

## 点线特征融合的激光雷达单目惯导SLAM系统

崔云轩,刘桂华,余东应,郭中远,张文凯

(西南科技大学 信息工程学院,四川 绵阳 621010)

**摘要:**多传感器融合的SLAM系统定位精度相比单一传感器的SLAM系统更高,但在低纹理场景或退化场景下的定位精度有待提高。提出一种点线特征融合的激光雷达视觉单目惯导紧耦合SLAM系统(PL2VI-SLAM),其由点线特征融合的视觉惯导系统(PLVIS)和激光雷达惯导系统(LIS)两个子系统组成。通过PLVIS系统实现点线特征的提取与匹配,使用滑动窗口选择性地引入关键帧,并将惯性导航器件与相机紧耦合以解算位姿。LIS系统将多个约束集成到因子图中进行联合优化,其初始化状态可以作为PLVIS的初始猜测,通过扫描匹配实现激光雷达里程计,并将点云深度分别与PLVIS系统的特征点以及特征线进行关联,为视觉特征提供精确的深度值,提升定位精度。此外,两个子系统将联合进行回环检测,并对位姿进行矫正。在jackal、handled以及自制的长走廊数据集上的实验结果表明,与LVI-SAM、VINS-MDNO及LIO-SAM系统相比,该系统的定位精度更高,适用于低纹理场景及退化场景,并能满足实时性要求。

**关键词:**低纹理场景;退化场景;点线特征提取;线特征匹配;特征深度关联

开放科学(资源服务)标志码(OSID):



**中文引用格式:**崔云轩,刘桂华,余东应,等.点线特征融合的激光雷达单目惯导SLAM系统[J].计算机工程,2022,48(7):254-263.

**英文引用格式:**CUI Y X, LIU G H, YU D Y, et al. Lidar-mono-inertial SLAM system with fusion of point-line features[J]. Computer Engineering, 2022, 48(7): 254-263.

## Lidar-Mono-Inertial SLAM System with Fusion of Point-Line Features

CUI Yunxuan, LIU Guihua, YU Dongying, GUO Zhongyuan, ZHANG Wenkai

(College of Information Engineering, Southwest University of Science and Technology, Mianyang, Sichuan 621010, China)

**[Abstract]** The multisensor fusion Simultaneous Localization and Mapping (SLAM) system has higher localization accuracy than the single-sensor SLAM system. However, its localization accuracy in low-texture or degraded scenes needs improvement. The Point-Line with lidar-Visual-mono-Inertial tightly coupling SLAM system (PL2VI-SLAM) is proposed. This system comprises two subsystems; the Point-Line with Visual-Inertial System (PLVIS) and the Lidar Inertial System (LIS). The PLVIS subsystem first extracts and matches the point-line features. Next, this subsystem closely couples the inertial measurement unit with a camera to enhance the position by selectively introducing keyframes through sliding windows. LIS integrates multiple constraints into the factor graph joint optimization, and its initial state can be used as the initial guess of PLVIS. The lidar odometry is achieved by scan-matching, and its point cloud depth is associated with the points and feature lines of PLVIS to provide a precise depth value for visual features. These procedures further improve positioning accuracy. Finally, the two subsystems jointly conduct loop-closure to correct the position. The experimental results for the jackal, handled, and self-made long-corridor data sets show that compared with VINS-MONO, LIO-SAM, and LVI-SAM systems, this system exhibits improved positioning accuracy, can satisfactorily handle low-texture and degraded scenes, and can meet the real-time requirements.

**[Key words]** low texture scene; degradation scene; point line feature extraction; line feature matching; feature depth association

**DOI:** 10.19678/j.issn.1000-3428.0062245

**基金项目:**国防科工局核能开发科研项目([2016]1295);四川省科技厅重点研发项目(2021YFG0380)。

**作者简介:**崔云轩(1998—),男,硕士研究生,主研方向为视觉SLAM、激光SLAM、三维重建;刘桂华,教授、博士;余东应、郭中远、张文凯,硕士研究生。

**收稿日期:**2021-08-02 **修回日期:**2021-09-06 **E-mail:**610481953@qq.com

## 0 概述

同步定位与建图(Simultaneous Localization and Mapping, SLAM)技术在移动机器人状态估计及环境感知领域起到重要作用。根据传感器的不同,SLAM技术分为视觉SLAM以及激光SLAM技术<sup>[1-2]</sup>。基于视觉SLAM技术的方法适用于位置识别,但容易受到低纹理、光照、剧烈运动的影响;而基于激光SLAM技术的方法能够捕捉远程环境的细节信息,精度较高,但会受退化场景的影响。多传感器融合的SLAM技术将相机、激光雷达、惯性传感器(Inertial Measurement Unit, IMU)相结合,能够进一步提高SLAM系统的鲁棒性和精度。因此,激光雷达、相机及IMU融合的SLAM系统也越来越受到人们的关注。

目前多传感器融合的SLAM系统主要分为视觉惯性系统(Visual Inertial System, VIS)、雷达惯性系统(Lidar Inertial System, LIS)及激光雷达视觉惯性系统(Lidar Visual Inertial System, LVIS)。根据传感器的直接或间接融合测量方法,相机与IMU的融合方式主要分为松耦合和紧耦合。松耦合方法<sup>[3-4]</sup>通过两种方式处理图像和IMU的测量数据,并分别估计相机和IMU的状态,融合得到最终结果。紧耦合方法<sup>[5-6]</sup>将相机与IMU紧耦合,联合优化估计状态,从而获得高精度的位姿。与松耦合方法相比,紧耦合方法的鲁棒性和精度更高。虽然采用紧耦合融合方式的系统能够实现高精度的位置估计,但依然会受到低纹理场景的影响。因此,研究人员提出PL-SLAM系统<sup>[7]</sup>、基于单目的点线特征系统<sup>[8-10]</sup>以及基于双目的点线特征系统<sup>[11-12]</sup>,其中PL-SLAM系统建立在ORB-SLAM系统<sup>[13-14]</sup>的基础上。HE等<sup>[15]</sup>提出PL-VIO系统,FU等<sup>[16]</sup>提出PL-VINS系统,通过在前端引入线特征,对点线特征联合优化,克服了低纹理环境下无法进行特征跟踪的问题。

同样,LIS系统也分为紧耦合和松耦合系统。ZHAO等<sup>[17]</sup>提出一种实时的激光雷达惯性里程计及建图系统LIOM,该系统通过卡尔曼滤波融合激光雷达和IMU数据,实现准确的位姿估计。文献[18]介绍了一种紧耦合系统,该系统将激光雷达里程计、IMU预积分因子、GPS因子以及回环检测因子引入因子图中,联合优化实现高精度、实时的移动机器人轨迹估计和地图构建。相比于LIOM系统,该系统通过ISAM2实现局部和全局的优化,其鲁棒性和精度更高,但仍然会受到退化场景的影响。

针对退化场景的问题,文献[19]对紧耦合系统进行改进,将激光雷达、相机以及IMU紧耦合,实现了LVIS系统,该系统包含VIS系统和LIS系统,VIS系统采用VINS-MONO系统的方法<sup>[20]</sup>,而LIS系统依据LIO-SAM系统的方法<sup>[18]</sup>,将2个子系统紧耦合,完成系统初始化、状态估计和回环检测,当遇到

低纹理或退化场景时会进行故障检测,每次故障检测均会重新进行初始化,这样虽然能够阻止系统跟踪失败,但会对定位精度产生影响。

本文结合点线特征的方法对LVI-SAM系统进行改进,在VIS系统前端提取与匹配点线特征,并在后端联合进行BA优化<sup>[16]</sup>,实现点线融合。通过短线剔除策略筛选长线特征,落实稳定有效的线特征跟踪。在LIS系统中,将线特征的端点与激光雷达深度进行关联,完成线特征的深度融合,从而获得准确的空间线特征,进一步提高系统的定位精度。

