

基于GNSS融合的单目视觉SLAM算法

何定洲*, 穆洪斌, 王 硕, 沈敦亮, 朱 勇

(北京宇航系统工程研究所, 北京 100076)

摘 要: 针对单目视觉SLAM(Simultaneous Localization and Mapping, 同时定位与建图)算法无法恢复地理尺度信息、位姿估计误差累积大等问题, 提出了一种基于GNSS融合的单目视觉SLAM算法。该算法基于图优化理论, 在视觉里程计中构建相似变换估计模型求解真实尺度, 在后端优化中加入GNSS全局地理位置节点, 设计优化求解策略, 提高关键帧位姿以及三维地图点位置估计的准确性, 最后在序列末尾对全局地图离线迭代优化, 保证地图构建的全局一致性。实验表明, 在室外环境下, 运动轨迹在5 km以内时, 所提出方法的尺度估计误差在0.1%以内, 轨迹误差在2 m以内, 满足实际应用需求。

关键词: 同时定位与建图; 多源融合; 单目视觉; 尺度恢复

中图分类号: TN965.8 **文献标志码:** A **文章编号:** 1005-488X(2022)02-0100-07

Monocular Vision SLAM Algorithm Based on GNSS Fusion

HE Dingzhou, MU Hongbin, WANG Shuo, SHEN Dunliang, ZHU Yong
(Beijing Institute of Space System Engineering, Beijing 100076, CHN)

Abstract: Aiming at the problems that the monocular visual SLAM (Simultaneous Localization and Mapping) algorithm could not recover the geographic scale information and the accumulation of pose estimation errors is large, a monocular vision SLAM algorithm based on GNSS fusion has been proposed. Based on the graph optimization theory, the algorithm built a similarity transformation estimation model in the visual odometry to solve the real scale, added the GNSS global location node in the back-end optimization, designed the optimization solution strategy, and improved the key frame pose and 3D map point position estimation. Finally, at the end of the sequence, the global map was iteratively optimized offline to ensure the global consistency of map construction. Experiments showed that in the outdoor environment, when the motion trajectory was within 5 km, the scale estimation error of the method proposed in this paper was within 0.1%, and the trajectory error was within 2 m, which could meet the needs of practical applications.

Key words: SLAM; multi-source fusion; monocular vision; scale recovery

收稿日期: 2022-03-02

作者简介: 何定洲(1976—), 男, 硕士, 高级工程师, 研究方向为精密仪器及机械; (E-mail: alex_heshen@163.com)

穆洪斌(1989—), 男, 博士, 高级工程师, 研究方向为车辆工程;

王 硕(1983—), 男, 博士, 高级工程师, 研究方向为非线性滤波。

* 通讯作者

引言

同时定位与建图技术(Simultaneous Localization and Mapping, SLAM)最早由 Smith、Self 等人在 20 世纪 80 年代提出,可以根据视觉信息对未知环境实时构建增量地图并进行定位^[1],在 AR/VR、无人机避障以及车辆导航等方面得到广泛应用^[2-3],其中单目传感器具有体积小、易安装、价格低廉的优势,得到众多 SLAM 研究学者的关注和研究。

单目视觉 SLAM 技术对连续帧图像进行特征提取与匹配,估计当前相机的位姿并恢复三维地图信息^[4-5]。目前基于单目视觉的同时定位与建图技术主要存在以下三个问题:(1)单目视觉传感器缺少地理位置基准,无法恢复真实尺度信息;(2)由于光照影响、运动模糊、或者相似场景如树叶过多等导致图像特征提取与匹配不准确,导致关键帧位姿恢复不准确,使得视觉跟踪容易发生丢失现象;(3)基于视觉的位姿估计是一个递推的过程,存在较大的误差累积,后端优化与闭环检测只能在一定程度上进行修正。因此需要融合惯性导航^[6-7]、GNSS^[8]、轮式里程计^[9]等等传感器信息,保证单目 SLAM 系统的鲁棒性,提高地图构建的精度和全局一致性。

