

控制与决策

Control and Decision

完全在线的双目直接法视觉SLAM算法

贾嫣晗, 邹凤山, 徐方, 杜振军, 刘明敏

引用本文:

贾嫣晗, 邹凤山, 徐方, 杜振军, 刘明敏. 完全在线的双目直接法视觉SLAM算法[J]. *控制与决策*, 2023, 38(11): 3093–3102.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2021.2248>

您可能感兴趣的其他文章

Articles you may be interested in

无人系统视觉SLAM技术发展现状简析

A survey of visual SLAM in unmanned systems

控制与决策. 2021, 36(3): 513–522 <https://doi.org/10.13195/j.kzyjc.2019.1149>

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

基于深度学习的仿生集群运动智能控制

Intelligent control of bionic collective motion based on deep learning

控制与决策. 2021, 36(9): 2195–2202 <https://doi.org/10.13195/j.kzyjc.2020.0071>

磁悬浮开关磁阻电机的自适应终端滑模控制

Adaptive terminal sliding mode control of bearingless switched reluctance motor

控制与决策. 2021, 36(6): 1449–1456 <https://doi.org/10.13195/j.kzyjc.2019.1064>

基于视觉引导多AGV系统的改进A*路径规划算法

Improved A* path planning algorithm for vision-guided multi-AGV system

控制与决策. 2021, 36(8): 1881–1890 <https://doi.org/10.13195/j.kzyjc.2019.1670>

完全在线的双目直接法视觉SLAM算法

贾嫣晗^{1,2,3†}, 邹凤山^{1,2,4}, 徐方^{1,2,4}, 杜振军^{1,2,4}, 刘明敏^{1,2,4}

- (1. 中国科学院沈阳自动化研究所 机器人学国家重点实验室, 沈阳 110016;
2. 中国科学院 机器人与智能制造创新研究院, 沈阳 110169; 3. 中国科学院大学, 北京 100049;
4. 沈阳新松机器人自动化股份有限公司, 沈阳 110168)

摘要: 视觉SLAM在机器人的室外作业如野外探索、定位侦察中扮演了重要角色. 为了使得机器人可以更好地进行室外作业, 提出一种不受词袋模型的固定词汇限制的完全在线实时双目直接法视觉SLAM算法. 作为直接法视觉SLAM, 所提到的系统可以利用任何具有足够强度梯度的图像像素, 使其在缺少特征点的区域仍具有很强的鲁棒性. 在系统算法中引入双目静态残差约束并去除遮挡的滑动窗优化来增强系统的跟踪精度, 增加闭环检测和位姿图优化模块, 并建立在线词袋模型, 使得系统在大规模且陌生的环境中依然可以进行工作. 将此算法在公开的EuRoC数据集和KITTI数据集上进行性能评估, 结果表明, 所提出的系统的定位精度优于最先进的直接法视觉SLAM系统, 且室内场景和室外场景均具有鲁棒性.

关键词: 双目直接法; 在线词袋模型; 闭环检测; 去除遮挡

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2021.2248

引用格式: 贾嫣晗, 邹凤山, 徐方, 等. 完全在线的双目直接法视觉SLAM算法[J]. 控制与决策, 2023, 38(11): 3093-3102.

Fully online stereo direct vision SLAM algorithm

JIA Yan-han^{1,2,3†}, ZOU Feng-shan^{1,2,4}, XU Fang^{1,2,4}, DU Zhen-jun^{1,2,4}, LIU Ming-min^{1,2,4}

- (1. State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; 2. Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang 110169, China; 3. University of Chinese Academy of Sciences, Beijing 100049, China; 4. Shenyang SIASUN Robot & Automation Co., LTD., Shenyang 110168, China)

Abstract: Visual simultaneous localization and mapping (SLAM) plays an important role in the outdoor work of robots, such as field exploration, positioning and reconnaissance. In order to make the robots work better outdoors, this paper proposes a fully online real-time binocular direct vision SLAM algorithm which is not limited by the fixed vocabulary of the bag of word. As a direct vision SLAM, the system mentioned in this paper can make use of any image pixel with sufficient intensity gradient to make it still have strong robustness in the area lacking feature points. The system algorithm introduces binocular static residual constraint and sliding window optimization to remove occlusion to enhance the tracking accuracy of the system, adds closed-loop detection and pose map optimization modules, and establishes an online bag of word, so that the system can still work in a large-scale and unfamiliar environment. The performance of this algorithm is evaluated on the public EuRoC dataset and KITTI dataset. The positioning accuracy of the proposed system is better than the most advanced direct visual SLAM system, and is robust to indoor scenes and outdoor scenes.

Keywords: stereo direct method; online bag of word; loop detection; remove occlusion

0 引言

随着科技技术的不断发展,机器人在替代人类完成复杂特殊任务方面有着越来越高的要求. 特别是在野外探索、太空探测、定位侦察等任务中,机器人扮演着极为重要的角色,这便要求机器人具有自主定位和建图的功能,且对于定位的精确性和鲁棒性有着

很高的要求.

视觉同时定位与建图(simultaneous localization and mapping, SLAM)中,主要使用的传感器有单目相机、深度相机、双目相机3种. 单目相机^[1]必须通过运动的过程估计深度,且具有尺度不确定性,需要进行专门的初始化. 深度相机^[2]可直接进行深度测量,但

收稿日期: 2021-12-27; 录用日期: 2022-07-04.

基金项目: 国家自然科学基金项目(U20A20197).

责任编辑: 高会军.

†通讯作者. E-mail: jiayanhan20@mails.ucas.ac.cn.

测量范围较窄,且容易受到日光的干扰,故只能应用于室内场景.相比之下,双目相机^[2]可根据左右视差计算深度,且固定基线的双目相机具有真实的尺度信息,受光线影响较小,可同时适用于室内室外场景.

视觉SLAM系统通常由前端视觉里程计、后端优化、闭环检测以及地图构建4部分组成.在前端跟踪丢失的情况下,闭环检测可以进行重定位,起到修正的作用,对于大范围定位导航的精度有重要意义.现在主流的视觉SLAM主要分两种,根据跟踪相机位姿方式不同分为特征点法^[3]和直接法^[4].特征点法通过提取图像中的特征点描述子进行特征点匹配,通过最小化重投影误差计算获得位姿,提取的特征点可用于重定位和闭环检测,但特征提取和匹配的过程需花费大量时间.直接法则是通过最小化光度误差进行位姿估计,无需特征提取和匹配过程,因此耗时短,但无法有效抑制累计误差造成的位姿漂移.

