doi:10.3772/j.issn.1002-0470.2021.07.001

基于新型多传感器融合策略的移动端双目视觉惯性 SLAM 闭环算法研究①

任金伟② 郑 鑫 李昱辰 朱建科③

(浙江大学计算机科学与技术学院 杭州 310027)

摘 要 研究了复杂环境下基于视觉同时定位与地图构建(SLAM)算法的移动端实时定位问题。该问题有如下几个难点:首先是移动端设备的计算资源受限,这对算法的优化与解算效率提出了更严格的要求;其次是测试场景的复杂多变,使得算法在低纹理及快速运动等情况下容易丢失目标;最后是实际应用时对系统的可拓展性要求较高,需要具备复杂场景下的适应性。针对上述问题,本文提出了面向移动端的双目视觉惯性 SLAM 算法,采用新型多传感器融合策略,通过将双目视觉图像和惯性测量数据进行紧耦合优化,设计了移动端回环检测算法,显著提升了系统的鲁棒性和可靠性。通过实验验证了所提方法的有效性,其在定位精度上超过了当前同类方法的最好结果,并开发了移动端的增强现实(AR)应用,以展示系统在真实场景中的效果。

关键词 同时定位与地图构建(SLAM);惯性测量单元(IMU);移动设备;回环检测;增强现实(AR)

0 引 言

智能机器人在工业生产、日常生活中发挥了越来越多的作用,它们往往需要在复杂的环境中实时获取自身的位置,以便于进一步作业。同时定位与地图构建(simultaneous localization and mapping, SLAM)算法[1]是解决机器人自主导航和交互的一个重要研究方向,成为了自主无人系统领域的研究热点,几十年来已经取得了丰富的成果,但受限于算法对复杂环境的鲁棒性,各类先进的研究工作都只能在限定环境或指定条件下工作。例如,单目 SLAM[2]总是需要繁琐的初始化,且有尺度不可直接观测和尺度漂移的问题,而双目相机和 RGB-D 相机可以解决初始化和尺度的问题,但是在快速移动和动态场景以及光照变化的场景下很难达到理想的效果。鉴于单一传感器的局限性,多传感器的融合已成为当前的研究热点。近些年来,视觉惯性里程计(visual inertial odometry, VIO)的方法逐渐成为主流[3]。

多状态约束下的卡尔曼滤波器[4-5](multi state constraint Kalman filter, MSCKF)是一种基于扩展卡尔曼滤波(extend Kalman filter, EKF)的视觉惯性融合 SLAM 算法,作为早期的视觉惯性导航系统(vision inertial navigation system, VINS)算法之一,目前被大量用于增强现实(augmented reality, AR)开发工具中,甚至也被用于航天器的降落和登陆中[6]。此后,各种基于滤波的 VINS 系统都是在 MSCKF 的基础上进行改进和开发的。基于光束平差法(bundle adjustment, BA)的 VINS 会把过去一段时间内的测量进行优化,减少错误的线性化点带来的误差,但也会带来高昂的计算代价。VINS-Mono[7]是一个基于 BA 优化的实时紧耦合单目 VINS 系统。为了限制算法的计算量,VINS-Mono 采用了滑动窗口优化

① 国家重点研发计划(2016YFB1001501)资助项目。

② 男,1995 年生,博士生;研究方向:计算机视觉;E-mail:zijinxuxu@zju.edu.cn

③ 通信作者,E-mail:jkzhu@zju.edu.cn

(收稿日期:2020-05-15)

(fixed-lag smoothing)的方式。同时为了获得一个全局一致的地图,算法还设计了一个非实时的闭环检测和重定位线程,并且后台一直优化全局的关键帧位姿图,在当时取得了较好的结果。

虽然上述多种算法已经取得了较高的精度,但往往局限于高算力的个人计算机(personal computer, PC)平台和高精度的传感器设备。随着硬件设计和生产水平的提高,低成本轻量级的惯性测量单元(inertial measurement unit, IMU)元件变得十分普遍,这使得在移动设备和小型飞行器上实现高精度的定位成为可能,这对大量的移动应用,尤其是增强现实和自动驾驶的发展产生了长远的影响。例如市面上成熟的商用移动设备 SLAM,谷歌 Tango 项目(目前称 ARCore),就是采用扩展卡尔曼滤波方式融合数据和惯性测量数据进行运动跟踪。香港科技大学基于 VINS-Mono,开源了一个运行在 iOS 设备上的单目实时 SLAM 应用 VINS-Mobile。然而,低成本 IMU 原件的数据误差与移动平台的低算力限制了此类应用的研发,目前依然存在较大缺口。

由于传感器误差和计算资源的限制,当前所有的 SLAM 系统都无法避免在大尺度和长距离的场景下产生轨迹漂移。为了限制上述误差从而获得一个鲁棒和精确的位姿估计,需要加入一些全局信息对系统进行限制,在视觉 SLAM 中常用的方法便是闭环检测。ORB-SLAM[8] 中使用词袋模型[9] 建立当前帧和地图点的共视图约束,采取全局 BA 方式联合优化相机约束和闭环约束。由于 VINS-Mono 采用了滑动窗口策略无法保持全局地图的一致性,因此将闭环检测视为一个相机重定位过程,检测关键帧和过去相似帧的相对变换,从而校正整个窗口的轨迹漂移。随着深度学习的发展,有大量的工作[10-12]能够高效地从图像数据中提取目标特征,近年来基于卷积神经网络的闭环检测方式 NetVLAD[13] 和 HF-Net[14] 可以在环境变化大的情况下获得较好效果,但耗时的特征计算过程很难满足移动端 SLAM 的实时应用需求。因此本文闭环检测方式还是采取了经典的词袋方式,同时针对移动端特点进行了相应优化。

针对双目设备传感器数据的特点以及移动计算平台的算力,本文对 SLAM 系统进行了针对性的改进,和其他同类工作相比,本文方法的优势主要在于额外设计的多传感器融合策略及移动端闭环检测两个部分。前者能够在不降低系统精度的情况下,为更多传感器的接入提供高效便捷的融合方式;后者在存在闭环路径的情况下,能够显著提升定位精度,这是由于滑动窗口的优化方式必然会带来累计误差,而闭环检测算法能够消除该次回环期间引入的误差值。综上所述,本文的主要贡献点如下。

