

doi:10.19306/j.cnki.2095-8110.2020.03.012

多智能体协同视觉 SLAM 技术研究进展

王璐, 杨功流, 蔡庆中, 陈威, 闫旭亮

(北京航空航天大学仪器科学与光电工程学院, 北京 100191)

摘要: 多智能体的协同视觉 SLAM 系统在地面机器人或空中飞行器梯队中应用得越来越广泛。由于不同的智能体之间会交换或共享信息, 协同视觉 SLAM 系统不同于应用在单个智能体上的典型的视觉 SLAM 系统, 具有更好的鲁棒性、准确性和高效性。阐述了多智能体协同视觉 SLAM 的发展历史和现有的研究方法, 描述了协同视觉 SLAM 系统实现所涉及的关键技术, 包括协同姿态估计和地图构建, 以及新兴的分布式架构。此外, 对多智能体协同视觉 SLAM 技术的发展趋势进行了展望。

关键词: 多智能体; 协同; 视觉 SLAM; 地图构建

中图分类号: V11 文献标志码: A 开放科学(资源服务)标识码(OSID):
文章编号: 2095-8110(2020)03-0084-09



Research Progress in Collaborative Visual SLAM for Multiple Agents

WANG Lu, YANG Gong-liu, CAI Qing-zhong, CHEN Wei, YAN Xu-liang

(School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100083, China)

Abstract: The collaborative visual SLAM system of multiple agents is more and more widely used in ground robot or air vehicle echelon. Because different agents exchange or share information, collaborative visual SLAM system is different from the typical visual SLAM system applied to a single agent. The collaborative visual SLAM system has better robustness, accuracy and efficiency. This paper describes the development history and existing research methods of multi-agent collaborative visual SLAM, and describes the key technologies involved in the realization of collaborative visual SLAM system, including collaborative attitude estimation and mapping, as well as the emerging distributed architecture. In addition, the development trend of multi-agent collaborative visual SLAM technology is prospected.

Key words: Multiple agents; Collaborative; Visual SLAM; Map Construction

0 引言

多智能体协同视觉即时定位与地图构建(Simultaneous Localization and Mapping, SLAM)^[1-2]指

搭载一个或多个相机等视觉传感器的运动主体, 通过对环境信息的感知估计出自身位姿的变化并重建未知环境的三维地图, 对自身位姿变化的估计以及三维地图的重建对增强现实和机器人的自主导

收稿日期: 2020-02-04; 修订日期: 2020-02-27

基金项目: 国家重点研发计划(2016YFB0501600); 装备预研基金重点项目(6140517020101)

作者简介: 王璐(1996-), 女, 硕士研究生, 主要从事导航与控制方面的研究。E-mail: wangluhbhs@163.com

通信作者: 杨功流(1967-), 男, 教授, 博士生导师, 主要从事惯性技术研究。E-mail: yanggongliu@buaa.edu.cn

航都至关重要^[3]。近年来,视觉 SLAM 技术日渐成为计算机视觉、机器人、深度学习、增强现实等领域的研究热点,与使用激光雷达等传感器的传统 SLAM 技术相比,视觉 SLAM 技术使用的相机具有轻量化、低成本并且包含丰富视觉信息的特点,因此适用于有成本限制、载荷有限的平台。

经过十余年的迅速发展,视觉 SLAM 技术已经成功地应用于很多商业产品中,例如微软的 HoloLens、Google 的 Project Tango,还有两大手机操作系统 Android^[4]和 iOS^[5]也利用视觉 SLAM 技术为消费者提供增强现实体验。另外,无人机上也使用了视觉 SLAM 技术,例如 DJI 的四旋翼无人机的几种型号——Mavic Air 和 Phantom 4,使得它们可以在全球定位系统(Global Positioning System, GPS)拒止环境下进行可靠的导航。因此,视觉 SLAM 技术凭借其廉价的传感器和巨大的感知潜力受到了越来越广泛的关注。

目前,国内外对视觉 SLAM 的研究形成了一些理论方法,但是对多智能体视觉 SLAM 研究的方法较少,这些智能体可能是无人机群或者使用不同 AR 设备的人^[6]。在这样的环境下,每个智能体在独立工作的同时,还要时刻关注其他智能体的运动信息,这就要求不同智能体中的视觉 SLAM 系统相互协作、共享信息^[7],从而为多个智能体同时解决定位和地图构建的问题。

1 多智能体协同视觉 SLAM 发展现状

在 SLAM 技术研究初期,主要的关注点在于以激光雷达为主要传感器的 SLAM 系统。随着计算机能力和三维重建(Structure from Motion, SfM)技术^[8]的发展,相机以其成本低、质量小、信息丰富等优点逐渐成为解决 SLAM 问题的理想传感器。随着单目视觉 SLAM 技术的迅速发展和日益强大,是否可以将视觉 SLAM 技术应用于多个相机,特别是一组独立运动的相机在学术界引起广泛关注。