视觉与其他传感器融合的方法主要分为基于滤波和基于图优化的方法^[10], Strasdat 等人已经证明,在相同的计算成本下,基于图优化比基于滤波的 SLAM 技术具有更高的效率和精度^[11]。R. Mur-Artal 等人提出了一种基于图优化的单目惯性紧耦合 SLAM 技术^[12],在前端对惯性偏差和视觉尺度进行估计,在后端以及闭环检测部分通过图优化的方式对惯性视觉信息融合优化,并且可以地图重用来实现零漂移定位,该方法的尺度估计精度为 1%,但为了获得良好的尺度和偏差估计,初始化需要 10 到 15 秒的时间。香港科技大学 Qin Tong 等人提出基于滑动窗口的方法对单目惯性紧耦合系统进行高精度的状态估计^[13],提高了初始化速度,通过引入 IMU 约束尺度、俯仰角和横滚角,闭环检测的位姿图优化由七自由度约束减小到四自由度,仅需优化三维平移和航向角,提高了实时性。清华大学的王京针对室内场景提出基于图优化的轮式里程计与单目视觉融合 SLAM 方法^[14],并将估计结果约束在三自由度平面空间,提高了地图构建精度。由于惯性导航系统、轮式里程计与视觉的位姿估计都是根据局部信息进行递推,融合也有一定的误差累积,无法提高真实地理位置信息准确度。2019 年, Qin

Tong 等人又提出对视觉全局位姿信息进行优化的 SLAM 算法^[15],将视觉的位姿估计结果与 GNSS 位置信息进行融合,减小全局定位的误差漂移。

GNSS 可以在室外提供绝对位置信息,与单目视觉有良好的互补性,为了提高视觉地图构建的鲁棒性和精度,文章提出一种基于 GNSS 融合的单目视觉 SLAM 算法,研究单目 ORB-SLAM 系统中的图优化模型并改进相关算法,在前端对视觉地图的真实尺度信息进行准确估计,在后端迭代优化局部关键帧位姿和 3D 地图点位置信息,最后在序列运行结束后,系统对全局地图离线优化并保存地图。此方法不仅为单目视觉提供真实地理参考基准,而且保证视觉状态量的局部精确和全局无漂移估计,在无回环场景下依然可以构建全局一致的视觉地图,实现视觉三维地图重构。

1 单目/GNSS 融合 SLAM 系统设计

单目 ORB-SLAM 系统采用图优化算法,对关键帧以及场景中的地标位姿进行筛选与迭代优化,可以对室外大规模场景构建具有几何特征的三维视觉地图,文章通过分析单目 ORB-SLAM 中的图优化算法并在此基础上提出改进,设计单目/GNSS 融合 SLAM 系统框架,保证地理尺度估计的准确性,提高地图构建精度。

1.1 基于图优化的单目 ORB-SLAM 算法分析

单目 ORB-SLAM 系统在追踪线程中对图像序列进行特征提取与匹配,采用高鲁棒性的运动模型估计当前帧位姿和地图点位置,不同于双目、深度视觉,单目视觉没有基线长度或者深度信息作为参考,需要基于 2D-2D 对极几何约束的初始化过程来恢复三维信息,通过将初始地图点中值深度归一化的方式来固定相对尺度,因此具有尺度不确定性,需要其他测距测姿传感器对视觉尺度进行约束估计。

ORB-SLAM 系统通过构建不同的图优化模型对关键帧以及地图点信息进行光束平差优化(Bundle Adjustment, BA)来提高地图构建精度,主要包括:(1)在跟踪线程中根据前、后帧的特征点匹配关系对当前帧进行位姿图优化;(2)在局部建图线程中对有共视关系的局部关键帧位姿和地图点位置进行迭代优化;(3)在回环检测成功后对全局地图中所有关键帧和地图点进行优化。

基于视觉的后端优化只能对局部地图信息进行一定程度的误差修正,在无闭环的场景下地图构建精度并不高。文中通过融合 GNSS 全局位置信息,对单目 ORB-SLAM 系统中的图优化算法进行改进,在前端增加相似变换参数估计线程对视觉尺度进行最优估计,在后端优化模型中增加 GNSS 地理位置节点,改进优化策略,提高视觉建图的鲁棒

性和精度。

1.2 单目/GNSS 融合 SLAM 系统框架

通过融合 GNSS 全局位置信息,对单目 ORB-SLAM 系统中的图优化算法进行改进,提出的基于图优化的单目/GNSS 融合 SLAM 系统框架如图 1 所示。