(1) 针对移动计算平台计算资源受限的特点,SLAM 算法后端优化采用滑动窗口的方式,同时采用预积分的方式处理高频 IMU 信息。为了减少近似线性化对系统能观性的改变,使用首次估计雅可比(first estimated Jacobian, FEJ)的方式固定线性化点减小累积误差。且根据 ARM(advanced RISC machines)平台结构特点,采用 Neon 指令对算法进行优化,最终系统能够在移动设备上达到实时且稳定的效果。

(2) 采用新型多传感器融合策略将双目摄像头、IMU 数据进行紧耦合的优化,在保证计算精度的同时保留了更好的系统拓展性。使用一种单线程的闭环检测方式,添加回环约束修正偏移的轨迹,显著提升了系统的定位精度。

(3) 通过仿真实验与真机测试,验证了所提方法的可行性与有效性,系统的定位结果超过了此前的移动端 SLAM 算法,并通过开发移动端增强现实应用验证了算法的实用性。

1 系统构成

本文提出了面向移动端的视觉惯性 SLAM 框架,整个系统融合双目相机数据和惯性测量单元(IMU)数据,采用新型多传感器融合策略,实现系统位姿的实时输出。同时为了减少非线性优化带来的累积误差,提出了一个适用于移动端的单线程闭环检测模块。系统结构如图 1 所示。

1.1 系统前端

本文的系统前端采用了 Basalt[15] 描述的方法,首先对双目相机采集到的连续帧进行FAST[16] 特征

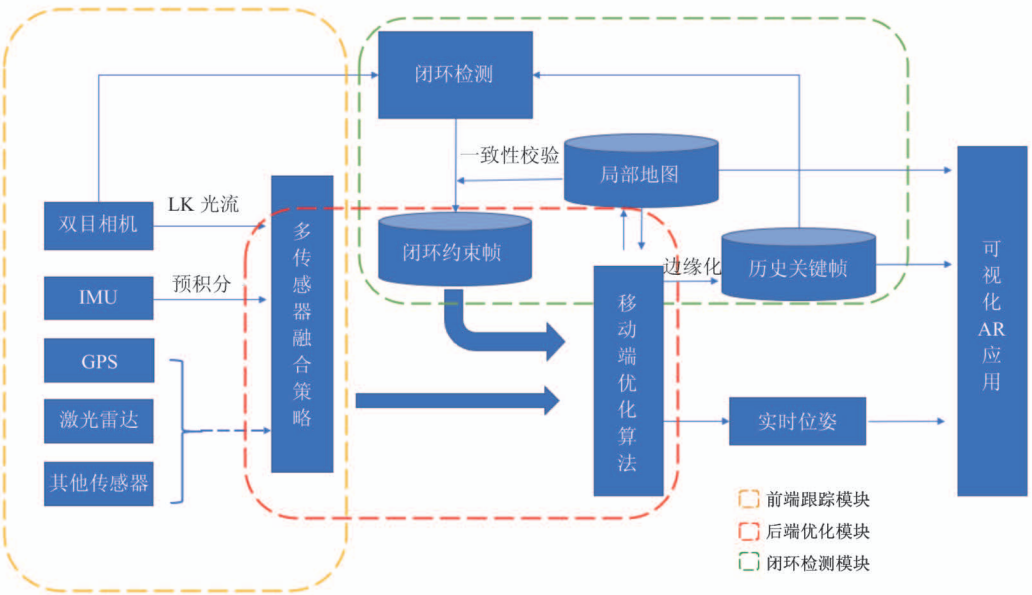


图 1　系统结构图

点检测,然后采用金字塔 LK 光流[17]法进行跟踪。稀疏光流法通过对两帧图像之间的对应小块进行匹配,最小化每一对点集的残差平方和,进而估计出两帧之间的位姿变换。金字塔的方法是为了应对较大位移的情况,对每一组点集进行从粗放到精细的多级跟踪操作。由于一般 SLAM 模型中的空间点都假设为符合高斯分布,但实际世界坐标系下的空间点分布往往是不满足的。为了避免此问题带来的误差,本文利用逆深度[18]来进行参数化操作。在初始化深度特征点之前,本文需要先估计出相机在连续帧之间的运动参数,采取融合 IMU 的方式,可以通过对 IMU 观测数据的预积分,得到连续帧之间的变换参数初始值。这种方式可以给出较为精确的相对位姿初始化值,加快后端优化的收敛速度。随后,本文将观测点的重投影误差连同基于预积分的 IMU 误差项,一起加入到 BA 优化框架中,得到视觉惯性里程计的状态估计。其中,视觉约束的重投影误差 $\boldsymbol{r}_{it}$ 定义为

$$\boldsymbol{r}_{it}=\boldsymbol{z}_{it}-\boldsymbol{\pi}_t(\boldsymbol{T}_{wt}^{-1}\boldsymbol{T}_{ws}\boldsymbol{q}_i(u,\ v,\ \lambda)) \tag{1}$$

式中,视觉特征点 i 在主导帧 s 中首次出现,在目标帧 t 中再次被观测,$\boldsymbol{z}^{it}$ 表示该点在 t 帧中的像素坐标,$\boldsymbol{q}_i(u,\ v,\ \lambda)$ 表示主导帧所在相机坐标系下的该三维特征点,其中 λ 表示逆深度,u、v 表示平面上的参数化坐标。$\boldsymbol{T}_{ws}$ 表示相机坐标系 s 到世界坐标系的变换,再通过 $\boldsymbol{T}_{wt}^{-1}$ 将其变换到相机坐标系 t 中,$\boldsymbol{\pi}_t$ 表示从相机坐标系 t 转换到像素坐标系,最后两者之间的差值作为该特征点的重投影误差用于优化。

1.2　多传感器融合策略

在基于非线性优化的状态估计中,不同传感器的观测信息被加入优化框架中,包括视觉观测、惯性测量等。传统的多传感器融合 SLAM 框架中,一般以相机帧作为主导帧,其余传感器在算法上要与其对齐,共同约束一个定位状态量。然而,由于不同传感器的频率存在差异,需要统一进行消息同步与近似,这为算法引入了更多数据误差,影响了定位精度。此外,系统的可拓展性也受到了限制。随着传感器数量的增加,会给算法带来更大的同步复杂度以及数据近似误差。如图 2 所示,方块表示各个传感器的观测约束,圆圈代表状态估计量。