多数协同视觉 SLAM 系统采用的是中心式架构,第一部分(智能体前端)利用视觉传感器信息对智能体的实时状态进行计算;第二部分(服务器后端)利用前端结果,根据多视图几何理论,对地图进行融合、简化和优化,输出位姿及三维地图,后端算法主要通过滤波^[9-11]和非线性最小二乘优化^[12-14]两种方法来实现。此外,回环检测(Loop Closure Detection)^[15]通过重复探测可以消除累积误差,从而

在视觉 SLAM 系统中构成闭环。

2003 年,第一个视觉 SLAM 系统^[16]问世,该系统主要是基于扩展卡尔曼滤波(Extend Kalman Filter, EKF)算法^[17],利用从图像上提取的特征点实现对运动主体状态的估计和三维地图的构建。Georg Klein 等采用了优化的方法,提出了基于加速分割检测(Features from Accelerated Segment Test, FAST)特征的并行跟踪和建图(Parallel Tracking and Mapping, PTAM)^[18]。作为 PTAM 的扩展,并行跟踪与多重映射(Parallel Tracking and Multiple Mapping, PTAMM)^[19]将单目视觉 SLAM 扩展到了 2 个独立移动的相机上,一个相机放在肩上,另一个在手上,二者同时运行,将关键帧发送到同一个地图发生器,从而优化了运动主体的位姿和特征点的 3D 坐标,同时,系统还存储了再临同一位置时使用的局部地图。

CoSLAM^[20]是一种应用于一组独立运动相机的协同视觉 SLAM。该系统支持 12 个独立运动的相机并把其分为不同的组,同一组中的相机需要具有共同的视域,通过采用相机间特征匹配、姿态估计和地图构建的方式,实现同一组内相机之间的协同。该系统通过不同视域中的特征点匹配来三角化 3D 运动点,实现相机间的建图,从而估计出运动特征点的 3D 轨迹^[21]。

由于 CoSLAM 中所有的图像信息需要进行实时的发送,因此通信负载会随着相机数量的增加而增大。为了降低通信负载,C2TAM^[22]中的智能体都配备相机并把提取的关键帧图像信息发送到云端;云端利用这些关键帧图像信息构建每个智能体所在的 3D 局部地图,并在识别出相同位置时将这些地图信息进行融合;最后利用光束平差法对构建好的地图进行优化。

为了进一步降低通信负载,CSfM^[23]中每个智能体独立运行单目视觉里程计,其发送的关键帧数据不仅包含关键帧图像上的特征点,而且还有这些关键帧之间的相对姿态。服务器接收到关键帧数据后为每个智能体构建 3D 局部地图,为了获得局部地图中新到达的关键帧的绝对姿态,CSfM 通过比较车载和服务器上计算的关键帧姿态之间的相对差异,从车载视觉 SLAM 系统中估计地图的未知比例因子。

在 CCM-SLAM^[24]中,每个智能体只运行具有有限关键帧数的视觉里程计,智能体将检测到的关键帧信息发送到服务器;服务器根据这些信息进行

局部地图的构建,并通过位置识别的方法将局部地图信息进行融合;姿态估计和光束平差^[25]在服务器中被应用于对地图的细化^[26]。

为了获得更好的鲁棒性和准确性,可以将惯性测量单元(Inertial Measurement Unit, IMU)与相

机等视觉传感器进行组合,使得两类传感器进行互补。IMU辅助的视觉SLAM系统通常被称为视觉惯性导航系统^[27-29],MOARSLAM^[30]就在智能体中采用了单目视觉惯性SLAM的框架。多智能体协同视觉SLAM方案对比如表1所示。

表1 多智能体协同视觉SLAM方案对比

Tab. 1 Comparison of collaborative visual SLAM for multiple agents methods

SLAM系统	年份	智能体个数	前端特点	后端特点	地图类型
PTAMM	2008	2	姿态估计	三角化;重定位;光束平差法	全局地图
CoSLAM	2012	12	检测图像信息	相机内/间姿态估计;地图构建;光束平差法	全局地图
CSfM	2013	2	视觉里程计	位置识别;地图融合;姿态优化;光束平差法	局部地图
C2TAM	2013	2	姿态估计	三角化;重定位;光束平差法	局部地图
MOARSLAM	2016	3	视觉-惯性里程计	位置识别	具有相对位姿关系的地图
CCM-SLAM	2018	3	视觉里程计	位置识别;地图融合;删掉冗余关键帧	局部地图

目前,虽然大多数视觉SLAM算法都集中在单个平台上的应用,但是随着无人机、机器人、多用户设备等多智能体应用的发展,尤其是5G^[31]移动网络的出现,越来越多的科研人员将投入到多智能体协同视觉SLAM技术的研究当中。

