbf{P}_o^k) = {}^{uv} \mathbf{p}_o^{k+1} - {}^{uv} \hat{\mathbf{p}}_o^{k+1} \\ e_6({}^w_i \mathbf{P}_o^{k+1}, {}^w_i \mathbf{P}_o^k, {}^w_i \mathbf{T}_k^{k+1,m}) = {}^w_i \mathbf{P}_o^{k+1} - {}^w_i \mathbf{T}_k^{k+1,m} {}^w_i \mathbf{P}_o^k \end{cases} \quad (20)$$

在不同时刻下动态目标的变换矩阵之间引入平滑因子对目标运动加以优化, 平滑因子 $e_7({}^w_i \mathbf{T}_k^{k+1,m}, {}^w_i \mathbf{T}_k^{k+2,m})$ 为:

$$e_7({}^w_i \mathbf{T}_k^{k+1,m}, {}^w_i \mathbf{T}_k^{k+2,m}) = ({}^w_i \mathbf{T}_k^{k+1,m})^{-1} {}^w_i \mathbf{T}_k^{k+2,m} \quad (21)$$

根据式(17)~(21)可得, 由各误差函数构建的系统总误差函数为:

$$\begin{aligned} e = & \sum_k \sum_s \gamma(e_1^T \Sigma^{-1} e_1) + \sum_k \gamma(e_2^T \Sigma^{-1} e_2) + \\ & \sum_k \sum_i \sum_m \gamma(e_3^T \Sigma^{-1} e_3 + e_4^T \Sigma^{-1} e_4 + e_6^T \Sigma^{-1} e_6) + \\ & \sum_k \sum_i \gamma(e_5^T \Sigma^{-1} e_5 + e_7^T \Sigma^{-1} e_7) \end{aligned} \quad (22)$$

式中: γ 表示 Huber 鲁棒核函数, 目的是较少错误匹配。 Σ^{-1} 对应误差函数中的观测协方差矩阵。本文使用列文伯格-马夸尔特方法对上述优化函数求解, 以获得最优的状态估计结果。

5 实验及结果分析

为了验证本文所提出的基于的动态目标检测方法、基于 EKF 轨迹预测的动态目标跟踪方法和联合因子图优化方法的有效性, 本文分别在 KITTI 跟踪数据集和真实场景中展开实验, 评估所提方法的有效性和提升相机定位精度的表现。

5.1 KITTI 跟踪数据集测试

KITTI 包含立体相机、图像标签和 IMU 等多种传感器信息, 涵盖从低动态到多动态目标及目标遮挡的高动态环境。本文在 KITTI 跟踪数据集对所提算法有效性进行了实验分析, 并与传统算法 ORB-SLAM3、跟踪算法 DynaSLAM II、VDO-SLAM 算法进行性能对比, 以及对本文算法进行耗时分析。采用的评价标准为绝对轨迹误差 (absolute trajectory error, ATE) 和相对位姿误差 (relative pose error, RPE), 其中 RPE 包括平移误差 (RPE_t) 和旋转误差 (RPE_r)。

1) 动态目标检测效果分析

在 KITTI 跟踪数据集上, 本文选取包含少量动态目标

以及存在遮挡的多动态目标场景的多个序列,对所提出的基于几何-运动误差联合模型的动态目标检测方法进行验证。实验以 ATE 为评价指标,比较 3 种算法性能:传统的 ORB-SLAM3 算法、基于几何误差的 ORB-SLAM3 算法(算法 1)以及本文所设计的基于几何-运动误差联合模型的 ORB-SLAM3 算法(算法 2)。为保证实验公平性,3 种算法使用相同的图像和 IMU 数据预处理策略,算法对比结果如表 1 所示,表中加粗字体表示对应序列中表现最优的误差结果。

表 1 相机绝对轨迹误差对比

Table 1 Comparison of camera absolute trajectory errors

序列	ATE/m		
	ORB-SLAM3	算法 1	算法 2
00	1.430	1.240	0.970
01	1.980	2.420	2.370
03	0.890	0.780	0.560
04	1.650	1.460	1.350
05	1.260	1.340	1.240
18	0.910	1.100	0.870
20	16.770	1.560	1.410
平均值	3.556	1.414	1.253

