

融合 RTK 约束的视觉路标构建方法

罗康 许舒晨 赵科东 孙永荣

(南京航空航天大学自动化学院 南京 211100)

摘要:在卫星信号受阻或拒止的复杂环境中,仅依赖全球导航卫星系统(global navigation satellite system, GNSS)的车载定位系统往往难以获得稳定且高精度的位置估计,视觉路标因此成为提升车辆自主导航鲁棒性的关键辅助信息。然而,传统视觉地图构建方法普遍存在建图耗时长、依赖连续跟踪、对光照与动态场景敏感等问题,导致路标精度与一致性不足。针对这些问题,提出一种融合 RTK(real-time kinematic)约束的高精度视觉路标构建方法。首先,基于车辆行驶航向角序列检测近直路段,并采用线性插值策略从中选取具有代表性的图像帧作为视觉路标,构建语义先验以增强路标的长期稳定性与可重识别性;随后,以路标帧为中心生成邻域图像集,结合 RTK 提供的全局位置信息对初始位姿进行校正,提升局部地图重建的初始化质量与鲁棒性;进一步在局部稀疏建图与全局优化中引入自适应 RTK 约束,实现尺度一致、地理对齐的高精度路标构建,并在 RTK 信号受干扰或短时失效时触发退化平滑约束以保持系统稳定。实验结果表明,相较传统方法,所提方法在地图精度、重建时间与跨季节、跨光照以及动态场景下的定位鲁棒性方面均显著提升,建图精度提高 82.2%,耗时降低 81.3%,定位精度最高提升 70.1%。

关键词: RTK 辅助建图;视觉路标;地图重建;视觉导航;数字地图匹配导航

中图分类号: TP242.6; TP391.4

文献标识码: A

国家标准学科分类代码: 510.8040

Visual roadmark construction method incorporating RTK constraints

Luo Kang Xu Shuchen Zhao Kedong Sun Yongrong

(College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211100, China)

Abstract: In complex environments where satellite signals are obstructed or denied, vehicle localization systems that rely solely on GNSS often fail to provide stable and high-precision position estimates. Visual landmarks therefore serve as essential auxiliary information for enhancing the robustness of autonomous navigation. However, traditional visual map construction methods typically suffer from long processing times, dependence on continuous tracking, and strong sensitivity to illumination variations and dynamic objects, resulting in insufficient landmark accuracy and spatial consistency. To address these limitations, this paper proposes a visual landmark construction method incorporating RTK constraints. The method first detects near-straight road segments based on the vehicle's heading-angle sequence and selects representative image frames via linear interpolation as candidate visual landmarks, while introducing semantic priors to enhance their long-term stability and re-identifiability. Subsequently, a neighborhood image set is constructed around each landmark frame, and the initial camera poses are corrected using globally referenced RTK measurements to improve the robustness and reliability of local map initialization. Furthermore, an adaptive RTK constraint is integrated into both local sparse reconstruction and global optimization, enabling scale-consistent and geographically aligned high-precision landmark construction. A degraded smoothing constraint is activated when RTK signals are disturbed or temporarily unavailable to maintain system stability. Experimental results demonstrate that, compared with conventional approaches, the proposed method achieves notable improvements in map accuracy, reconstruction efficiency, and localization robustness under varying illumination, seasonal changes, and dynamic scenes, yielding an 82.2% increase in mapping accuracy, an 81.3% reduction in processing time, and up to a 70.1% improvement in localization precision.

Keywords: RTK-assisted map construction; visual landmarks; map reconstruction; visual navigation; digital map matching navigation

0 引言

在自动驾驶场景中,高精度车载定位技术是保障系统安全决策与路径规划的关键支撑。目前,全球导航卫星系统(global navigation satellite system, GNSS)被广泛用于车辆的全局定位。然而,在复杂城市街区、隧道等卫星信号遮挡或拒止环境中,卫星导航定位精度与可用性严重下降,难以满足自动驾驶对定位鲁棒性的要求^[1-2]。

视觉同时定位与建图(visual simultaneous localization and mapping, VSLAM)技术可通过图像序列构建环境地图,并估计载体自身位姿,在卫星拒止情况下可独立运行^[3-4]。然而,VSLAM 本质上基于里程计航位推算,长时间运行易产生累积误差导致定位精度下降^[5]。同时,其依赖连续帧间特征匹配,难以保持地理全局一致性,且在光照变化、遮挡或场景重复时易跟踪失败^[6]。

为解决上述问题,基于视觉地图的重定位技术被引入视觉导航体系。该类技术通过将车载相机图像与预构建的视觉地图进行匹配,实现对 VSLAM 系统漂移的全局约束与修正,从而有效提升系统长航时定位精度^[7-8]。薛杨等^[9]将高精度卫星数据引入 VSLAM 系统,构建密集视觉地图,在短时定位任务中取得良好效果;薛源等^[10]利用 RTK(real-time kinematic)辅助实现 VSLAM 跟踪失败后的地图恢复与融合,提升了建图过程鲁棒性与完整性;Jiang 等^[11]提出一种融合惯性姿态信息与 Wiener 去模糊的改进型 ORB-SLAM 算法,有效缓解运动模糊对定位精度的影响;然而,上述方法多聚焦于提升传统 SLAM 的跟踪鲁棒性与建图效率,虽具实时性优势,但地图精度受限,且缺乏面向任务或定位需求驱动的构图机制。Schönberger 等^[12]提出一种改进的增量式 SfM(structure-from-motion)方法,显著提升了传统 SfM 的鲁棒性与精度;Jia 等^[13]提出一种平面标记辅助的全局 SfM 方法,通过旋转一致性检查解决姿态歧义问题,优化三维重建精度;Pan 等^[14]提出 GLOMAP 方法,兼顾全局 SfM 的效率与增量 SfM 的精度与鲁棒性;上述基于 SfM 的视觉建图方法通常在全局或大范围图像集上执行重建,计算量大