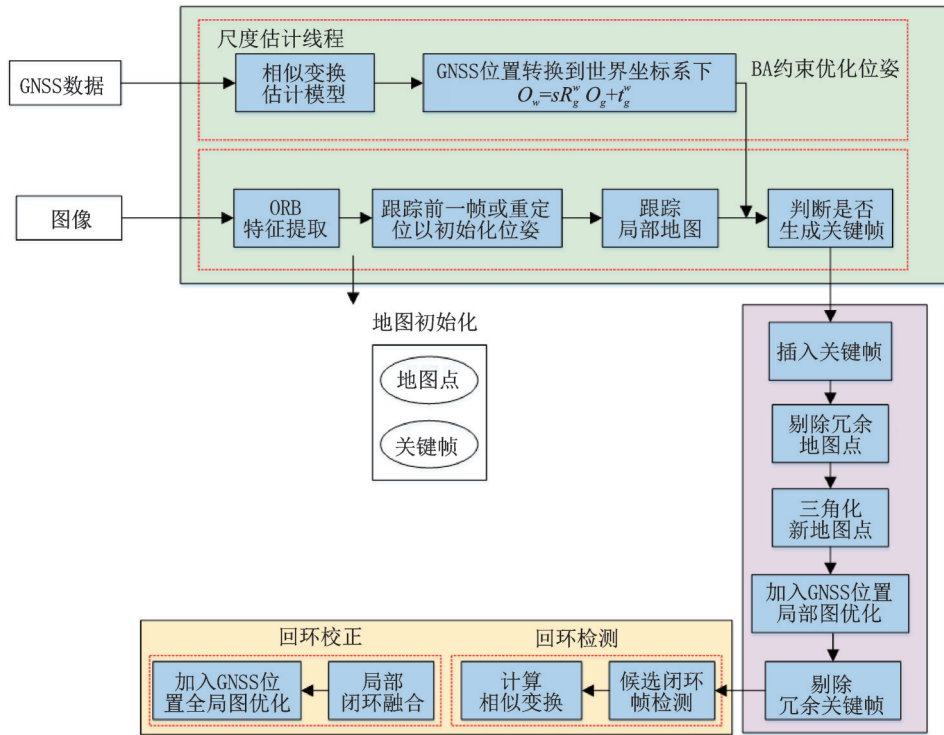


图 1 单目视觉 GNSS 融合 SLAM 系统框架示意图

Fig.1 Schematic of the monocular vision GNSS fusion SLAM system framework

主要改进内容包括以下两点:

(1)增加尺度估计线程,构建相似变换估计模型,保持视觉地图尺度一致性

在跟踪线程生成一定数量关键帧和相关地图点信息后,开始求解相似变换参数。由于关键帧和 GNSS 数据采集初始时间不同,采集频率不同,采用邻域线性插值法将两者对齐。假设其中某一关键帧的采集时间为 t_k ,距离 t_k 左右最近的 2 个 GNSS 采集时刻分别为 t_j 和 t_{j+1} ,则关键帧在 t_k 的 GNSS 位置如式 1 所示。

$$P_g(t_k) = P_g(t_j) + \frac{P_g(t_{j+1}) - P_g(t_j)}{t_{j+1} - t_j} (t_k - t_j) \quad (1)$$

采用三点法求解地理坐标系到视觉世界坐标系下的相似变换参数 $T_g^w = [s, R_g^w, t_g^w]$,以此作为初值构建最小二乘问题,基于所有对齐的 GNSS 位置与关键帧位置约束,求解全局最优的相似变换

参数。

最后将 GNSS 位置信息相似变换到世界坐标系下,基于 GNSS 位置与最小化重投影误差的量测约束,对所有关键帧位姿与地图点位置进行非线性迭代优化,提高当前全局地图的构建精度。

(2)融合 GNSS 位置信息,设计新的代价函数,迭代优化关键帧的位姿信息。

跟踪局部地图后,最小化三维点 $P_w^i \in R^3$ 到当前帧所匹配的特征点 $P_{uv}^i \in R^2$ 的重投影误差,优化当前帧位姿,代价函数如式 2 所示。

$$e_{ij} = P_{uv}^i - \frac{KT_j P_w^i}{z_i} \quad (2)$$

其中 K 表示相机的外参, T_j 表示关键帧 j 的变换矩阵,包含旋转变换 R_j 和平移矩阵 t_j 。

同样的,根据相似变换参数将采集的 GNSS 位置坐标转化到世界坐标系下,利用 GNSS 位置信息

为平移分量定义代价函数如下式所示。

$$e'_{ij} = O_w^{ij} - SO_w^{gi} \quad (3)$$

其中, O_w^{ij} 和 O_w^{gj} 分别表示世界坐标系下由 GNSS 位置坐标和 T_j 转化的关键帧位置, S 表示相似变换系数, 包含比例系数 s , 旋转变换 R_g^w 和平移变换 t_g^w 。