## 1 本文系统框架

本文在LVI-SAM系统的基础上实现PL2VI-SLAM系统,系统的整体框架如图1所示,该系统融合了视觉惯导紧耦合系统(Point-Line with Visual Inertial System, PLVIS)和LIS系统,PLVIS系统包括相机IMU初始化、特征提取与匹配、位姿跟踪、局部地图构建及回环检测。LIS系统包括雷达、IMU初始化、点云提取与匹配、局部建图及回环检测。相比于LVI-SAM系统,本文系统在PLVIS系统初始化阶段以及前端里程计部分融合了点线特征的提取与匹配,后端采用滑动窗口优化算法限制匹配的点特征与线特征的数量。

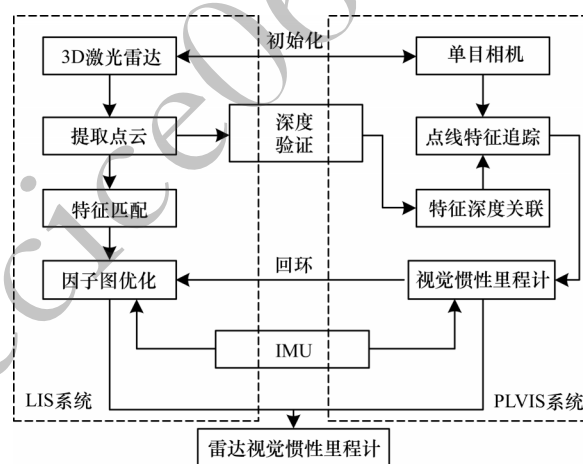


图1 PL2VI-SLAM系统框架

Fig.1 Framework of PL2VI-SLAM system

相较于VINS-MONO系统,本文系统在初始化阶段增加了线特征,促进系统的初始化进程。通过重投影误差对点线特征进行优化,从而获得三维点线特征,同时以LIS系统的初始化成功与否作为整体系统初始化是否成功的判断依据。初始化成功后根据构建的静态局部地图实现连续帧的位姿估计。

在PLVIS系统跟踪的过程中,同时进行点线特征与关键帧的选择,如果点线特征跟踪的关键帧数量超过设定的阈值 $a$ ,即可判定为跟踪成功,并将其添加到全局地图中,点线特征跟踪数量的阈值表达式如式(1)所示:

$$\min [p, l] = a \quad (1)$$

其中: $\mathbf{p}$ 是观测到的点特征; $\mathbf{l}$ 是线特征。

关键帧的数量则使用一个滑动窗口来维护,通过以下方式能够实现PLVIS系统的局部跟踪过程。

最后使用DBOW2词袋模型<sup>[21]</sup>实现PLVIS系统回环检测,并联合LIS系统实现整个PL2VI-SLAM系统的回环检测。

## 2 PL2VI-SLAM系统

PL2VI-SLAM系统由PLVIS视觉惯性系统以及LIS雷达惯性系统组成。其中:PLVIS系统包括点线特征的提取与匹配、IMU预积分、点线特征深度关联以及回环检测;LIS系统包括点云特征的提取与匹配、IMU预积分以及回环检测。

### 2.1 视觉点线特征提取与匹配

首先,每一帧图像采用KLT稀疏光流法<sup>[22]</sup>跟踪前一帧点特征,并提取当前帧新的点特征,使每帧图像中的点特征数量不少于100个,然后在2个相邻特征之间设置最小像素分离值来实现统一的特征分布。根据RANSAC算法的基本矩阵模型<sup>[23]</sup>,能够检验特征的离群值,并剔除离群值大的特征点。通过离群值检验的点将投影到单位球体上,然后延伸到深度归一化平面上,获得特征点的最终位置。

由于OpenCV库中的LSD算法包含大量的短线特征,计算代价较大,若单纯使用LSD算法,很难满足实时性要求。因此采用线特征长度筛选策略,保留主线特征而剔除不必要的短线特征,既能满足实时性要求,又能提高位姿精度。

长度拒绝策略是通过定义特征线的最小长度来拒绝短线特征,线特征长度的最低阈值表达式如式(2)所示:

$$L_{\min} = \left[ u \times \min(W_i, H_i) \right] \quad (2)$$

其中: $(W_i, H_i)$ 表示输入图像的最小长度或高度; $u$ 是比例因子, $u$ 越大,拒绝的线特征长度就越长,本文采用PL-VINS系统的拒绝策略,使 $u=0.125$ 。线特征提取完成后,分别采用LBD算法<sup>[24]</sup>和KNN算法<sup>[25]</sup>来描述和匹配线特征。对于不必要的线特征,用以下2种方法剔除:

- 1)通过KNN算法计算线特征之间的汉明距离,如果大于30,则剔除,否则保留。
- 2)测试所匹配的线特征间的角度是否小于0.1 rad,如果是,则保留,否则剔除。

经过以上2种方法对线特征进行处理后,发现其速度至少是LSD算法的3倍。

### 2.2 点线特征残差

匹配完点线特征后,需要构建点线特征的残差模型。本文依据VINS-MONO系统的方法构建点的残差。定义一个单位球体上的相机测量残差,对于第 $i$ 帧首次观察到的第 $n$ 个特征,该特征对应的第 $j$ 帧观测的特征残差为:

$$\begin{aligned} r_c(\mathbf{z}_n^{c_j}, \mathbf{X}) &= [\mathbf{b}_1 \mathbf{b}_2]^T \times \left( \tilde{\mathbf{p}}_n^{c_j} - \frac{\mathbf{p}_n^{c_j}}{\|\mathbf{p}_n^{c_j}\|} \right) \\ \tilde{\mathbf{p}}_n^{c_j} &= \pi_c^{-1} \left( \begin{bmatrix} \hat{u}_n^{c_j} \\ \hat{v}_n^{c_j} \end{bmatrix} \right) \\ \mathbf{p}_n^{c_j} &= \mathbf{R}_c^c \left( \mathbf{R}_w^{b_j} \left( \mathbf{R}_c^b \left( \mathbf{R}_c^w \frac{1}{\lambda_n} \pi_c^{-1} \left( \begin{bmatrix} u_n^{c_j} \\ v_n^{c_j} \end{bmatrix} \right) + \mathbf{p}_c^b \right) + \mathbf{p}_{b_i}^w - \mathbf{p}_{b_j}^w \right) - \mathbf{p}_c^b \right) \end{aligned} \quad (3)$$

其中: $\mathbf{X}$ 是不同时刻下IMU的状态,具体参见公式(7)。 $\mathbf{R}_b^c$ 、 $\mathbf{R}_c^b$ 、 $\mathbf{R}_w^{b_j}$ 、 $\mathbf{R}_{b_i}^w$ 分别为IMU到相机、相机到IMU、世界坐标系到IMU、IMU到世界坐标系的旋转矩阵, $\mathbf{p}_c^b$ 、 $\mathbf{p}_{b_i}^w$ 、 $\mathbf{p}_{b_j}^w$ 分别为相机到IMU、世界坐标系到IMU、IMU到世界坐标系的平移。 $[u_n^{c_j}, v_n^{c_j}]$ 是第 $i$ 帧观测到的特征点的像素坐标,则其在第 $j$ 帧的投影为 $[\hat{u}_n^{c_j}, \hat{v}_n^{c_j}]$ , $\pi_c^{-1}$ 将相机内参转换为单位向量。