图 2　传统的多传感器融合框架[19-20]示意图

针对上述问题,本文采用一种新型的多传感器融合策略,不同的传感器观测都以优化窗口中的通用帧来表示。每一个通用帧都对应一个定位状态

量，而不是多个观测共同约束一个状态量。如图 3 所示，圆圈可以表示相机帧，方块及其他形状代表全球定位系统（global positioning system，GPS）、激光雷达、IMU 等观测帧，在框架中能够方便地进行传感器增减。本系统中，双目相机帧之间构成视觉约束，通用帧之间利用高频 IMU 预积分进行约束。该设计模式避免了多个传感器之间时间戳对齐及数据近似等问题，考虑到 IMU 的频率极高，其中的对齐误差能够忽略不计。在室外导航等场景中，GPS 数据能够很好地提高系统可靠性，后续的研发中利用该融合策略能够在保证定位精度的前提下更方便地进行系统拓展。

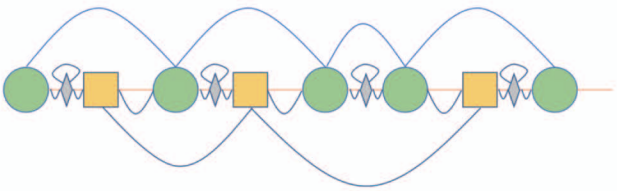


图 3 通用帧模式融合框架

这种通用帧模式的设计针对多传感器时间戳对齐的问题，简化了多传感器之间的特异性，抽象成统一的观测帧形式，在实际开发应用过程中更为高效便捷，而其定位精度本身并未受到影响。在传感器设备及应用场景存在变化的开发背景下，这种模式具有更大的优势。

1.3 移动端优化

本文中的 SLAM 算法前端采用了角点检测和光流的方式，程序运行时间主要集中在后端非线性优化中。该算法最终是为了面向 ARM 架构的移动计算平台，在移动端会受到计算资源限制的挑战，因此从软件开发和算法设计两个方面进行移动端的优化。在软件开发过程中主要利用多核的 ARM 架构多核中央处理器（central processing unit，CPU）的特点，进行多线程程序编写。同时对涉及到使用 Eigen 线性代数库进行矩阵计算的代码，开启 Neon 指令优化，上述软件开发过程和工程实践结合紧密。而算法设计主要利用了滑动窗口优化，虽然全局 BA 优化可以保持信息的全局一致性，而且也被证明在纯视觉 SLAM 中要优于基于滤波的优化算法，但对于融合 IMU 的 SLAM 算法，IMU 状态估计带来的参数成倍增加，全局方式的 BA 优化在计算上变得不可行[15]。为了将计算范围限制在一定的区域里，本文将选取一些在时序上接近当前时刻的关键帧进行优化，利用滑动窗口优化提高计算效率同时又保证计算精度，对关键帧的处理仅保留该帧在 IMU 参考系下的位姿。随着系统不断在未知环境中运行，新的相机帧很快会超过滑动窗口的数量，此时需要处理窗口中旧的信息，从而加入新观测的信息。本文处理的方式是对要丢弃的变量采取边缘化处理。根据前面视觉约束和 IMU 约束的介绍，本文已经把所有的噪声模型建模成了高斯分布。因此在多元高斯分布中去掉一个变量最合理的方式就是将该变量从多元高斯分布中边缘化掉（marginalization）。本文中采用的边缘化策略是当最近帧数目大于阈值时，进行边缘化操作。对最近帧队列中时序上最早的一帧进行判断，有如下两种情况。（1）若不是关键帧，则边缘化掉该帧所有的状态；（2）若是关键帧，边缘化掉速度分量 v 和 IMU 偏置 b，保留位姿分量和该帧三角化的路标点。同时选取关键帧队列中和地图联系最少的一帧，边缘化该帧位姿和对应的路标点。优化算法如图 4 所示。通过引入滑动窗口机制，避免了关键帧和路标点的无限制增长，将计算限制在移动设备可以承受的范围内。但因为滑动窗口的边缘化操作，其边缘化掉的变量所带来的先验信息会加入到信息矩阵中，使得 BA 问题中原本条件独立的状态变得相关。这将会带来系统不一致的问题，本文采用 FEJ 方法[21-22]解决该问题。当状态变量被边缘化后线性化点便固定，所有当前窗口和先验中有联系变量的线性化点，都采用上述固定的线性化点，在状态更新过程中相对固定的线性化点进行参数更新。

本文融合 IMU 和双目图像的滑动窗口算法框架如图 4 所示，整个滑动窗口需要估计的状态参数用 $\boldsymbol{s}$ 表示，包括一定数量的关键帧和最近帧。每一帧状态表示相应的位姿和 IMU 状态。目标优化函数如下：

$$\min_s\{\|\boldsymbol{r}_p-\boldsymbol{J}_p\boldsymbol{s}\|^2+\sum_{i,j\in B}\|\boldsymbol{r}_B(\hat{\boldsymbol{z}}_{ij}^b,\boldsymbol{s})\|^2_{\Sigma ij}+\sum_{i\in P,j\in obs(i)}\|\boldsymbol{r}_C(\hat{\boldsymbol{z}}_{ij}^c,\boldsymbol{s})\|^2_{\Sigma ij}\} \tag{2}$$

其中能量函数第 1 项为先验残差,$\boldsymbol{r}_p$ 和 $\boldsymbol{J}_p$ 分别为前一帧边缘化后的先验误差和先验雅可比矩阵。第 2 项表示预积分观测量和前后帧位姿的残差项,$\hat{\boldsymbol{z}}_{ij}^{b}$ 为连续帧 i, j 之间的预积分观测量,具体做法可参考文献[23],$\hat{\boldsymbol{z}}_{ij}^{c}$ 表示特征点像素坐标,其中 i 为路标点序号,j 为对应观测帧序号,参见式(1)。上述能量函数第 3 项表示了视觉残差。