且缺乏实时性,同时传统 SfM 框架未考虑高精度 GNSS 信号的约束,重建精度易受累积误差影响,地图缺乏真实地理尺度以及几何信息。

针对上述问题,本文提出一种融合 RTK 约束的视觉路标构建方法。该方法首先通过视觉 SLAM 系统的位姿序列检测近直路段图像帧,并基于插值策略选取具代表性的图像帧作为视觉路标帧;随后,以路标帧为窗口中心构建邻域图像集,基于该图像集进行局部地图重建,并融合 RTK 约束校正图像帧初始位姿,确保建图初始化的鲁棒性;最后,在建图优化过程中,添加全局 RTK 约束进一步提高地图的精度。通过该路标构建方法,本文有效提升了建图精度和实时性,并且利用实际导航场景实验验证了路标对于定位精度以及鲁棒性的提升。

1 视觉路标构建系统设计

1.1 视觉路标构建系统框架

本文的系统框架如图 1 所示,主要包括预处理、视觉路标点选取、路标邻域图像帧位姿校正模块以及路标邻域局部地图重建模块。系统以车载相机采集的图像序列及其对应时刻的 RTK 定位信息为输入。首先,预处理模块提取图像中的 SuperPoint^[15] 局部特征与 NetVLAD^[16] 全局描述符,用于后续的路标数据库的构建;同时,采用 ORB-SLAM2^[17] 的前端模块恢复图像帧位姿。视觉路标点选取模块基于 SLAM 系统输出的航向角序列,选取车辆处于近直路段的图像帧,并通过等距插值策略提取目标视觉路标图像帧。随后,路标邻域图像集位姿校正模块结合 RTK 提供的全局定位信息,对视觉路标图像帧及其邻域帧的递推位姿进行姿态修正与坐标系转换,以统一至地理坐标系下。在此基础上,路标邻域局部地图重建模块利用提取的视觉特征与校正后的位姿,对视觉路标邻域帧进行局部地图重建,并在全局 BA(bundle adjustment)优化过程中引入 RTK 约束,以提升重建精度。最终,将视觉路标帧所观测到的地图点、局部特征、全局描述符及其观测关联关系共同存入视觉路标数据库,供后续视觉定位使用。

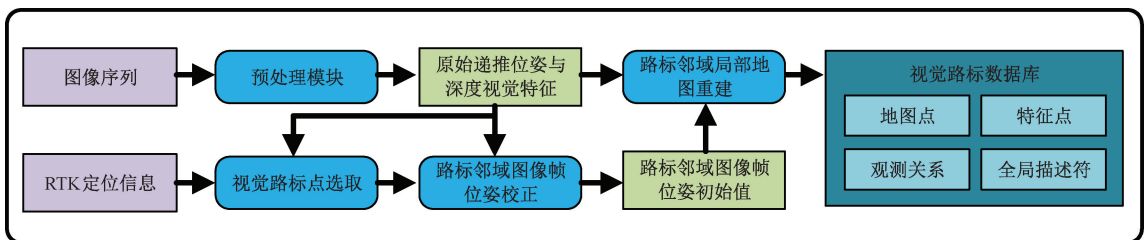


图 1 系统框架示意图

Fig. 1 System framework

1.2 视觉路标建模

为使视觉路标更为高效紧凑,本文对视觉路标数据单元进行了建模,并在路标邻域地图重建后对数据元素进行精简,仅保留与定位强相关的关键元素。具体而言,对于每一个被筛选为视觉路标的图像帧 I_k ,仅保留以下信息。

1) Superpoint 局部特征点集合

$$F_k = \{(\mathbf{u}_i, \mathbf{d}_i) \mid i = 1, 2, \dots, N_k\} \quad (1)$$

其中, $\mathbf{u}_i \in \mathbf{R}^2$ 为图像坐标位置, $\mathbf{d}_i \in \mathbf{R}^{D_i}$ 为对应的局部特征描述符,通常 $D_i = 256, N_k$ 为图像帧 I_k 包含的特征点数量。

2) NetVLAD 全局图像描述符

$$\mathbf{g}_k \in \mathbf{R}^{D_g} \quad (2)$$

表征整张图像的全局外观信息,用于实现图像间粗匹配和重定位检索,其中, $D_g = 4\ 096$ 。

3) 可观测三维地图点集

$$M_k = \{\mathbf{X}_j \in \mathbf{R}^3\} \quad (3)$$

包括所有能被图像帧 I_k 成功观测到的稀疏三维地图点,这些地图点通常对应于重建过程中成功匹配的图像特征,且能够满足三角化条件。

4) 地图点观测关系

将图像帧 I_k 与其可观测地图点 M_k 之间的观测关系存储为索引列表,用于后续的优化与数据关联。

最终保留的视觉路标单元可形式化表示为:

$$V_k = (I_k, F_k, \mathbf{g}_k, M_k, O_k) \quad (4)$$

其中, O_k 代表地图点观测关系。在多个路标邻域地图构建完成后,仅保留由筛选算法选出的视觉路标帧集合 $\{V_k\}_{k=1}^K$,并剔除冗余图像帧与不被任何路标观测到的孤立地图点,从而形成轻量级、高可用性的视觉路标数据库。

2 视觉路标构建方法

与现有方法通常先对完整图像序列执行全局三维重建,再从中筛选定位关键帧不同,本文提出一种以路标选择为驱动的局部地图构建策略。如图 2 所示,该策略优先选取具备代表性的图像帧作为视觉路标候选,随后在其时间邻域内提取有限图像帧构成局部图像窗口,结合 RTK 高精度位置信息执行局部稀疏重建,仅保留目标帧可观测的稀疏三维点用于构建视觉路标数据库。

2.1 视觉路标图像帧选取

考虑到车辆在直线路段上的航向角变化较小,且对环境的观测更具一致性,该类路段适合作为视觉路标插入区域。为此,本文基于视觉 SLAM 系统输出的位姿序列构造滑动窗口检测直线路段,并通过线性插值策略实