由于视觉残差的自由度为2,因此,将残差向量投影到切平面上,如图2所示。其中, $\mathbf{b}_1$ 、 $\mathbf{b}_2$ 是任意选择的正交基,它们跨越 $\tilde{\mathbf{p}}_n^{c_j}$ 正切平面。

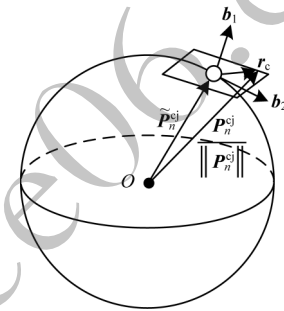


图2 点特征残差示意图

Fig.2 Schematic diagram of point feature residuals

首先建立图像平面上的线特征对应关系,然后通过三角化这些对应关系来初始化世界坐标系中的空间线 $\mathbf{l}_w$ ,其端点为 $M$ 、 $N$ ,如图3所示。

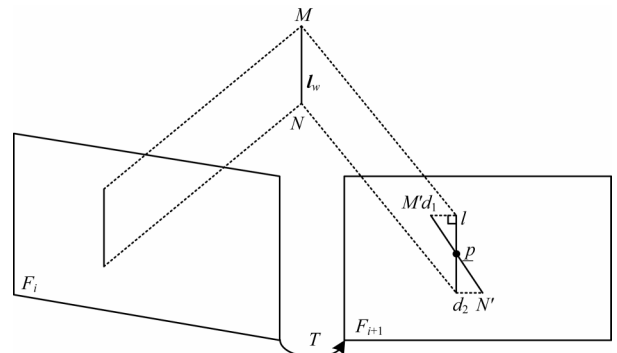


图3 线特征的重投影误差示意图

Fig.3 Schematic diagram of reprojection error of line features

$\mathbf{l}_w$ 用Plücker坐标系表示为 $(\mathbf{n}_w^T, \mathbf{d}_w^T)^T$ ,其中: $\mathbf{n}_w^T$ 和 $\mathbf{d}_w^T$ 都是三维向量; $\mathbf{n}_w^T$ 是由世界坐标系原点和 $\mathbf{l}_w$ 确定的平面的法向量; $\mathbf{d}_w^T$ 是由 $\mathbf{l}_w$ 的2个端点确定的方向



向量。

线特征的重投影误差是根据点线距离进行建模的。首先定义空间线,然后定义线的几何变换,给定一个世界坐标系到相机坐标系的变换矩阵  $\mathbf{Q}_{cw} = [\mathbf{R}_{cw}, \mathbf{t}_{cw}]$ ,  $\mathbf{R}_{cw}$  和  $\mathbf{t}_{cw}$  分别表示旋转矩阵与平移向量。根据变换矩阵  $\mathbf{Q}_{cw}$ ,可以得到  $\mathbf{l}_w$  从世界坐标系变换到相机坐标系下的坐标  $\mathbf{l}_c$  表达式:

$$\mathbf{l}_c = \begin{bmatrix} n_c \\ d_c \end{bmatrix} = \mathbf{Q}_{cw} \times \mathbf{l}_w = \begin{bmatrix} \mathbf{R}_{cw} & [\mathbf{t}_{cw}]_{\times} \mathbf{R}_{cw} \\ 0 & \mathbf{R}_{cw} \end{bmatrix} \begin{bmatrix} \mathbf{n}_w \\ d_w \end{bmatrix} \quad (4)$$

将  $\mathbf{l}_c$  转换到图像平面上,获得投影线  $l$  [26]:

$$l = [l_1, l_2, l_3]^T = \mathbf{K}_l \mathbf{n}_c \quad (5)$$

其中:  $\mathbf{K}_l$  表示线投影矩阵;  $\mathbf{n}_c$  可从式(4)获得。假设  $\mathbf{l}_w$  表示由第  $i$  帧  $\mathbf{c}_i$  观测到的第  $j$  个空间线  $\mathbf{l}_j$ , 则线特征的重投影误差可定义为:

$$r_l(\mathbf{z}_i^c, \mathbf{X}) = d(p, l) = \frac{p^T \times l}{\sqrt{l_1^2 + l_2^2}} \quad (6)$$

其中:  $d(p, l)$  表示点到直线的距离函数;  $p$  为端点  $M'$ 、 $N'$  的坐标;  $l_1$ 、 $l_2$ 、 $l_3$  均为直线方程系数;  $p$  是线特征中点的齐次坐标。

### 2.3 PLVIS 子系统

无论是子系统的初始化还是状态估计,都离不开IMU积分模型。通过IMU积分可以获得机器人在任意时刻的位置、速度以及姿态,从而为本文系统提供高精度的状态估计。而在IMU积分的过程中,系统会受到加速度和陀螺仪偏置的影响,如果每次计算都要将IMU坐标系转换到世界坐标系,计算量将非常大,严重影响系统的实时性,因此本文采用IMU预积分模型[27]。

对于PLVIS子系统,当初始化完成后,将激光雷达测量的特征深度值关联到视觉里程计中,从而获得一个用于特征深度估计的关键帧。将关键帧对应的不同时刻的状态定义为:

$$\mathbf{X} = [\mathbf{R}, \mathbf{p}, \mathbf{v}, \mathbf{b}] \quad (7)$$

其中:  $\mathbf{R}$  是李群表示的旋转矩阵;  $\mathbf{p}$ 、 $\mathbf{v}$ 、 $\mathbf{b}$  是经过IMU预积分获得的位置、速度以及偏置。在IMU预积分过程中,将关键帧的状态加入到滑动窗口中,执行BA优化,从而获得IMU坐标系到世界坐标系的变换  $\mathbf{Q} = [\mathbf{R} | \mathbf{p}]$ 。

1) 初始化。对于PLVIS系统,初始化的质量在很大程度上取决于IMU传感器的运动和精度,当IMU传感器以较小速度或匀速移动时,VINS-MONO系统无法进行初始化,这是因为当加速度不大时,尺度无法观测,而IMU传感器的偏置和噪声影响了原始加速度的测量,所以初始化能够对这些参数有良好的猜测,会使后续的迭代优化更快收敛。

相比于VINS-MONO系统,本文系统在初始化时,首先初始化LIS系统,并获得状态  $\mathbf{x}$  和偏置  $\mathbf{b}$ ,通过线性插值法将状态  $\mathbf{x}$  和偏置  $\mathbf{b}$  插入关键帧中,并作为PLVIS系统的初始化猜测,从而保证初始化的速度和鲁棒性。其次,为进一步提高初始化的鲁棒性,增加了线特征的初始化,使系统能更好地应对低纹理情况。

2) IMU预积分。根据IMU传感器的原始信息,可以定义IMU传感器的测量值为:

$$\hat{\mathbf{w}}_t = \mathbf{w}_t + \mathbf{b}_t^w + \mathbf{n}_t^w \quad (8)$$

$$\hat{\mathbf{a}}_t = \mathbf{R}_t^{\text{BW}} (\mathbf{a}_t - \mathbf{g}) + \mathbf{b}_t^a + \mathbf{n}_t^a \quad (9)$$

其中:  $\hat{\mathbf{w}}_t$  和  $\hat{\mathbf{a}}_t$  分别为  $t$  时刻在IMU坐标系下的角速度和加速度;  $\mathbf{g}$  是世界坐标系下的重力向量;  $\mathbf{R}_t^{\text{BW}}$  是世界坐标系到IMU坐标系的旋转变换矩阵。

