

DOI: 10.13382/j.jemi.B2306771

动态特征剔除的无人系统视觉/惯性导航方法^{*}

多靖赞¹ 赵 龙² 赵毅琳² 李俊韬¹

(1. 北京物资学院智能物流系统北京市重点实验室 北京 101149; 2. 北京航空航天大学数字导航中心 北京 100191)

摘 要:为降低动态环境对视觉/惯性导航系统定位精度与稳定性的影响,提出了一种动态特征剔除的视觉/惯性导航方法。该方法在视觉/惯性导航系统 VINS 框架基础上,以结构相似度作为成本量生成端到端网络,检测环境中的动态区域;通过特征光流矢量对已检测到的动态区域进行对称光流筛选,剔除该区域内的动态特征;融合视觉和惯性测量构造代价函数,通过非线性优化方法有效估计无人系统状态。实验结果表明,动态特征剔除后的视觉/惯性导航方法具有良好的定位精度和稳定性,其位置均方根误差在 EuRoC 公开数据集和实际场景采集数据上分别为 0.081 和 1.982 m,仅为 VINS 的 35.5% 和 24.9%。该方法可在复杂应用环境中提供精确的位置信息,且在低成本无人系统导航定位方面具有良好的实用价值。

关键词: 视觉/惯性导航; 动态特征剔除; 结构相似度; 端到端网络; 对称光流

中图分类号: TP242.6 **文献标识码:** A **国家标准学科分类代码:** 510.8

Visual/inertial navigation method for unmanned system with dynamic feature removal

Duo Jingyun¹ Zhao Long² Zhao Yilin² Li Juntao¹

(1. Beijing Key Laboratory of Intelligent Logistics System, Beijing Wuzi University, Beijing 101149, China;

2. Digital Navigation Center, Beihang University, Beijing 100191, China)

Abstract:To reduce the impact of dynamic environment on the localization accuracy and stability of visual/inertial navigation system, a dynamic feature removal visual/inertial navigation method is proposed in this paper. Based on the VINS framework, this method uses structural similarity as the cost volume to generate an end-to-end network for dynamic regions detection. Symmetric optical flow screening is then performed on the identified dynamic regions to remove non-consistent outliers and further eliminate dynamic features that affect localization. The cost function is constructed by fusing visual and inertial measurements, and the nonlinear optimization method is used to estimate the unmanned system states effectively. The experimental results show that the visual/inertial navigation method with dynamic feature removal has good localization accuracy and stability, the position root mean square error is 0.081 and 1.982 m on EuRoC publicly available datasets and real scenario data respectively, which is only 35.5% and 24.9% of VINS. This method can provide accurate position information in complex application environment, and has good practical value in the navigation of low-cost unmanned systems.

Keywords: visual/inertial navigation; dynamic feature removal; structural similarity; end-to-end network; symmetric optical flow

0 引 言

无人系统具备小型化、低能耗和高安全性等优点,在森林防火、电力巡线和事故搜救等民用领域以及任务侦

察、定点巡航和物资输送等军用领域均得到了广泛应用^[1-3]。随着“中国制造 2025”和人工智能国家战略计划的推进,无人系统已向人工智能方向发展,其智能自主及应用已成为我国社会发展中的重大科学技术问题。

在复杂应用环境中,无人系统自主探测、感知环境信

收稿日期: 2023-07-24 Received Date: 2023-07-24

^{*} 基金项目: 国家自然科学基金(42274037)、航空科学基金(2022Z022051001)、国家重点研发计划(2020YFB0505804)、北京市教委科技计划重点项目(KZ202210037046)、北京市通州区科技创新人才项目(JCQN2023030)、北京物资学院青年科研基金(2022XJQN22)项目资助

息和运动状态,获取连续可靠的定位、导航和授时 (positioning, navigation and timing, PNT) 信息是实现任务自主的关键。目前,无人系统多采用 GNSS (global navigation satellite system) 校准 INS (inertial navigation system) 实现无漂移状态估计。然而在室内、树林和城市峡谷等应用环境中,GNSS 信号易受遮挡,无法提供稳定可靠的导航定位信息。一旦 GNSS 信号丢失,仅由 INS 系统提供的状态估计将快速漂移。因此,需研究自主性强、定位精度高的其他导航方法,在 GNSS 信号拒止环境下为无人系统提供连续可靠的 PNT 信息。

近年来,随着计算机视觉技术发展以及微处理器运算能力增强,采用相机作为传感器构建视觉导航系统,提高在 GNSS 信号拒止环境下定位精度与可靠性的方法逐渐成为研究热点^[3-6]。与其他传感器相比,相机具备两大优势:首先,相机是一种被动传感器,隐蔽性强且不易被干扰,该特性在军事应用领域优势明显;其次,相机具有低成本和小型化优点,且采集信息内容丰富,易实现环境感知。惯性测量单元 (inertial measurement unit, IMU) 可以通过加速度和角速度测量实现短时状态估计,虽然长时间使用存在较大累积误差,但短时相对位移精度较高。因此,相机和 IMU 具备优势互补特点,且较单一导航系统而言,视觉/惯性组合导航系统^[7-9]可在一定程度上提升导航定位精度和可靠性。然而,视觉/惯性组合导航系统难以适应宽动态、长航时、大范围导航定位任务^[10],原因在于相机采集的视觉信息易受动态环境变化影响,导致图像帧间递推误差增大,最终产生较大累计误差。为降低动态环境对导航定位精度与稳定性的影响,Maity 等^[11]引入边缘点特征提出了 Edge SLAM (simultaneous localization and mapping) 算法,Pumarola 等^[12]在特征点基础上引入了直线特征,提出了 PL-SLAM 算法,新特征的引入在一定程度上增加了图像特征的丰富度,但未考虑新特征类型仍存在动态问题;Zhao 等^[13]在视觉/惯性测量基础上引入了轮式里程计测量方法,Shan 等^[14]引入了图像深度测量方法,通过引入更多传感器数据在一定程度上约束了动态环境对导航定位精度和稳定性的影响,但效果并不显著;Dem 等^[15]使用 YOLO 网络尝试检测并剔除动态特征,然而 YOLO 网络只能对特定类型物体进行识别,无法提取无标签的动态物体;Xiao 等^[16]使用与 YOLO 相似的 SSD 网络进行目标检测,并通过目标跟踪算法减小漏检概率,但仍未有效解决无标签的动态目标提取问题。

针对宽动态、长航时、大范围的无人系统导航定位任务,本文提出了一种基于动态特征剔除的视觉/惯性导航方法。该方法在视觉/惯性导航系统 (visual/inertial navigation system, VINS)^[8] 框架基础上进行改进,构建了以结构相似度作为成本量的端到端网络,感知并检测环