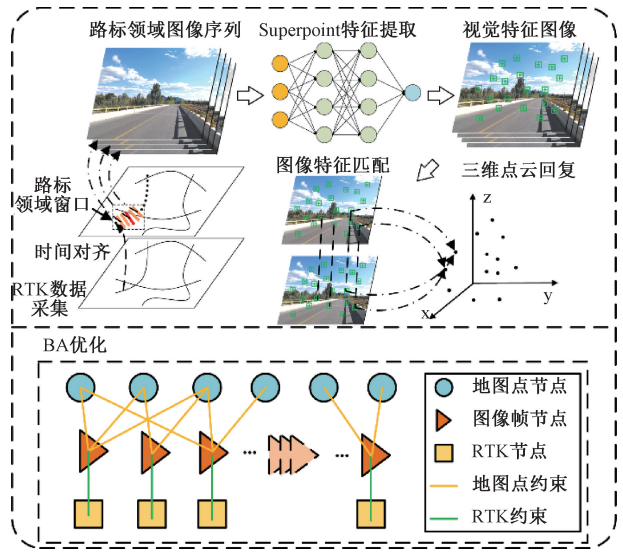


图 2 视觉路标构建策略

Fig. 2 Visual landmark construction strategy

现对视觉路标图像帧的选取。

记视觉 SLAM 系统在 t 时刻输出的位姿为 T_t ,相机与机体系之间的固联外参为 T_c^b ,则车辆的位姿可表示为:

$${}^b T_t = [{}^b R_t \mid {}^b t_t] = T_c^b T_t (T_c^b)^{-1} \quad (5)$$

利用该位姿可计算车辆此时相较于初始时刻机体系的航向,记为 θ_t 。同时刻航向变化值可计算为:

$$\Delta\theta_t = |\theta_t - \theta_{t-1}| \quad (6)$$

定义长度为 M 的滑动窗口,设定航向阈值 ε ,若窗口内的航向变化值满足以下条件:

$$\forall k \in \{t - M + 1, \dots, t\}, \theta_k < \varepsilon \quad (7)$$

则认为车辆此时行驶于近似直线路段,随后顺序移动窗口,将后续图像帧记录在同一近直路段图像集合中。当窗口内的图像帧均不满足上述条件时,认为车辆开始转向,并停止向当前近直路段图像集添加新的图像帧,同时开始下一个近直路段图像集的检测。记某个近直路段图像集的索引列表为 $\mathcal{F} = \{t_s, t_s + 1, \dots, t_e\}$,对应起止帧 t_s 与 t_e 的 RTK 坐标分别为 p_s 与 p_e ,则这些图像帧所处路段长度可估算为:

$$L = \|p_e - p_s\| \quad (8)$$

设定路标距离间隔阈值 d ,根据线性插值方法选取插入路标点的位置为:

$$p_i = p_s + \frac{i \cdot d}{L} (p_e - p_s), i = 1, 2, \dots, \lfloor \frac{L}{d} \rfloor \quad (9)$$

为将插值点 p_i 与图像帧关联,设 P_t 为图像帧 $t \in \mathcal{F}$ 对应的 RTK 坐标,则插值点对应的视觉路标帧为:

$$t_i^* = \operatorname{argmin}_{t \in \mathcal{F}} \|P_t - p_i\| \quad (10)$$

最终, t_i^* 对应的图像帧即作为该直线路段的视觉路

标帧。

为进一步提升视觉路标在长期导航场景中的可重识别性,在路标图像帧选取过程中引入语义先验信息,对候选路标帧进行语义稳定性评估,从而增强路标的判别性与长期一致性。

采用文献[18-19]方法对语义数据集进行标注,基于数据集对文献[18]提出的轻量级语义分割网络进行训练,随后对图像进行逐帧语义解析,获得像素级语义标签图,最终数据集语义分割结果示例如图3所示。考虑到不同语义类别在跨季节、跨光照条件下的稳定性具有显著差异,本文构建语义稳定性得分函数如下:

$$S_{\text{sem}} = \sum_{c \in \mathcal{C}} \omega_c \cdot r_c \quad (11)$$

其中, c 表示语义类别, r_c 为该类别在图像中的像素占比, ω_c 为其语义稳定性权重。对于跨域变化较小、形状一致性强的类别,给予较高权重;对于随季节和光照变化显著的类别,给予较低权重。本文将路标判别相关语义类别设置为建筑、植被、天空、道路,其权重分别设置为 0.55、0.1、0.01 和 0.34。



图3 语义分割示意图

Fig. 3 Illustration of semantic segmentation

在完成近直路段检测与插值候选点生成后,本文对候选路标帧依据语义得分进行二次筛选。

$$I_k^* = \operatorname{argmax}_{I_k \in L} S_{\text{sem}}(I_k) \quad (12)$$

其中, L 为与插值位置相邻的候选帧集合。该策略能够优先选择语义结构稳定且几何特征丰富的关键帧作为路标帧,从而提高路标的长期可复用性与重定位成功率。

此外,语义先验不仅能够用于路标帧的选取,还可在路标邻域地图构建过程中辅助剔除来自动态的不稳定特征点。因此,在构建视觉路标的局部地图时,在特征提取后利用语义分割结果对特征点进行语义过滤。将动态剔除相关语义类别设置为车辆、行人、植被、骑行者,在提取特征点后对属于动态语义区域的特征点进行剔除处理,从而避免动态目标上的特征进入三角化与优化流程,剔除效果如图4所示。

2.2 路标邻域图像帧位姿校正

为重建视觉路标帧所观测的三维点,选取每个路标帧前后各 K 帧图像组成邻域窗口,并构建对应局部图像集用于局部建图。由于小规模图像集难以通过传统 SfM

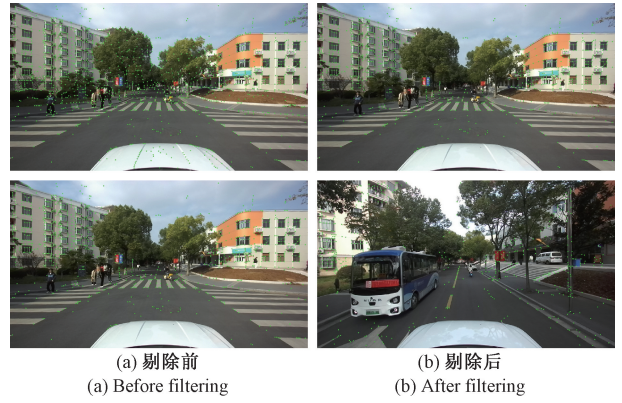


图4 动态特征语义剔除

Fig. 4 Semantic removal of dynamic features

完成初始化,通过视觉 SLAM 前端估算图像帧位姿,并引入 RTK 坐标作为地理先验,构造优化问题校正图像帧位姿,并估计图像帧在地理系下的粗略位姿。

为获得车辆在 ENU 坐标系下的初始姿态,利用连线平行于车辆中轴的双 RTK 天线测量结果进行估计。记前、后 RTK 天线在初始时刻的 ENU 坐标分别为 $\mathbf{p}_f = (x_f, y_f, z_f)^T$, $\mathbf{p}_r = (x_r, y_r, z_r)^T$, 则车辆的航向角与俯仰角可分别估计为:

$$\psi = \arctan(y_f - y_r, x_f - x_r) \quad (13)$$

$$\theta = \arctan(z_f - z_r, \sqrt{(y_f - y_r)^2 + (x_f - x_r)^2}) \quad (14)$$

车辆静止时姿态稳定,横滚角近似为零,车辆在 ENU 系下的初始粗略位姿 $\mathbf{T}_b^{\#g}$ 可由上述姿态角转化求得。

对于路标帧 I_i , 其相邻图像集为 $W_i = \{I_{i-K}, \dots, I_i, \dots, I_{i+K}\}$, 对应的 SLAM 参考系下初始位姿为 $\{T_i \in SE(3)\}$, 采集时刻对应的 ENU 坐标为 $\{t_i\}$ 。以车体初始姿态 $\mathbf{T}_b^{\#g}$ 为初始条件,结合相机与机体系外参 \mathbf{T}_c^b , 可构造如下优化问题求解地理坐标系下的转换关系。

$$\mathbf{T}_b^{*g} = \operatorname{argmin}_{T_b^{\#g}} \sum_{i=1}^{2K+1} \| \operatorname{trans}(\mathbf{T}_b^{\#g} \mathbf{T}_c^b \mathbf{T}_i (\mathbf{T}_c^b)^{-1}) - t_i \| \quad (15)$$

其中, $\operatorname{trans}(\cdot)$ 代表提取位姿矩阵的平移向量。据此,窗口内所有图像在地理系下的粗略初始位姿可进一步计算为:

$$\mathbf{T}_i^* = \mathbf{T}_b^{*g} \mathbf{T}_c^b \mathbf{T}_i \quad (16)$$

上述估计结果作为后续稀疏重建的初始条件,用于增强地图重建的鲁棒性。

2.3 路标局部建图与 RTK 约束优化

基于 2.2 节估计的图像初始位姿,对每个视觉路标帧及其相邻图像帧构建局部图像集,进行稀疏三维建图。图像间匹配、初始三角化与位姿估计采用传统 SfM 流程实现,在此基础上,为增强重建的尺度一致性与地理对齐精度,引入 RTK 定位数据作为平移约束参与全局 BA

优化。

记 BA 过程中每帧相机的位姿为 $T_i = [R_i | t_i] \in SE(3)$, 三维地图点坐标为 $X_j \in R^3$, 则重投影误差可描述为:

$$e_{proj} = \sum_{i,j} \rho(\|u_{ij} - \pi(R_i X_j + t_i)\|) \quad (17)$$

其中, u_{ij} 为图像中第 i 帧对第 j 个点的观测, $\pi(\cdot)$ 为相机投影函数, $\rho(\cdot)$ 是优化过程中的鲁棒核函数。将 ENU 系下的 RTK 观测 P_i 引入优化目标函数, 对每一帧相机的平移向量 t_i 加入 RTK 先验项, 则相机平移部分的误差可描述为:

$$e_{RTK} = \sum_i \rho(\|t_i - P_i\|^2) \quad (18)$$

加入 RTK 约束项后全局 BA 优化问题可描述为:

$$\min_{R_i, t_i, X_j} \sum_{i,j} \rho(\|u_{ij} - \pi(R_i X_j + t_i)\|) + \sum_i \lambda_i \rho(\|t_i - P_i\|^2) \quad (19)$$

其中, λ 为权重系数, 用于平衡 RTK 约束与视觉重投影约束的影响, 基于 Ceres 框架求解该优化问题。最终, 仅储存视觉路标图像帧所观测的三维地图点用于视觉定位。

为增强系统在复杂环境中的自适应性, 引入基于不确定性估计的自适应权重策略, 使 λ 能随 RTK 定位质量动态调整。记 RTK 接收机输出的三维位置标准差分别为 σ_x 、 σ_y 以及 σ_z , 定义 RTK 的不确定性度量为:

$$\sigma_{RTK} = \sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2} \quad (20)$$

构造 RTK 权重的自适应表达式为:

$$\lambda_i = \frac{1}{\sigma_{RTK,i}^2 + \varepsilon} \quad (21)$$

其中, ε 为防止分母趋近于零的常数。由此可知, 当 RTK 信号稳定, 误差较小时, λ 增大, 使优化过程更多依赖高可信度的 RTK 数据; 当 RTK 信号受遮挡或精度下降时, λ 自动减小, 避免受不可靠先验影响。

2.4 RTK 信号失效情况的退化处理策略

在实际道路环境中, RTK 信号可能受到遮挡、反射或多路径影响, 导致定位结果出现跳变或短时失效。若在此情况下仍将其作为强约束加入优化, 将可能引入不一致的先验信息, 影响局部建图的精度与稳定性。本文设计了 RTK 失效判别策略与退化约束切换的鲁棒处理机制, 使建图过程在 RTK 部分失效的情况下仍能保持稳定。

1) 失效判别策略

对于每帧 RTK 数据, 利用接收机输出的定位协方差、固定解状态分析其定位质量, 当接收机信息满足以下任一条件时认为 RTK 短时失效: 固定解状态非“NARROW_INT”, 即窄巷模糊度未固定为整数, 定位精度为非厘米级; 任意一个维度标准差超过阈值, 即满

足式(22)。

$$\max(\sigma_x, \sigma_y, \sigma_z) > \tau \quad (22)$$

其中, τ 为标准差阈值, 本文设定为 0.05。

2) 退化约束切换

为保持局部地图结构的一致性, 当 RTK 信息低可信度或失效时, 将式(19)中的 RTK 约束项替换为弱约束即:

$$e_{RTK}^* = \sum_i \eta \rho(\|t_i - t_{i-1}\|^2) \quad (23)$$

其中, $\eta \ll \lambda$ 为弱平滑因子, 该项仅对相邻帧位姿提供连续性约束, 不强制约对齐地理坐标, 本文设定为 0.01。

3 实验结果与分析

3.1 实验平台与数据集

为了验证本文算法的有效性, 本文使用车载实验平台采集图像以及 RTK 定位数据进行实验验证。如图 5 所示, 平台所搭载的传感器为一台 Zed2 双目相机、两台 NovAtel OEM7 RTK 接收机, 其中一台 RTK 设备用于采集定位真值, 另一台仅用于辅助测量初始航向角与俯仰角。实验运行算法所使用的计算机 CPU 型号为 Intel Core i9-13900HX, 内存大小为 64 GB, 该计算机安装的操作系统为 Ubuntu18.04。