从表 1 可得,传统 ORB-SLAM3 未对动态目标进行处理,导致在多数序列中尤其是序列 20 这类高动态场景下出现较大的轨迹误差。算法 1 通过几何误差剔除部分动态特征点,提升了一定的定位精度,但在特征点遮挡或沿极线运动等情况下效果有限。相比之下,算法 2 通过引入 IMU 信息构建几何-运动误差联合模型,更有效地识别动态目标,提升鲁棒性。实验结果显示,算法 2 的 ATE 平均值相较于 ORB-SLAM3 和算法 1 分别下降了 60.23% 和 64.76%,在动态场景下表现出更高的准确性与鲁棒性。

2) 动态目标跟踪效果分析

为进一步验证本文算法的动态目标追踪精度,对相同序列追踪目标的 RPE 进行评价,并将结果与 VDO-SLAM 进行对比,实验结果如表 2 所示。从表 2 中可得,与 VDO-SLAM 相比,本文算法在追踪目标的 RPE_t 和 RPE_r 的平均值分别降低了 28.86% 和 17.77%,表明基于预测的动态目标跟踪相比较 VDO-SLAM 取得更好的跟踪精度。

图 7 展示了序列 00 全局地图中相机和追踪目标轨迹,矩形表示相机位姿,圆点和三角形表示两个不同追踪目标轨迹,序列 00 中目标 1(Object1)存在被目标 2(Object2)遮挡以及被其他静态物体遮挡,本文算法在目标 1 和 2 跟踪过程中能够稳定地估计其连续轨迹,具有较好动态目标跟踪效果和稳定性。

3) 联合因子图优化效果分析

为验证本文提出的联合因子图优化模型的有效性,在同样序列下对算法在不同因子图模型的轨迹精度进行对

表 2 目标轨迹误差对比

Table 2 Object trajectory error comparison

序列	VDO-SLAM		本文算法	
	RPE _t /m	RPE _r / (°)	RPE _t /m	RPE _r / (°)
00	0.107 70	1.052 00	0.072 80	0.944 2
01	0.157 30	0.905 10	0.097 50	0.707 8
03	0.096 50	0.291 90	0.083 40	0.236 2
04	0.193 70	0.828 80	0.114 70	0.668 1
05	0.114 00	0.370 50	0.065 90	0.381 3
18	0.082 50	0.245 30	0.063 40	0.161 7
20	0.082 40	0.366 30	0.095 70	0.239 2
平均值	0.119 16	0.579 99	0.084 77	0.476 9

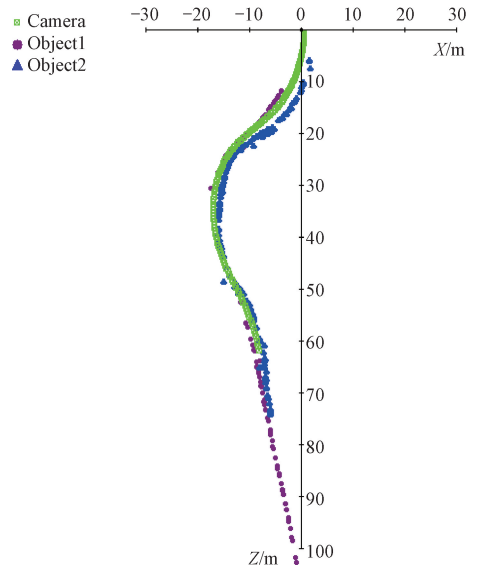


图 7 序列 00 动态目标跟踪效果

Fig. 7 Sequence 00 dynamic object tracking effect

比评估。其中设置未改进因子图模型和本文提出的联合因子图模型,评价结果如表 3 所示,从表 3 中可以看出联合因子图模型后相机的轨迹误差相比较于改进前 RPE_t 和 RPE_r 的平均值分别降低了 12.50% 和 11.11%,因此经过联合因子图模型优化后能够进一步降低轨迹误差。

表 3 因子图优化效果对比

Table 3 Comparison of factor graph optimization results

序列	未改进因子图模型		联合因子图模型	
	RPE _t /m	RPE _r / (°)	RPE _t /m	RPE _r / (°)
00	0.040	0.050	0.030	0.050
01	0.060	0.040	0.050	0.040
03	0.060	0.040	0.060	0.030
04	0.070	0.050	0.060	0.050
05	0.050	0.030	0.050	0.020
18	0.050	0.030	0.040	0.020
20	0.070	0.030	0.060	0.030
平均值	0.057	0.039	0.050	0.034