PTAM(parallel tracking and mapping)算法^[5]是早期提出的基于特征法的视觉SLAM算法,它可以实现跟踪建图两线程的同时进行,并使用BA(bundle adjustment)进行实时约束.现在主流的特征点法视觉SLAM以ORB-SLAM(oriented fast and rotated brief SLAM)^[3]为代表,在此基础上进行延伸,ORB-SLAM对于剧烈运动具有鲁棒性,使用多线程结构实现跟踪、建图、回环检测同时进行.在ORB-SLAM的基础上,提出了可用于单目、双目和深度相机的ORB-SLAM2^[2],该系统更加完善和容易使用.基于特征点的方法对于光照变化不敏感,但在弱纹理区或特征点缺失的部分无法使用,且只能进行稀疏特征点建图,对于实际导航意义不大.

经典的直接法视觉SLAM框架有Engel等^[4]提出的基于直接法的半稠密算法LSD-SLAM(large-scale direct monocular SLAM),它可以创建出更好地表示场景纹理信息的半稠密地图,并且可以检测修正尺度漂移,有利于大场景的应用;SVO(fast semi-direct monocular visual odometry)是Forster等^[6]于2014年提出的半直接法视觉里程计,先提取特征点,然后利用光流法估计位姿,但缺少后端优化和闭环检测的部分.单目DSO(direct sparse odometry)^[7]是在LSD-SLAM基础上提出的基于稀疏直接法的视觉里程计,该算法将相机内参和位姿、地图点云的逆深度值等所有参数都放入优化算法中,并采用相机光度标定模型,充分考虑了透镜晕减、相机曝光参数等因素的影响,使得算法精确度更高、鲁棒性更强,但在定位过程中还是有较大的漂移;双目DSO算法^[8]的提出进一步提高了视觉里程计的定位精度,在一个边缘

化框架内使用Schur补体将时间多视角立体视觉与静态立体视觉整合来达到全局优化,但是由于没有闭环检测功能,随着时间增加仍然会产生漂移误差;LDSO(direct sparse odometry with loop closure)^[9]提出将特征点法的闭环检测模块融入到DSO视觉里程计中,解决了这一问题,但在无闭环的路段中具有较大的漂移误差,且在重合度较差的闭环路段,LDSO中的闭环检测算法会失效,LDSO采用的闭环检测算法主要延续的是ORB-SLAM2的闭环检测中的方法,采用已训练好的离线的词袋模型,这种方法并不适用于所有的环境;OV2-SLAM^[10]中选择的一种iBoW-LCD^[11]的变更版本作为词袋模型,可以创建适合当前环境的词汇树,实现完全在线定位.

本文提出一种可以适合所有场景的带有闭环检测的双目直接法视觉SLAM算法.主要贡献在于:1)通过双目基线引入静态残差,并且去除滑动窗口优化中无法被观测到的遮挡点,解决滑窗优化中的遮挡问题,提高跟踪精度;2)加入闭环检测模块和位姿图优化模块,通过Shi-Tomasi分数^[12]改进ORB特征提取少量特征点,对闭环匹配帧的特征点进行位姿估计,通过最小化3D和2D的能量误差进行位姿图优化;3)引入iBoW-LCD^[11]词袋模型,通过图像中的描述子构建在线词汇树,使得词汇树适合当前环境,并修改iBoW-LCD中对初始集群中心选择的方法,使初始集群中心分散.

1 完全在线的双目直接法视觉SLAM系统框架

完全在线的双目直接视觉SLAM系统是在单目DSO视觉里程计^[7]原始框架的基础上改进的.完全在线是相对于使用离线词袋模型的视觉SLAM系统而言.绝大部分的视觉SLAM算法在闭环检测部分使用已训练好的词袋模型,字典是固定的,只在一定的环境中适用.而完全在线的视觉SLAM算法在闭环检测模块使用在线词袋模型,可以根据周围环境的变化在词袋模型中加入词汇,从而实现算法的完全在线.

在本算法的实现中,首先,在系统的前端部分,使用双目相机向系统中输入双目图像,将双目图像曝光处理,而后进行双目立体匹配得到图像的深度信息.将系统初始化后,利用直接图像对齐的方法,引入两对光度参数减小光照强度变化对系统的影响,通过最小化光度误差对点进行追踪,从而获取图像的位姿信息.然后判断图像是否为关键帧,在关键帧中选择一定数目的点进行滑动窗口优化.

在滑动窗口优化时引入双目静态残差并通过去

除滑动窗口中被遮挡的点提高视觉里程计的定位精度. 系统从选择的点中, 通过使用对于角落信息更加敏感的 Shi-Tomasi 分数提取角点, 从角点中提取特征用于闭环检测的匹配. 接着, 在闭环检测部分, 使用改进的 iBoW-LCD 模型建立在线的词汇树对系统闭环检测, 在这个过程中, 需要提取额外的特征, 这些特征

仅用来建立和更新词汇树, 然后对检测的图像进行相似性比对. 图像检测到闭环匹配时, 通过最小化代价函数优化位姿, 与滑动窗口优化结合达到全局位姿优化的效果, 最终得到优化后的地图.

图 1 为完全在线的双目直接法视觉 SLAM 系统框架, 框出的部分是系统较双目 DSO 改进的部分.

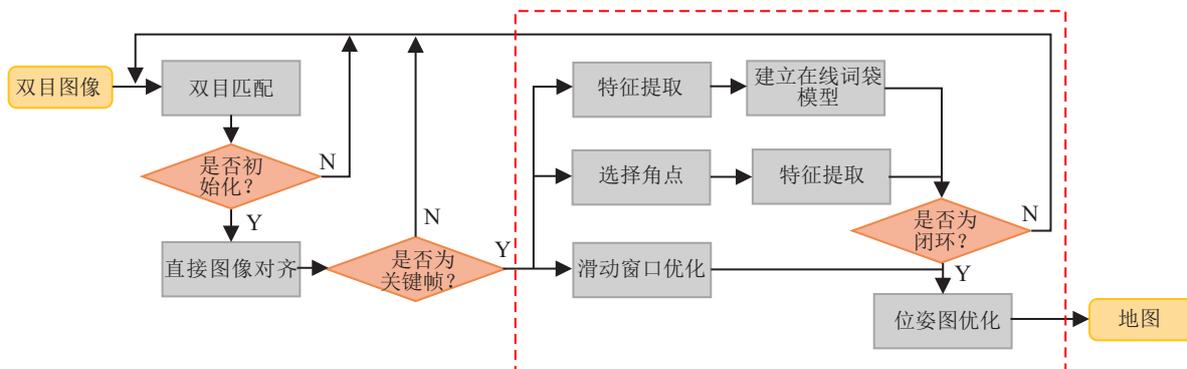


图 1 完全在线的双目直接法视觉 SLAM 系统框架

2 完全在线的双目直接法视觉 SLAM 算法实现

2.1 位姿追踪

完全在线的双目直接视觉 SLAM 算法是将双目相机作为传感器传入成对的图像, 从双目相机的基线可以直接计算出绝对尺度信息.