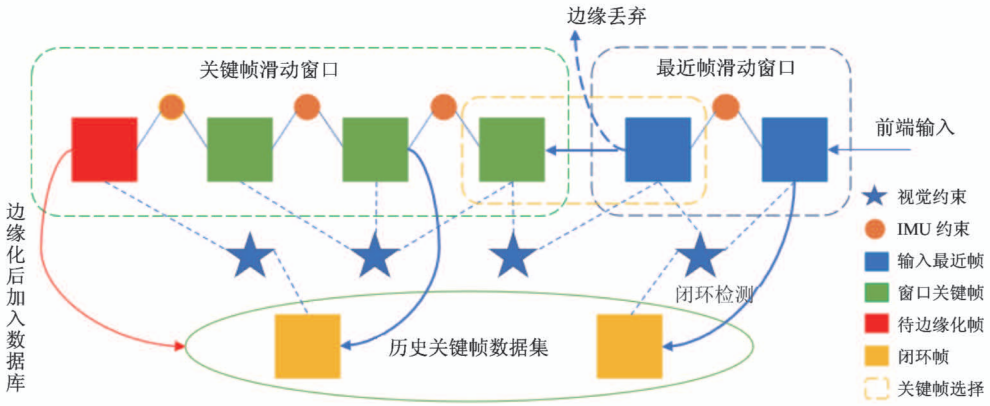


图 4　闭环检测算法流程图

1.4　回环检测

闭环检测的核心目的是检测并利用当前状态和过去位置的约束。当前主流的方法是建立相机轨迹上不同关键帧之间的共视关系,形成众多伪观测数据加入后端优化中。大多数实时 SLAM 系统会采取双线程的方式限制轨迹漂移,一个线程用于当前窗口闭环相似帧的检测,另一个线程进行全局带闭环约束的关键帧位姿图优化校正轨迹漂移。本文系统是面向移动设备开发的,不可避免涉及到计算资源受限的问题。虽然像 VINS-Mobile 这种双线程闭环检测的方式在理论上可以获得更准确的位姿估计,但关键帧会随着长距离的运动线性地增长,后台全局位姿图的优化不可避免会对计算效率带来严峻的挑战。因此本文闭环检测算法为了保证移动端实时性,做了如下假设:将窗口中被边缘化的关键帧视为估计准确,在后续优化过程中不对其进行调整。很显然上述假设在一定程度上破坏了地图一致性,但根据最终实验结果的比较,轨迹估计的精确程度会优于不添加闭环检测模块的算法。

针对移动端运动特点以及计算资源受限的特征,本文提出了一种单线程的闭环检测算法。为了减少计算开销,免去全局位姿图优化的线程,将当前关键帧检测到的闭环约束添加入滑动窗口,并保持约束至该关键帧被边缘化。采用上述方式在时间开销上只会增加关键帧检测闭环候选帧的时间,采用词袋方式可以快速地完成。优化过程中新增加的闭环约束对于观测约束而言不在一个数量级上,时间开销可忽略不计。虽然这种单线程的闭环检测算法假设了地图数据库中的关键帧位姿为真值,带来了一定的不一致性风险,但由于本文中用到的滑动窗口算法给出的原始位姿估计十分准确,从而使得这种方法在时间效率和精度上都取得了很好的效果。单线程闭环检测算法详见算法 1。

本文中闭环检测模块是采用二进制特征描述子(binary robust independent elementary features,BRIEF)描述子的 DBoW3 词袋进行闭环检测。前端光流跟踪到的 FAST 角点平均每帧为 100 个左右,虽然这种数量的特征点用于特征跟踪已经可以取得很好的效果,但用于闭环检测却远远不够。因此在检测到新的关键帧后,会额外计算一定数量的 FAST 角点(本文实验中取 800 个)用于闭环检测,并同时计算新旧角点的 BRIEF 描述子。关键帧的词袋描述向量是用额外获取的描述子计算的,用该向量在词袋中寻找相似度分数最高的几个候选帧。当从词袋中获得一些候选帧时,需要进行闭环有效性的检测,排除一些异常情况造成的误检,获得最优的闭环帧。本文中主要从时序和空间几何关系两个方面对闭环候选帧进行校验。在时序关系上,保证闭环候选帧

和当前关键帧之间相差一定的时间阈值，减小不必要的优化开销。时序接近的关键帧场景纹理存在很大的相似情况，在短时间内窗口并不会产生明显的漂移，而且地图点之间的共视关系已经很好地约束了这些邻近关键帧。满足时序校验后，需要再次检验两个图像帧的空间关系，这里存在一个朴素的假设，闭环候选帧和当前关键帧在空间上存在于一个非常接近的位置。首先利用当前帧光流跟踪的 FAST 角点描述子去匹配闭环候选帧额外计算的角点描述子，当匹配点大于给定阈值后，利用随机抽样一致算法对异常匹配点进行剔除获得符合要求的内点匹配关系。当内点匹配数目大于阈值后即认为是合理的闭环候选帧，选取内点匹配数目最大的一帧成为当前关键帧的闭环帧，用于后续的优化步骤。闭环检测模块仅给出了当前关键帧和闭环帧之间的特征匹配关系，本文提出的闭环检测算法采用紧耦合的方式，将闭环约束直接添加至优化的目标函数进行联合的滑动窗口优化。对于关键帧数据库中用来检测的闭环候选帧，当关键帧被移出滑动窗口后才会被固定位姿并用于闭环检测，这种方式保证添加进关键帧的位姿基本正确，并在后续的闭环检测过程中被视为真值。同时在特征匹配过程中会使用当前光流跟踪的角点描述子去匹配闭环候选帧额外计算的角点描述子，这是因为当前跟踪的关键帧特征点在三角化后会获得相应的深度信息并加入局部地图中。相当于建立了一组 3D-2D 的匹配关系，可以方便地利用视觉重投影方式建立闭环约束，假设获得的一组闭环约束中对应地图中序号为 i 的路标点，可以查询到窗口中对应的主导帧 $h(i)$，那么根据式(1)计算对应的闭环残差，整个带闭环优化的目标函数如式(3)所示。其中前 3 项含义同式(2)，最后一项表示了窗口内所有关键帧的闭环约束，$\hat{\boldsymbol{z}}_{ij}^{l}$ 为闭环检测构建的伪观测。$loop(i)$ 为窗口内关键帧对应闭环帧的集合，i 为闭环约束合格匹配内点对应的路标点。

$$\begin{aligned}\min_{s}\{&\|\boldsymbol{r}_p-\boldsymbol{J}_p\boldsymbol{s}\|^2+\sum_{i,j\in B}\|\boldsymbol{r}_B(\hat{\boldsymbol{z}}_{ij}^{b},\boldsymbol{s})\|_{\Sigma ij}^2\\&+\sum_{i\in P,j\in obs(i)}\|\boldsymbol{r}_C(\hat{\boldsymbol{z}}_{ij}^{c},\boldsymbol{s})\|_{\Sigma ij}^2\\&+\sum_{i\in P,j\in loop(i)}\|\boldsymbol{r}_L(\hat{\boldsymbol{z}}_{ij}^{l},\boldsymbol{s})\|_{\Sigma ij}^2\}\end{aligned}\tag{3}$$