境中的动态区域;采用对称光流筛选并剔除已识别动态区域中的动态特征;通过非线性优化策略构造代价函数,融合视觉和惯性测量,有效估计位姿信息。实验结果表明,该方法可有效降低动态环境对导航定位结果的影响,提高定位精度及稳定性,适用于 GNSS 信号拒止条件下的复杂导航任务。

1 动态特征剔除

环境动态变化产生的动态特征,直接影响图像特征跟踪效果,进而降低了视觉/惯性导航系统定位精度与稳定性。本文以结构相似度作为成本量,生成端到端网络进行动态环境区域筛选;通过特征光流矢量对已检测到的动态区域进行对称光流筛选,剔除该区域内的动态特征,减小了动态环境变化对导航定位精度和稳定性的影响。

1.1 结构相似度

结构相似度是以图像亮度、对比度与结构差异 3 项作为度量指标对图像相似度或失真程度进行衡量,相较均方根误差和峰值信噪比等以数值绝对差异作为度量标准,结构相似度构建的数学模型更符合人类视觉系统直观差异感受,更容易筛选出实际场景中的动态变化。

图像亮度通常使用平均灰度进行衡量,即:

$$\mu = \frac{1}{n \times m} \sum_{x=1}^n \sum_{y=1}^m I(x, y) \quad (1)$$

式中: $I(x, y)$ 为像素坐标 (x, y) 处图像的灰度值; n 和 m 分别为图像宽度和高度。两张图像 A 与 B 的亮度对比函数可表示为:

$$l(A, B) = \frac{2\mu_A\mu_B + t_1}{\mu_A^2 + \mu_B^2 + t_1} \quad (2)$$

式中: μ_A 和 μ_B 分别为图像 A 与 B 的平均灰度; $t_1 = (K_1 \times 255)^2$ 为避免奇异性引入的经验常量,本文取 $K_1 = 0.01$ 。

图像对比度使用像素标准差进行衡量,即:

$$\sigma = \left(\frac{1}{n \times m - 1} \sum_{x=1}^n \sum_{y=1}^m (I(x, y) - \mu)^2 \right)^{\frac{1}{2}} \quad (3)$$

式中: $I(x, y)$ 为像素坐标 (x, y) 处图像的灰度值; n 和 m 分别为图像宽度和高度; μ 为图像的平均灰度。两张图像 A 与 B 的对比度对比函数可表示为:

$$c(A, B) = \frac{2\sigma_A\sigma_B + t_2}{\sigma_A^2 + \sigma_B^2 + t_2} \quad (4)$$

式中: σ_A 和 σ_B 分别为图像 A 与 B 的像素标准差; $t_2 = (K_2 \times 255)^2$ 为避免奇异性引入的经验常量,本文取 $K_2 = 0.03$ 。

两张图像 A 与 B 的结构对比函数可表示为:

$$s(A, B) = \frac{\sigma_{AB} + t_3}{\sigma_A\sigma_B + t_3} \quad (5)$$

式中: σ_A 和 σ_B 分别为图像 A 与 B 的像素标准差; $t_3 = t_2/2$ 为避免奇异性引入的经验常量; σ_{AB} 定义为:

$$\sigma_{AB} = \frac{1}{n \times m - 1} \sum_{x=1}^n \sum_{y=1}^m (I_A(x, y) - \mu_A)(I_B(x, y) - \mu_B) \quad (6)$$

式中: $I_A(x, y)$ 和 $I_B(x, y)$ 分别为图像 A 与 B 在像素坐标 (x, y) 处的灰度值; μ_A 和 μ_B 分别为图像 A 与 B 的平均灰度; n 和 m 分别为图像的宽度和高度。

两张图像 A 与 B 的结构相似度指数可表示为:

$$SSIM(A, B) = l(A, B) \cdot c(A, B) \cdot s(A, B) \quad (7)$$

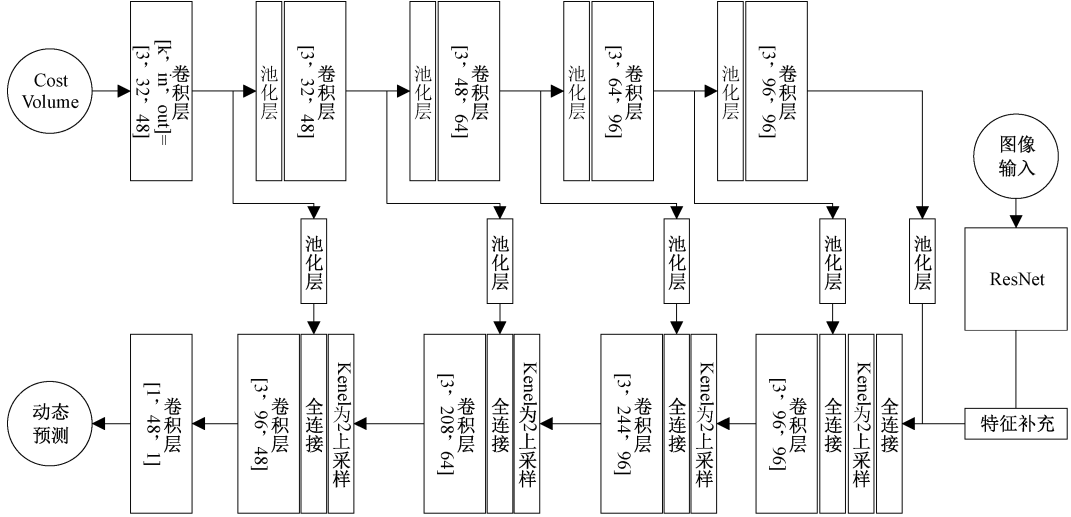


图 1 动态区域检测网络模型

Fig. 1 Network model of dynamic regions detection

先验结构差异由结构相似度指标计算获得,可表示为:

$$pe(x, y) = \frac{1 - SSIM(T_{t'}(x, y)T_t(x, y))}{2} \quad (8)$$

式中: $T(x, y)$ 为像素坐标 (x, y) 附近大小为 3×3 pixels 的滑窗; t 和 t' 分别表示参考帧和计算帧; $SSIM(\cdot)$ 为两帧之间的结构相似指标。

根据先验结构差异定义代价体系为:

$$C(x, y) = 1 - 2 \cdot \sum_{t' \in \{1, 2, \dots, N\}, t' \neq t} pe_{t'}^t(x, y) \cdot w_{t'} \quad (9)$$

式中: $w_{t'}$ 为计算帧 t' 对应的权重,可根据先验结构差异自适应调整,差异越大权重越小。据此定义的先验结构差异包含了连续图像帧组 $t' \in \{1, 2, \dots, N\}$ 在以 t 为参考帧像素坐标 (x, y) 处的差异,可检测出先验结构差异过大的动态目标区域。

在网络模型训练时,由于缺少直接由结构差异性定义的语义标签数据,本文将常见动态物体以及亮度对比强烈区域作为待剔除对象构建训练真值。训练数据由采集道路数据获取,选取 2 min 内具有较多动态目标的 1 800 帧图像,通过计算图像像素对比度生成亮度对比强