根据IMU测量值,可以得到相机在不同时刻的位置、速度以及旋转状态的表达式如下所示:

$$\mathbf{v}_{t+\Delta t} = \mathbf{v}_t + \mathbf{g}\Delta t + \mathbf{R}_t (\hat{\mathbf{a}}_t - \mathbf{b}_t^a - \mathbf{n}_t^a) \Delta t \quad (10)$$

$$\mathbf{p}_{t+\Delta t} = \mathbf{p}_t + \mathbf{v}_t \Delta t + \frac{1}{2} \mathbf{g} \Delta t^2 + \frac{1}{2} \mathbf{R}_t (\hat{\mathbf{a}}_t - \mathbf{b}_t^a - \mathbf{n}_t^a) \Delta t^2 \quad (11)$$

$$\mathbf{R}_{t+\Delta t} = \mathbf{R}_t \exp \left( \left( \hat{\mathbf{w}}_t - \mathbf{b}_t^w - \mathbf{n}_t^w \right) \Delta t \right) \quad (12)$$

其中:  $\mathbf{R}_t$  是IMU坐标系到世界坐标系的旋转变换矩阵,可以理解为  $\mathbf{R}_t^{\text{BW}}$  的转置。

由于  $\hat{\mathbf{w}}_t$  和  $\hat{\mathbf{a}}_t$  会受到IMU的偏置  $\mathbf{b}_t$  以及白噪声  $\mathbf{n}_t$  的影响,为了获得相邻时刻IMU的变换,使用文献[27]中的预积分模型,将  $i$ 、 $j$  时刻的预积分测量值表示为:

$$\Delta \mathbf{v}_{ij} = \mathbf{R}_i^T (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g} \Delta t_{ij}) \quad (13)$$

$$\Delta \mathbf{p}_{ij} = \mathbf{R}_i^T \left( \mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i \Delta t_{ij} - \frac{1}{2} \mathbf{g} \Delta t_{ij}^2 \right) \quad (14)$$

$$\Delta \mathbf{R}_{ij} = \mathbf{R}_i^T \mathbf{R}_j \quad (15)$$

3) 特征深度关联。在PLVIS系统初始化以及视觉里程计中,本文将LIS系统中激光雷达测量的深度值作为视觉特征的深度。由于激光雷达产生的是稀疏的点云,因此需要通过叠加多个激光雷达帧,才能获得特征密集的深度图。相比于LVI-SAM系统,本文系统不仅将点特征与深度值关联起来,而且将线特征端点与深度值联系起来。将点特征、线特征端点、激光雷达深度点投影到一个单位球体上。此时,降采样激光雷达深度点云在单位球体上利用视觉特征的极坐标搜索二维K-D树。找到球体上对应距离最近的3个深度特征,此特征深度即为相机中心到视觉特征的距离。同理,可以获得线特征端点的深度。视觉点线特征的深度关联示意图如图4所示,其中,虚线的长度代表特征深度。

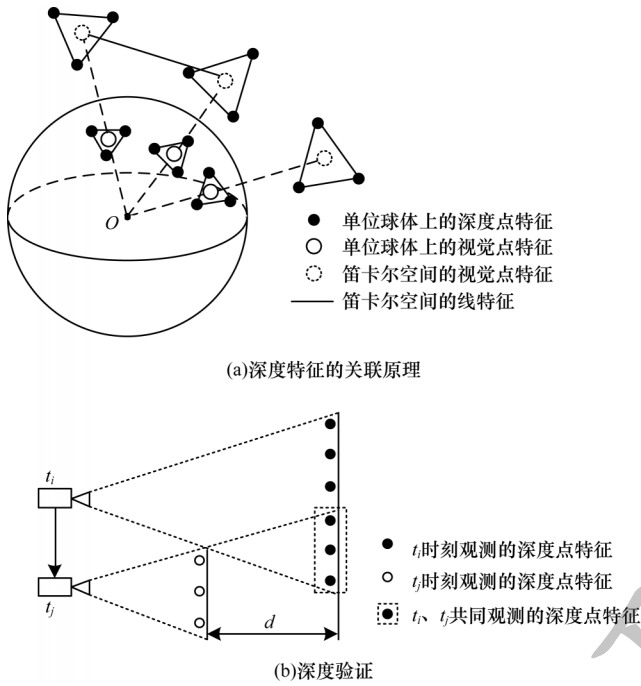


图4 视觉点线特征的深度关联

Fig.4 Visual point-line feature depth association

投影到球体上的3个深度特征进行逆投影后,它们的距离可能相差甚远,如图4(b)所示, $t_i$ 时刻和 $t_j$ 时刻叠加的激光雷达帧,它们的深度特征间距 $d$ 很大,这样会严重影响视觉特征的深度估计。因此,需要检查逆投影后的深度特征之间的距离,当其中某个深度特征的距离超过了本文设定的阈值2 m时,即可判断对应的视觉特征没有深度,从而更有利于还原真实场景。视觉特征深度关联的效果如图5所示。在图5(b)中,绿色代表关联的点特征,蓝色代表关联的线特征(彩色效果见《计算机工程》官网HTML版本)。上述的深度特征均指激光雷达观测到的特征。

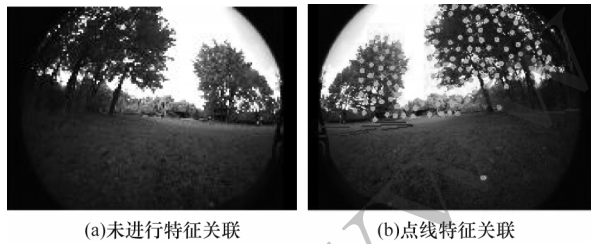


图5 深度特征关联前后的效果图

Fig.5 Effect drawing before and after depth feature association

4)故障检测。PLVIS系统容易受到剧烈运动、光照变化和低纹理环境影响。当机器人经历剧烈运动或进入低纹理环境时,被跟踪特征的数量将大大减少,可能会导致优化失败,并产生较大的IMU偏差。LVI-SAM系统的解决办法是当跟踪的特征数量低于阈值时,系统进行故障检测并重新进行初始化。但每次初始化时,如果PLVIS系统遇到低纹理场景,初始化将失败,导致定位失败,定位精度降低。因

此,本文加入线特征,只有当跟踪的线特征数量低于阈值时才进行故障检测。

5)视觉回环检测。回环检测模块利用了DBoW2词袋模型,当新的关键帧到来时,将提取简短的二进制BRIEF描述子<sup>[28]</sup>,并且将其与之前的描述子进行匹配,如果相似度超过了设定的阈值,即可判定该帧是一个回环帧。此外,PLVIS系统一旦检测到回环,就会将此信息传达给LIS系统,并将DBoW2词袋模型得到的回环帧的时间戳发送给LIS系统进行验证,如果该时间戳对应的关键帧在LIS系统里也是回环帧,则可验证该帧为回环帧。

## 2.4 LIS子系统

图1左侧所示为基于LIO-SAM系统实现的LIS子系统框架。LIO-SAM系统中的因子图包含了IMU预积分因子、激光雷达里程计因子、回环检测因子以及可选的GPS因子。而LIS子系统还包含了视觉里程计因子,并将它们添加到因子图中,通过ISAM2算法<sup>[29]</sup>优化,获得一个高精度的里程计。IMU预积分过程在2.3节已介绍过,在因子图中,只需添加IMU预积分和视觉里程计估计的状态,并与其他因子进行联合优化。因此,下文仅介绍激光雷达里程计和回环检测。