2 多智能体协同视觉SLAM的关键技术

多智能体就是由多个相互作用的单智能体构成的系统,每个智能体都是多智能体系统的重要组成部分。在多智能体协同视觉SLAM系统中,智能体之间相互通信,相互协调,并行地求解问题,不仅可以有效地利用空间分布的信息资源,提高问题求解效率;而且系统中单个智能体的损坏一般不会影响到其他个体的运行,与单智能体相比具有更好的容错性和抗干扰性。

然而,多智能体协同视觉SLAM的研究还存在诸多技术挑战,相对于对单智能体的控制,多智能体之间交换的信息量会增加,大量的计算与通信会拥塞有限的通信信道,造成延迟;而有限的通信资源在智能体成员之间如何分配也是多智能体协同视觉SLAM系统需要解决的一个重要问题。智能体在运动过程中,重复经过某场景时,该场景能否被其他智能体快速匹配也会对匹配正确率与位姿精度产生影响。

与经典单目视觉SLAM相似,多智能体协同视觉SLAM也可以大致分为位姿跟踪和地图构建两部分。然而不同的是,多智能体协同视觉SLAM主要是对来自不同智能体的信息的处理,下面对其关

键技术进行讨论。

2.1 视觉前端

视觉前端是多智能体协同视觉SLAM中不可缺少的一部分,一般分为直接法和特征点法。直接法^[32-33]不需要进行特征提取而是直接对像素进行操作,近年来,因其在精度和效率上的优势而受到广泛关注。虽然在单个智能体上使用直接法会收到很好的效果,但是很难直接匹配不同智能体上的相机之间的像素以建立对应关系。

2.1.1 特征检测与匹配

特征点的检测和匹配是特征点法中的关键一步,特征点由关键点和描述子组成。关键点的检测算法有很多种,其中,FAST角点^[34]是早期视觉SLAM^[35-36]中最常用的关键点检测算法之一,在提取出关键点后对其描述子进行计算,通过描述子的匹配来完成特征匹配。为了在不同的视角之间匹配相应的特征点,需要使用对大视角变化鲁棒的描述子^[37]。其中,ORB描述子^[38]是BRIEF描述子^[39]的一个扩展,具有旋转不变性和对大视角变化的鲁棒性,对位姿估计、地图构建和环路检测来说具有较高的可靠性;SURF描述子^[40]因具有较高的效率也得到了广泛的应用。

引导搜索是一种提高特征匹配效率的方法^[41],通过计算由当前状态变量预测的路标的可能位置范围来降低匹配时的计算量。这一思想自早期的视觉SLAM系统以来就一直被采用,尤其是一些基于滤波器的视觉SLAM系统^[42]中,这一范围通常由预测不确定性或投影协方差矩阵决定;而在基于

关键帧的系统中,这一范围为极线。在 PTAM 和 ORB-SLAM^[43]中,执行极线搜索以在不同关键帧之间进行三角化而得到新点。在 CoSLAM 中,对不同的相机应用极线搜索来生成新的地图点。

2.1.2 外点剔除

匹配过程中会产生外点,通常有两类,第一类主要是由错误的特征匹配引起的;而第二类则来自环境中的运动特征点,如果把运动特征点作为静态点来处理,就会导致位姿估计的不一致。

外点剔除一般通过随机抽样一致(Random Sample Consensus, RANSAC)算法^[44-45]或者鲁棒估计^[46]来实现。其中 RANSAC 算法通过随机抽取最小的 2D-3D 对应点集,采用 PnP 算法计算运动主体的位姿,并用当前的位姿估计值检验剩余对应点的一致性,将对应点一致性最好的一组位姿估计结果作为最优解,而不一致的对应点则作为外点被剔除。该算法的优点是鲁棒性好,但是计算量大。

鲁棒估计通过对现有非线性最小二乘优化中的残差函数稍作修改来实现,相比于 RANSAC 算法,鲁棒估计计算量小,但是当外点数量较大时,其鲁棒性会变差。因此,为了在不影响系统鲁棒性的前提下减小计算量,可以采用将 RANSAC 算法和鲁棒估计相结合的方法。

2.2 协同位姿的估计

对六自由度位姿的实时估计是多智能体协同视觉 SLAM 的一项基本任务,相机位姿估计可以通过融合来自多个相机的信息来实现,主要有基于滤波的(filter-based)和基于优化的(optimization-based)两种不同方法。