设计构建一个统一的代价函数, 此函数将每个误差项进行整合:

$$C = \sum_{i,j} \rho(e_{i,j}^T \Omega_{i,j}^{-1} e_{i,j} + e'_{i,j}{}^T \Omega'_{i,j}{}^{-1} e'_{i,j}) \quad (4)$$

其中 ρ 表示鲁棒核函数, $\Omega_{i,j}$ 和 $\Omega'_{i,j}$ 分别表示与检测关键点和平移分量的尺度相关联的协方差矩阵。

2 基于图优化的单目/GNSS融合算法

基于图优化理论, 对单目/GNSS融合SLAM系统中的尺度估计算法以及位姿优化算法进行具体设计, 根据GNSS位置约束准确估计视觉尺度, 通过构建后端图优化模型以及设计迭代优化策略, 弥补视觉由于长时间位姿估计存在误差累积的缺陷。

2.1 单目尺度估计

以初始关键帧的相机坐标系作为世界坐标系, 后续关键帧的位姿和地图空间点位置均在世界坐标系下表示, 设第 i 个关键帧的变换矩阵为 T_w^{ci} , 所包含的旋转矩阵和平移矩阵分别表示为 R_w^{ci} 和 t_w^{ci} , 第 i 个关键帧原心在世界坐标系下的位置坐标为:

$$O_w^{ci} = -(R_w^{ci})^T t_w^{ci} \quad (5)$$

将GNSS数据从地球坐标系转化为东北天地理坐标系, 记第 i 个关键帧原心在地理坐标系下的坐标位置为 O_w^{gi} 。记关键帧由变换矩阵求得的位置序列为 $\{O_w^c\}$, 关键帧的地理坐标系位置序列为 $\{O_w^g\}$, 理想情况下两序列存在相似变换关系 S , 即 $\{O_w^c\} = S\{O_w^g\}$, 其中 S 属于相似变换群 $Sim(3)$, 如式6所示。

$$S \in Sim(3) = \left\{ \begin{bmatrix} sR & t \\ 0 & 1 \end{bmatrix} \in R^{4 \times 4} \right\} \quad (6)$$

图优化问题是一个非凸问题, 良好的状态量初值不仅可以加速优化的过程, 还可以避免优化陷入局部最优。因此文中采用RANSAC算法随机抽取3点, 根据Horn的三点法^[16]对相似变换参数进行初始求解, 再根据内点个数以及总体误差等指标筛选出较为可靠的相似变换初值。

求得初值后, 将误差求和, 构建最小二乘问题, 然后寻找最好的相似变换参数, 使它最小化。

$$S^* = \arg \min_s \frac{1}{2} \sum_{i=1}^n \|O_w^{ci} - SO_w^{gi}\|_2^2 \quad (7)$$

对 S 左乘扰动量 $\delta\zeta$, 然后考虑误差 e 的变化关于扰动量的导数, 利用链式法则, 如式8所示。

$$\frac{\partial e}{\partial \delta\zeta} = \lim_{\delta\zeta \rightarrow 0} \frac{e(\delta\zeta \oplus \zeta) - e(\zeta)}{\delta\zeta} = \frac{\partial e}{\partial SO_w^{gi}} \frac{\partial SO_w^{gi}}{\partial \delta\zeta} = -\frac{\partial SO_w^{gi}}{\partial \delta\zeta} \quad (8)$$

根据相似变换李群的扰动模型, 其中:

$$\frac{\partial SO_w^{gi}}{\partial \delta\zeta} = [I \quad -SO_w^{gi \wedge} \quad SO_w^{gi}] \quad (9)$$