1)激光雷达里程计约束。在LIS子系统中,若要实现一个完整的激光雷达里程计,需要对激光雷达检测到的点云特征进行提取与匹配。本文采用LOAM系统的特征提取方式,根据局部区域点的曲率来提取边缘和平面的特征。若将第 $i$ 帧提取的边缘和平面特征分别表示为 $F_i^e$ 和 $F_i^p$ ,则该时刻提取的所有特征组成的激光雷达帧为 $F_i$ ,可以表示为:

$$F_i = \{F_i^e, F_i^p\} \quad (16)$$

由于LIS系统是基于LIO-SAM系统实现的,其关键帧的选择是通过一种简单有效的启发式方法:与之前的状态 $x_i$ 相比,当机器人姿态变化超过用户定义的阈值时,则选择激光雷达帧 $F_{i+1}$ 作为关键帧。在因子图中,新保存的关键帧 $F_{i+1}$ 与一个新的机器人状态节点 $x_{i+1}$ 关联。2个关键帧之间的激光雷达帧被丢弃。通过这种方式添加关键帧,既能平衡建图的密度和内存消耗,又能保持相对稀疏的因子图,从而实现实时的非线性优化。每增加一个新的关键帧,其距离变化和旋转角度变化的阈值选分别为1m和 $10^\circ$ 。在此基础上,使用1个滑动窗口来包含固定数量的子关键帧 $\{F_{i-n}, F_{i-n+1}, \dots, F_i\}$ ,子关键帧在世界坐标系下的变换为 $\{Q_{i-n}, Q_{i-n+1}, \dots, Q_i\}$ ,将其转换到世界坐标系下,最终合并到一个体素地图 $M_i$ 中。根据所提取的边缘特征和面特征,对应的边缘特征体素地图和平面特征体素地图可以表示为 $M_i^e$ 和 $M_i^p$ 。激光雷达帧和体素地图的关系可以表示为:



$$\begin{aligned}
M_i &= \{M_i^e, M_i^p\} \\
M_i^e &= \hat{F}_i^e \cup \hat{F}_{i-1}^e \cup \dots \cup \hat{F}_{i-n}^e \\
M_i^p &= \hat{F}_i^p \cup \hat{F}_{i-1}^p \cup \dots \cup \hat{F}_{i-n}^p
\end{aligned} \quad (17)$$

其中:  $\hat{F}_i^e$ 、 $\hat{F}_i^p$  分别为世界坐标系下表示的第  $i$  帧边缘、面特征。对边缘特征和面特征进行匹配,并利用点线、点面之间的距离关系建立位姿求解方程:

$$d_{e_s} = \frac{\left| (p_{i+1,k}^e - p_{i,u}^e) \times (p_{i+1,k}^e - p_{i,v}^e) \right|}{\left| p_{i,u}^e - p_{i,v}^e \right|} \quad (18)$$

$$d_{p_s} = \frac{\left| (p_{i,u}^p - p_{i,w}^p) \times (p_{i,u}^p - p_{i,v}^p) \times (p_{i+1,k}^p - p_{i,u}^p) \right|}{\left| (p_{i,u}^p - p_{i,v}^p) \times (p_{i,u}^p - p_{i,w}^p) \right|} \quad (19)$$

尽管激光雷达可以很好地提供深度特征,然而如果遇到退化场景,也会误匹配。不仅会影响 LIS 系统的状态估计,也会影响 PLVIS 系统的状态估计。因此,本文利用文献[30]提出的方法,将扫描匹配中的非线性问题表示成迭代求解的线性问题:

$$\min_{\mathbf{Q}} \left\| \mathbf{A}\mathbf{Q} - \mathbf{b} \right\|^2 \quad (20)$$

其中:  $\mathbf{A}$  和  $\mathbf{b}$  由线性化位姿  $\mathbf{Q}$  得到,如果  $\mathbf{A}^T \times \mathbf{A}$  获得的最小特征值超过了设定的阈值,说明分解  $\mathbf{Q}$  后的矩阵  $\mathbf{A}$  是非奇异的,此时 LIS 系统会发送故障,使当前帧的激光雷达里程计因子不能添加到因子图中。

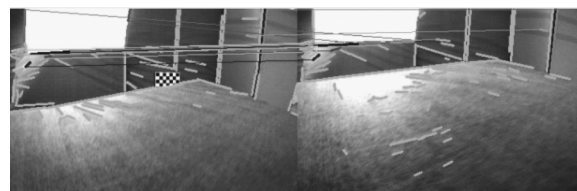
2) 回环检测。回环检测是根据点云之间的欧几里得距离实现的,能通过 ISAM2 算法集成到因子图中。当一个新的状态  $x_i$  到来时,首先通过欧式距离找到与  $x_i$  状态最为接近的候选状态  $x_k$  ( $k=1, 2, \dots, i-1$ ),然后将关键帧  $F_i$  与  $F_k$  转换到世界坐标系下,得到相对变换  $\Delta T_{i,k}$ ,并将其添加到因子图中,最后通过前端扫描匹配法对  $F_i$  和  $\{F_{k-j}, F_{k-j+1}, \dots, F_{k+j-1}, \dots, F_{k+j}\}$  进行匹配。在 LIS 系统中,  $j=12$ , 欧氏距离设置为 15 m。

### 3 实验与结果分析

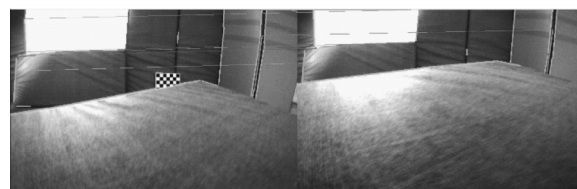
首先选用 EUROC 数据集对线特征匹配效果进行分析,然后使用 LVI-SAM 系统提供的手持式 handled 数据集以及移动小车 jackal 数据集进行定位精度分析。最后,为验证本文系统的有效性,将 realsensD435i 相机、xsens 惯导传感器以及 VLP16 线激光雷达刚性连接在一起,并标定它们之间的外参,手持传感器录制一个长走廊的数据集,并对端到端的误差进行评估。本文选用的数据集包含低纹理和退化场景。实验设备配置为 4 核 i7 处理器,8 GB 内存, GTX1060, ubuntu16.04 操作系统。

#### 3.1 线特征匹配效果

基于 EUROC 的 V1\_03 数据集,选取相邻 2 个关键帧对线特征进行匹配,匹配结果如图 6 所示(彩色效果见《计算机工程》官网 HTML 版本)。



(a) LSD 算法匹配结果



(b) 本文算法匹配结果

图 6 不同算法的匹配结果对比

Fig.6 Comparison of matching results of different algorithms

从图 6 可以看出,当场景中的纹理较少时, LSD 算法虽然能够进行线特征匹配,但会存在大量的短线特征与误匹配,如图中的黄线所示。而本文算法能剔除短线特征,在提高计算速度的同时也减少了误匹配,效果明显比 LSD 算法好。对线特征匹配数量及耗时进行对比分析,结果如表 1 所示。

表 1 匹配耗时及匹配数量对比

Table 1 Comparison of matching time and matching quantity

匹配算法	匹配耗时/s	匹配线个数	有效匹配数
LSD 算法	0.018 1	9	5
本文算法	0.005 7	5	5

由表 1 可知, LSD 算法匹配耗时最长,而本文算法保留了主线特征,且匹配速度是 LSD 算法的 3 倍。

本文系统在运行 jackal 数据集的过程中,其三维可视化效果如图 7 所示(彩色效果见《计算机工程》官网 HTML 版本)。其中,红色的线是三维线特征,紫色和白色的点是局部视觉特征,周围的彩色点云是不同距离上的激光雷达扫描的点云。