在基于滤波的方法中,目前有基于 EKF 和基于粒子滤波的方法。在基于 EKF 的多智能体协同视觉 SLAM 中,未知变量通过扩展卡尔曼滤波器进行估计,通常把运动主体的自身运动参数和路标的三维坐标选为状态变量^[47],通过 EKF 的迭代解决位姿估计和路标三维坐标的更新问题。由此可见,该方法的效率随着路标数量的增多而降低,因此,在实际的应用中,为了保证系统的实时性,路标的数量应该被限制在一定的范围内^[48]。其中,一种降低计算量的方法是从状态变量中移除路标,但是保留运动主体在当前时刻和过去时刻的位姿,使得观测模型只依赖于运动主体的位姿,这种方法被称为多状态约束卡尔曼滤波(Multi-State Constraint Kalman Filter, MSCKF),是设计视觉惯性里程计

的主要方法。

在基于优化的方法中,则主要是基于关键帧优化的方法。关键帧是一种数据结构,通常存储图像上检测到的特征点及其对应的 3D 点,使用最近的关键帧中得到的 3D 点,及其在当前图像中匹配的特征点,能够通过 3D-2D 对准的方式,如 PnP 算法或者非线性最小二乘优化^[49-50]来求解当前姿态。通过对匹配特征点的三角化,在帧间进行地图构建,再利用光束平差法对关键帧位姿和三维地图进行优化。为了使视觉 SLAM 具有较高的效率,一般将姿态计算与地图构建任务交替求解,也就是说,姿态估计是在地图构建的基础上完成的,反之亦然。

2.3 协同地图的构建

关键帧是多智能体协同视觉 SLAM 中表示地图的最佳方法,它可以紧凑地存储和地图相关的任务(如地图融合和优化)所需的必要信息,包括特征点坐标及其描述子、特征点三维坐标、关键帧的相机位姿及其拓扑关系(通常称为位姿图^[51])等。

构建地图的核心工作是对特征点进行三角化,从而得到新的 3D 点。在多智能体协同视觉 SLAM 中,地图构建可以进一步分为相机内地图构建和相机间地图构建。其中前者与单目视觉 SLAM 中的建图类似,仅使用同一个相机的图像;后者在使用不同相机的图像构建地图时,首先要确定相机视野是否相同,然后再利用特征匹配建立不同相机之间的点对应关系。

2.4 回环检测

回环检测是视觉 SLAM 系统中的一个重要组成部分,在多智能体协同视觉 SLAM 中,有两种回环检测的方式:相机内闭环和相机间闭环。其中前者通过检测单个相机曾经到达过的场景,减小漂移误差;后者通过检测多个独立地图之间的重叠区域,进行地图融合。它们在实现时有着相同的步骤,即位置识别和位姿图优化,从而优化关键帧的位姿和特征点的 3D 坐标。

2.4.1 位置识别与相对位姿的计算

位置识别(Place recognition)^[52]用来识别曾经到达过的场景,是闭环或地图融合的基础,一般通过下采样图像等作为全局描述子来检测 2 幅图像是否匹配。在小场景中,使用存储的关键帧的小图像即可完成匹配^[53];当场景变大时,为确保系统的实时性,通常使用词袋技术进行位置识别,如 DBoW2^[54]等。DBoW2 方法使用二进制描述子

(BRIEF 或 ORB) 和 FAST 角点作为特征点, 具有很好的高效性和实时性, 是目前单目视觉 SLAM 位置识别的主流方法。在多智能体协同视觉 SLAM 中, 通常采用基于神经网络的方法^[55]来实现。

在完成位置识别后, 为了进一步优化, 需要获取 2 幅图像之间的相对位姿, 目前, 有 2D-2D 匹配、2D-3D 匹配以及 3D-3D 匹配三种方法来实现相对位姿的计算。在基于 2D-2D 匹配的方法中, 只需要用到相匹配特征点的 2D 坐标信息, 将平移向量的尺度作为未知变量, 利用五点算法^[56]对本质矩阵进行估计, 利用 RANSAC 方法来处理异常值, 再通过矩阵分解从基本矩阵中提取两帧之间的相对位姿; 在基于 2D-3D 匹配^[57]的方法中, 使用 PnP 算法或非线性最小二乘优化的方法, 利用当前帧中 2D 特征点与另一帧中 3D 特征点的匹配关系获得相对位姿; 而在基于 3D-3D 匹配的方法中, 可以直接利用三维点云计算得到相对位姿, 这一方法通常用于 RGB-D SLAM^[58] 和半稠密视觉 SLAM 系统。

2.4.2 位姿和场景三维坐标的优化

在完成位置识别后, 一般使用光束平差法对运动主体的位姿和场景 3D 结构进行优化, 光束平差法通过 Gaussian-Newton 迭代或 Levenberg-Marquardt 迭代对残差进行非线性最小二乘的求解, 但是当问题规模较大或初始值与真实值差距较大时, 很容易得到局部极小值。因此, 为了获得更好的初始值, 可以在使用光束平差法之前进行位姿图的优化^[59]。