式中: $l(A, B)$ 、 $c(A, B)$ 与 $s(A, B)$ 分别为两张图像 A 与 B 的亮度对比、对比度对比和结构对比函数。

1.2 端到端网络

以结构相似度作为成本量,构建 Unet 端到端网络,检测环境中的动态区域,其网络结构如图 1 所示。该网络以关键帧前后的连续图像帧组作为输入,计算图像帧组与关键帧之间的先验结构差异,进行动态目标预测标注,输出该像素为动态像素的可能性,形成用于剔除动态特征的掩膜。该网络可减少中间变量转换次数,降低多模型串联引入的累积误差,具有良好的泛化性。

烈区域标签,通过 Mask-R-CNN^[17] 识别动态物体并二值化生成动态物体区域标签,采用预测值与构建真值的二元交叉熵作为损失函数,获得动态剔除网络参数。

此外,为降低使用结构差异对运动目标附近背景图像的影响,使用 U 型网络结构 ResNet-18^[18] 对图像进行特征编码,引入先验语义信息实现特征补充,通过共享权重对网络成本量进行编码连接,该网络可用于输入图像帧数量不同的场景,确保了不同相机接入的泛化能力。

1.3 对称光流动态点剔除

VINS-Mono^[8] 在提取特征点后通过 LK (Lucas Kanade) 光流进行特征跟踪,并通过 RANSAC (random sample consensus) 方法剔除外点 (误匹配点与动态点)。为提高外点筛选的准确性,本文在已检测的动态区域内,加入基于对称光流的二次筛选机制,即分别计算以 t_k 时刻为参考, t_{k+1} 时刻特征点的光流矢量,以及以 t_{k+1} 时刻为参考, t_k 时刻特征点的光流矢量,其计算结果须满足模长相等,方向相反的特性,如不满足上述特性,可将其视为外点进行剔除。

为提高对称光流的计算效率,在计算以 t_{k+1} 时刻为

参考, t_k 时刻特征点光流矢量时, 采用逆向光流法^[19], 定义优化目标函数为:

$$\text{error} = \sum_x [I_k(\mathbf{W}(\mathbf{x}; \mathbf{0})) - I_{k+1}(\mathbf{W}(\mathbf{x}; \mathbf{p}))]^2 \quad (10)$$

式中: $\mathbf{x}$ 表示坐标为 (x, y) 的像素; $\mathbf{p} = (p_1, p_2)$ 表示光流矢量; $\mathbf{W}(\mathbf{x}; \mathbf{p}) = (x + p_1, y + p_2)$; $I_k(\cdot)$ 和 $I_{k+1}(\cdot)$ 分别表示 t_k 和 t_{k+1} 时刻对应图像的灰度值。由于 t_{k+1} 时刻为参考时刻, 则 $\mathbf{W}(\mathbf{x}; \mathbf{p})$ 为已知量, 采用高斯牛顿法求解式 (10), 其雅可比矩阵为:

$$J(\mathbf{p}) = \sum_x \nabla I_k \frac{\partial \mathbf{W}}{\partial \mathbf{p}} \quad (11)$$

式中: ∇I_k 为 t_k 时刻对应图像的梯度。由式 (10) 可知, t_k 时刻对应运动前的图像, 其对应的 $\mathbf{p}$ 值为向量 $\mathbf{0}$, 则优化迭代量

$$\Delta \mathbf{p} = - [J(\mathbf{p})^T J(\mathbf{p})]^{-1} J(\mathbf{p})^T [I_k(\mathbf{x}) - I_{k+1}(\mathbf{W}(\mathbf{x}; \mathbf{p}))] \quad (12)$$

相较于正向光流法需多次计算雅可比矩阵进行迭代求解, 逆向光流法的雅可比矩阵与 $\mathbf{p}$ 无关, 因此, 仅需计算初始雅可比矩阵, 后续优化过程可重复使用初始雅可比矩阵, 有效降低了优化求解的计算量。

2 算法框架

本文算法在 VINS-Mono^[8] 框架基础上进行了优化改进, 针对弱纹理环境中图像特征数量有限的问题, 加入直线特征, 对特征类型和数量进行补充; 引入动态特征剔除策略降低动态环境变化对导航定位精度的影响, 算法框架如图 2 所示, 图中虚线框图为本文优化改进部分。

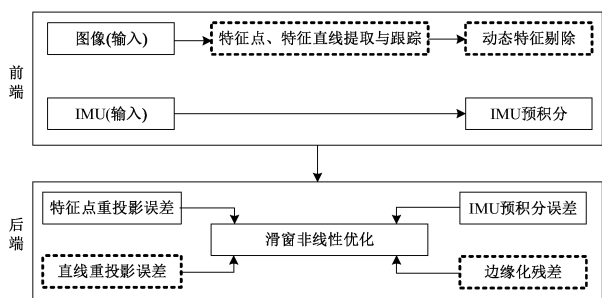


图 2 动态特征剔除的视觉/惯性导航算法框架

Fig. 2 Visual/inertial navigation algorithm framework with dynamic feature removal

算法前端在特征点检测与跟踪基础上加入了直线特征检测与匹配算法, 采用 LSD (line segment detector)^[20] 算法提取图像中的直线特征。该方法基于图像梯度向量, 利用区域生长判断直线区域的一致性, 实现直线检测, 融合点与直线的特征检测结果如图 3 所示。将检测到的直线, 采用 LBD (line band descriptor)^[21] 描述子进行

表示, 并采用曼哈顿距离作为度量方式进行匹配。

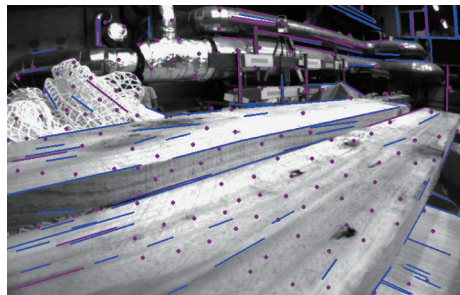


图 3 融合点与直线的特征检测结果

Fig. 3 Feature detection results of fusion points and lines

算法后端融合视觉和惯性测量, 选用基于小孔成像模型的视觉特征点重投影误差、直线重投影误差以及 IMU 预积分误差构建目标函数, 采用高斯牛顿法优化求解状态量, 其目标函数可表示为:

$$J(\chi) = \sum_{k=1}^n \sum_{i=1}^m \| \mathbf{r}(\hat{\mathbf{z}}_{kf_i}^c, \chi) \|^2 + \sum_{k=1}^n \sum_{j=1}^M \| \mathbf{r}(\hat{\mathbf{z}}_{kl_j}^c, \chi) \|^2 + \sum_{k=1}^{n-1} \| \mathbf{r}(\hat{\mathbf{z}}_{kb_{k+1}}^c, \chi) \|^2 + J_{\text{mar}}(\chi) \quad (13)$$