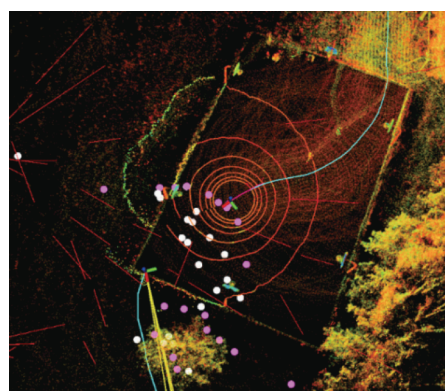


图 7 线特征三维可视化的效果

Fig.7 Effect of 3D visualization of line features

#### 3.2 jackal 数据集

在 jackal 数据集中,以 GPS 轨迹作为 groundtruth,将本文系统分别和 LVI-SAM、LIO-SAM 及 VINS-MONO 系统进行轨迹对比(在进行 jackal 和 handled 数据集实验时, LIO-SAM 系统和 LVI-SAM 系统均加了 GPS 因子),结果如图 8 所示。

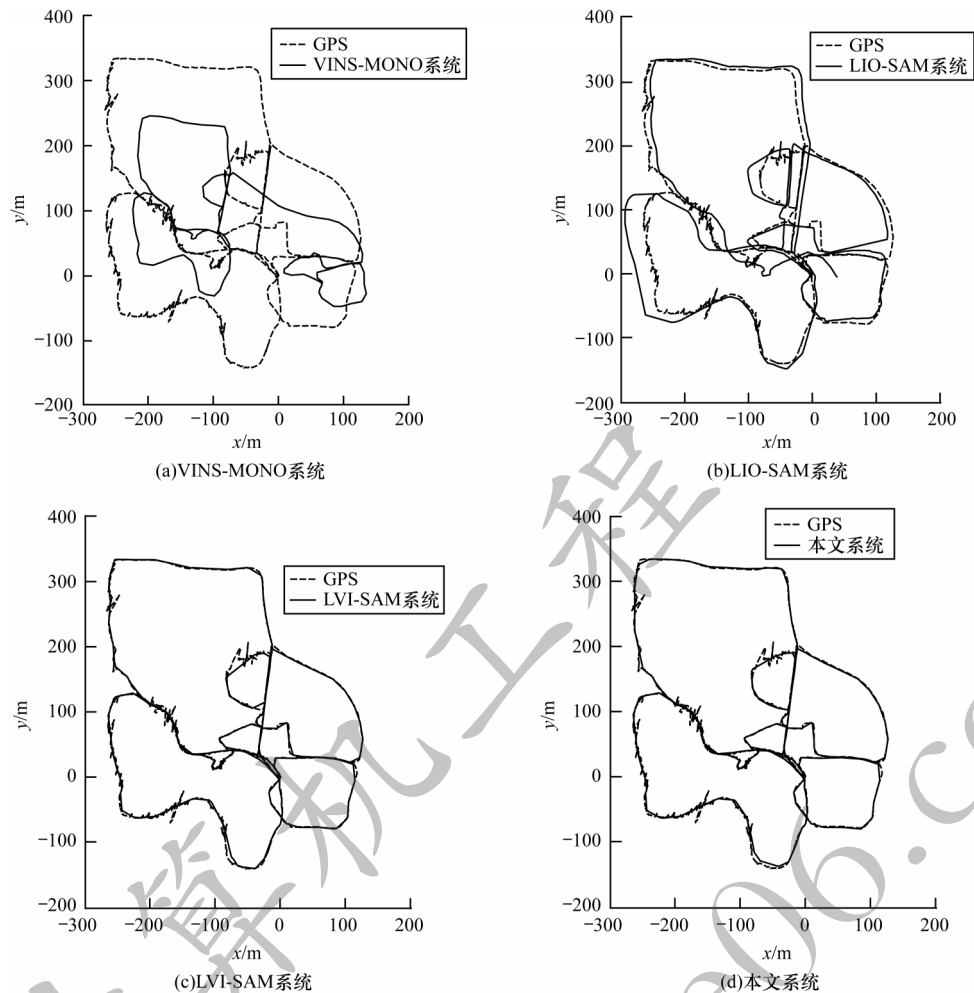


图8 不同系统在jackal数据集下的轨迹对比

Fig.8 Trajectory comparison of different systems under jackal date set

图8对应的GPS轨迹中有部分震荡,原因是该部分在实际情况中受到了树木的遮挡。由图8可以看出,VINS-MONO系统在运行一段时间后跟踪丢失,而本文系统能够有效实现位姿跟踪,且相比于LIO-SAM系统和LVI-SAM系统,本文系统的轨迹更接近GPS groundtruth。

jackal数据集由移动小车收集,且起点与终点位置一样,运动稳定。本文分析了不同系统间端到端的误差,使用位置误差和旋转误差进行评估,两个误差越接近0,定位精度越高,结果如表2所示。

表2 不同系统的端到端误差对比

系统	位置误差/m	旋转误差/(°)
VINS-MONO 系统	43.53	12.41
LIO-SAM 系统	1.52	26.89
LVI-SAM 系统	0.13	1.47
本文系统	0.08	0.97

由表2可知,与其他系统相比,本文系统的位置误差和旋转误差均最小。

3.3 handled数据集

handled数据集是操作者手持传感器录制的,相

比于jackal数据集,操作者的运动更剧烈,且有多处是退化场景。如图9所示,当激光雷达与地面垂直或处于空旷地带时,激光雷达将无法提取足够的点云,导致扫描匹配失败,从而使里程计跟踪失败。