在位姿图中, 节点表示关键帧的相机位姿, 边表示不同位姿之间的约束关系, 通常为刚性变换。将位置识别引入的新边插入到位姿图中, 然后把新边和旧边的所有约束集合在一起, 通过非线性最小二乘迭代求解, 使得约束误差平方和最小。事实证明, 采用刚性变换作为位姿约束对旋转和平移漂移的校正效果较好, 但对尺度漂移的校正效果较差。因此, 更好的方法是使用相似变换作为位姿约束。

位姿图优化中的一个重要问题是优化过程中, 可能存在一些错误的回环检测, 也就是将不同位置错误地识别为相同位置。这些错误大多是由于描述不完善或感知混淆造成的, 从而导致了错误的相对位姿约束, 如果处理不当会对地图构建造成较大的影响。为了避免错误的发生, 可以通过匹配帧之间相对变换的先验信息, 或者一组回环检测的一致性检验^[60]结果来识别异常值。一致性检验过程如

图 1 所示, 红线表示闭环测量值, 黑色实线表示连续帧之间的相对变换。

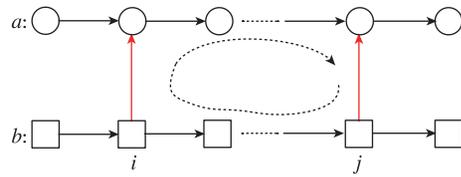


图 1 a、b 两机器人对闭环测量的一致性检验
Fig. 1 Consistency test of closed-loop measurement between robot a & b

经过位姿图优化后, 通过光束平差法进一步优化关键帧的相机位姿和点云三维坐标^[61]。其中, 光束平差问题可通过开源工具^[62-64]进行求解, 这一过程与单目视觉 SLAM 大致相同。一些系统^[65-66]使用视觉惯性 SLAM 作为前端, 在这种情况下, 光束平差的代价函数中还需要包含惯性测量值, 为了避免重复积分和大量计算, 惯性测量值可以通过预积分获得^[67]。

2.5 系统初始化

目前, 多智能体协同视觉 SLAM 有两种初始化的方法, 一种典型的初始化方法是对前 2 个关键帧之间的相对位姿使用五点算法, 并对相应的特征点进行三角化, 得到初始的 3D 点, 为每个智能体构建独立的初始化局部地图, 一旦通过位置识别发现局部地图之间有重叠时, 就将局部地图进行融合; 第二种方法要求所有的相机能够在场景中看到同一个地方, 从不同相机捕获的图像中估计出运动主体的相对位姿, 由所有的智能体一起完成。

3 分布式架构

现有的多智能体视觉 SLAM 多采用中心式架构, 该架构中包含能够处理与地图优化有关的多种计算任务的服务器, 例如: 回环检测、姿态优化、光束平差等。但是当服务器出现故障或通信受限时, 这种中心式架构是不可取的, 因此对分布式架构^[68]的研究成为一大热点。

分布式架构增加了单个智能体的功能, 相应智能体的软硬件结构更为复杂。但由于这种任务协调机制, 使得系统的鲁棒性更强, 几乎不受单个智能体的约束, 且可根据不同的条件需求扩充或减少智能体的个数, 系统灵活可调。

分布式架构一般有两种: 一种是每一个智能体都

需要与其他智能体进行信息交流,如图 2 所示;另一种是每一个智能体只需要与其附近的智能体进行信息交流,这种分布式架构效率更高,如图 3 所示。

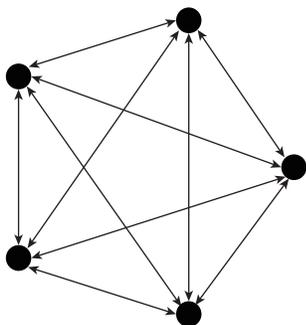


图 2 一种普通的分布式架构

Fig. 2 A common kind of distributed architecture

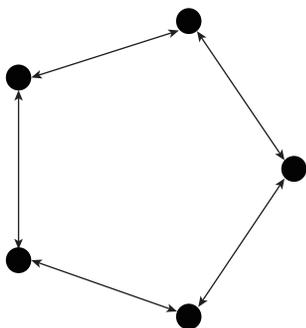


图 3 一种效率更高的分布式结构

Fig. 3 A more efficient distributed architecture

DDF-SAM^[69]中的分布式架构可以使机器人交换概要地图。概要地图是通过变量消除得到的且只包含环境特征信息,每个机器人利用其自身和相邻机器人的概要地图,以一种非线性约束最优化的方式进行邻域地图的构建,直至邻域地图中包括所有机器人的概要地图信息,从而得到全局地图。为了避免循环更新的问题,DDF-SAM 中采用了基于 RANSAC 算法的数据关联方法。DDF-SAM2.0^[70]中引入了一种抗因子(Anti-factor),使得机器人在避免循环更新的同时,降低了计算复杂度。目前,DDF-SAM 和 DDF-SAM2.0 中的分布式架构已经在二维平面场景中进行了测试。