使用鲁棒性较好的列文伯格-马夸尔特方法对目标函数进行多次非线性迭代优化, 将GNSS全局位置信息相似变换到世界坐标系下, 可以有效解决单目视觉SLAM尺度漂移问题, 同时可以约束优化关键帧位姿和地图点的累积误差。

根据比例系数将局部地图中的关键帧位姿和地图点空间位置进行等比例扩展, 形成真实比例局部地图, 再次进行相似变换非线性迭代, 得到最优的尺度估计。

2.2 后端局部优化

在前端完成相似变换估计后, 先建立重投影观测方程和GNSS位置观测方程, 分析待优化的状态量顶点, 构建代价函数, 对局部关键帧和地图点进行最优估计。

考虑局部地图中的第 j 个3D地图点 $P_w^j \in R^3$, 其投影在第 i 个关键帧的像素坐标为 $P_{uv}^i \in R^2$, 像素位置与空间位置的关系如下:

$$z_i P_{uv}^i = K T_w^{ci} P_w^j \quad (10)$$

根据 P_w^j 的特征描述子在 P_{uv}^i 的一定范围内匹配像素坐标 P_{uv}^{i-best} 并构建重投影观测方程, 观测方程如下:

$$P_{uv}^{i-best} = \frac{1}{z_i} K T_w^{ci} P_w^j + \omega_{i,j} \quad (11)$$

根据2.1小节估计所得最优相似变换参数将GNSS位置测量值转化到世界坐标系下, 构建GNSS位置观测方程如下:

$$O_w^{ci} = SO_w^{gi} + v_i \quad (12)$$

其中 $O_w^{ci} = -(R_w^{ci})^T t_w^{ci}$, 表示当前第 i 关键帧原心在世界坐标系下的位置信息。

如图2所示, 将与当前关键帧 $KF_{Current}$ 共视当前帧地图点的关键帧定义为一级相邻关键帧 KF_1 (包含 $KF_{Current}$), 其位姿会被优化; 将与一级相邻关键帧有共视关系的关键帧称为二级相邻关键帧 KF_2 , 将

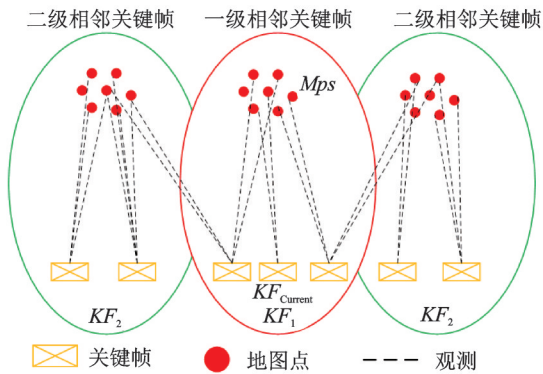


图2 后端图优化模型示意图

Fig.2 Schematic of the back-end graph optimization model

其加入优化图,但其位姿不会被优化。构建后端局部优化模型,图优化顶点包括 $KF_{Current}$ 、 KF_1 以及局部地图点 Mps ,图优化误差边包括重投影观测约束与 GNSS 位置观测约束。

则后端图优化的代价函数为:

$$\begin{aligned} & \{P_w^i, T_j | T_j \in KF_1, P_w^i \in Mps\} = \\ & \arg \min_{P_w^i, T_j} \sum_{T_j \in KF_1 \cup KF_2} \sum_{P_w^i \in Mps} \rho_1(e_{i,j}^T \Omega_{i,j} e_{i,j}) \quad (13) \\ & + \arg \min_{T_j} \sum_{T_j \in KF_1} \rho_2(e_{i,j}^T \Omega_{i,j}^{-1} e_{i,j}) \end{aligned}$$

卡方检验用于统计样本的实际观测值与理论推断值之间的偏离程度,实际观测值与理论推断值之间的偏离程度就决定卡方值的大小,如果卡方值越大,二者偏差程度越大;反之,二者偏差越小;若两个值完全相等时,卡方值就为0,表明理论值完全符合。

完成局部地图的图优化模型后,设计优化求解策略,提出外点,提高视觉状态量估计的准确性。首先获取视觉状态量后构建图优化模型,将状态量形成待优化位姿顶点和误差边加入优化模型中;接着进行第一次优化,对于不符合卡方检测的误差边从求解器中剔除,并设置其不参与下一次的优化;然后进行第二次优化,剔除连接误差较大的关键帧和地图点;最后优化后更新关键帧位姿以及地图点的位置、平均观测方向等属性。