2.1.1 初始化

单目视觉初始化利用的是随机深度, 而双目视觉可以通过匹配获得估计值. 在双目图像进行曝光处理后, 根据左右两图像之间的误差计算每个像素的深度值. 一般的, 根据左侧相机设置相机坐标系. 双目相机的几何成像模型如图 2 所示.

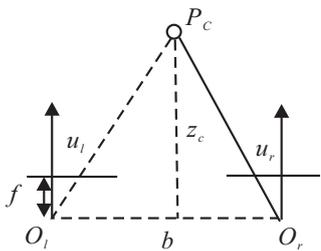


图 2 双目相机几何模型

在图 2 中: 点 $P_c = (x_c, y_c, z_c)$ 为地图点 P 在相机坐标系下的坐标; O_l 为左相机坐标系的原点; O_r 为右相机坐标系原点; f 为相机焦距; b 为左右相机间固定基线的长度; u_l 、 u_r 分别为左右成像平面的坐标, $d = u_l - u_r$ 为左右相机的视差; 根据三角形相似性可以得到地图点的深度值为

$$z_c = \frac{fb}{d}. \quad (1)$$

直接双目视觉 SLAM 不需要通过一定的初始运

动来使用前帧的三角化进行深度估计, 而是通过双目图像的立体匹配建立第 1 帧的深度地图, 系统使用 3×5 邻域的归一化 NCC (normalized cross-correlation) 搜索水平对极线进行匹配. 通过双目匹配获得的较好的深度值可以增加跟踪的精度.

2.1.2 直接图像对齐

直接双目视觉 SLAM 系统的位姿估计是通过直接图像对齐的方式实现的, 如图 3 所示.

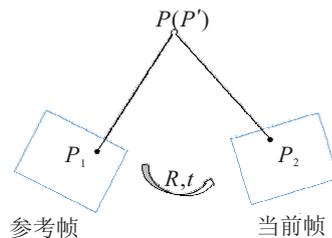


图 3 直接图像对齐投影关系

直接图像对齐需要根据图像中的像素强度计算, 这就使得直接法系统受光照变化的影响很大. 因此通过图像引入参数 a_i 和 b_i 对像素强度变化进行建模, 直接图像对齐时的光度误差能量模型可以表达为

$$E_{ij}^p = \sum_{\tilde{p} \in N_P} \omega_{\tilde{p}} \left\| (I_j[\tilde{p}'] - b_j) - \frac{\tau_j e^{a_j}}{\tau_i e^{a_i}} (I_i[\tilde{p}] - b_i) \right\|_\gamma, \quad (2)$$

$$E_{ij} = \sum_{p \in P_i} E_{ij}^p. \quad (3)$$

其中: 点集 P 为参考帧中可以被当前帧观测到的一系列地图点, 点集 N_P 为定义的 8 点 Pattern; \tilde{p} 为 Pattern 中的投影点; I_i 为最近关键帧的图像; I_j 为当前帧的图像; τ_i 和 τ_j 分别为最近关键帧和当前帧的曝光时

间; $\|\cdot\|$ 为Huber范数; ω 为对高图像梯度进行降权的权重.

2.1.3 去除一定遮挡的滑动窗口优化

算法中的滑动窗口优化在DSO滑窗优化^[7]的策略上进行改进, 当时间较长的点从活动窗口边缘化时, 候选点激活并加入联合优化. 每个激活点存储在一个关键帧中并且可被活动窗口内的一些其他关键帧观测. 每次一个激活点被另外关键帧观测到, 就会创造光度损失函数内部的光度能量因子. 系统将双目约束引入到能量因子图中, 每个因子依赖于该点的逆深度、主关键帧和观测到该点的关键帧及其亮度校正因子即2.1.2节的光度仿射参数 a_i 和 b_i , 提高了滑动窗口优化的性能.

式(2)可以被称为动态残差, 对于双目相机而言, 可以引入两对光度仿射参数 a_i^L 和 b_i^L 、 a_i^R 和 b_i^R , 并通过双目固定极限引入左右相机的静态残差

$$E_{is}^p = \omega_p \left\| \left(I_i^R[p'] - b_i^R \right) - \frac{\tau_i^R e^{a_i^R}}{\tau_i^L e^{a_i^L}} \left(I_i^L[p] - b_i^L \right) \right\|_\gamma. \quad (4)$$

算法中使用的滑动窗口的代价函数为

$$E = \sum_{p \in P_i} E_{ij}^p + \lambda E_{is}^p. \quad (5)$$

令 $x = [x_p^T \ x_d^T]^T$. 其中: x_p 为畸变后的当前活动窗口的关键帧中点的相机位姿, x_d 为点的逆深度, DSO中最小化光度误差根据列文伯格-马夸尔特(L-M)迭代求解可表达为

$$H \Delta x = -J^T W r. \quad (6)$$

其中: H 为Hessian矩阵, Δx 为整体自变量的最优增量, J 为雅可比矩阵, W 为权重矩阵, r 为残差. 由于 H 为类似箭头的稀疏模式, 可以执行BA优化.

然而, 滑窗优化的关键帧之间可能存在大的视角变化或者平移变换, 使得主导帧优化时存在遮挡点, 这些点无法被观测帧观测到, 从而导致上述的代价函数模型不准确, 使得滑窗优化后得到的位姿精度大大降低. 因此, 在计算对应像素的光度误差之前, 需要先剔除掉被观测到的关键帧所遮挡的点. 对于图中的点 p , 经位姿变换后得到点 p' , 将经过变换的 p' 点 z 轴坐标与双目匹配后得到深度图中对应的该点深度值进行比较, 如果两值相差过大, 则说明该点被遮挡, 将其从最小化光度误差模型中删除.