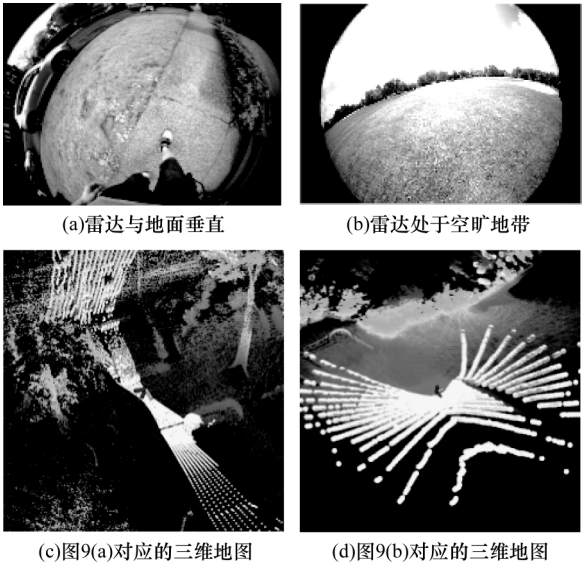


图9 两种退化场景的示意图

Fig.9 Schematic diagram of two degradation scenarios

本文对比了不同系统在 handled 数据集下的轨迹,结果如图 10 所示。由图 10 可知,与其他系统相

比,本文系统和 LVI-SAM 系统的轨迹更接近 groundtruth。

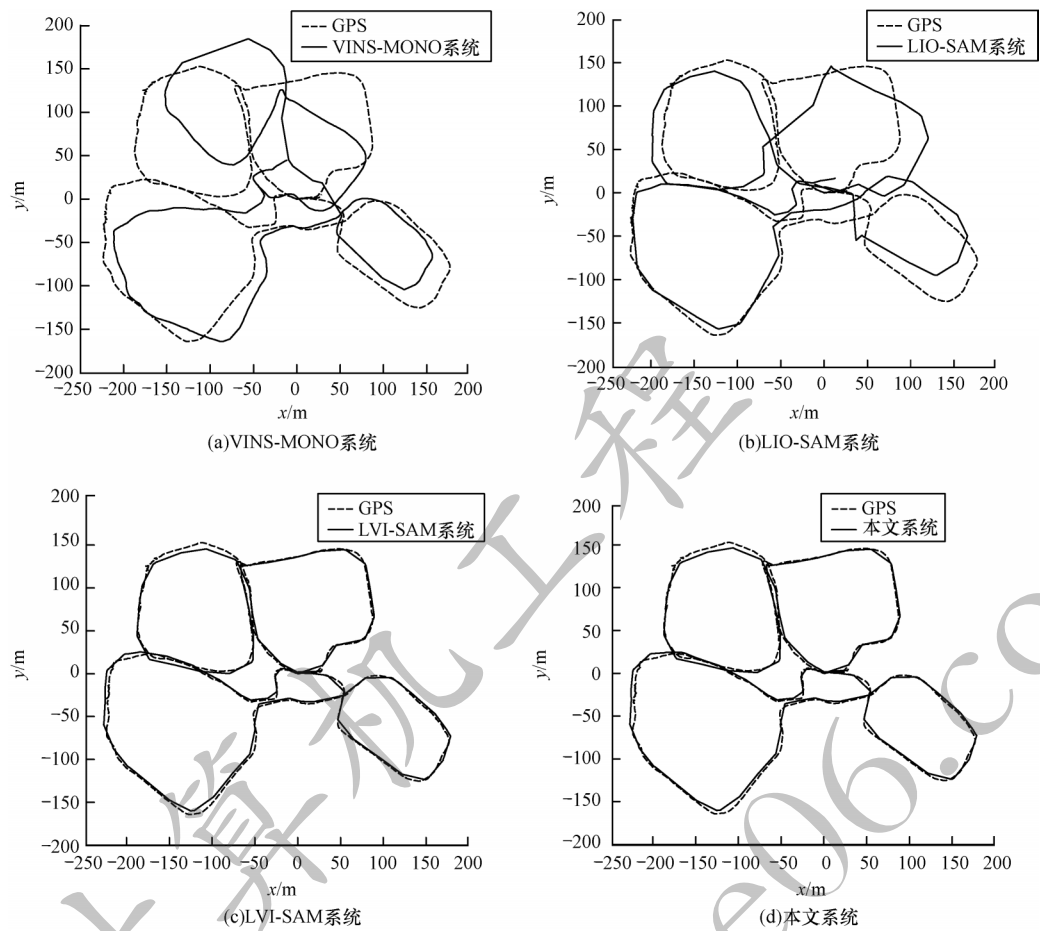


图 10 不同系统在 handled 数据集下的轨迹对比

Fig.10 Trajectory comparison of different systems under handled date set

使用绝对轨迹误差 (Absolute Trajectory Error, ATE) 和相对轨迹误差 (Relative Pose Error, RPE) 中的均方根误差 (Root Mean Squared Error, RMSE) 来评估不同系统的全局定位精度,结果如表 3 所示。

表 3 不同系统的定位精度对比

Table 3 Comparison of positioning accuracy of different systems

系统	ATE (max)	ATE (mean)	ATE (RMSE)	RPE (max)	RPE (mean)	RPE (RMSE)
VINS-MONO 系统	116.305	40.365	49.415	197.965	0.868	5.560
LIO-SAM 系统	82.496	25.755	32.787	73.809	3.552	11.436
LVI-SAM 系统	7.954	3.436	4.089	59.467	3.597	11.425
本文系统	7.530	3.401	3.843	56.456	3.616	11.227

由表 3 可知,在有退化场景的 handled 数据集下,本文系统的定位精度优于其他先进的开源 SLAM 系统,验证了本文系统在退化场景下有更好的效果。

3.4 长走廊数据集

为验证本文系统的有效性,本文录制了具有低

纹理与退化场景的长走廊数据集。在长走廊的另一端,特意对传感器进行旋转,将激光雷达扫描线与地面垂直,并让相机观测低纹理区域,低纹理区域如图 11 所示。

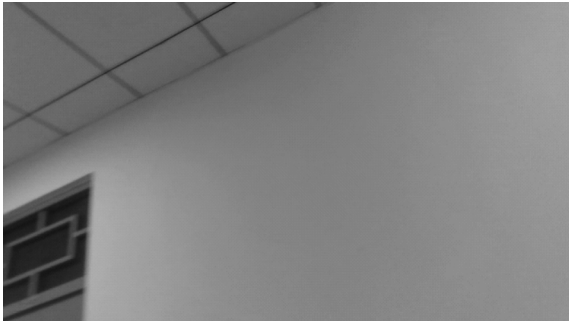


图 11 低纹理区域

Fig.11 Low textured area

此外,本文进行了线特征匹配,结果如图 12(a) 和图 12(b) 所示。其中,图 12(a) 对应未退化场景,图 12(b) 对应低纹理退化场景,退化场景下的激光雷达扫描线是与地面垂直的。



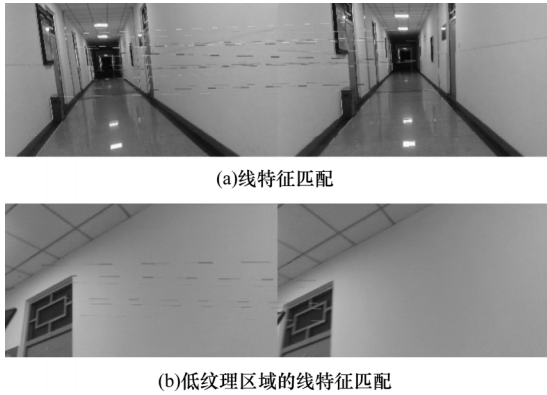


图12 低纹理场景下的线特征匹配结果

Fig.12 Line feature matching results under low texture scenes

由于长走廊数据集中传感器的起始位置和终端位置一样,且终端的激光雷达坐标系在初始的激光雷达坐标系的 $x$ 轴方向上旋转了 $180^\circ$ ,因此本文通过评估各个系统的端到端误差,从而进一步验证本文系统的有效性。设起点位置为 $(0,0,0)$ ,终点位置为 $(x,y,z)$ ,则终点位置与起点位置之间的误差可以使用如式(21)所示的三维空间欧氏距离表示:

$$e(t) = \sqrt{x^2 + y^2 + z^2} \quad (21)$$

角度误差为起点位置和终点位置的欧拉角之差:

$$r(\mathbf{R}) = \theta_{\text{end}} - \theta_{\text{begin}} \quad (22)$$

此外,在长走廊数据集下对比了各个系统的轨迹,结果如图13所示。可以看出,本文系统可以很好地还原真实轨迹情况,能够保持起点和终点位置一致,而LVI-SAM和LIO-SAM系统在激光雷达进行垂直翻转时均发生了漂移。虽然VINS-MONO系统的轨迹完整,但在相机翻转的时候扫到了低纹理区域,其轨迹发生了偏移,导致终点和起点位置不一致。

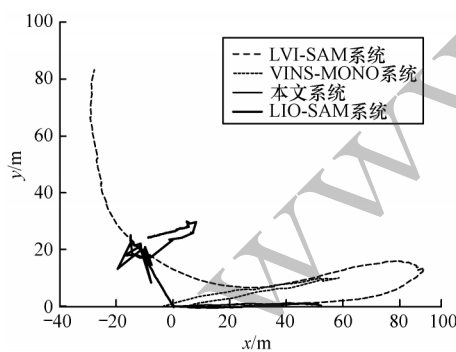


图13 不同系统在长走廊数据集下的轨迹

Fig.13 Trajectories of different systems under long corridor data set

不同系统的端到端误差对比结果如表4所示。可以看出,LIO-SAM和LVI-SAM系统的轨迹均跟踪失败。虽然VINS-MONO系统跟踪成功,但最终的位置和起点的位置相差8.87 m,角度误差为 $183.36^\circ$ 。