式中: n 为滑窗中图像的帧数; m 为滑窗中观测到特征点的数目; M 为滑窗中观测到特征直线的数目; χ 为滑窗中的状态量集合, $\chi = [\mathbf{x}_{b_1}, \dots, \mathbf{x}_{b_n}, \mathbf{p}_{f_1}^w, \dots, \mathbf{p}_{f_m}^w]$; 滑窗中第 k 帧图像对应的状态量 $\mathbf{x}_{b_k} = [\mathbf{p}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{R}_{wb_k}, \mathbf{b}_{ab_k}, \mathbf{b}_{gb_k}]$ 包含了 IMU 坐标系相对世界坐标系的位置 $\mathbf{p}_{b_k}^w$ 、速度 $\mathbf{v}_{b_k}^w$ 、姿态 $\mathbf{R}_{wb_k}$ 、加速度计零偏 $\mathbf{b}_{ab_k}$ 和陀螺仪漂移 $\mathbf{b}_{gb_k}$; $\mathbf{p}_{f_i}^w$ 为特征点 f_i 在世界坐标系中的位置; $\mathbf{r}(\hat{\mathbf{z}}_{f_i}^c, \chi)$ 为特征点 f_i 在滑窗中第 k 帧图像上的重投影误差; $\mathbf{r}(\hat{\mathbf{z}}_{l_j}^c, \chi)$ 为特征直线 l_j 在滑窗中第 k 帧图像中的重投影误差; $\mathbf{r}(\hat{\mathbf{z}}_{b_{k+1}}^c, \chi)$ 为滑窗中第 k 和 $k+1$ 帧图像间对应的 IMU 预积分误差项; $J_{\text{mar}}(\chi)$ 为由状态量边缘化产生的先验信息。

2.1 特征点重投影误差

将特征点从世界坐标系投影到像素坐标系, 计算该值与视觉测量值之间的差异, 即为特征点的重投影误差。特征点 f_i 在滑窗内第 k 帧图像的重投影误差可表示为:

$$\mathbf{r}(\hat{\mathbf{z}}_{f_i}^c, \chi) = \pi(\mathbf{R}_{bc}^{-1}(\mathbf{R}_{wb_k}^{-1}(\mathbf{p}_{f_i}^w - \mathbf{p}_{b_k}^w) - \mathbf{p}_c^b)) - \hat{\mathbf{z}}_{f_i}^c \quad (14)$$

式中: $\hat{\mathbf{z}}_{f_i}^c$ 为视觉特征点 f_i 在第 k 帧像素坐标中的观测值; $\mathbf{R}_{bc}$ 和 $\mathbf{p}_c^b$ 表示从相机坐标系到 IMU 坐标系的旋转矩阵和平移向量; 记 $\mathbf{p}_{f_i}^c = \mathbf{R}_{bc}^{-1}(\mathbf{R}_{wb_k}^{-1}(\mathbf{p}_{f_i}^w - \mathbf{p}_{b_k}^w) - \mathbf{p}_c^b)$ 表示视觉特征点 f_i 在第 k 帧相机坐标系 c_k 中的 3 D 坐标, $\pi(\mathbf{p}_{f_i}^c)$ 表示将 $\mathbf{p}_{f_i}^c$ 通过内参数矩阵从相机坐标系转换到像素坐标系的函数。

2.2 直线重投影误差

直线的普吕克坐标可表示为 $\mathbf{L} = (\mathbf{n}^T, \mathbf{d}^T)^T$, 其中, $\mathbf{d}$ 为直线的方向向量, $\mathbf{n}$ 为直线与坐标原点构成平面的法向量。将三维空间中的直线投影到二维图像平面, 可得直线在像平面中的一般表达式为:

$$l_1 x + l_2 y + l_3 = 0 \quad (15)$$

式中: l_1, l_2 与 l_3 为平面直线表达式的参数, 满足:

$$\begin{bmatrix} l_1 \\ l_2 \\ l_3 \end{bmatrix} = \begin{bmatrix} f_y & 0 & 0 \\ 0 & f_x & 0 \\ -f_y c_x & -f_x c_y & f_x f_y \end{bmatrix} \mathbf{n} \quad (16)$$

式中: f_x 与 f_y 为相机 x 和 y 轴的焦距; (c_x, c_y) 为相机主点坐标。

直线重投影误差定义为将空间直线投影到像平面后与图像中观测到的直线端点的距离, 直线 l_j 在滑窗内第 k 帧图像中的重投影误差可表示为:

$$\mathbf{r}(\hat{\mathbf{z}}_{l_j}^k, \mathbf{X}) = \begin{bmatrix} \frac{\mathbf{s}^T \mathbf{l}_j}{\sqrt{l_{j1}^2 + l_{j2}^2}} \\ \frac{\mathbf{e}^T \mathbf{l}_j}{\sqrt{l_{j1}^2 + l_{j2}^2}} \end{bmatrix} \quad (17)$$

式中: $\mathbf{s}$ 与 $\mathbf{e}$ 分别为第 k 帧图像中观测到直线起始点和终止点的齐次坐标; $\mathbf{l}_j = [l_{j1}, l_{j2}, l_{j3}]$ 为直线 l_j 在第 k 帧图像平面中的表达式系数矩阵。

2.3 IMU 预积分误差

基于非线性优化的状态估计需对状态量进行多次迭代更新, 而 IMU 采用递推方法估计状态量。为避免优化时因初始状态改变引起的 IMU 重复积分运算, 本文采用了 IMU 预积分策略^[22], 引入与初始状态量无关的相对运动增量方程, 滑窗中第 k 和 $k+1$ 帧图像间对应的 IMU 预积分误差项可表示为:

$$\mathbf{r}(\hat{\mathbf{z}}_{b_k b_{k+1}}, \mathbf{X}) = \begin{bmatrix} \mathbf{R}_{wb_k}^{-1}(\mathbf{p}_{b_{k+1}}^w - \mathbf{p}_{b_k}^w - \mathbf{v}_{b_k}^w \Delta t_{b_{k+1}b_k} + \frac{1}{2} \mathbf{g}^w \Delta t_{b_{k+1}b_k}^2) - \Delta \hat{\mathbf{p}}_{b_k b_{k+1}} \\ \mathbf{R}_{wb_k}^{-1}(\mathbf{v}_{b_{k+1}}^w - \mathbf{v}_{b_k}^w + \mathbf{g}^w \Delta t_{b_{k+1}b_k}) - \Delta \hat{\mathbf{v}}_{b_k b_{k+1}} \\ 2[(\Delta \hat{\mathbf{q}}_{b_k b_{k+1}})^{-1} \otimes \mathbf{q}_{wb_k}^{-1} \otimes \mathbf{q}_{wb_{k+1}}]_{xyz} \\ \mathbf{b}_{a_{b_{k+1}}} - \mathbf{b}_{a_{b_k}} \\ \mathbf{b}_{g_{b_{k+1}}} - \mathbf{b}_{g_{b_k}} \end{bmatrix} \quad (18)$$