2.1.4 角点选择

直接法中, 依然需要选择点来进行追踪, 但不需要点的可重复性, 这就使得像DSO这样的直接法不能实现闭环检测. 因此, 在本系统中, 在挑选像素点的过程中, 通过使用对于角落信息更加敏感的 Shi-

Tomasi分数^[12]将具有高梯度的像素分为角点和非角点, 为了后续的闭环检测和位姿图优化, 算法保留角点, 对于非角点则不需要它们的可重复性. 从选取的少量的角点中提取特征信息, 将每幅图像中提取的特征与其他图像进行匹配, 对闭环检测进行判断.

2.2 闭环检测

虽然双目DSO算法已经具备了较好的定位精度, 但在大规模地图中, 依然会存在尺度漂移, 而无法进行合适的修正. 加入闭环检测模块可以使得定位算法在大规模地图中保持精准定位. 闭环检测可以实现对轨迹的重定位, 对当前帧的位姿与实际位姿的差距进行修正.

2.2.1 在线词袋模型

词袋模型(BoW)^[13]已被证明是非常有效的快速闭环检测. 通过建立特征字典和从特征字典中进行搜索来快速计算不同关键帧之间的相似度得分. 但是, 大多数SLAM算法都是使用离线训练的词袋模型, 因此依赖于训练数据库, 存在偏差. 另一种策略是利用目前获得的图像建立在线词袋模型. 本文算法创建一个适合当前环境的词袋模型, 从而避免离线受训的词袋模型潜在的拟合不足的问题.

iBoW-LCD^[11]是一种在线建立词袋模型的方法, 采用文献[14]介绍的一种层次树的方法处理增量的可视化学字典, 这种层次树可以实时地添加和删除描述符. 这个结构包含在一个树中, 其中非叶节点包含集群中心, 叶节点存储要匹配的视觉描述符. 因此, 增量词汇表的可视化单词存储在叶节点中. 为了构建这些树, 算法从初始点集中随机选择 K 个描述子作为集群中心. 接下来, 根据汉明距离将每个剩余的输入描述符分配到其最近的集群中心. 这个过程会递归地重复, 直到集群中的描述符数量低于某个阈值. 这种方法可以构建几棵树, 在搜索过程中并行使用它们可以获得更高的性能. 但这种方法随机选取集群中心, 可能导致结果与实际分布差距较大.

系统借鉴AFK-MC2(fast and provably good seedings for k -means)聚类算法^[15]的思想, 使用无假设的马尔科夫链蒙特卡罗采样法改进层次树中的选取集群中心的方法, 将距离有关的分布加入到原始分布, 从初始点集 X 中随机选择一个描述子 c_1 作为集群中心, 对点 $x \in X$ 使用汉明距离进行非均匀定义

$$q(x|c_1) = \frac{d(x, c_1)^2}{2 \sum_{x' \in X} d(x', c_1)^2} + \frac{1}{2|X|}. \quad (7)$$

使用 $q(x)$ 从点集 X 中采样出长为 M 的数列, 取最后 $k-1$ 个作为集群中心, 得到 K 个分布不集中的集群

中心,其余步骤与iBoW-LCD模型相同。

图4显示了在kitti00序列上本文算法中的在线词袋模型词汇量以及每张图片的平均响应时间随着时间的变化曲线,每100ms进行一次统计。词汇量的变化主要与图片的数量有关。尽管词汇量随着处理图像的增多而增加,但随着图片中相似区域的增加,新添加的词汇会相对减少,因此图片的平均响应时间保持更稳定,并存在适度的增加。

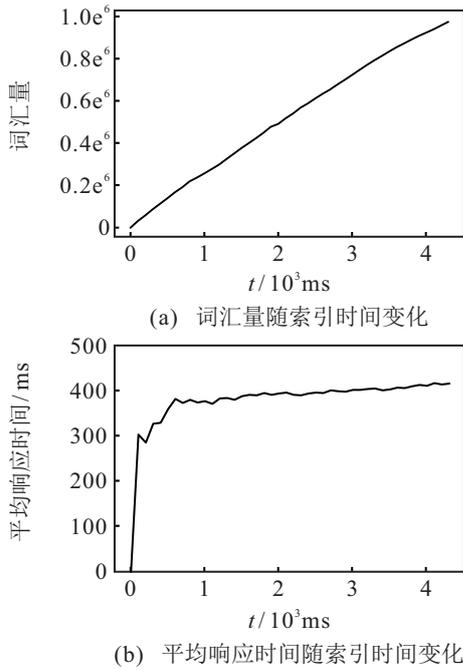


图4 在线词袋模型在kitti00序列上词汇量和平均响应时间变化

2.2.2 词袋模型特征

在接收到一个新的关键帧后,执行一个特征提取步骤。前面提到系统只使用角点进行特征相似性匹配和位姿优化,因此选用的特征较少,就需要在更新词袋模型之前提取额外的特性。使用FAST检测器和BRIEF描述符提取特征,为了避免算法占用内存的提升与算法速度降低,这里提取的特征只用于更新词袋模型,系统依然只使用前面的角点提取到的BoW特征计算与之前关键帧的相似度得分。

2.2.3 位姿图优化

使用基于g2o的位姿图优化方法调整全局位姿。当图像检测到闭环匹配时,开始进行全局位姿调整。由于滑动窗口中包含绝对位姿信息,系统使用的是已边缘化的关键帧进行位姿图优化。首先对闭环匹配的关键帧使用PNP做初始的位姿变化估计,然后通过最小化3D和2D几何约束对其他关键帧的位姿进行优化,从而达到位姿图优化的结果。

令 P' 为闭环匹配帧中角点的集合, P 为当前关键帧中角点的集合, d 为匹配点的逆深度值, P_1 为 P

中深度值在之前所求得的深度图中未知的点的集合, P_2 为 P 中深度值在之前所求得的深度图中已知的点的集合,通过最小化下列代价函数 E 求解闭环匹配帧到关键帧的位姿变化 S_{cr} ,即

$$E = \sum_{p_i \in P_1} \omega_1 \|S_{cr} \Pi^{-1}(p'_i, d_{p'_i}) - \Pi^{-1}(p_i, d_{p_i})\|_2 + \sum_{p_j \in P_2} \omega_2 \|\Pi(S_{cr} \Pi^{-1}(p'_j, d_{p'_j})) - p_j\|_2. \quad (8)$$