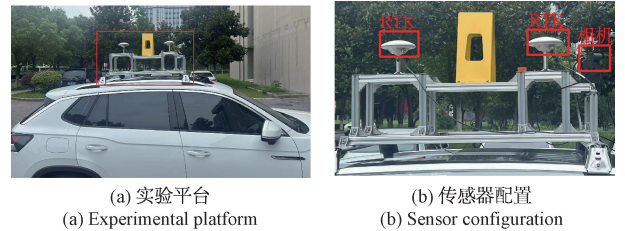


图5 车载实验平台示意图

Fig. 5 Illustration of the on-board experimental platform

实验在某校园内东西两个校区进行, 采集的数据集基本情况如表 1 所示。由表 1 可知, 道路数据采集时间分布在 4 个不同的季节以及 2 个行驶区域, 同时部分数据采集于一天中的不同时段, 可用于验证算法对季节、光照变化的鲁棒程度, 数据集典型场景变化如图 6 所示。在此基础上, 为进一步评估算法在动态环境中的鲁棒性, 本文从原始采集序列中选取了序列 11、12 两组包含明显动态干扰的典型场景。这两组数据均采集于校园交通高峰期, 场景中可能存在较高密度的行人、车辆及骑行者等动态障碍物, 其运动会造成部分静态背景的遮挡以及大量非稳定特征点的引入。该部分数据被专门用于验证本文方法在动态障碍物干扰条件下的视觉重定位稳定性与导航鲁棒性。

表 1 数据集基本情况

Table 1 Overview of the dataset

序列	采集日期	行驶长度/m	时段	行驶区域
01	2024-04-16	1 587.97	上午	东区
02	2024-04-16	1 358.26	上午	西区
03	2024-07-17	1 612.22	下午	东区
04	2024-07-17	1 196.88	下午	西区
05	2024-11-21	1 162.50	下午	东区
06	2024-11-21	1 179.39	傍晚	西区
07	2025-01-14	1 185.78	傍晚	西区
08	2025-01-14	1 184.22	下午	西区
09	2025-01-14	1 542.11	下午	东区
10	2025-01-14	1 576.04	傍晚	东区
11	2024-07-17	1 600.62	下午	东区
12	2025-01-14	1 200.86	下午	西区



图 6 数据集场景示意图

Fig. 6 Illustration of dataset scenarios

表 2 视觉路标构建精度与耗时对比

Table 2 Comparison of visual landmark construction accuracy and time consumption

方法	08			09		
	平均误差/m	最大误差/m	构图耗时/min	平均误差/m	最大误差/m	构图耗时/min
本文	0.029	0.564	12	0.049	0.236	7
文献[11]	0.534	2.321	8	0.416	2.174	4
文献[14]	0.027	0.583	60	0.043	0.217	42

表 3 路标初始化成功数量与成功重建精度对比

Table 3 Comparison of landmark initialization success count and reconstruction accuracy

序列	目标路标数量	本文		文献[9]	
		成功数量	平均误差/m	成功数量	平均精度/m
08	31	31	0.027	7	0.053
09	19	19	0.041	5	0.067

3.2 视觉路标构建实验分析

本实验首先验证视觉路标构建方法在建图时长与鲁棒性方面的优势,使用本文方法以及文献[11,14]方法分别构建视觉路标。文献[11]为先进的 SLAM 方法,而文献[14]为先进的 SfM 方法,本实验使用上述方法对原始数据进行整体稀疏重建,再从其中挑选与本文路标相同的图像帧进行精度对比以及整体重建时间的对比。滑动窗口长度 M 、航向阈值 ε 以及距离间隔阈值 d 分别设置为 10、0.4 以及 50。

选取序列 08、09 分别构建西区、东区的区域视觉路标,统计路标的平均二维位置误差以及路标构建全流程的总时间,以路标二维位置误差评估建图精度,结果如表 2 所示。由表 2 可知,本文方法与文献[14]方法构建的视觉路标都达到了厘米级的精度,且两种方法的精度相差无几,但相较于文献[14]的方法,本文方法在路标构建耗时缩减了 81.37%,相较于传统整体高精度重建的方法具有较好的实时性。同时相较于文献[11]的方法,即使本文在构建路标的实时性上有略微的劣势,但在平均误差与最大误差指标上分别达到 91.78% 以及 82.21% 的提升。

接着验证本文算法在局部地图构建任务上的鲁棒性,该部分实验使用与本文方法相同流程构建路标,即先筛选路标邻域图像帧,再进行局部地图重建,使用路标局部地图重建成功数量评估鲁棒性,局部地图初始化成功数量与成功重建的路标精度对比结果如表 3 所示。由结果可知,传统 SfM 方法在极小规模图像集重建任务上的鲁棒性较差,而本文由于引入了路标邻域图像帧位姿校正方法,为重建过程提供了较好的图像位姿初值,使得建图过程具有较好的鲁棒性。

3.3 RTK 失效实验分析

为验证所提出的 RTK 自适应权重与退化策略,在不改变原有轨迹结构的前提下,构造 RTK 干扰以及 RTK 丢失两种典型异常场景。记原始 RTK 数据在时刻 t_k 的三维位置与标准差如式(24)。

$$\mathbf{p}_k = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix}, \boldsymbol{\sigma}_k = \begin{bmatrix} \sigma_{x,k} \\ \sigma_{y,k} \\ \sigma_{z,k} \end{bmatrix} \quad (24)$$