苏黎世大学采用了一种新的分布式算法来解决位姿优化的问题^[71],其思想是在非线性最小二乘优化的迭代过程中,不采用矩阵分解的直接求解方法,而是采用逐次超松弛(Successive Over-Relaxation, SOR)和雅可比超松弛(Jacobi Over-Relaxa-

tion, JOR)等迭代方法求解正规方程,减少了数据传输量,推动了分布式架构在多智能体协同视觉 SLAM 领域的发展。

分布式架构因其独特的任务协调机制,在车联网应用场景中,可以对道路环境等数据进行全网实时更新,具有广泛的应用前景;在多机器人系统、编队控制等应用场景中,对提高系统鲁棒性和智能性具有重要研究价值。与中心式架构相比,将分布式架构应用到多智能体视觉 SLAM 系统中要复杂得多,主要的技术难点可以归结如下 2 个方面:

1) 分布式架构需要将计算任务分配给每一个具有有限计算能力的智能体,并设计相应的分布式算法来完成计算任务;另外,由于不同的智能体之间交换的信息量不完全相同,所以在设计分布式算法时需要考虑通信负载和带宽限制。

2) 分布式架构需要考虑不同智能体之间数据的同步和一致性,这个问题随着智能体数量的增加而变得更加复杂。

4 总结与展望

随着计算机和人工智能等技术的发展,多智能体协同视觉 SLAM 得到了广泛的关注。通过实现运动主体的实时精确定位和地图构建,多智能体协同视觉 SLAM 对机器人领域和增强现实领域产生了积极的影响。目前,国内外对多智能体协同视觉 SLAM 的研究已经取得了很大进展,但是仍有一些问题需要解决,例如:如何设计有效的分布式算法来解决相机间的协同和地图优化问题;如何在算法上解决高速运动情况下导致的视觉信息鲁棒性降低的问题。而且,我国在多智能体协同视觉 SLAM 的研究中与发达国家仍有较大差距,所以要进一步加强与先进国家的技术交流,缩小与国外先进水平的差距。

参考文献

- [1] Fuentes-Pacheco J, Ruiz-Ascencio J, Rendón-Mancha J M. Visual simultaneous localization and mapping: a survey[J]. Artificial Intelligence Review, 2015, 43(1): 55-81.
- [2] Scaramuzza D, Fraundorfer F. Visual odometry [Tutorial][J]. IEEE Robotics & Automation Magazine, 2011, 18(4): 80-92.
- [3] Cui Z, Tan P. Global structure-from-motion by similarity averaging[C]//Proceedings of 2015 IEEE Inter-