其中: $\Pi: \mathbf{R}^3 \rightarrow \Omega$ 为三维空间点到二维空间点直接图像对齐的投影函数, $\Pi^{-1}: \Omega \times \mathbf{R} \rightarrow \mathbf{R}^3$ 为二维空间点到三维空间点直接图像对齐的反投影函数, ω_1 和 ω_2 是用来使得公式中单位统一的权值。一般而言,可以仅通过使用最小化3D约束估计位姿,系统加入2D点最小化重投影误差来降低计算深度值时的噪声对于系统结果的影响。

3 算法性能分析

为了验证系统算法的各项性能,本文分别在室内场景的EuRoC数据集^[16]和室外场景的KITTI数据集^[17]上进行算法功能的验证。

在算法评估中,将本文中的算法与基于直接法的双目DSO、双目LSD-SLAM、双目SVO和基于特征点法的双目ORB-SLAM2以及实时的OV2-SLAM进行比较来验证本文算法的优势。在评估过程中主要通过取平均相关平移误差 t_{rel} 和旋转误差 r_{rel} ^[17]以及均方根误差RMSE(使用 t_{abs} 或ATE^[18]表示)进行对比,将结果最好的数据用粗体标出。为了保持数据的准确性,本文在评价其他算法时使用原论文中的数据或者在GitHub上发布的数据。

3.1 EuRoC数据集

EuRoc数据集提供了两个室内房间和一个工厂内部的序列图像,包含简单、中等、复杂3种难度的数据集。在此数据集上,进行视觉里程计和视觉SLAM算法的评估。从EuRoc数据集上进行精度对比,与经典的双目开源算法SVO、双目DSO、ORB-SLAM2以及OV2-SLAM进行对比,结果如表1和表2所示。

从表1结果可以看出,本文提到的算法在视觉里程计部分要整体优于其他视觉里程计算法,因此在室内场景下去除一定遮挡点可以提升视觉里程计算法的精度。

从表2可以看出,完全在线的双目视觉SLAM算法在与离线模型契合的环境中与使用离线词袋模型的ORB-SLAM2算法相当,但效果明显好于使用在线词袋模型的ORB-SLAM2算法和OV2-SLAM算法。因此,完全在线的双目视觉SLAM算法可以适用于室内环境,但由于室内范围较小,且环境与离线词

袋模型较为契合,算法提升效果并不明显,因此,本文着重于在大规模的室外场景下进行精度效果对比。

表1 视觉里程计精度对比ATE 单位: m

Seq.	length	SVO ^[10]	stereoDSO	OV2-SLAM ^[10]	ours
MH 01	79.84	0.04	0.05	0.05	0.04
MH 02	72.75	0.05	0.05	0.04	0.05
MH 03	130.58	0.06	0.07	0.05	0.06
MH 04	91.55	×	0.16	0.12	0.11
MH 05	97.32	0.12	0.08	0.10	0.07
V1 01	58.51	0.05	0.05	0.09	0.05
V1 02	75.72	0.05	0.06	0.08	0.06
V1 03	78.77	×	0.22	0.17	0.18
V2 01	36.34	0.05	0.04	0.10	0.05
V2 02	83.01	×	0.10	0.10	0.08

表2 视觉SLAM精度对比ATE 单位: m

Seq.	ORB-SLAM2 ^[6]	ORB-SLAM2(RT) ^[10]	OV2-SLAM ^[10]	ours
MH 01	0.04	0.14	0.04	0.04
MH 02	0.02	0.12	0.04	0.05
MH 03	0.03	0.31	0.04	0.04
MH 04	0.12	0.25	0.06	0.05
MH 05	0.06	0.28	0.07	0.06
V1 01	0.04	0.18	0.09	0.05
V1 02	0.02	×	0.07	0.06
V1 03	0.05	×	0.09	0.07
V2 01	0.04	0.29	0.07	0.05
V2 02	0.04	×	0.06	0.06

3.2 KITTI数据集

3.2.1 消融实验

KITTI数据集提供了室外大规模场景的双目序列矫正高分辨率图像。在KITTI数据集上将算法中的

改进部分(滑动窗口优化时去除遮挡点、加入闭环检测和位姿图优化模块、使用在线词袋模型)进行消融实验,验证各部分改进的算法效果。

表3 完全在线的双目直接法视觉SLAM算法消融实验

Seq.	ORB-SLAM2 ^[6]	ORB-SLAM2(RT) ^[10]	OV2-SLAM ^[10]	ours
MH 01	0.04	0.14	0.04	0.04
MH 02	0.02	0.12	0.04	0.05
MH 03	0.03	0.31	0.04	0.04
MH 04	0.12	0.25	0.06	0.05
MH 05	0.06	0.28	0.07	0.06
V1 01	0.04	0.18	0.09	0.05
V1 02	0.02	×	0.07	0.06
V1 03	0.05	×	0.09	0.07
V2 01	0.04	0.29	0.07	0.05
V2 02	0.04	×	0.06	0.06

从表3可以看出,对双目DSO去除一定遮挡点后视觉里程计定位精度的效果得到了一定的提升。在加入闭环检测模块和位姿图优化之后,在可以检测到闭环的序列上,系统的定位精度得到了较大的提升,降低了由漂移带来的累计误差。在离线词袋模型无法检测到闭环的08,09序列上,使用离线词袋模型的系统精度效果与视觉里程计相比无明显变化。而使用在线词袋模型的情况下,在08,09这种不适用于离线词袋模型的闭环序列上依然可以检测到闭环并进行位姿图优化,从而提升精度效果。

3.2.2 与现有的双目相机SLAM算法的比较

关闭系统的闭环检测模块和位姿图优化模块,将视觉里程计部分与双目DSO、ORB-SLAM2的视觉里程计、LSD的视觉里程计进行比较,精度评估结果如表4所示。图5为本文算法与双目DSO、ORB-SLAM2在00、04、05、09序列上的轨迹图对比。

表4 视觉里程计精度对比

Seq.	stereoDSO ^[8]		ORB-SLAM2 ^[8]		LSD-VO ^[3]		ours	
	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$
00	0.84	0.26	0.83	0.29	1.09	0.42	0.79	0.27
01	1.43	0.09	1.38	0.20	2.13	0.37	1.41	0.16
02	0.78	0.21	0.81	0.28	1.09	0.37	0.73	0.25
03	0.92	0.16	0.71	0.17	1.16	0.32	0.91	0.15
04	0.65	0.15	0.45	0.18	0.42	0.34	0.68	0.12
05	0.68	0.19	0.64	0.26	0.90	0.34	0.61	0.22
06	0.67	0.20	0.82	0.25	1.28	0.53	0.72	0.18
07	0.83	0.36	0.78	0.42	1.25	0.79	0.76	0.32
08	0.98	0.25	1.07	0.31	1.24	0.38	1.03	0.22
09	0.98	0.18	0.82	0.25	1.22	0.28	0.92	0.18
10	0.49	0.18	0.58	0.28	0.75	0.34	0.65	0.22