进行闭环检测的目的是为了校正非线性优化中不可避免的轨迹漂移，其矫正流程如图 5 所示。本文算法中认为数据库中的第 i 帧位姿是一个真值，可以表示为相对于真实世界坐标系 w1 的位姿 $\boldsymbol{T}_{bi}^{w1}$。轨迹漂移利用数学模型可以表示为：在不断地非线性优化过程中，窗口相对的世界坐标系 w1 发生了漂移，变为了窗口参考系 w2，那么闭环检测模块需要做的便是求解两个参考系的漂移 $\boldsymbol{T}_{w2}^{w1}$。在紧耦合闭环优化过程中，本文只是将 w1 系下的真值作为优化初始化值而固定其姿态，在不断地迭代优化后可以获得闭环帧在窗口参考系下的位姿 $\boldsymbol{T}_{bi}^{w2}$，从而窗口的漂移计算如下：

$$\boldsymbol{T}_{w2}^{w1}=\boldsymbol{T}_{bi}^{w1}(\boldsymbol{T}_{bi}^{w2})^{-1}\tag{4}$$

根据漂移校正对应关键帧后所有相机帧的位姿，从而获得更加准确的结果。

算法 1　带闭环检测的滑动窗口优化算法

输入：前端光流跟踪结果和 IMU 初始化的位姿

1：loop

2：若为关键帧，额外计算一定数量的特征点，得到对应的词袋向量

3：根据词带向量检索出一组候选帧

4：校验候选帧，取匹配数目最大的合格候选帧为闭环帧

5：闭环帧和窗口内地图点建立闭环约束，约束保持至改关键帧被边缘化出优化窗口

6：带闭环的多传感器融合优化

7：窗口边缘化。若边缘化的关键帧有对应的闭环帧，计算窗口漂移并校正该帧后的位姿

8：end loop

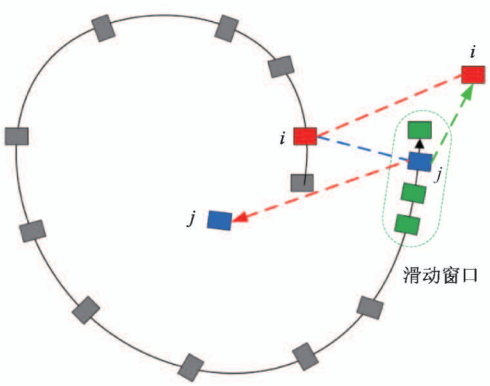


图 5　轨迹漂移矫正

2　实验

2.1　实验配置

系统运行的软硬件配置为：搭载 Kirin 980 的智

能手机、Android 10.0 系统、Intel Realsense D435i 相机,算法在移动设备上进行实时运算,主要包含如下三方面实验。

(1)验证本文提出的 SLAM 算法在 Euroc[24] 数据集上的定位精度,并与其他主流 SLAM 算法进行对比实验,即整体算法框架的测试,并进一步通过实验对比回环检测模块带来的定位精度提升。

(2)验证本文提出的 SLAM 算法在移植到移动嵌入式设备后的有效性,对比和 PC 端算法在 Euroc 数据集上运行结果的时间开销,即移动端测试。

(3)验证以本文 SLAM 算法为基础开发的移动端应用,连接真实传感器设备进行移动定位的可行性,即移动端定位应用的效果测试。

2.2 公开数据集测试

本文在 Euroc 数据集的所有序列上将本文提出的带闭环检测的双目视觉惯性 SLAM 算法同主流的 SLAM 算法进行对比,比较算法包含 VI-DSO[25]、OKVIS[26]、VINS-FUSION[20] 和 Basalt[15]。其中 VI-DSO 仅使用单目数据,Basalt 仅使用双目数据,OKVIS 和 VINS-FUSION 分别有单目和双目版本。在对比实验中,误差度量方式采用绝对轨迹误差(absolute trajectory error, ATE),计算数据集中每一个图像帧的轨迹误差。部分算法无法取得开源代码,实验数据从相应论文中获取,实验结果如表 1 所示。

表 1 本文算法在 Euroc 数据集上的绝对轨迹误差对比/m

序列	单/双目	MH-01	MH-02	MH-03	MH-04	MH-05	V1-01	V1-02	V1-03	V2-01	V2-02
VI-DSO[25]	单目	**0.06**	**0.04**	0.12	0.13	0.12	0.06	0.07	0.10	**0.04**	0.06
OKVIS[26]	单目	0.34	0.36	0.30	0.48	0.47	0.12	0.16	0.24	0.12	0.22
VINS-FUSION[20]	单目	0.18	0.09	0.17	0.21	0.25	0.06	0.09	0.18	0.06	0.11
OKVIS[26]	双目	0.23	0.15	0.23	0.32	0.36	**0.04**	0.08	0.13	0.10	0.17
VINS-FUSION[20]	双目	0.24	0.18	0.23	0.39	0.19	0.10	0.10	0.11	0.12	0.10
Basalt[15]	双目	0.07	0.05	0.06	0.12	0.12	0.05	0.05	0.10	**0.04**	**0.05**
本文算法框架	双目	0.08	**0.04**	**0.05**	**0.10**	**0.09**	**0.04**	**0.04**	**0.05**	**0.04**	**0.05**

从上述量化的定位结果分析可得,本文提出的双目惯性 SLAM 算法与主流的 SLAM 算法相比,除 MH-01 序列外都取得了最好的效果。由于 MH-01 序列较为简单,最好的几个结果之间精度相差无几(0.06 ~ 0.08 m)。而在运动激烈的 MH-03(中等)、MH-04(困难)、MH-05(困难)和 V1-03(困难)等序列获得了更高的定位精度,其中 MH-05 序列相对于 Basalt[15] 误差降低了 0.03 m, V1-03 序列降低了 0.05 m,在之前方法的最好结果上有了显著提升。