1) RTK 干扰。在整个轨迹 30% 的时间区间内,向每个维度的 RTK 数据添加偏置与噪声,并增大其标准差。扰动后的位置可定义为:

$$\tilde{\mathbf{p}}_k = \mathbf{p}_k + \mathbf{b}_k + \mathbf{n}_k \quad (25)$$

其中, \mathbf{b}_k 为周期性偏置项,用于模拟多路径反射现象,可表示为:

$$\mathbf{b}_k = [A \sin(2\pi f t_k) \quad 0 \quad 0]^T \quad (26)$$

其中,偏置幅值 A 取 0.5 m,频率 f 取 0.1 Hz;随机噪声项 \mathbf{n}_k 用于模拟 RTK 定位精度下降,可表示为:

$$\mathbf{n}_k \sim \mathcal{N}(\mathbf{0}, \text{diag}(\sigma_n^2, \sigma_n^2, \sigma_n^2)) \quad (27)$$

其中, σ_n 取值为 0.3。为了反映该段 RTK 输出可靠性下降,将 3 个方向的定位标准差增加为:

$$\tilde{\sigma}_{d,k} = \max(\sigma_d, \sigma_{i,k}), i = x, y, z \quad (28)$$

其中, σ_d 设定为 0.04。

2) RTK 丢失。随机将整个轨迹 30% 的时间区间内的 RTK 定位解状态设定为“NARROW_FLOAT”,即载波相位的整周模糊度没有被固定为整数。依据所提出的失效判别策略,该段 RTK 定位数据视作失效。

基于上述模拟场景,在序列 08 上开展本文方法与文献[11,14]整体稀疏地图重建的精度对比实验。重复 5 次实验,记录每次实验的最大误差与平均误差并取平均值,结果如表 4 所示。由结果可知,本文方法相较于 RTK 完整的情况下有部分微小精度损失,但整体的视觉路标构建精度仍然可保持在亚米级,可见所提出的退化策略在上述情况中维持了路标构建过程的鲁棒性;文献[11]在 RTK 失效的情况下路标构建精度出现了较为严重的下滑,这是由于其地图稀疏重建依赖于完整的 RTK 约束,仅依靠视觉特征点的约束效力较弱,未能维持路标构建过程的鲁棒性;文献[14]凭借其精细化的重建过程,在上述模拟场景中能够维持一定路标构建精度,但总体精度仍劣于本文算法。

表 4 RTK 失效情况下的视觉路标构建精度对比

Table 4 Comparison of visual landmark construction accuracy under RTK failure

方法	RTK 干扰		RTK 丢失	
	平均误差/m	最大误差/m	平均误差/m	最大误差/m
本文	0.056	1.189	0.137	0.761
文献[11]	2.738	5.129	1.724	4.376
文献[14]	1.081	2.274	0.423	0.878

3.4 视觉导航定位算法实验分析

实验选用光照变化、季节变化、动态障碍 3 个典型的变化场景验证本文视觉路标在真实导航场景下的有效性。实验选取序列 08、09 分别构建西区、东区的区域视觉路标,使用其余对应区域的数据作为定位数据集验证算法定位效果,与代表传统手工特征法的文献[11]进行对比。实验中,本文使用文献[20]的重定位算法。当重定位成功时,车辆的位姿即修正为重定位的结果,而未发生重定位时刻的位姿则由 ORB-SLAM2 中的视觉里程计递推。

1) 光照变化实验

首先针对场景光照变化进行实验,选取序列 01、02、07、10 进行视觉导航定位实验,其中序列 01、02 采集时间为上午,而序列 07、10 采集时间为傍晚,与建图序列 08、09 所处时段有明显的光照差异。实验结果如图 7 以及表 5 所示,由图 7 可以看出本文方法在场景光照变化时,仍然具有较好的定位精度,相较于手工特征法,本文方法定位轨迹与轨迹真值更为重合。同时由于视觉路标的修正作用,本文方法的定位误差能够一直维持在较低水平,而文献[11]方法重定位点数量较少,定位鲁棒性较

差,部分路标未能成功修正定位结果,使得其误差发散较为严重。由表 5 可知,本文方法平均定位误差与最大误差均小于文献[11]方法,相较于文献[11]方法定位精度提升了 70.08%,同时重定位成功的点位数量约为文献[11]方法的 2.4 倍。

2) 季节变化实验

接着针对场景季节变化进行实验,选取序列 03、04、05、06 进行视觉导航定位实验,其中序列 03、04 采集时间为 7 月份,而序列 05、06 采集时间为 11 月份,与序列 08、09 采集时间所处季节不同。实验结果如图 8 以及表 6 所示,与光照变化实验结果相同,本文方法在场景发生季节变化时,仍然具有较好的定位精度。同时定位误差能够一直维持在较低水平,视觉定位鲁棒性较好。相较于文献[11]方法,本文方法定位精度提升了 83.62%,同时重定位成功的点位数量约为文献[11]方法的 2.4 倍。

3) 动态障碍实验

最后针对动态物干扰的典型场景进行实验,选取序列 11、12 进行视觉导航定位实验。实验结果如图 9 与表 7 所示。在密集动态障碍物干扰条件下,文献[11]方法的重定位成功帧数显著下降,导致定位结果在较长时间