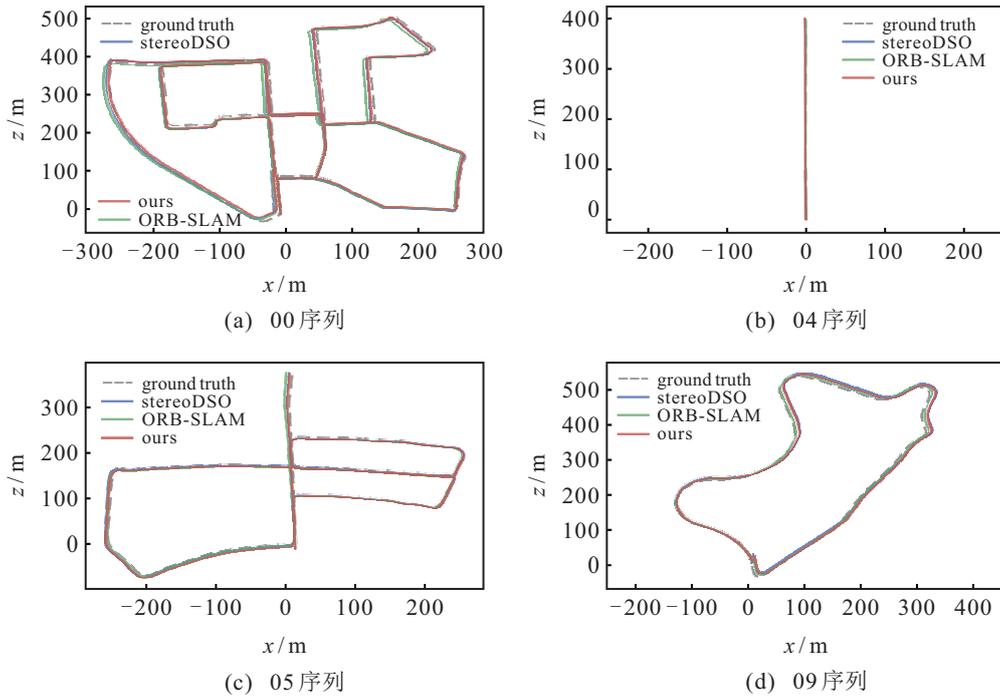


图5 本文算法的视觉里程计与双目DSO、ORB-SLAM2的视觉里程计在00、04、05、09序列上的轨迹对比

整体而言,本文算法的定位精度效果在有较大视角变化的数据集中要优于双目DSO算法,与ORB-SLAM2算法的精度效果相当,但ORB-SLAM2算法中存在可以使定位精度大幅提升的局部BA优化.由此可以看出,在滑动窗口优化的过程中,去除掉一定

数目的被遮挡点,可以提高优化效果,增加定位精度.

先将离线的词袋模型加入到系统中,与同为使用离线词袋模型的两个分别为特征点法和直接法中最先进的双目视觉SLAM算法进行精度的比较,用于评估系统在离线时的整体性能,评估结果如表5所示.

表5 使用离线模型的SLAM系统精度对比

Seq.	ORB-SLAM2 ^[6]			LSD-SLAM ^[3]			ours (离线)		
	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$	t_{abs}/m	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$	t_{abs}/m	$t_{rel}/\%$	$r_{rel}/(^{\circ}/100\text{m})$	t_{abs}/m
00	0.70	0.25	1.3	0.63	0.2	1.0	0.62	0.23	0.9
01	1.39	0.21	10.4	2.36	0.6	9.0	1.42	0.10	8.7
02	0.76	0.23	5.7	0.79	0.2	2.6	0.74	0.22	5.0
03	0.71	0.18	0.6	1.01	0.2	1.2	0.69	0.14	0.6
04	0.48	0.13	0.2	0.38	0.3	0.2	0.51	0.16	0.3
05	0.40	0.16	0.8	0.64	0.1	1.5	0.44	0.15	0.8
06	0.51	0.15	0.8	0.71	0.1	1.3	0.53	0.17	1.0
07	0.50	0.28	0.5	0.56	0.2	0.5	0.48	0.30	0.4
08	1.05	0.32	3.6	1.11	0.3	3.9	0.96	0.22	3.2
09	1.05	0.32	3.6	1.11	0.3	3.9	0.96	0.22	3.2
10	0.60	0.27	1.0	0.72	0.3	1.5	0.68	0.24	1.2

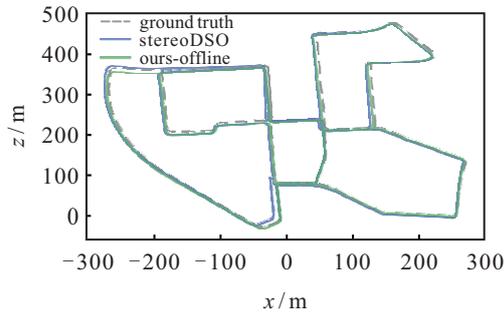
从表5中可以看出,本文中的算法效果要优于ORB-SLAM2和LSD-SLAM算法,且ORB-SLAM2对于特征点要求较高,在弱纹理或空白区域中算法会失效,而本文算法视觉里程计采用最小化光度误差的方法,可以避免弱纹理区域失效的问题.而且结合表

4可以看出,在具有闭环回路的数据集上,本文中的系统效果要远优于双目DSO算法,图6显示了00序列和05序列时使用离线词袋模型的系统与双目DSO在轨迹图、平移误差和旋转误差上的差距.随着距离的增大,由于双目DSO无法闭环检测,漂移误差会逐

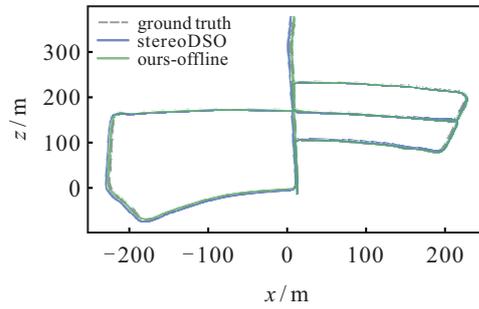
渐积累,因此并不适合长距离定位.