式中: $\mathbf{g}^w$ 为重力加速度在惯性坐标系中的投影; $\Delta t_{b_{k+1}b_k}$ 为第 k 和 $k+1$ 帧的时间间隔; $\Delta \hat{\mathbf{p}}_{b_k b_{k+1}}, \Delta \hat{\mathbf{v}}_{b_k b_{k+1}}$ 和 $\Delta \hat{\mathbf{q}}_{b_k b_{k+1}}$ 分别为位置、速度和姿态的预积分估计值; $\mathbf{q}_{wb_k}$ 为姿态的单位四元数表示; $[\cdot]_{xyz}$ 表示四元数的虚部。

2.4 边缘化先验信息

随着运行时间推移, 视觉/惯性导航系统状态量数目将显著增加, 算法复杂度将随状态量数目增加而增大。如果求解所有状态量, 将无法实时满足需求; 如果仅求解最新一帧图像对应的状态量将忽略各状态量间的关联, 降低系统的精度。本文通过构建滑窗, 限制了待求解状态量数目; 采用状态量边缘化策略, 将其余状态量转换为待求解状态量的先验观测信息, 实现了算法复杂度和精确性之间的平衡。

假设 $\mathbf{x}_\mu$ 为 $\mathbf{X}$ 中要边缘化的状态量, $\mathbf{x}_\lambda$ 为 $\mathbf{X}$ 中与 $\mathbf{x}_\mu$ 存在误差关联的状态量, $\mathbf{x}_p$ 为 $\mathbf{X}$ 中除去 $\mathbf{x}_\mu$ 和 $\mathbf{x}_\lambda$ 后剩余的状态量, 即 $\mathbf{X} = [\mathbf{x}_\mu, \mathbf{x}_\lambda, \mathbf{x}_p]$ 。由于 $\mathbf{x}_p$ 相对于 $\mathbf{x}_\mu$ 独立, 可将高斯牛顿法的增量方程简化为:

$$\begin{bmatrix} \mathbf{H}_{\mu\mu} & \mathbf{H}_{\mu\lambda} \\ \mathbf{H}_{\lambda\mu} & \mathbf{H}_{\lambda\lambda} \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_\mu \\ \delta \mathbf{x}_\lambda \end{bmatrix} = \begin{bmatrix} \mathbf{b}_\mu \\ \mathbf{b}_\lambda \end{bmatrix} \quad (19)$$

采用舒尔补 (Schur Complement)^[23] 原理简化式 (19), 得:

$$\begin{bmatrix} \mathbf{H}_{\mu\mu} & \mathbf{H}_{\mu\lambda} \\ \mathbf{0} & \mathbf{H}_{\lambda\lambda}^* \end{bmatrix} \begin{bmatrix} \delta \mathbf{x}_\mu \\ \delta \mathbf{x}_\lambda \end{bmatrix} = \begin{bmatrix} \mathbf{b}_\mu \\ \mathbf{b}_\lambda^* \end{bmatrix} \quad (20)$$

式中: $\mathbf{H}_{\lambda\lambda}^* = \mathbf{H}_{\lambda\lambda} - \mathbf{H}_{\lambda\mu} \mathbf{H}_{\mu\mu}^{-1} \mathbf{H}_{\mu\lambda}$, $\mathbf{b}_\lambda^* = \mathbf{b}_\lambda - \mathbf{H}_{\lambda\mu} \mathbf{H}_{\mu\mu}^{-1} \mathbf{b}_\mu$ 。

求解方程:

$$\mathbf{H}_{\lambda\lambda}^* \delta \mathbf{x}_\lambda = \mathbf{b}_\lambda^* \quad (21)$$

可得增量 $\delta \mathbf{x}_\lambda$ 值, 进而更新状态量 $\mathbf{x}_\lambda$ 。

3 实验结果

分别在公开数据集和实际场景采集数据上, 将本文算法与两种现有先进开源视觉/惯性导航系统: VINS-Mono^[8] 和 ORB-SLAM3^[9] 进行比较, 评估算法的准确性和有效性。本文实验均在硬件配置为 Intel i7-11700K 处理器、16 GB 内存及 NVIDIA GeForce RTX 3060 显卡终端上运行。

3.1 公开数据集测试实验

选用欧洲机器人技术挑战赛 (European Robotics Challenge, EuRoC) 公开数据集^[24] 进行室内测试实验。EuRoC 数据集基于 AscTec 无人机平台配置了视觉/惯性传感器系统, 采集具有精确时间戳的双目图像和 IMU 数据, 且基于 Leica MS50 激光器和 Vicon 运动捕捉系统提供精确的 6 D 运动数据作为真值参考, 该数据集已被不同国家和地区的视觉导航算法开发者广泛用于算法性能指标测试。EuRoC 数据集提供了工业厂房 (简称 MH 场景) 和 Vicon Room (简称 V 场景) 两种场景下的 11 组数据, 其中, 图像帧率为 20 Hz, 分辨率为 752×480 pixels; IMU 采集频率为 200 Hz。

区别于基于目标识别的动态区域检测方法^[15-16], 本

文方法以结构差异程度作为基准进行像素分割,由于网络输入为多连续帧间的差异对比,而非基于目标识别方法中的单帧图像,因此本文方法不但可以检测由动态环境变化引起的结构差异,还可以泛化至检测因自动曝光或相机运动引起的亮度、对比度变化过大的区域。对比图 4(a)与(b)以及图 5(a)与(b)中实验结果可知,该实验场景为静态场景不包含动态目标,但本文方法检测出图像中的高亮以及对比度过大区域,将其判定为结构差异较大区域,而该区域中的特征不稳定,存在较多误匹配点,因此剔除该区域内的特征可以有效提高算法的精度。

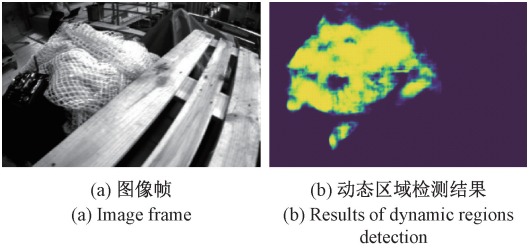


图 4 动态区域检测结果(“MH_01”数据集第 1 110 帧)
Fig. 4 Results of dynamic regions detection
(frame 1 110 in “MH_01” datasets)