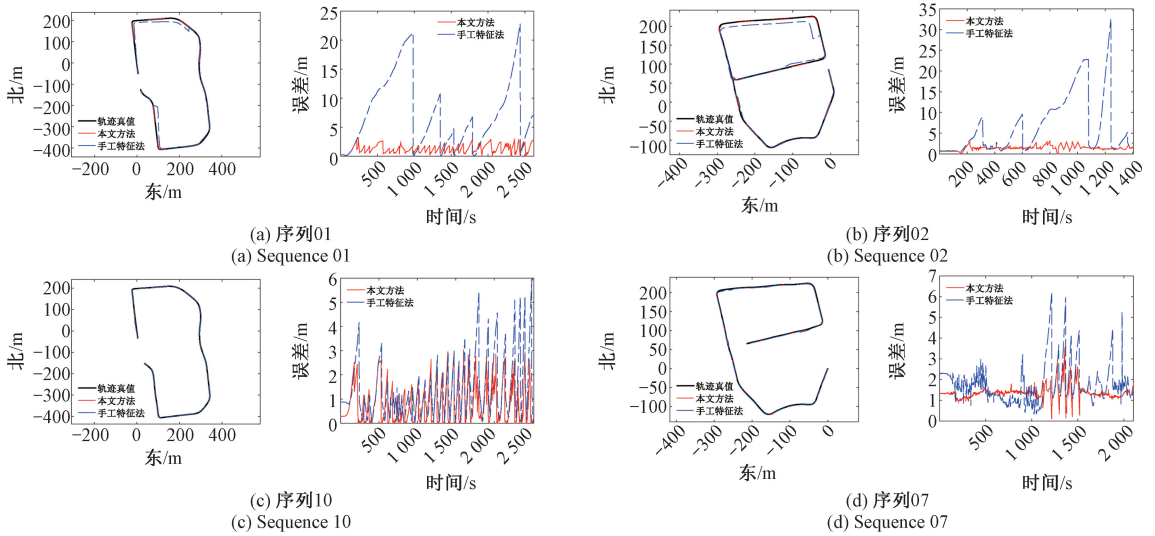


图 7 光照变化实验定位轨迹与误差误差曲线

Fig. 7 Localization trajectories and error curves under illumination variations

表 5 光照变化实验定位数据

Table 5 Localization data under illumination variation experiments

序列	本文			文献[11]		
	平均误差/m	最大误差/m	重定位点数	平均误差/m	最大误差/m	重定位点数
01	1.626	7.334	356	7.777	25.183	16
10	0.828	3.037	846	1.361	5.977	401
02	1.418	3.152	585	6.598	32.626	31
07	1.369	3.562	1687	1.778	6.227	990

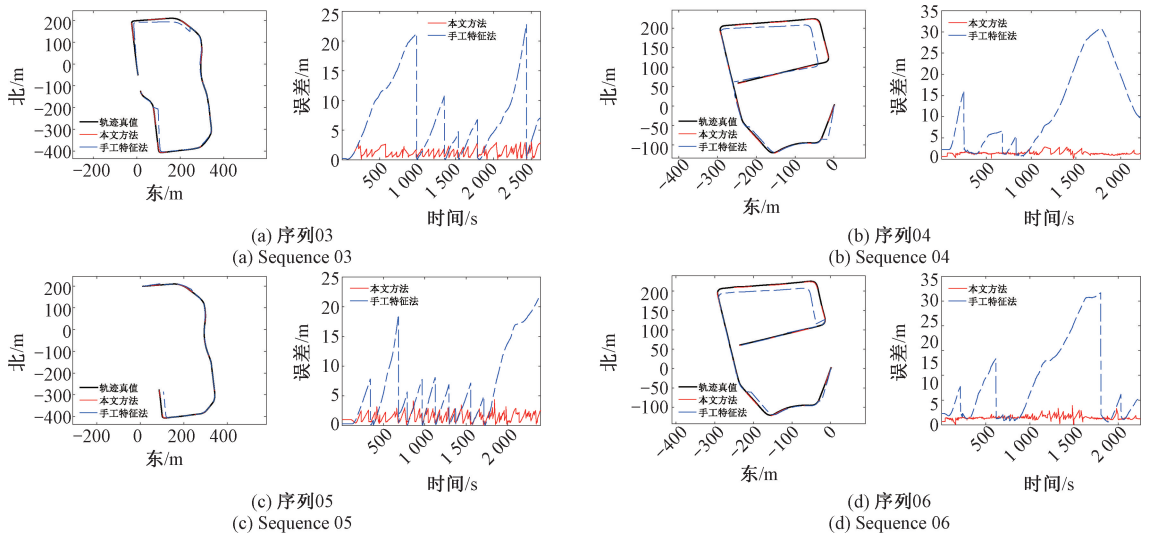


图 8 季节变化实验定位轨迹与误差误差曲线

Fig. 8 Localization trajectories and error curves under seasonal variations

内得不到有效修正, 轨迹漂移明显, 最大定位误差出现大幅增大。相比之下, 尽管本文方法在该场景中也受到一定影响, 表现为定位精度略有下降、重定位成功帧数有所减少, 但整体仍能保持米级精度, 定位轨

迹稳定且无明显发散, 表明所构建的视觉路标对动态目标具有良好的鲁棒性。定量结果显示, 相较于文献[11]方法, 本文方法的平均定位精度提升了 91.63%。

表 6 季节变化实验定位数据
Table 6 Localization data under seasonal variation experiments

序列	本文			文献[11]		
	平均误差/m	最大误差/m	重定位点数	平均误差/m	最大误差/m	重定位点数
03	1.252	3.226	357	7.242	22.819	12
05	1.432	4.355	437	6.517	22.013	21
04	1.437	2.858	916	11.719	30.831	41
06	1.472	3.897	912	8.680	26.727	48

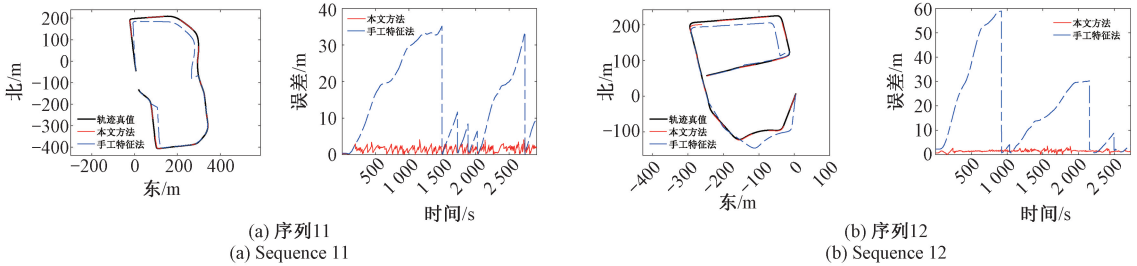


图 9 动态障碍实验定位轨迹与误差误差曲线
Fig. 9 Localization trajectories and error curves under dynamic obstacles

表 7 动态障碍实验定位数据
Table 7 Localization data under dynamic obstacles experiments

序列	本文			文献[11]		
	平均误差/m	最大误差/m	重定位点数	平均误差/m	最大误差/m	重定位点数
11	1.447	4.086	315	15.531	35.185	4
12	1.449	3.505	723	19.101	58.959	10

4 结 论

本文针对现有视觉地图构建方法在建图耗时、鲁棒性、精度方面的问题,提出了一种融合 RTK 约束的视觉路标构建方法。该方法通过滑动窗口法分析航向差值序列检测直线路段,并以线性插值策略筛选代表性图像帧作为视觉路标;以目标路标帧为中心构建路标邻域图像帧集合,引入 RTK 约束校正其初始位姿,并在路标邻域图像集合的局部建图过程中构造全局优化问题,并在优化过程中引入全局 RTK 约束以提高地图重建精度。实验结果表明,该方法有效提升了建图过程的鲁棒性与实时性,同时提升了地图构建精度与重定位精度。未来工作可进一步探索结合深度学习的特征提取与地图更新机制,以提升系统在长期运行中的可持续性与智能性。

参考文献

[1] HE G J, YUAN X D, ZHUANG Y, et al. An integrated GNSS/LiDAR-SLAM pose estimation framework for large-scale map building in partially GNSS-denied environments[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-9.