图 8~11 分别展示了序列 18 在未改进因子图模型与联合因子图模型优化下的相机位姿的 RPE_t 和 RPE_r。由图中对比结果可知,相较于未改进因子图模型的轨迹曲线,联合因子图模型的轨迹曲线有效消除了尖峰误差,且与真值之间的偏差进一步减小。这表明,本文提出的联合因子图优化方法能够显著降低相机位姿估计误差,减弱动态目标对相机定位精度的影响,从而整体提升系统的定位精度。

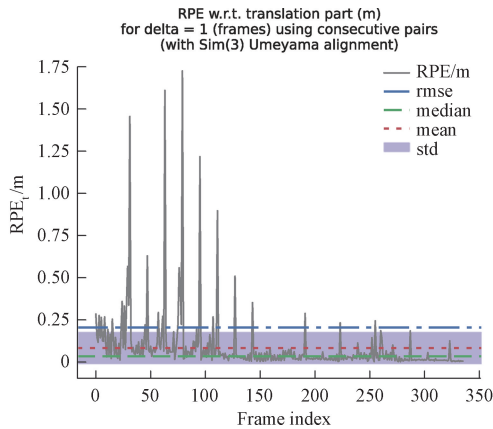


图 8 未改进因子图模型的 RPE_t

Fig. 8 RPE_t of the unimproved factor graph model

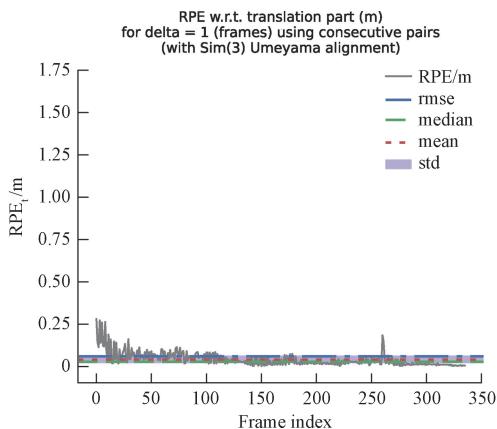


图 9 联合因子图模型的 RPE_t

Fig. 9 RPE_t of the joint factor graph model

4) 算法性能对比

同样在相同序列对本文算法的相机相对位姿误差进行评估,并与 ORB-SLAM3、DynaSLAM II、VDO-SLAM 和 VIMOT^[15] 算法进行性能比较,实验对比结果如表 4 所示。从表 4 中可以看出:ORB-SLAM3 算法相比较于其他算法相对位姿估计误差较大;相比较 ORB-SLAM3,跟踪算法 DynaSLAM II、VDO-SLAM、VIMOT 通过对动态目标进行检测和跟踪,能够在动态环境下提升定位精度;本文算法使用几何-运动误差联合模型在选取序列上的 RPE_t 和 RPE_r 平均值分别为相比较 ORB-SLAM3 降低了 28.57% 和 20%,同时整体误差平均值优于跟踪算法