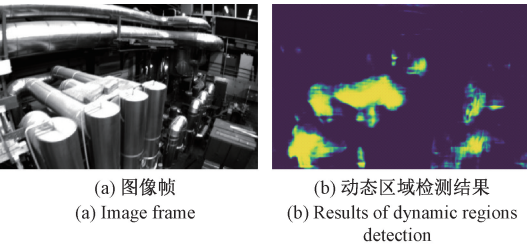


图 5 动态区域检测结果(“MH_01”数据集第 1 801 帧)
Fig. 5 Results of dynamic regions detection
(frame 1 801 in “MH_01” datasets)

由于缺少像素结构差异分类真值进行动态剔除网络性能直接评估,本文通过定位精度评估动态剔除网络对导航结果的优化效果。将 VINS-Mono、添加直线特征的 VINS-Mono、ORB-SLAM3 和本文算法进行定位精度对比测试,选取绝对位置误差(absolute position error, APE)作为算法精度的度量指标,限于篇幅,本文仅展示了“MH_04”和“V1_02”两组数据的实验结果,如图 6 和图 7 所示,实验结果包括轨迹曲线和绝对位置误差。整个数据集上的位置均方根误差(root mean squared error, RMSE)统计结果如表 1 所示。

由图 6(a)和图 7(a)可以看出,VINS-Mono、添加直线特征的 VINS-Mono、ORB-SLAM3 和本文算法都可以有效估计无人机位置,且不发生显著漂移。由图 6(b)和图 7(b)可以看出,在 95% 以上的时间段内本文算法的绝对

位置误差都小于添加直线特征的 VINS-Mono,证明了采用本文动态剔除网络对单帧“低质量”特征剔除的有效性。由表 1 可以看出,上述 4 种方法在 EuRoC 数据集上位置均方根误差均值分别为 0.228、0.153、0.116 和 0.081 m。相较于 VINS-Mono,添加直线特征的 VINS-Mono 在 EuRoC 数据集上的位置均方根误差均值下降了 0.075 m,由此可见,添加特征直线在一定程度上解决了弱纹理环境中的特征退化问题,提高了导航定位的精度。但是其效果并不稳定,在 V1_01 和 V2_01 场景中,添加直线特征后的位置均方根误差分别增大了 0.017 和 0.015 m,出现了负优化现象。而本文算法的定位精度明显优于其他 3 种算法。其原因主要在于,本文算法在添加直线特征的基础上,引入了动态特征剔除策略,检测图像中的结构差异较大的区域,剔除了该区域中的不稳定特征,提供的视觉观测精度更高。在 EuRoC 数据集上 ORB-SLAM3 的精度明显高于 VINS-Mono,这是因为 ORB-SLAM3 采用 ORB 特征进行匹配跟踪,VINS-Mono 使用金字塔光流实现匹配跟踪,ORB 特征的匹配精度优于光流跟踪,且 VINS-Mono 未对地图信息进行有效处理,而 ORB-SLAM3 对地图点云进行了有效的删除、合并与优化,在一定程度上提高了算法精度。

表 1 EuRoC 数据集上 4 种算法位置均方根误差统计结果

数据集	Table 1 The position RMSE of four different methods on EuRoC datasets (m)			
	算法			
	VINS-Mono	VINS-Mono+ 直线特征	ORB-SLAM3	本文算法
MH_01	0.183	0.112	0.083	0.058
MH_02	0.241	0.229	0.101	0.097
MH_03	0.314	0.133	0.146	0.094
MH_04	0.272	0.163	0.119	0.089
MH_05	0.361	0.225	0.122	0.062
V1_01	0.167	0.184	0.068	0.074
V1_02	0.185	0.096	0.139	0.043
V1_03	0.202	0.188	0.158	0.097
V2_01	0.086	0.101	0.053	0.039
V2_02	0.199	0.117	0.110	0.076
V2_03	0.302	0.161	0.182	0.126
平均值	0.228	0.153	0.116	0.081

3.2 车载导航数据测试实验

为验证在宽动态、长航时、大范围环境中算法的有效性,选择北航校园作为实验场景,开展室外车载导航测试实验。实验中操控小车沿北航校园马路做矩形轨迹运动,行驶距离为 1 005.19 m,数据持续 919.25 s。如图 8 所示,实验平台为松灵 HUNTER 遥控车,该车采用阿克曼前轮转向线控底盘,具备类汽车特征,适用于水泥、柏油等硬质路面长时间运行。将 MYNTEYE-D1000-50 相

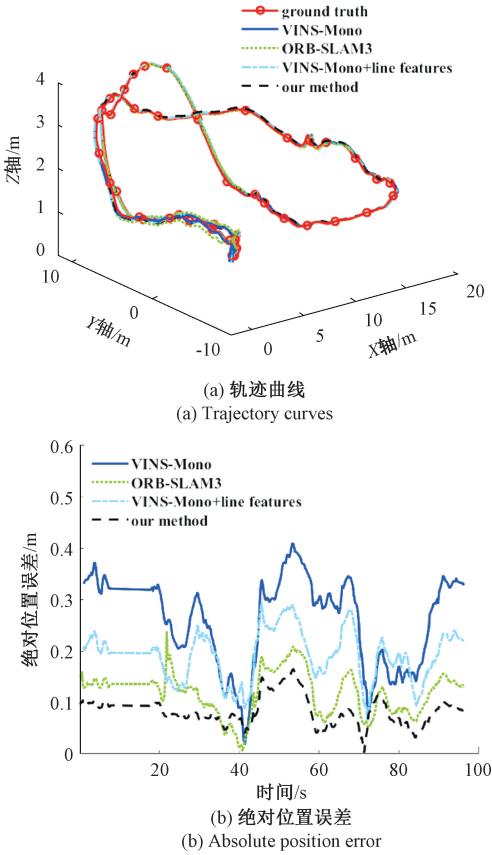


图 6 “MH_04”数据集实验结果

Fig. 6 Experimental results on “MH_04” datasets

机固联在遥控车上,前下视拍摄,采集数据。MYNTEYE-D1000-50 相机为紧凑型双目相机,基线长 120 mm,可提供帧率为 30 Hz,分辨率为 640×360 pixels 的双目图像数据,该相机集成了采集频率为 200 Hz 的 IMU,并实现了数据外触发同步功能。实验采用 NovAtel-PwrPak7-E1 轻便型多功能接收机获取位置参考真值,进行算法性能评估,其定位精度可达 1 cm。

采用本文动态区域检测方法检测图像中的动态目标,其结果如图 9 和 10 所示。对比图 9(a)与(b)以及图 10(a)与(b)中实验结果可知,算法采用结构相似度作为成本量,可以有效检测出环境中的动态目标,并将其标记为动态像素。

将本文算法、VINS-Mono、添加直线特征的 VINS-Mono 和 ORB-SLAM3 进行定位精度对比测试,选取 APE 作为算法精度的度量指标,结果如图 11 所示,实验结果包括轨迹曲线和绝对位置误差,位置均方根误差统计结果如表 2 所示。