[2] IOANNIDES R T, PANY T, GIBBONS G. Known vulnerabilities of global navigation satellite systems, status, and potential mitigation techniques [J]. Proceedings of the IEEE, 2016, 104(6) : 1174-1194.

[3] KAZEROONI I A, FITZGERALD L, DOOLY G, et al. A survey of state-of-the-art on visual SLAM [J]. Expert Systems with Applications, 2022, 205: 117734.

[4] TAHERI H, XIA Z C. SLAM; definition and evolution [J]. Engineering Applications of Artificial Intelligence, 2021, 97: 104032.

[5] 杨雪梅, 李帅永. 移动机器人视觉 SLAM 回环检测原理、现状及趋势 [J]. 电子测量与仪器学报, 2022, 36(8) : 1-12.

YANG X M, LI SH Y. Principle, current situation and trend of visual SLAM loop closure detection for mobile robot [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(8) : 1-12.

[6] 黄超, 黄予昕, 杨泽彬, 等. VIG-SLAM: 基于自适应多传感器融合的 SLAM 算法 [J]. 电子测量与仪器学报, 2025, 39(5) : 67-74.

HUANG CH, HUANG Y X, YANG Z B, et al. VIG-SLAM: Adaptive multi-sensor fusion-based SLAM

- algorithm[J]. *Journal of Electronic Measurement and Instrumentation*, 2025, 39(5): 67-74.
- [7] 蔡旭东, 王永才, 白雪薇, 等. 基于先验地图的视觉重定位方法综述[J]. *软件学报*, 2024, 35(2): 975-1009.
- CAI X D, WANG Y C, BAI X W, et al. Survey on visual relocalization in prior map [J]. *Journal of Software*, 2024, 35(2): 975-1009.
- [8] CHALVATZARAS A, PRATIKAKIS I, AMANATIADIS A A. A survey on map-based localization techniques for autonomous vehicles [J]. *IEEE Transactions Intelligent Vehicles*, 2022, 8(2): 1574-1596.
- [9] 薛杨, 孙永荣, 赵科东, 等. 基准地图测绘下的视觉导航算法[J]. *兵工自动化*, 2019, 38(10): 22-27.
- XUE Y, SUN Y R, ZHAO K D, et al. Visual navigation algorithm based on standard mapping [J]. *Ordnance Industry Automation*, 2019, 38(10): 22-27.
- [10] 薛源, 孙永荣, 赵科东, 等. RTK 辅助的单目视觉地图构建与恢复方法[J]. *导航定位与授时*, 2020, 7(3): 70-76.
- XUE Y, SUN Y R, ZHAO K D, et al. Map construction and recovery method based on RTK-aided monocular vision [J]. *Navigation Positioning and Timing*, 2020, 7(3): 70-76.
- [11] JIANG L, TANG X, LI X, et al. Improved ORB-SLAM algorithm with deblurring image [C]. 2024 4th International Conference on Electrical Engineering and Control Science (IC2ECS). IEEE, 2024: 770-774.
- [12] SCHÖNBERGER J L, FRAHM J M. Structure-from-Motion Revisited [C]. *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 4104-4113.
- [13] JIA ZH L, RAO Y, FAN H, et al. An efficient visual SfM framework using planar markers [J]. *IEEE Transactions on Instrumentation and Measurement*, 2023, 72: 1-12.
- [14] PAN L, BARÁTH D, POLLEFEYS M, et al. Global structure-from-motion revisited [C]. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024: 58-77.
- [15] DETONE D, MALISIEWICZ T, RABINOVICH A, et al. SuperPoint: Self-supervised interest point detection and description [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018: 224-236.
- [16] ARANDJELOVIC R, GRONAT P, TORII A, et al. NetVLAD: CNN architecture for weakly supervised place recognition [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 5297-5307.
- [17] MUR-ARTAL R, TARDÓS J D. ORB-SLAM2: An open-source slam system for monocular, stereo, and rgb-d cameras [J]. *IEEE Transactions on Robotics*, 2017, 33(5): 1255-1262.
- [18] LIU SH L, ZENG ZH Y, REN T H, et al. Grounding dino: Marrying dino with grounded pre-training for open-set object detection [C]. *European Conference on Computer Vision*. Cham: Springer Nature Switzerland, 2024: 38-55.
- [19] KIRILLOV A, MINTUN E, RAVI N, et al. Segment anything [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 4015-4026.
- [20] LI D J, SHI X S, LONG Q W, et al. DXSLAM: A robust and efficient visual SLAM system with deep features [J]. *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020: 4958-4965.

作者简介



罗康, 2023 年于北方工业大学获得学士学位, 现为南京航空航天大学硕士研究生, 主要研究方向为视觉导航、数字地图构建、地图匹配定位。

E-mail: luokang0814@nuaa.edu.cn

Luo Kang received his B.Sc. degree from North China University of Technology in 2023. Now he is a M.Sc. candidate at Nanjing University of Aeronautics and Astronautics. His main research interests include visual navigation, digital map construction and map-matching localization.



许舒晨(通信作者), 2017 年于南京航空航天大学获得学士学位, 2020 年于南京航空航天大学获得硕士学位, 现为南京航空航天大学博士研究生, 主要研究方向为视觉导航、数字地图构建、地图匹配定位。

E-mail: xsc@nuaa.edu.cn

Xu Shuchen (Corresponding author) received his B.Sc. and M.Sc. degrees both from Nanjing University of Aeronautics and Astronautics in 2017 and 2020, respectively. Now he is a Ph.D. candidate at Nanjing University of Aeronautics and Astronautics. His main research interests include visual navigation, digital map construction and map-matching localization.