本文中的系统使用改进初始集群中心的IBOW-LCD模型实现实时的系统算法,将算法与其他实时

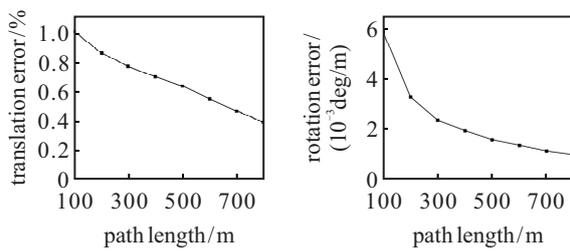
的OV2-SLAM算法和实时的ORB-SLAM2算法以及本文系统在离线时的算法进行比较,评估改进后的词袋模型的效果.



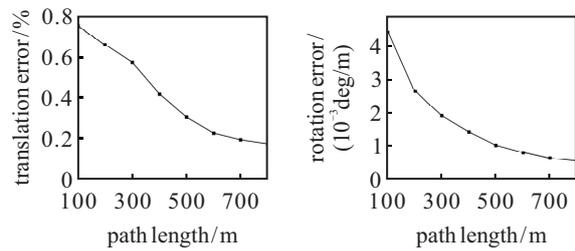
(a) 00序列上ours-offline与stereoDSO轨迹对比



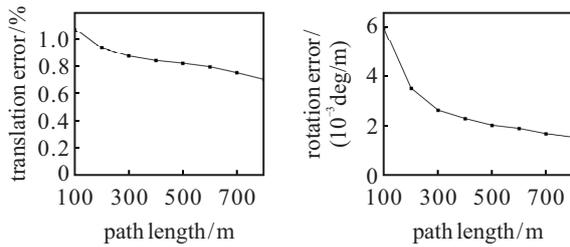
(b) 05序列上ours-offline与stereoDSO轨迹对比



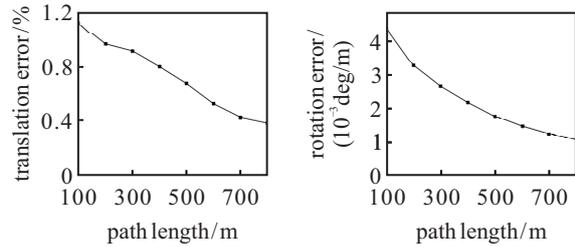
(c) 00序列ours-offline随距离变化的平移误差和旋转误差



(d) 05序列ours-offline随距离变化的平移误差和旋转误差



(e) 00序列stereoDSO随距离变化的平移误差和旋转误差



(f) 05序列stereoDSO随距离变化的平移误差和旋转误差

图6 00、05序列上ours-offline与stereoDSO对比

表6 使用在线模型的SLAM系统ATE(m)精度对比

Seq.	ours(离线)	ORB-SLAM2 ^[10]		OV2-SLAM ^[10]		ours	
		RT(10Hz)	RT(5Hz)	RT(10Hz)	RT(5Hz)	RT(10Hz)	RT(5Hz)
00	0.9	10.74*	3.79	1.17	10	0.94	0.88
01	8.7	341.81	71.62	31.95	26.34	22.62	18.34
02	5.0	×	34.26	6.24	6.45	6.04	5.68
03	0.6	2.55	0.58	1.25	1.2	0.55	0.60
04	0.3	5.76*	3.86	1.18	1.7	0.95	0.65
05	0.8	9.15	3.70	1.44	1.3	0.73	1.25
06	1.0	×	4.92	1.27	0.17	1.34	0.64
07	0.4	1.07	0.53	0.37	0.3	0.25	0.32
08	3.2	5.41*	3.81	3.68	3.7	3.10	2.56
09	3.2	5.11*	2.92	1.59	1.8	1.32	1.22
10	1.2	2.91*	1.11	0.66	0.64	0.78	0.84

从表 6 中可以看出, 本文的算法效果较其他在线 SLAM 算法(除了在离线的词袋模型更适用的场景以及距离较短的数据集外), 实时的系统算法效果均优于离线的系统算法, 特别是在 08, 09 两个序列中, 由于闭环相似度低, 离线的系统算法无法检测到闭环结果, 而在实时的算法中, 系统可以完成闭环检测并进行位姿优化. 图 7 显示了 09 序列上 ours 与 ours-offline 轨迹图对比, 图 8 显示了本文中的算法在 09 序列上的各项数据.

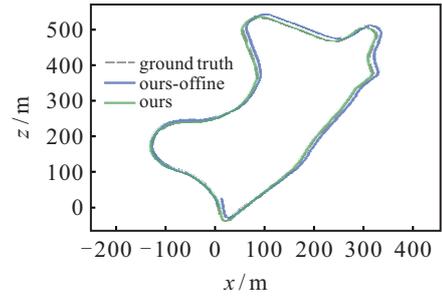
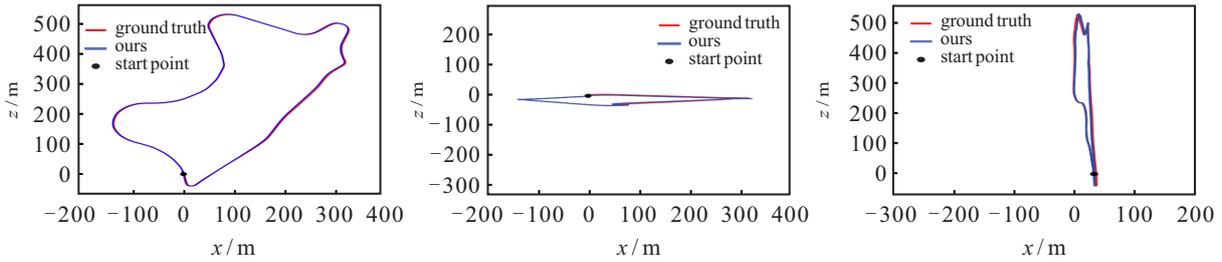
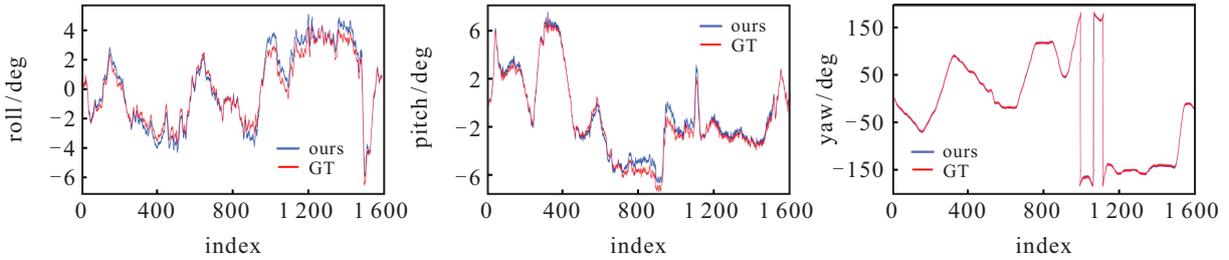