而本文系统的位置误差仅为0.88 m,角度误差为 $5.79^\circ$ ,说明本文系统的跟踪轨迹更接近真实轨迹。

表4 不同系统的端到端误差对比

Table 4 Comparison of end-to-end error of different system

系统	位置误差/m	旋转误差/ $^\circ$
VINS-MONO 系统	8.87	183.36
LIO-SAM 系统	Fail	Fail
LVI-SAM 系统	Fail	Fail
本文系统	0.88	5.79

最后,本文采用文献[18]的耗时评估方法,在长走廊数据集下评估各系统处理一帧数据所消耗的平均时间,结果如表5所示。可以看出,当雷达的旋转频率为10 Hz时,本文系统处理一帧数据所消耗的平均时间为56.83 ms,比LIO-SAM系统快19.22 ms。因此,本文系统更能满足实时性要求。

表5 不同系统处理一帧数据消耗的平均时间对比

Table 5 Comparison of average time consumed by different systems to process one frame date

系统	平均时间/ms
LIO-SAM 系统	76.05
LVI-SAM 系统	49.15
本文系统	56.83

## 4 结束语

本文提出一种视觉点线特征融合的激光雷达单目惯导SLAM系统。该系统主要由LIS系统及点线融合的PLVIS系统组成,2个子系统以紧耦合的方式构建了一个完整的激光雷达视觉惯性里程计,能够有效应对低纹理和退化场景。在初始化阶段,将LIS子系统的初始化成功与否作为判断依据,提升系统的鲁棒性,并在状态跟踪阶段观测线特征,提升整个系统在退化场景的跟踪能力。最后,PLVIS子系统将回环检测帧传递给LIS子系统,进一步修正系统轨迹。在jackal、handled以及自制的长走廊数据集上的实验结果表明,与VINS-MONO、LIO-SAM、LVI-SAM系统相比,本文系统具有更高的定位精度和鲁棒性。但该系统会由于场景中存在动态物体而导致相机跟踪失败。因此,下一步将考虑结合深度学习,剔除场景中的动态物体,实现更加鲁棒的位姿跟踪,使该系统适用于复杂环境。

## 参考文献

- [1] CADENA C, CARLONE L, CARRILLO H, et al. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age [J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [2] HUANG B C, ZHAO J, LIU J B. A survey of simultaneous

- localization and mapping[EB/OL]. [2021-07-01]. <https://www.xueshufan.com/publication/2972879745>.
- [3] WEISS S, SIEGWART R. Real-time metric state estimation for modular vision-inertial systems[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2011: 4531-4537.
  - [4] KNEIP L, WEISS S, SIEGWART R. Deterministic initialization of metric state estimation filters for loosely-coupled monocular vision-inertial systems[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C., USA: IEEE Press, 2011: 2235-2241.
  - [5] BLOESCH M, BURRI M, OMARI S, et al. Iterated extended Kalman filter based visual-inertial odometry using direct photometric feedback[J]. The International Journal of Robotics Research, 2017, 36(10): 1053-1072.
  - [6] LEUTENEGGER S, LYNEN S, BOSSE M, et al. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3): 314-334.
  - [7] PUMAROLA A, VAKHITOV A, AGUDO A, et al. PL-SLAM: real-time monocular visual SLAM with points and lines[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2017: 4503-4508.
  - [8] WEI X Y, HUANG J, MA X Y. Real-time monocular visual SLAM by combining points and lines[C]//Proceedings of IEEE International Conference on Multimedia and Expo. Washington D. C., USA: IEEE Press, 2019: 103-108.
  - [9] ZHANG F K, RUI T, YANG C S, et al. LAP-SLAM: a line-assisted point-based monocular VSLAM[J]. Electronics, 2019, 8(2): 243-247.
  - [10] CHENG S, YANG J, KANG Z, et al. A scene-assisted point-line feature based visual slam method for autonomous flight in unknown indoor environments[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2019, 13: 777-783.
  - [11] ZUO X X, XIE X J, LIU Y, et al. Robust visual SLAM with point and line features[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C., USA: IEEE Press, 2017: 1775-1782.
  - [12] GOMEZ OJEDA R, MORENO F A, ZUÑIGA-NOËL D, et al. PL-SLAM: a stereo SLAM system through the combination of points and line segments[J]. IEEE Transactions on Robotics, 2019, 35(3): 734-746.
  - [13] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.
  - [14] 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.
  - [15] HE Y J, ZHAO J, GUO Y, et al. PL-VIO: tightly-coupled monocular visual-inertial odometry using point and line features[J]. Sensors, 2018, 18(4): 1159.
  - [16] FU Q, WANG J L, YU H S, et al. PL-VINS: real-time monocular visual-inertial SLAM with point and line features[EB/OL]. [2021-07-01]. <https://arxiv.org/abs/2009.07462>.
  - [17] ZHAO S B, FANG Z, LI H L, et al. A robust laser-inertial odometry and mapping method for large-scale highway environments[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C., USA: IEEE Press, 2019: 1285-1292.
  - [18] SHAN T X, ENGLLOT B, MEYERS D, et al. LIO-SAM: tightly-coupled lidar inertial odometry via smoothing and mapping[C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Washington D. C., USA: IEEE Press, 2021: 5135-5142.
  - [19] SHAN T X, ENGLLOT B, RATTI C, et al. LVI-SAM: tightly-coupled lidar-visual-inertial odometry via smoothing and mapping[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2021: 5692-5698.
  - [20] 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.
  - [21] GALVEZ LÓPEZ D, TARDOS J D. Bags of binary words for fast place recognition in image sequences[J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.
  - [22] LUCAS B D, KANADE T. An iterative image registration technique with an application to stereo vision[EB/OL]. [2021-07-01]. <https://www.semanticscholar.org/paper/An-Iterative-Image-Registration-Technique-with-an-Lucas-Kanade/a06547951c97b2a32f23a6c2b5f79c8c75c9b9bd>.
  - [23] PONCE, HALL P. Multiple view geometry in computer vision[M]. Cambridge, U K: Cambridge University Press, 2003.
  - [24] 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.
  - [25] 戴志诚, 李小年, 陈增照, 等. 基于KNN算法的可变权值室内指纹定位算法[J]. 计算机工程, 2019, 45(6): 310-314.
  - [26] DAI Z C, LI X N, CHEN Z Z, et al. Variable-weight Indoor fingerprinting localization algorithm based on KNN algorithm[J]. Computer Engineering, 2019, 45(6): 310-314. (in Chinese)
  - [27] ZHANG G X, LEE J H, LIM J, et al. Building a 3D line-based map using stereo SLAM[J]. IEEE Transactions on Robotics, 2015, 31(6): 1364-1377.
  - [28] FORSTER C, CARLONE L, DELLAERT F, et al. On-manifold preintegration for real-time visual-inertial odometry[J]. IEEE Transactions on Robotics, 2016, 33(1): 1-21.
  - [29] CALONDER M, LEPETIT V, STRECHA C, et al. Brief: binary robust independent elementary features[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2010: 778-792.
  - [30] KAESSE M, JOHANSSON H, ROBERTS R, et al. iSAM2: incremental smoothing and mapping using the bayes tree[J]. The International Journal of Robotics Research, 2012, 31(2): 216-235.
  - [31] ZHANG J, KAESSE M, SINGH S. On degeneracy of optimization-based state estimation problems[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2016: 809-816.