为了验证闭环检测算法的有效性,本文对该模块做了单独的分解实验,从而验证闭环检测的有效性。两种模式下的绝对轨迹误差如表 2 所示,每组序列的轨迹图如图 6 所示。在 MH-03、MH-05 和 V1-03 这 3 组轨迹中存在大量的闭环场景,根据表 2 中的误差数据,对应的绝对轨迹误差在有无开启闭环检测模块会出现较大差别(误差分别减低了 0.034 m、0.019 m 和 0.027 m)。在剩余闭环场景不明显的数据序列上,有无闭环检测模块对轨迹精度基本无影响。同时,本文对闭环检测模块做了进一步的量化分析,图 7 给出了两种模式在 Euroc 数据序列上随着运动距离的增加,位置误差和旋转误差的变化。图中显示随着运动距离的增加,带有闭环检测的算法平均误差会小于不带闭环的算法

表 2 闭环检测模块绝对轨迹误差

序列	有闭环/m	无闭环/m
MH-01	0.095	**0.092**
MH-03	**0.052**	0.086
MH-05	**0.092**	0.111
V1- 01	**0.041**	0.043
V1- 03	**0.050**	0.077
V2- 02	**0.060**	0.064

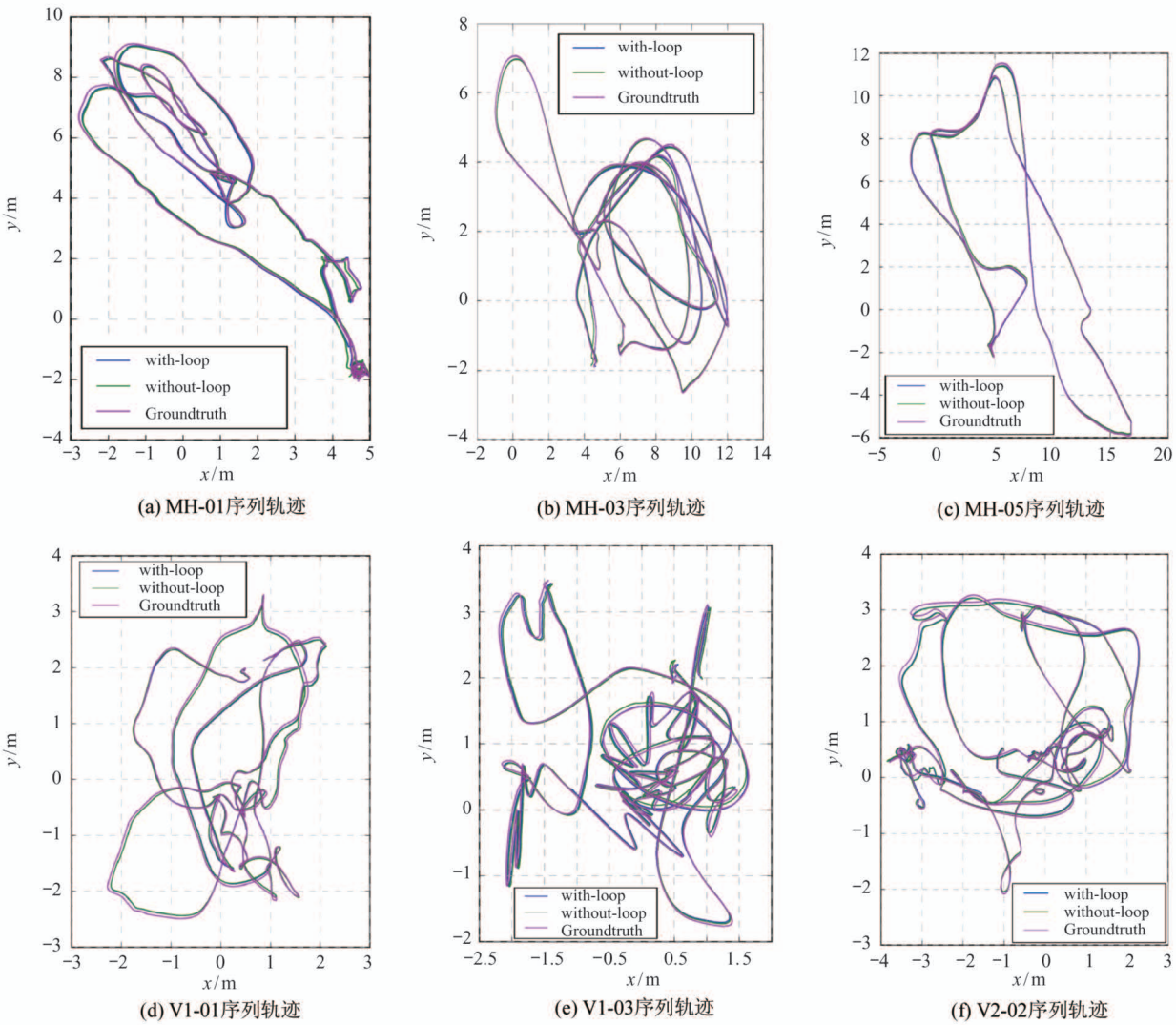


图 6　Euroc 数据集下闭环(with-loop)和无闭环(without-loop)算法估计的轨迹和真值(Groundtruth)对齐的显示图

(图 7(b)26.0 m 后、图 7(c)19.0 m 后、图 7(e)23 m 后对比效果较为显著)。本实验证明了带有闭环检测的算法可以有效地在闭环发生的场景下,校正轨迹漂移,提高算法估计的精度。

2.3　时间开销分析

本文统计了系统各个模块的时间开销,如表 3 所示。从表中可见,边缘化的时间开销极低,所有模块平均时间总计约为 35 ms,考虑到回环检测不是逐帧运行,能够满足移动端实时性的要求。