- national Conference on Computer Vision (ICCV). IEEE, 2015.
- [4] Bohren J, Foote T, Keller J, et al. Little ben: the ben franklin racing team's entry in the 2007 DARPA urban challenge[J]. *Journal of Field Robotics*, 2008, 25(9): 598-614.
- [5] Urmson C, Anhalt J, Bagnell D, et al. Autonomous driving in urban environments: boss and the urban challenge[J]. *Journal of Field Robotics*, 2008, 25(8): 425-466.
- [6] Howard A. Multi-robot simultaneous localization and mapping using particle filters[J]. *The International Journal of Robotics Research*, 2006, 25(12): 1243-1256.
- [7] Choudhary S, Carlone L, Nieto C, et al. Distributed mapping with privacy and communication constraints: lightweight algorithms and object-based models[J]. *The International Journal of Robotics Research*, 2017, 36(12): 1286-1311.
- [8] Schonberger J L, Frahm J M. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)-Las Vegas, NV, USA (2016. 6. 27-2016. 6. 30)[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)-Structure-from-Motion Revisited. IEEE, 2016: 4104-4113.
- [9] 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.
- [10] Mourikis A, Roumeliotis S. A multi-state constraint Kalman filter for vision-aided inertial navigation[C]// Proceedings of IEEE International Conference on Robotics & Automation. IEEE, 2007.
- [11] Li M, Mourikis A I. High-precision, consistent EKF-based visual-inertial odometry[J]. *The International Journal of Robotics Research*, 2013, 32(6): 690-711.
- [12] Triggs B. Bundle adjustment—a modern synthesis [M]// *Vision Algorithms: Theory and Practice*. Springer Berlin Heidelberg, 1999.
- [13] Kaess M, Johannsson 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.
- [14] Straub J, Hilsenbeck S, Schroth G, et al. Fast relocalization for visual odometry using binary features [C]// Proceedings of IEEE International Conference on Image Processing (ICIP). IEEE, 2013.
- [15] Williams B P, Klein G, Reid I D. Real-time SLAM relocalisation[C]// Proceedings of IEEE 11th International Conference on Computer Vision (ICCV). Rio De Janeiro, Brazil, 2007.
- [16] Montemerlo M, Becker J, Bhat S, et al. Junior: the stanford entry in the urban challenge[J]. *Journal of Field Robotics*, 2008, 25(9): 569-597.
- [17] Links I K F. An introduction to the Kalman filter [R]. University of North Carolina at Chapel Hill, United States, 1995.
- [18] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]// Proceedings of 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2007.
- [19] Castle R O, Klein G, Murray D W. Video-rate localization in multiple maps for wearable augmented reality[C]// Proceedings of 2008 12th IEEE International Symposium on Wearable Computers (ISWC). Pittsburgh, PA, USA. IEEE, 2008.
- [20] Zou D, Tan P. CoSLAM: collaborative visual SLAM in dynamic environments[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 35(2): 354-366.
- [21] Castle R O, Klein G, Murray D W. Video-rate localization in multiple maps for wearable augmented reality[C]// Proceedings of 2008 12th IEEE International Symposium on Wearable Computers (ISWC). Pittsburgh, PA, USA. IEEE, 2008.
- [22] Riazuelo L, Civera J, Montiel J M M. C2TAM: a cloud framework for cooperative tracking and mapping[J]. *Robotics and Autonomous Systems*, 2014, 62(4): 401-413.
- [23] Forster C, Lynen S, Kneip L, et al. Collaborative monocular SLAM with multiple micro aerial vehicles [C]// Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2013.
- [24] Schmuck P, Chli M. CCM-SLAM: robust and efficient centralized collaborative monocular simultaneous localization and mapping for robotic teams[J]. *Journal of Field Robotics*, 2019, 36(4): 763-781.
- [25] 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.
- [26] Qin T, Li P, Shen S, et al. VINS-Mono: a robust and versatile monocular visual-inertial state estimator [J]. *IEEE Transactions on Robotics*, 2018, 34(4):

- 1004-1020.
- [27] Yu F, Xian W, Chen Y, et al. BDD100K: a diverse driving video database with scalable annotation tooling [J]. arXiv: Computer Vision and Pattern Recognition, 2018.
- [28] Zou D, Wu Y, Pei L, et al. StructVIO: visual-inertial odometry with structural regularity of man-made-environments[J]. IEEE Transactions on Robotics, 2019, 35(4): 999-1013.
- [29] Feng Z, Li J, Dai T. Sensor synchronization for android phone tightly-coupled visual-inertial SLAM [C]// Proceedings of China Satellite Navigation Conference. Springer, Singapore, 2018.
- [30] Morrison J G, Gálvez-López D, Sibley G. MOARSLAM: multiple operator augmented RSLAM [M]// Distributed Autonomous Robotic Systems. Springer Japan, 2016.
- [31] Shafi M, Molisch A F, Smith P J, et al. 5G: a tutorial overview of standards, trials, challenges, deployment and practice [J]. IEEE Journal on Selected Areas in Communications, 2017, 35(6): 1201-1221.
- [32] Engel J, Schöps T, Cremers D. LSD-SLAM: large-scale direct monocular SLAM [C]// Proceedings of European Conference on Computer Vision. Springer, Cham, 2014.
- [33] Engel J, Koltun V, Cremers D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 40(3): 611-625.
- [34] Rosten E, Drummond T. Machine learning for high-speed corner detection [C]// Proceedings of European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2006: 430-443.
- [35] Castle R O, Klein G, Murray D W. Video-rate localization in multiple maps for wearable augmented reality [C]// Proceedings of 2008 12th IEEE International Symposium on Wearable Computers (ISWC). Pittsburgh, PA, USA. IEEE, 2008.
- [36] Riazuelo L, Civera J, Montiel J M M. C2TAM: a cloud framework for cooperative tracking and mapping [J]. Robotics and Autonomous Systems, 2014, 62(4): 401-413.
- [37] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [38] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C]// Proceedings of 2011 IEEE International Conference on Computer Vision (ICCV). Barcelona, Spain. IEEE, 2011.
- [39] Calonder M, Lepetit V, Strecha C, et al. BRIEF: binary robust independent elementary features [C]// Proceedings of 2010 11th European Conference on Computer Vision (ECCV). Heraklion, Crete, Greece. Springer, Berlin, Heidelberg, 2010.
- [40] Bay H, Ess A, Tuytelaars T, et al. Speeded-up robust features (SURF) [J]. Computer Vision and Image Understanding, 2008, 110(3): 346-359.
- [41] Shi J, Tomasi C. Good features to track [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 1994: 593-600.
- [42] Gee A P, Chekhlov D, Calway A, et al. Discovering higher level structure in visual SLAM [J]. IEEE Transactions on Robotics, 2008, 24(5): 980-990.
- [43] 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.
- [44] Andrew A M. Multiple view geometry in computer vision [J]. Kybernetes, 2001, 30(9/10): 1865-1872.
- [45] Nistér D. Preemptive RANSAC for live structure and motion estimation [J]. Machine Vision and Applications, 2005, 16(5): 321-329.
- [46] Robotics C. Introduction to robust estimation and hypothesis testing [J]. Technometrics, 2005, 40(1): 77-78.
- [47] 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.
- [48] Strasdat H, Montiel J M M, Davison A J. Real-time monocular SLAM: why filter? [C]// Proceedings of 2010 IEEE International Conference on Robotics and Automation (ICRA). Anchorage, Alaska, USA. IEEE, 2010.
- [49] Quan L, Lan Z. Linear N-point camera pose determination [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 774-780.
- [50] Lepetit V, Moreno-Noguer F, Fua P. EPnP: an accurate O(n) solution to the PnP problem [J]. International Journal of Computer Vision, 2009, 81(2): 155-166.
- [51] Grisetti G, Kummerle R, Stachniss C, et al. A tutorial on graph-based SLAM [J]. IEEE Intelligent Transportation Systems Magazine, 2010, 2(4): 31-43.
- [52] Lowry S, Sunderhauf N, Newman P, et al. Visual