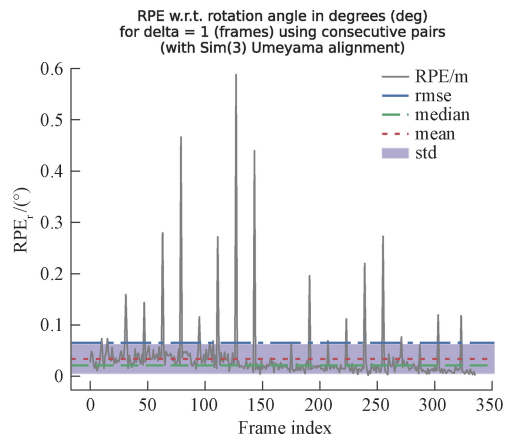


图 10 未改进因子图模型的 RPE_r

Fig. 10 RPE_r of the unimproved factor graph model

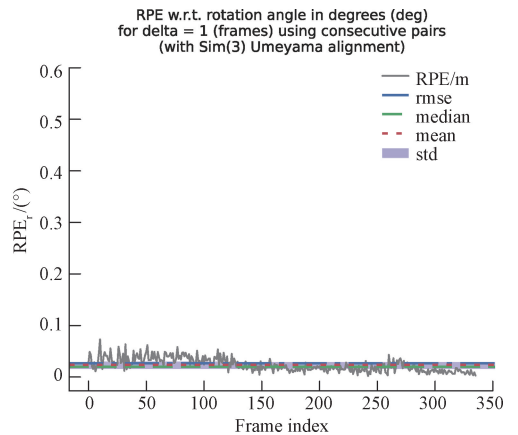


图 11 联合因子图模型的 RPE_r

Fig. 11 RPE_r of the joint factor graph model

DynaSLAM II、VDO-SLAM,表现出更好的定位精度。

5) KITTI 耗时评估

在选取的 KITTI 跟踪数据集序列中,本文算法平均耗时如表 5 所示。实验结果表明,动态目标跟踪、预测估计以及位姿优化耗时高。其中动态目标数量会对动态目标跟踪和动态目标预测估计的耗时产生影响,在序列 20 中存在大量动态目标的场景下,系统需要处理更多的动态目标状态估计与预测更新,耗时会显著上升。与其他跟踪算法的耗时比较如表 6 所示,本文在所选序列中能够以 9~11 fps 的平均速度进行运行,尽管相较于 DynaSLAM II 略低,但由于算法引入了 IMU 信息融合、基于 EKF 的轨迹预测及联合因子图优化等额外计算模块,因此在提升定位精度和动态目标跟踪鲁棒性的同时,实现了运行效率与精度的平衡。

5.2 真实场景实验

使用 RGB 相机和 IMU 传感器对实验室环境的真实场景进行数据采集,实验真实场景和场景设计平面图分别如图 12 和 13 所示。场景设计平面图中三角形为设备起

表 4 相机相对位姿误差对比

Table 4 Comparison of relative camera pose errors

序列	ORB-SLAM3		DynaSLAM II		VDO-SLAM		VIMOT		本文算法	
	RPE _t /m	RPE _r /($^{\circ}$)	RPE _t /m	RPE _r /($^{\circ}$)	RPE _t /m	RPE _r /($^{\circ}$)	RPE _t /m	RPE _r /($^{\circ}$)	RPE _t /m	RPE _r /($^{\circ}$)
00	0.070	0.060	0.040	0.060	0.050	0.050	0.040	0.060	0.030	0.050
01	0.070	0.040	0.050	0.040	0.120	0.040	0.050	0.090	0.050	0.040
03	0.080	0.050	0.060	0.040	0.090	0.040	0.060	0.040	0.060	0.030
04	0.080	0.060	0.070	0.060	0.110	0.050	0.060	0.050	0.060	0.050
05	0.070	0.020	0.060	0.030	0.100	0.020	0.070	0.030	0.050	0.020
18	0.060	0.030	0.030	0.020	0.070	0.020	0.050	0.040	0.040	0.020
20	0.060	0.040	0.070	0.040	0.160	0.030	0.050	0.030	0.060	0.030
平均值	0.070	0.043	0.054	0.041	0.100	0.036	0.054	0.049	0.050	0.034

表 5 KITTI 数据集上各模块耗时

Table 5 Time consumption of each module on KITTI dataset

模块	平均耗时/ms
特征检测	15.3
IMU 融合	2.4
几何-运动误差联合模型	12.8
相机位姿估计	24.6
动态目标跟踪	32.8
位姿优化	38.4

表 6 耗时比较

Table 6 Time consuming comparison

算法	FPS
DynaSLAM II	10~12
VDO-SLAM	5~8
本文算法	9~11

点位置,正方形分别为目标 1 和 2 的起点位置,箭头为运动方向和理想轨迹。实验中目标 1 和 2 在运动过程中会发生短暂遮挡,从而验证本文算法在动态遮挡下动态目标检测方法和动态目标跟踪方法性能。

1) 动态目标检测效果分析

动态目标检测效果如图 14 所示,从图 14 中可以看出本文的动态目标识别算法在数据帧中目标 1 运动时和目标 2 站立静止状态和运动状态时均能精准识别。

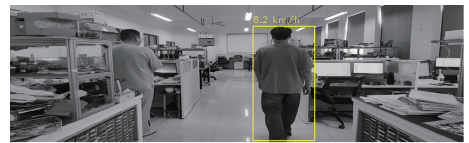


图 14 真实场景动态目标检测效果

Fig. 14 Real scene dynamic object detection effect

2) 动态目标检测效果分析

真实场景动态目标跟踪效果如图 15 所示,图中曲线为全局地图中相机和动态目标轨迹,矩形表示相机的位姿估计轨迹圆点表示目标 1(Object1)和目标 2(Object2)的运