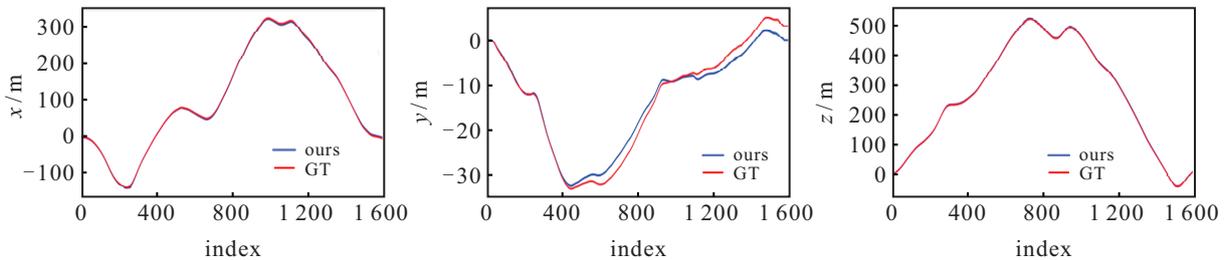
图 7 09 序列上 ours 与 ours-offline 轨迹对比



(a) 09 序列在 xz 、 xy 、 yz 平面的轨迹



(b) 09 序列轨迹的 rpy 视角对比



(c) 09 序列轨迹的 xyz 视角对比

图 8 09 序列上 ours 各项数据

3.3 运行时间评估

由于闭环检测和位姿图优化在单独线程中运行, 不会对主线程的运行时间造成影响. 然而, 由于在主线程中需要提取额外的特征点, 主线程的运行时间较双目 DSO 会有一定的增长. 在主线程、闭环线程、建图线程上本文算法与双目 DSO 算法中平均每帧的运行时间以及点选择时间如表 7 所示. 尽管闭环线程运行时间较长, 但整体上对系统的输出频率影响不大.

表 7 运行时间对比 单位: ms

	点选择时间	主线程	闭环线程	建图线程
stereoDSO	14	52	*	47
ours	27	71	435	83

4 结论

本文介绍了一种可以用于室内场景和室外场景的实时的双目直接法视觉 SLAM 系统. 本文详细介绍了该系统的实现过程, 并在 EuRoC 数据集上与 SVO、双目 DSO、ORB-SLAM2 算法进行对比, 在 KITTI 数据集中与双目 DSO、ORB-SLAM2、LSD-SLAM 以及 OV2-SLAM 算法进行全面的评估. 评估结果表明, 本文系统的精度要优于当前先进的视觉 SLAM 算法, 而且本文中提到的系统在室内场景和室外场景下都显示出较好的鲁棒性.

在未来的工作中, 计划将动态对象考虑到系统中来, 去除系统中运动的物体, 进一步提高定位和建图

的精度和鲁棒性,并计划通过适量降低在线模型中词汇量的数目进一步提高算法的运行效率。

参考文献(References)

- [1] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]. The 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, 2008: 225-234.
- [2] 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.
- [3] 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.
- [4] Engel J, Schöps T, Cremers D. LSD-SLAM: Large-scale direct monocular SLAM[C]. European Conference on Computer Vision. Berlin, 2014: 834-849.
- [5] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]. The 6th IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, 2007: 225-234.
- [6] Forster C, Pizzoli M, Scaramuzza D. SVO: Fast semi-direct monocular visual odometry[C]. 2014 IEEE International Conference on Robotics and Automation. Hong Kong, 2014: 15-22.
- [7] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.
- [8] Wang R, Schwörer M, Cremers D. Stereo DSO: Large-scale direct sparse visual odometry with stereo cameras[C]. 2017 IEEE International Conference on Computer Vision. Venice, 2017: 3923-3931.
- [9] Gao X, Wang R, Demmel N, et al. LDSO: Direct sparse odometry with loop closure[C]. 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, 2018: 2198-2204.
- [10] Ferrera M, Eudes A, Moras J, et al. OV2SLAM: A fully online and versatile visual SLAM for real-time applications[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 1399-1406.
- [11] Garcia-Fidalgo E, Ortiz A. iBoW-LCD: An appearance-based loop-closure detection approach using incremental bags of binary words[J]. IEEE Robotics and Automation Letters, 2018, 3(4): 3051-3057.
- [12] Shi J B, Tomasi. Good features to track[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, 1994: 593-600.
- [13] 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.
- [14] Muja M, Lowe D G. Fast matching of binary features[C]. The 9th Conference on Computer and Robot Vision. Toronto, 2012: 404-410.
- [15] Bachem O, Lucic M, Hassani H, et al. Fast and provably good seedings for k -means[C]. Advances in Neural Information Processing Systems. 2016: 55-63.
- [16] Burri M, Nikolic J, Gohl P, et al. The EuRoC micro aerial vehicle datasets[J]. The International Journal of Robotics Research, 2016, 35(10): 1157-1163.
- [17] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: The KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [18] Sturm J, Engelhard N, Endres F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]. IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura-Algarve, 2012: 573-580.

作者简介

贾嫣晗(1998—),女,硕士生,从事机器人视觉定位导航的研究, E-mail: jiayanhan20@mails.ucas.ac.cn;

邹凤山(1980—),男,高级工程师,博士,从事服务机器人导航与控制、智能机器人控制等研究, E-mail: zoufengshan@siasun.com;

徐方(1962—),男,教授,硕士,从事机器人视觉和智能感知的研究, E-mail: xufang@siasun.com;

杜振军(1981—),男,高级工程师,博士,从事服务机器人导航、机器人感知算法、图像处理等研究, E-mail: duzhenjun@siasun.com;

刘明敏(1992—),男,工程师,博士,从事机器人控制的研究, E-mail: liumingmin@siasun.com.