地图构建结束后,可以将地图信息基于估计的相似变换参数变换到地理坐标系下并保存,为系统在导航模式下长时间、高精度的匹配定位提供具有真实地理信息的参考源。

3 实验结果与分析

实验数据采用KITTI数据集的里程计序列^[17],

这些序列来自一辆在居民区周围行驶的汽车,车顶装有摄像头、64线激光雷达、GNSS、IMU等导航传感器,图像采集频率为10 Hz,GNSS数据采集频率为5 Hz,实验条件为一台16 GB内存的Intel Core i7 2.3 GHz笔记本。

KITTI数据集中的若干图像如图3所示,可以看出有些场景树叶较多,且有移动的车辆和行人,因此仅靠视觉估计是不准确的,实验主要将文中提出的融合算法与单目ORB-SLAM算法进行比较,对地图构建的尺度估计和轨迹估计精度进行定性和定量分析。



图3 KITTI序列中的若干图像示意图

Fig.3 Schematic of several images in the KITTI sequence

3.1 尺度估计分析

图4为文中提出的融合算法和单目ORB-SLAM算法对序列03场景恢复的地图,红色虚线为相机的真实运动轨迹;黑色直线、绿色点线和蓝色虚线分别表示单目ORB-SLAM算法和文中方法恢复的运动轨迹,其中单目ORB-SLAM构建的视觉轨迹可以通过求解一个全局相似变换参数,与真实地理坐标进行对齐;可以看出在无闭环修正的情况下,仅由单目视觉估计全局相似变换恢复的道路轨迹有较大的尺度漂移,而文中提出的单目与GNSS融合SLAM算

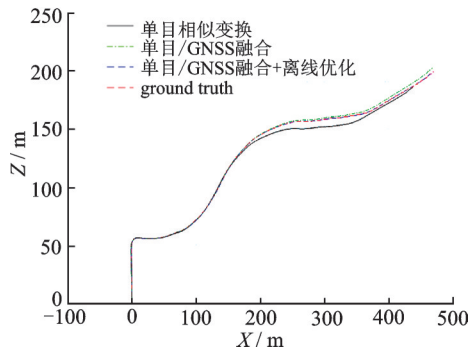


图4 序列03恢复的无闭环地图比较示意图

Fig.4 Schematic of the closed-loop map comparison without the recovery of sequence 03

法,可以较准确地将GNSS信息转换到世界坐标系下,在后端对关键帧位姿和地图点位置信息进行修正,获得与地理信息分布一致的全局视觉地图。

3.2 轨迹精确度分析

为了定量分析融合算法对地图构建精度的改

进,选择有代表性的序列00、02、05、07作为实验数据集,将融合算法构建的轨迹、单目ORB-SLAM构建的轨迹以及真实轨迹进行比较,其中单目ORB-SLAM构建的视觉轨迹可以通过求解一个全局相似变换参数,与真实地理坐标进行对齐。如图5~8所示,左列为算法构建的轨迹图,黑线表示单目ORB-SLAM恢复的视觉轨迹,绿线表示文章提出的单目/GNSS融合算法构建的视觉轨迹,蓝线跟绿线相比,是在图像序列处理结束后对全局地图再离线迭代优化50次构建的视觉轨迹,右列为三个轨迹的误差比较。

如图5、8右列所示,单目ORB-SLAM相似变换构建的轨迹有较大的误差累积,而融合算法可以较好的修正累积误差。进一步对地图构建精度进行定量分析,表1展示了三种算法对4个序列多次执行后的尺度误差和关键帧轨迹均方根误差。尺度误差是将文中算法估计的最优尺度因子 s_1 作为真值,将全局地图轨迹与GNSS数据对齐求解的尺度