图 12 实验室的真实场景

Fig. 12 Real laboratory scene

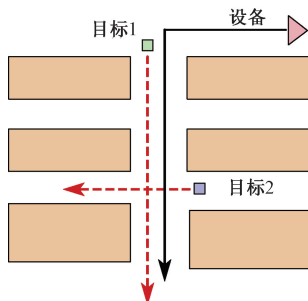


图 13 真实场景平面图

Fig. 13 Real scene floor plan

动轨迹。实验场景中目标 1 在运动过程中存在被目标 2 遮挡的情况。从图 15 可以看出,本文算法在目标 1 被遮挡期间能够预测其位姿状态,并在遮挡恢复后实现轨迹的修正,从而获得动目标 1 连续的运动轨迹。

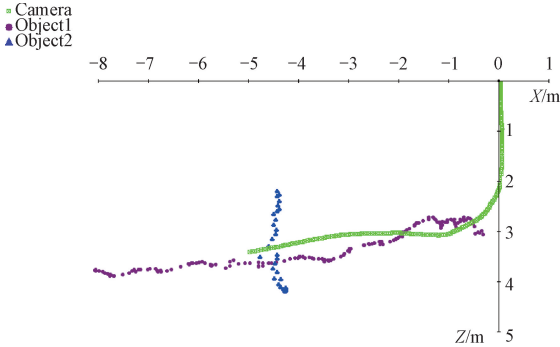


图 15 真实场景的动态目标跟踪效果

Fig. 15 Dynamic object tracking effect in real scenest

3) 联合因子图优化效果分析

图 16 展示了本文算法在真实场景下的相机位姿轨迹,实线和虚线分别表示联合因子图优化前(before)和优化后(after)的轨迹。从图 16 中可得,优化前在起始阶段设备抖动及转向时位姿估计误差较大;优化后轨迹更加平滑,有效减少了误差。

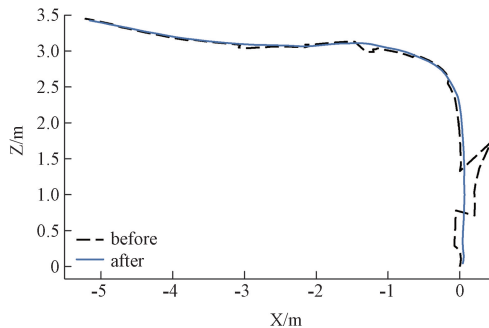


图 16 真实场景的相机位姿轨迹图

Fig. 16 Camera pose trajectory diagram of real scene

图 17 则对比了不同模型前后的 ATE 变化情况,结果

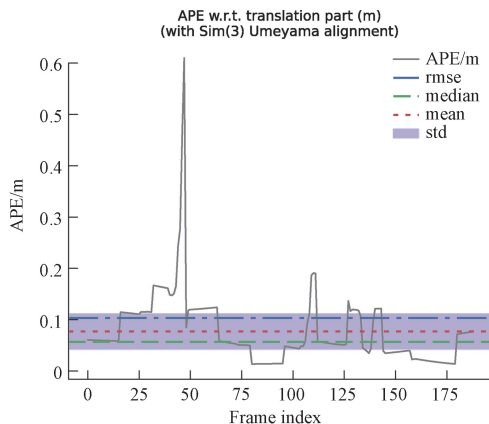


图 17 绝对轨迹误差对比

Fig. 17 Absolute trajectory error comparison

显示优化后曲线在起始阶段和转向过程中显著降低了定位误差,最大 ATE 降低幅度达 0.61,验证了联合因子图优化对提升定位精度的有效性。

6 结 论

针对高动态场景中目标遮挡造成 VSLAM 算法定位精度下降的问题,本文在 ORB-SLAM3 算法框架下引入目标检测机制,提出了一种基于几何-运动误差联合模型及目标轨迹预测的动态 VSLAM 算法,实现了动态目标的精准检测、持续跟踪与全局状态优化,提升了复杂动态环境下的系统鲁棒性。具体工作为:

针对目标遮挡对动态目标检测的影响,结合 IMU 预积分获取的运动信息对几何误差进行补充,提出几何-运动误差联合模型,通过概率模型判断特征点的遮挡情况和运动状态,提升动态目标检测的准确性。

针对动态目标跟踪受到遮挡影响,引入基于 EKF 的动态目标轨迹预测方法,实现动态目标运动状态的持续估计,降低不确定性,提升动态目标跟踪的精度和稳定性。

构建联合因子图模型,将相机位姿、地图点、动态目标位置及其运动轨迹统一优化,减少累积误差,提高算法的定位精度与地图构建质量。

在 KITTI 跟踪数据集和真实场景进行对本文所提算法的可行性和有效性进行实验验证,结果表明在 KITTI 选取序列上相比较 ORB-SLAM3 相机 RPE_r 和 RPE_t 相分别降低了 28.57% 和 20%。相比较 VDO-SLAM,追踪目标 RPE_r 和 RPE_t 平均值分别降低了 28.86% 和 17.77%。在高动态环境中出现遮挡情况下能够有效提升算法定位精度以及动态目标跟踪精度和稳定性。

参考文献

- [1] ZHANG Y, WU Y Q, TONG K, et al. Review of visual simultaneous localization and mapping based on deep learning[J]. Remote Sensing, 2023, 15(11): 2740.
- [2] LEBEDEV E, UCHAEV D, UCHAEV D. Stability assessment and performance comparison of VSLAM frameworks on open indoor datasets[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2024, 48: 101-107.
- [3] 邓鹏,罗静.复杂环境下机器人多传感器融合定位方法[J].电子测量与仪器学报,2023,37(12):48-57.
DENG P, LUO J. Multi-sensor fusion localization methods for robots in complex environments [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(12): 48-57.
- [4] 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.

- [5] WANG Y N, TIAN Y B, CHEN J W, et al. A survey of visual SLAM in dynamic environment: The evolution from geometric to semantic approaches[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-21.
- [6] 周非, 陈帅, 吴凯, 等. 快速跟踪分割辅助的动态 SLAM[J]. 仪器仪表学报, 2023, 44(5): 313-321.
ZHOU F, CHEN SH, WU K, et al. Dynamic SLAM assisted by fast tracking and segmentation[J]. Chinese Journal of Scientific Instrument, 2023, 44(5): 313-321.
- [7] PU H Y, LUO J, WANG G, et al. Visual SLAM integration with semantic segmentation and deep learning: A review[J]. IEEE Sensors Journal, 2023, 23(19): 22119-22138.
- [8] CHEN ZH, ZANG Q Y, ZHANG K H. DZ-SLAM: A SAM-based SLAM algorithm oriented to dynamic environments[J]. Displays, 2024, 85: 102846.
- [9] ZHAO X, DING W CH, AN Y Q, et al. Fast segment anything [J]. ArXiv preprint arXiv: 2306.12156, 2023.
- [10] CONG P CH, LI J X, LIU J J, et al. SEG-SLAM: Dynamic indoor RGB-D visual SLAM integrating geometric and YOLOv5-based semantic information[J]. Sensors, 2024, 24(7): 2102.
- [11] ZHANG J, HENEIN M, MAHONY R, et al. VDO-SLAM: A visual dynamic object-aware SLAM system[J]. ArXiv preprint arXiv:2005.11052, 2020.
- [12] 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.
- [13] SHI X Y, HUANG ZH Y, BIAN W K, et al. Videoflow: Exploiting temporal cues for multi-frame optical flow estimation[C]. IEEE/CVF International Conference on Computer Vision, 2023: 12469-12480.
- [14] ROSINOL A, ABATE M, CHANG Y, et al. Kimera: An open-source library for real-time metric-semantic localization and mapping [C]. 2020 IEEE International Conference on Robotics and Automation(ICRA). IEEE, 2020: 1689-1696.
- [15] FENG SH Q, LI X X, AI CH X, et al. VIMOT: A tightly coupled estimator for stereo visual-inertial navigation and multi-object tracking [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-14.

作者简介

梁若雨, 硕士研究生, 主要研究方向为机器视觉。

E-mail: 994054653@qq.com

黄卫华(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为视觉、组合定位。

E-mail: 64731259@qq.com