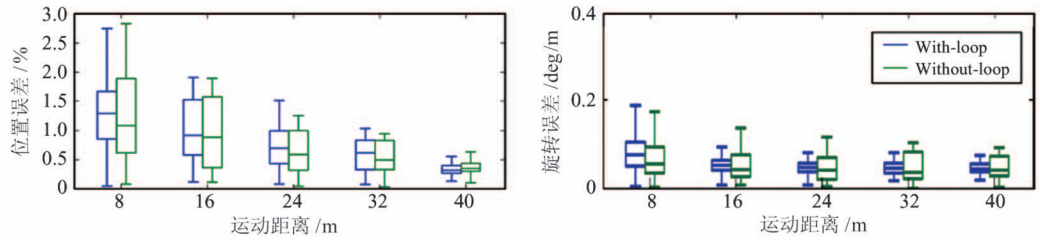
表 3　算法各模块时间开销

系统模块	光流	回环检测	窗口优化	边缘化
平均时间/ms	9.70	13.00	11.47	0.32
样本数量/帧	4464	550	4446	4443

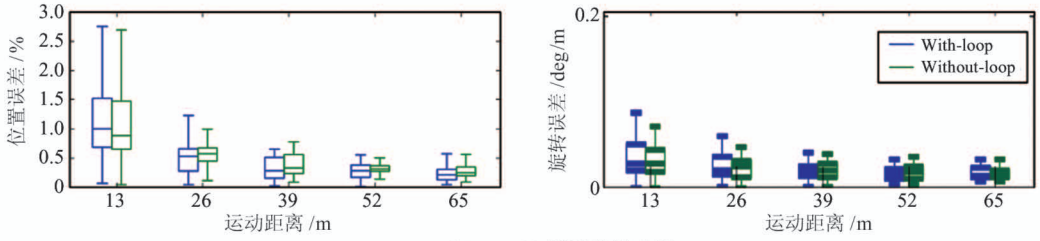
2.4　增强现实应用验证

基于本文提出的算法,构建了一个移动端 AR 应用,为了验证算法在移动端的实时性,本文在 HUAWEI P30 手机上安装了软件应用,配合 Realsense D435i 进行功能测试与可视化,如图 8 所示。

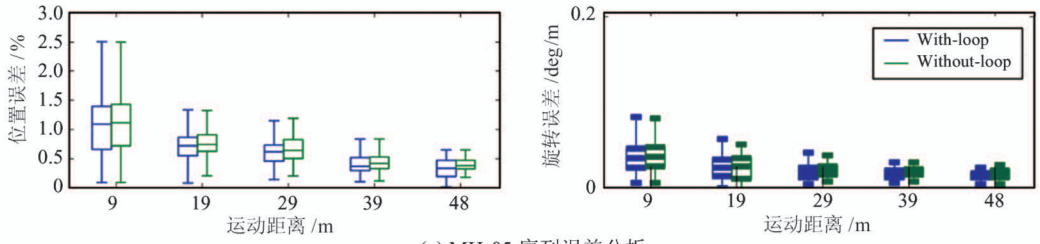
图 8(a)展示了移动端系统的轨迹图,通过回到出发点来评估定位的累计误差,坐标轴中一格代表 1 m 距离,起点与终点处位置基本重合,定位效果理想。图 8(b)展示了 AR 应用的测试,检测到画面中的平面后能够放置 AR 小方块,完成三维注册以加入到地图坐标中。图 8(c)是放置 AR 方块一段时间后再回到原处,能够发现小方块位置不变,验证了系统的可靠性。