由图 11(a)可以看出,VINS-Mono、添加直线特征的 VINS-Mono、ORB-SLAM3 的轨迹估计曲线存在较大误差累积,而本文算法可有效估计载体位置,且不发生显著漂移,解算结果更贴合轨迹真值。由图 11(b)可以看出,本

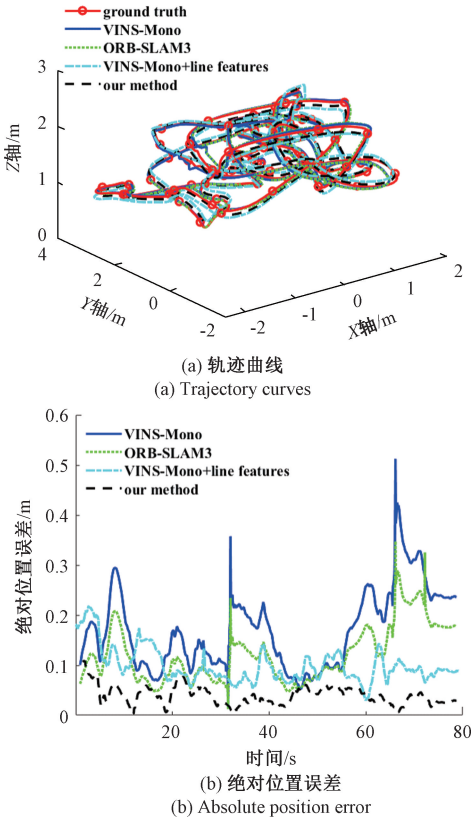


图 7 “V1_02”数据集实验结果

Fig. 7 Experimental results on “V1_02” datasets



图 8 车载导航数据采集平台

Fig. 8 Data acquisition platform of vehicle navigation

表 2 实际场景数据上 4 种算法位置均方根误差统计结果

Table 2 The position RMSE of four different methods on real scenario data (m)

算法	VINS-Mono	VINS-Mono+ 直线特征	ORB-SLAM3	本文算法
均方根误差	7.929	4.446	6.101	1.982

文算法在所有时刻的绝对位置误差都小于其他 3 种方法,未产生负优化效果,证明了在室外场景下采用本文动

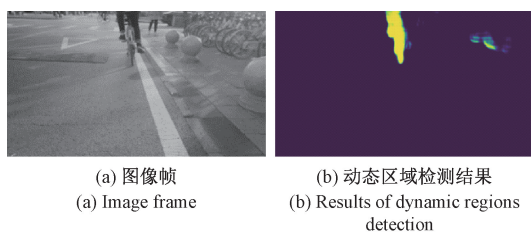


图 9 动态区域检测结果(无遮挡)

Fig. 9 Results of dynamic regions detection (occlusion free)

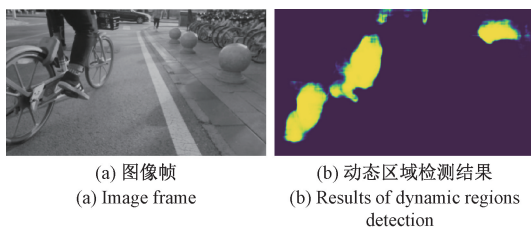
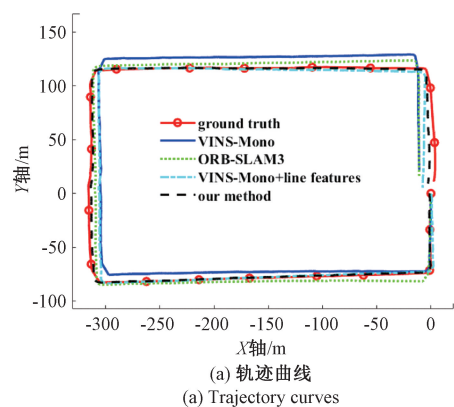
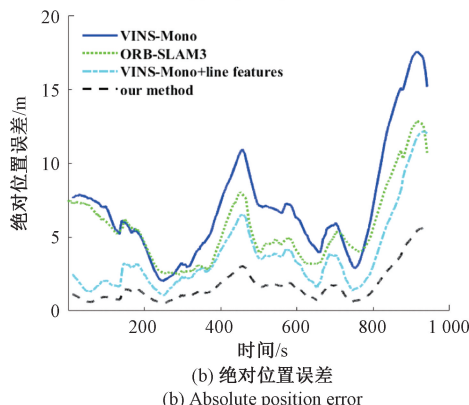


图 10 动态区域检测结果(部分遮挡)

Fig. 10 Results of dynamic regions detection (partial occlusion)



(a) 轨迹曲线



(b) 绝对位置误差

图 11 实际场景数据实验结果

Fig. 11 Experimental results on real scenario data

误差分别为 7.929、4.446、6.101 和 1.982 m,本文算法的定位精度明显优于其他 3 种方法。在室外车载导航测试实验中,上述 4 种方法的定位精度都低于 EuRoC 数据集上的定位精度,这是因为复杂室外环境给视觉/惯性导航系统的精度和稳定性带来了较大的挑战,而较长的运动轨迹带来了更大的积累误差。此外, EuRoC 数据集的传感器(相机和 IMU)精度要优于 MYNTEYE-D1000-50,这也是室外车载导航测试实验的定位精度低于 EuRoC 数据集精度的另一原因。即使在具有挑战性的环境中,本文算法仍可以有效检测到环境中的动态目标区域,剔除该区域内的动态特征,提供精确、有效的视觉观测,实现准确定位,这进一步说明本文算法对低成本传感器以及在复杂环境的适应性更强。实验中本文算法的单帧图像 GPU 平均处理时间为 0.091 s,输出频率可达 10 fps,且仅占用 2 GB 显存空间,说明该方法可应用于低算力的移动设备及嵌入式设备。

4 结 论

针对宽动态、长航时、大范围任务作业环境,文章提出了一种动态特征剔除的视觉/惯性导航方法。实验结果表明:

结合以结构相似性为成本量的端到端网络以及对称光流可以有效筛选动态特征与不稳定特征点;

动态特征剔除的视觉/惯性导航方法有效降低了动态环境对定位精度和稳定性的影响,较现有视觉/惯性导航方法定位精度更高,能够在卫星信号拒止环境下提供高精度导航定位结果;

该方法计算效率高、实时性强,可应用于移动设备与嵌入式设备,适用于低成本无人系统。

由于端到端网络基于结构相似性建立,可检测相同场景下环境的差异性,论文进一步工作将研究长航时无人系统重定位与回环检测,提高环境变化时重定位与回环检测的准确性。

参考文献

- [1] 陈丽, 陈洋, 杨艳华. 面向三维结构视觉检测的无人机覆盖路径规划[J]. 电子测量与仪器学报, 2023, 37(2): 1-10.
CHEN L, CHEN Y, YANG Y H. UAV coverage path planning for 3D structure visual inspection[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(2): 1-10.
- [2] 黄郑, 卞尹蕾, 王红星, 等. 无先验地图条件下电力管廊无人机自主巡检方法研究[J]. 电子测量技术, 2022, 45(19): 30-35.
HUANG ZH, BIAN Y L, WANG H X, et al. Inspection