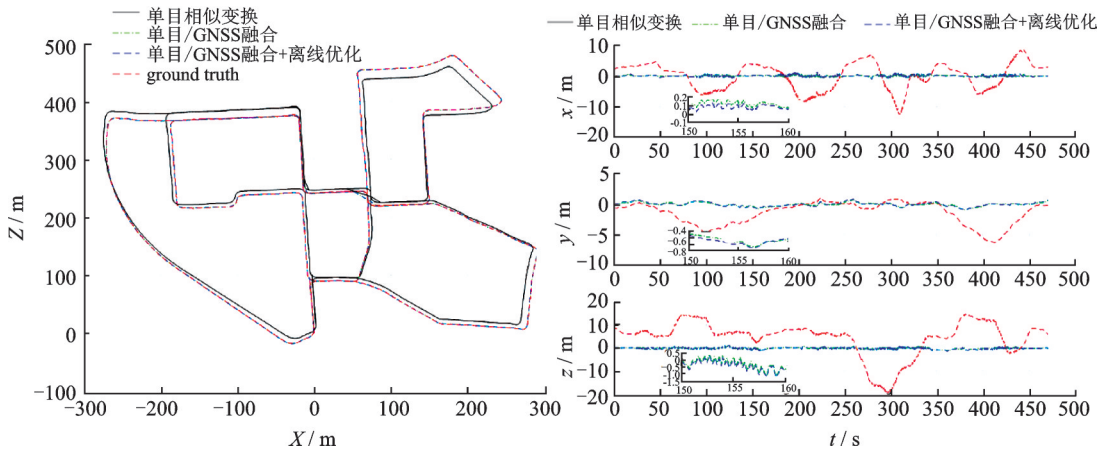


图5 序列00轨迹以及误差比较示意图

Fig.5 Schematic of sequence 00 trajectory and error comparison

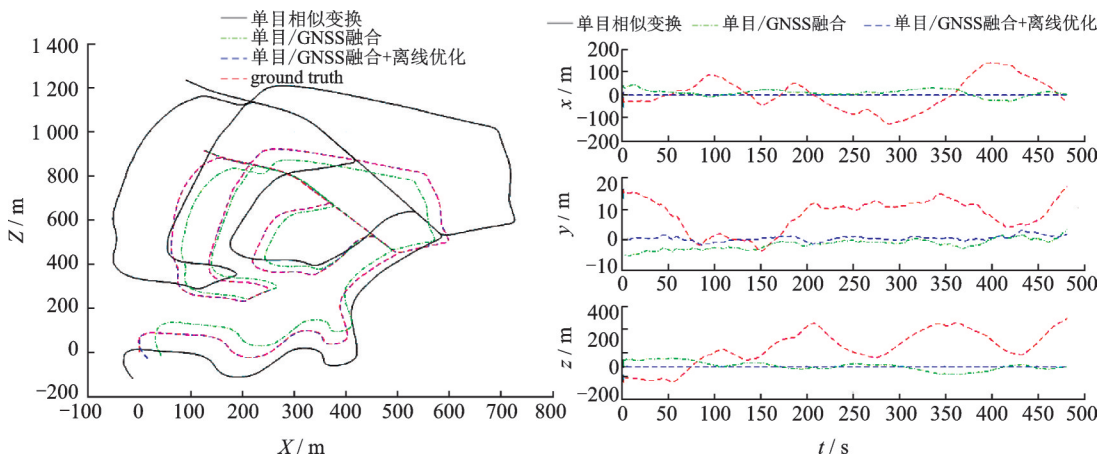


图6 序列02轨迹以及误差比较示意图

Fig.6 Schematic of sequence 02 trajectory and error comparison

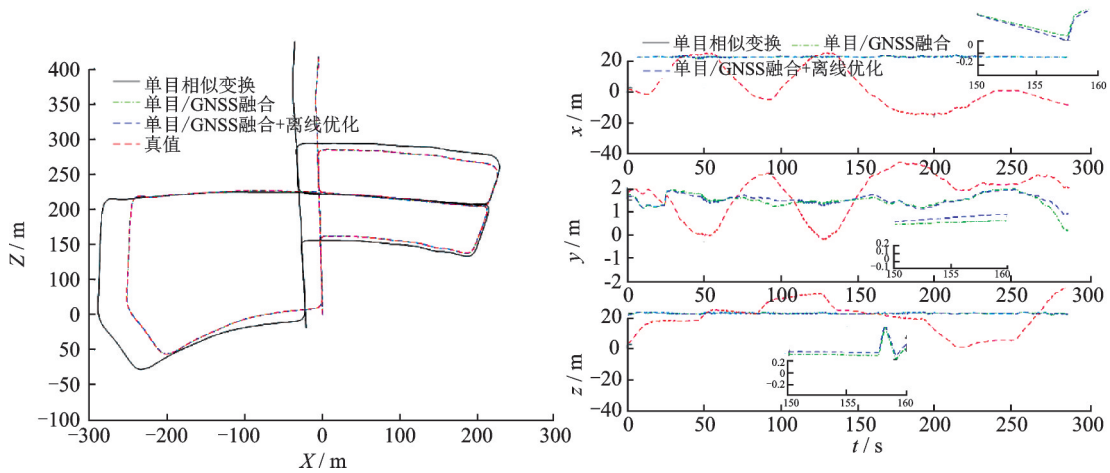


图 7 序列 05 轨迹以及误差比较示意图

Fig.7 Schematic of sequence 05 trajectory and error comparison

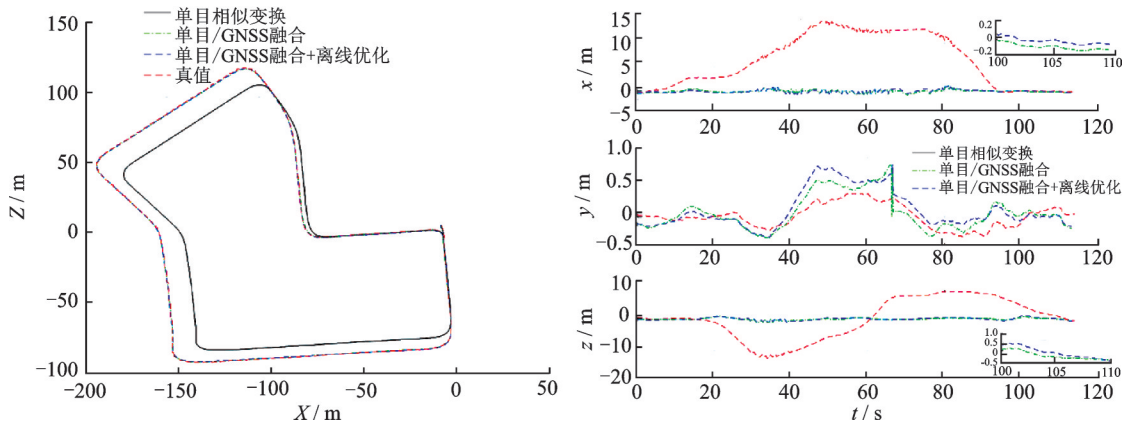


图 8 序列 07 轨迹以及误差比较示意图

Fig.8 Schematic of sequence 07 trajectory and error comparison

表 1 基于KITTI序列构建的视觉轨迹误差分析与比较表

Tab.1 Visual trajectory error analysis and comparison table based on KITTI sequence

序列	长度 / km	关键帧 个数/个	单目ORB-SLAM		单目/GNSS融合SLAM		单目/GNSS融合SLAM + 离线优化	