- place recognition: a survey[J]. *IEEE Transactions on Robotics*, 2015, 32(1): 1-19.
- [53] Klein G, Murray D W. Improving the agility of key-frame-based SLAM[C]// *Proceedings of European Conference on Computer Vision*. Springer, Berlin, Heidelberg, 2008; 802-815.
- [54] Galvez-Lopez 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.
- [55] Arandjelovic R, Gronat P, Torii A, et al. NetVLAD: CNN architecture for weakly supervised place recognition[C]// *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016; 5297-5307.
- [56] Nister D. An efficient solution to the five-point relative pose problem[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2004, 26(6): 756-770.
- [57] Strasdat H, Davison A J, Montiel J M M, et al. Double window optimisation for constant time visual SLAM[C]// *Proceedings of 2011 IEEE International Conference on Computer Vision (ICCV)*. Barcelona, Spain. IEEE, 2011.
- [58] Endres F, Hess J, Engelhard N, et al. An evaluation of the RGB-D SLAM system[C]// *Proceedings of 2012 IEEE International Conference on Robotics and Automation*. IEEE, 2012; 1691-1696.
- [59] Kummerle R, Grisetti G, Strasdat H, et al. G2o: a general framework for graph optimization[C]// *Proceedings of 2011 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2011.
- [60] Li H. Consensus set maximization with guaranteed global optimality for robust geometry estimation [C]// *Proceedings of 2009 IEEE 12th International Conference on Computer Vision*. IEEE, 2009; 1074-1080.
- [61] 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*. Morgan Kaufmann Publishers Inc. 1997.
- [62] Lourakis M I A, Argyros A A. SBA: a software package for generic sparse bundle adjustment[J]. *ACM Transactions Mathematical Software (TOMS)*, 2009, 36(1): 1-30.
- [63] Wu C, Agarwal S, Curless B, et al. Multicore bundle adjustment[C]// *Proceedings of 2011 24th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Colorado Springs, CO, USA. IEEE, 2011.
- [64] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters[J]. *IEEE Transactions on Robotics*, 2007, 23(1): 34-46.
- [65] Karrer M, Schmuck P, Chli M. CVI-SLAM—collaborative visual-inertial SLAM[J]. *IEEE Robotics and Automation Letters*, 2018, 3(4): 2762-2769.
- [66] Loianno G, Mulgaonkar Y, Brunner C, et al. A swarm of flying smartphones[C]// *Proceedings of 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016.
- [67] Forster C, Carlone L, Dellaert F, et al. On-manifold preintegration for real-time visual-inertial odometry [J]. *IEEE Transactions on Robotics*, 2015, 33(1): 1-21.
- [68] Grisetti G, Stachniss C, Burgard W. Improved techniques for grid mapping with rao-blackwellized particle filters[J]. *IEEE Transactions on Robotics*, 2007, 23(1): 34-46.
- [69] Cunningham A, Paluri M, Dellaert F. DDF-SAM: fully distributed SLAM using constrained factor graphs[C]// *Proceedings of 2010 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2010.
- [70] Cunningham A, Indelman V, Dellaert F. DDF-SAM 2.0: consistent distributed smoothing and mapping [C]// *Proceedings of 2013 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2013; 5220-5227.
- [71] Cieslewski T, Choudhary S, Scaramuzza D. Data-efficient decentralized visual SLAM[C]// *Proceedings of 2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018; 2466-2473.