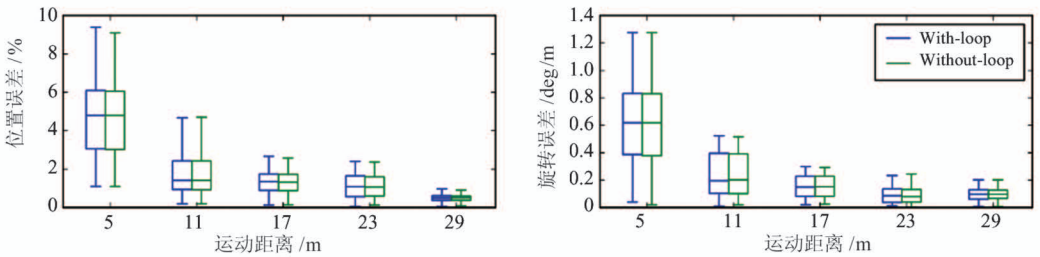
(a) MH-01 序列误差分析



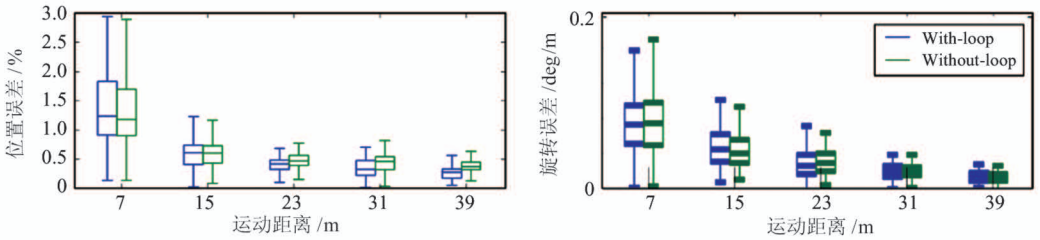
(b) MH-03 序列误差分析



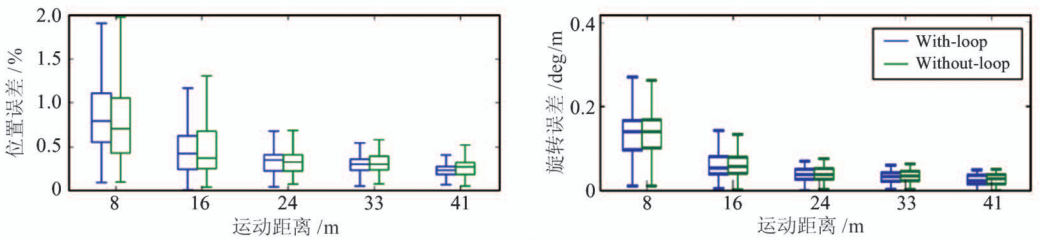
(c) MH-05 序列误差分析



(d) V1-01 序　吴差分析

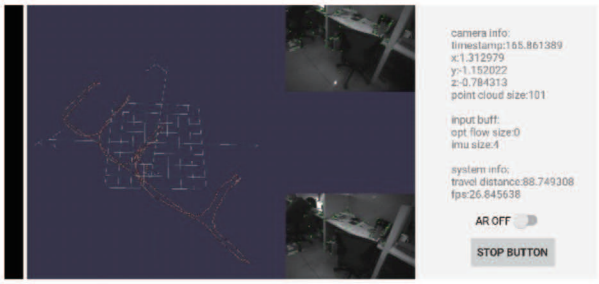


(e) V1-03 序列误差分析

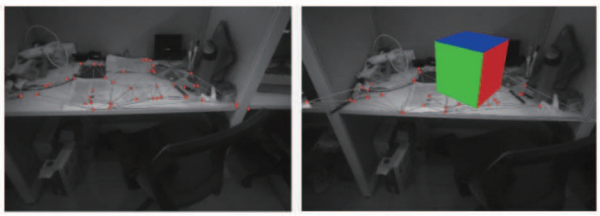


(f) V2-02 序列误差分析

图 7　Euroc 数据集下闭环(左,with-loop)和无闭环(右,without-loop)算法随运动距离变化,位置误差和旋转误差的箱型图分析



(a) 里程计轨迹图



(b) 平面检测并放置AR cube

(c) 全局轨迹与AR cube展示

图 8　增强现实效果

3　结　论

本文提出了基于新型多传感器融合策略的移动端双目视觉惯性 SLAM 闭环算法，全文主要从 3 个方面论述工作。首先是设计了一个新颖的多传感器融合策略，对不同频率的传感器设备免去了时间同步或是数据近似的问题，更好地发挥了传感器的精度，在面对变化的场景需求和硬件配置时具有更好的实用性与可拓展性。其次是针对移动设备需求改进了视觉惯性导航算法，设计了面向移动端优化的回环检测算法，极大降低了系统在闭环路径中产生的累积误差，显著提高了 SLAM 系统的定位精度，同时通过优化使其在移动平台上达到实时的效果。最后基于系统框架开发了增强现实的应用，验证了本算法的现实可行性及效果。实验证明，本系统的定位精度超过了当前最好的同类方法，得到了较为理想的效果。在未来工作中，考虑结合更多的传感器，如 GPS、磁力计、轮速计等设备，发挥新型多传感器融合的优势，进一步提升系统的鲁棒性。并针对重建的地图做进一步的应用开发，使 SLAM 算法能够融合到路径规划、导航等具体应用中去。

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] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(6): 1052-1067

[3] Huang G. Visual-inertial navigation: a concise review [C]//2019 International Conference on Robotics and Automation, Montreal, Canada, 2019:9572-9582

[4] Mourikis A I, Roumeliotis S I. A multi-state constraint Kalman filter for vision-aided inertial navigation [C]// Proceedings of 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, 2007: 3565-3572

[5] Li M Y, Mourikis A I. High-precision, consistent EKF-based visual-inertial odometry [J]. *The International Journal of Robotics Research*, 2013, 32(6): 690-711

[6] Mourikis A I, Trawny N, Roumeliotis S I, et al. Vision-aided inertial navigation for spacecraft entry, descent, and landing[J]. *IEEE Transactions on Robotics*, 2009, 25(2): 264-280

[7] 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

[8] Mur-Artal R, Montiel J M M, Tardos J D. ORB-SLAM: a versatile and accurate monocular SLAM system [J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147-1163

[9] Gálvez-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

[10] Zhang X Y, Wang S P, Yun X C. Bidirectional active learning: a two-way exploration into unlabeled and labeled data set[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(12):3034-3044

[11] Zhang X Y, Li C S, Shi H C, et al. AdapNet: adaptability decomposing encoder-decoder network for weakly supervised action recognition and localization [J]. *IEEE Transactions on Neural Networks and Learning Systems (TNNLS)*, 2020, doi:10.1109/TNNLS.2019.2962815

[12] Zhang X Y, Shi H C, Li C S, et al. Multi-instance multi-label action recognition and localization based on spatio-temporal pre-trimming for untrimmed videos[C]// AAAI Conference on Artificial Intelligence, New York,

USA, 2020:1-8

[13] 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, Las Vegas, USA, 2016: 5297-5307

[14] Sarlin P E, Cadena C, Siegwart R, et al. From coarse to fine: robust hierarchical localization at large scale[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 12716-12725

[15] Usenko V, Demmel N, Schubert D, et al. Visual-inertial mapping with non-linear factor recovery[J]. *IEEE Robotics and Automation Letters*, 2020,5(2): 422-429

[16] Rosten E, Porter R, Drummond T. Faster and better: a machine learning approach to corner detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 32(1): 105-119

[17] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]// Proceedings of the 7th International Joint Conference on Artificial Intelligence, Vancouver, Canada, 1981: 674-679

[18] Civera J, Davison A J, Montiel J M M. Inverse depth parametrization for monocular SLAM[J]. *IEEE Transactions on Robotics*, 2008, 24(5): 932-945

[19] Qin T, Cao S Z, Pan J, et al. A general optimization-based framework for global pose estimation with multiple sensors[J]. *arXiv*:1901.03642, 2019

[20] Qin T, Pan J, Cao S Z, et al. A general optimization-based framework for local odometry estimation with multiple sensors[J]. *arXiv*:1901.03638, 2019

[21] Engel J, Koltun V, Cremers D. Direct sparse odometry [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(3): 611-625

[22] Huang G P, Mourikis A I, Roumeliotis S I. Analysis and improvement of the consistency of extended Kalman filter based SLAM[C]// 2008 IEEE International Conference on Robotics and Automation, Pasadena, USA, 2008: 473-479

[23] Forster C, Carlone L, Dellaert F, et al. On-manifold pre-integration for real-time visual-inertial odometry [J]. *IEEE Transactions on Robotics*, 2017, 33(1):1-21

[24] 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

[25] Von Stumberg L, Usenko V, Cremers D. Direct sparse visual-inertial odometry using dynamic marginalization [C]// 2018 IEEE International Conference on Robotics and Automation (ICRA), Prague, Czech Republic, 2018:2510-2517

[26] 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

Mobile visual-inertial SLAM loopclosure algorithm based on novel multi-sensor fusion strategy

Ren Jinwei, Zheng Xin, Li Yuchen, Zhu Jianke

(College of Computer Science and Technology, Zhejiang University, Hangzhou 310027)

Abstract

This paper deals with the real-time localization of mobile visual simultaneous localization and mapping (SLAM) in cluttered environments. There are several difficulties in this task. Firstly, the limited computational resources lead to the requirements of the optimization and efficiency on the algorithm. Secondly, there are the cluttered and dynamic scenario. How to avoid the drift in low texture area and fast motion is the main difficulty. Finally, it requires good scalability, which can be accurately landed and has applicability in certain application domains. To tackle the above challenges, this paper proposes a stereo visual-inertial SLAM algorithm for mobile devices. A new multi-sensor fusion strategy that optimizes stereo visual terms and inertial measurement error in a tightly-coupled way is introduced. And the loop detection algorithm on the mobile devices significantly improves the robustness and reliability of the system. The effectiveness of the proposed method is evaluated through the intensive experiments, and the localization accuracy outperforms the state-of-the-art methods. Moreover, an augmented reality (AR) application is developed as an application of the system in the real scene.

Key words: simultaneous localization and mapping (SLAM), inertial measurement unit (IMU), mobile device, loop closure detection, augmented reality (AR)