			尺度误差/(%)	RMSE/m	尺度误差/(%)	RMSE/m	尺度误差/(%)	RMSE/m
00	3.72	1 562	1.394	10.136	0.011	0.550	0.003	0.510
02	5.06	2 073	29.792	180.633	9.822	33.275	0.023	1.882
05	2.21	948	8.677	25.734	0.048	0.591	0.042	0.370
07	0.69	390	9.484	10.260	0.142	0.555	0.10	0.450

因子作为地图估计尺度 s_2 , 通过求解 $e_s = |s_1 - s_2|/s_1$ 反映全局尺度估计的漂移程度。可以看出文中提出的融合算法在室外环境下, 轨迹在 5 km 以内, 平均尺度估计误差在 0.1% 以内, 轨迹误差为 2 m 以内, 融合构建精度比单目和 GNSS 两者都要高, 满足需求。除序列 02 之外, 文中提出的融合算法构建误差约为单目相似变换算法的 5%。序列 02 场景较为复杂, 期间单目尺度发生多次变化导致融合效果欠佳, 经过离线优化后其误差降低为单目相似变换

算法的 1%, 且不存在尺度漂移现象, 说明文章所提方法有助于提高视觉建图的鲁棒性和精度。

4 结 论

单目视觉 SLAM 系统具有尺度不确定性和较大的误差累积, 需要其他传感器进行补偿与修正。文章提出了基于 GNSS 融合的单目视觉 SLAM 算

(下转第 111 页)