态剔除网络对单帧“低质量”特征剔除的有效性。由表 2 可以看出, VINS-Mono、添加直线特征的 VINS-Mono、ORB-SLAM3 和本文算法在实际场景数据上位置均方根

- method employing UAV of electricity pipe gallery based on LiDAR without prior map [J]. *Electronic Measurement Technology*, 2022, 45(19): 30-35.
- [3] COOK K L. An evaluation of the effectiveness of low-cost UAVs and structure from motion for geomorphic change detection [J]. *Geomorphology*, 2017, 278 (1): 195-208.
- [4] 栾添添, 王皓, 尤波, 等. 狭窄空间的位姿辅助点 TEB 无人车导航方法[J]. *仪器仪表学报*, 2023, 44(4): 121-128.
- LUAN T T, WANG H, YOU B, et al. TEB unmanned vehicle navigation method for position and attitude auxiliary points in narrow space[J]. *Chinese Journal of Scientific Instrument*, 2023, 44(4): 121-128.
- [5] 张洪, 于源卓, 邱晓天. 基于改进关键帧选择的 ORB-SLAM2 算法[J]. *北京航空航天大学学报*, 2023, 49(1): 45-52.
- ZHANG H, YU Y ZH, QIU X T. ORB-SLAM2 algorithm based on improved key frame selection[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2023, 49(1): 45-52.
- [6] DUO J Y, ZHAO L, MAO J N. Sliding window based monocular SLAM using nonlinear optimization [C]. *Chinese Intelligent Systems Conference*, 2019: 519-529.
- [7] 谢波, 张国良, 李歆, 等. 一种单目 VIO 定位精度与跟踪稳定性优化方法[J]. *国外电子测量技术*, 2023, 42(4): 23-30.
- XIE B, ZHANG G L, LI X, et al. Optimization method for positioning accuracy and tracking stability of monocular VIO [J]. *Foreign Electronic Measurement Technology*, 2023, 42(4): 23-30.
- [8] QIN T, LI P L, SHEN S J. VINS-mono: A robust and versatile monocular visual-inertial state estimator [J]. *IEEE Transactions on Robotics*, 2018, 34 (4): 1004-1020.
- [9] CAMPOS C, ELVIRA R, JUAN J, et al. ORB-SLAM3: An accurate open-source library for visual, visual-inertial and multi-map SLAM [J]. *IEEE Transactions on Robotics*, 2021, 37(6): 1874-1890.
- [10] SHI X S, LI D J, ZHAO P P, et al. Are we ready for service robots? The OpenLORIS-scene datasets for lifelong slam [C]. *IEEE International Conference on Robotics and Automation*, 2020: 3139-3145.
- [11] MAITY S, SAHA A, BHOWMICK B. Edge SLAM: Edge points based monocular visual slam [C]. *IEEE International Conference on Computer Vision Workshops*, 2017: 2408-2417.
- [12] PUMAROLA A, VAKHITOV A, AGUDO A, et al. PL-SLAM: Real-time monocular visual SLAM with points and lines [C]. *IEEE International Conference on Robotics and Automation*, 2017: 4503-4508.
- [13] ZHAO H, JI X C, WEI D Y, et al. Online IMU-odometer extrinsic calibration based on visual-inertial-odometer fusion for ground vehicles [C]. *IEEE International Conference on Indoor Positioning and Indoor Navigation*, 2022: 1-8.
- [14] SHAN Z Y, LI R J, SCHWERTFEGER S. RGBD-inertial trajectory estimation and mapping for ground robots[J]. *Sensors*, 2019, 19(10): 2251.
- [15] DEM B L, NAKAZAWA K. Exploiting the accuracy-acceleration tradeoff: VINS-assisted real-time object detection on moving systems [C]. *IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 2019: 483-488.
- [16] XIAO L H, WANG J G, QIU X S, et al. Dynamic-SLAM: Semantic monocular visual localization and mapping based on deep learning in dynamic environment[J]. *Robotics and Autonomous Systems*, 2019, 117: 1-16.
- [17] HE K M, GKIOXARI G, PIOTR D, et al. Mask R-CNN[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2), 386-397.
- [18] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.
- [19] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi- direct monocular visual odometry[C]. *IEEE International Conference on Robotics and Automation*, 2014: 15-22.
- [20] GIOI R G V, JÉRÉMIE J, MOREL J M, et al. LSD: A fast line segment detector with a false detection control [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(4): 722-732.
- [21] ZHANG L L, KOCH R. An efficient and robust line segment matching approach based on LBD descriptor and pairwise geometric consistency [J]. *Journal of Visual Communication and Image Representation*, 2013, 24(7): 794-805.
- [22] FORSTER C, CARLONE L, DELLAERT F, et al. On-manifold preintegration for real-time visual-inertial Odometry[J]. *IEEE Transactions on Robotics*, 2017, 33(1): 1-21.
- [23] SIBLEY G, MATTHIES L, SUKHATME G. Sliding window filter with application to planetary landing [J]. *Journal of Field Robotics*, 2010, 27(5): 587-608.
- [24] BURRI M, NIKOLIC J, GOHL P, et al. The EuRoC

micro aerial vehicle datasets[J]. International Journal of Robotics Research, 2016, 35(10): 1157-1163.

作者简介



多靖贇, 2009 年于北京航空航天大学获得学士学位, 2013 年于华北电力大学获得硕士学位, 2018 年于北京航空航天大学获得博士学位, 现为北京物资学院讲师, 主要研究方向为计算机视觉导航与多传感器信息融合导航。

E-mail: duojingyun@bwu.edu.cn

Duo Jingyun received his B. Sc. degree from Beihang University in 2009, M. Sc. degree from North China Electric Power University in 2013, and Ph. D. degree from Beihang University in 2018, respectively. Now he is a lecturer in Beijing Wuzi University. His main research interests include computer

vision navigation and multi-sensor information fusion navigation.



李俊韬 (通信作者), 1999 年于吉林大学获得学士学位, 2002 年于吉林大学获得硕士学位, 2007 年于北京航空航天大学获得博士学位, 现为北京物资学院教授, 主要研究方向为多传感器信息融合导航及无人化物流。

E-mail: Ljtletter@126.com

Li Juntao (Corresponding author) received his B. Sc. degree from Jilin University in 1999, M. Sc. degree from Jilin University in 2002, and Ph. D. degree from BeiHang University in 2007, respectively. Now he is a professor in Beijing Wuzi University. His main research interests include multi-sensor information fusion navigation and unmanned logistics.