DOI:10.11992/tis.201607026
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20170111.1705.006.html

视觉 SLAM 综述

权美香[1],朴松昊[1,2],李国[1]
(1.哈尔滨工业大学 多智能体机器人实验室,黑龙江 哈尔滨 150000;2. 哈尔滨工业大学 多智能体机器人实验室,黑龙江 哈尔滨 150000)

摘 要:视觉 SLAM 指的是相机作为唯一的外部传感器,在进行自身定位的同时创建环境地图。SLAM 创建的地图的好坏对之后自主的定位、路径规划以及壁障的性能起到一个决定性的作用。本文对基于特征的视觉 SLAM 方法和直接的 SLAM 方法,视觉 SLAM 的主要标志性成果,SLAM 的主要研究实验室进行了介绍,并介绍了 SIFT,SURF,ORB 特征的检测与匹配,关键帧选择方法,并对消除累积误差的闭环检测及地图优化的方法进行了总结。最后,对视觉 SLAM 的主要发展趋势及研究热点进行了讨论,并对单目视觉 SLAM,双目视觉 SLAM,RGB_D SLAM 进行了优缺点分析。

关键词:视觉同步定位与创建地图;单目视觉;RGB_D SLAM;特征检测与匹配;闭环检测
中图分类号: TP391 **文献标志码**:A **文章编号**:1673-4785(2016)06-0768-09

中文引用格式:权美香,朴松昊,李国. 视觉 SLAM 综述[J]. 智能系统学报, 2016, 11(6): 768-776.
英文引用格式:QUAN Meixiang, PIAO Songhao, LI Guo. An overview of visual SLAM[J]. CAAI transactions on intelligent systems, 2016, 11(6): 768-776.

An overview of visual SLAM

QUAN Meixiang[1], PIAO Songhao[1,2], LI Guo[1]
(1. Multi-agent Robot Research Center, Harbin Institute of Technology, Harbin 150000, China; 2. Multi-agent Robot Research Center, Harbin Institute of Technology, Harbin 150000, China)

Abstract: Visual SLAM refers to simultaneously localizing itself and reconstructing the environment map using cameras as the only external sensor. The quality of the map created by SLAM plays a decisive role in the performance of the subsequent automatic localization, path planning, and obstacle avoidance. This paper introduced the feature-based visual SLAM and direct visual SLAM methods; the major symbolic achievement of visual SLAM; the main research laboratory of SLAM; and the method of SIFT, SURF, and ORB feature detection and matching, key frame selection. In addition, this paper summarized the loop closure detection and map optimization that removed the accumulated error. In the end, the development tendency and research highlights of SLAM were discussed and the advantages and disadvantages of monocular SLAM, binocular SLAM, and RGB_D SLAM were analyzed.

Keywords: visual simultaneous localization and mapping; monocular vision; RGB_D; feature detection and matching; loop closure detection

移动机器人的一个基本任务是在给定环境地图的条件下确定其所在的位置,然而环境地图并不是一开始就有的,当移动机器人进入未知的环境时,需要通过自身的传感器构建 3-D 环境地图,并且同时确定自身在地图中的位置,这就是 SLAM(simultaneous localization and mapping)问题,它是运动恢复结构(structure from motion,SfM)的实时版本。

收稿日期:2016-07-25. 网络出版日期:.
基金项目:.国家自然科学基金面上项目(61375081).
通信作者:朴松昊. E-mail: piaosh@hit.edu.cn.

相机作为唯一外部传感器的 SLAM 被称为视觉 SLAM。由于相机具有成本低,轻 ,很容易放到商品硬件上的优点,且图像含有丰富的信息,视觉 SLAM 得到了巨大的发展。根据采用的视觉传感器不同,可以将视觉 SLAM 主要分为三类:仅用一个相机作为唯一外部传感器的单目视觉 SLAM;使用多个相机作为传感器的立体视觉 SLAM,其中双目立体视觉的应用最多;基于单目相机与红外传感器结合构成的传感器的 RGB-D SLAM。

1　视觉 SLAM 方法介绍

视觉 SLAM 算法可根据利用图像信息的不同分为基于特征的 SLAM 方法和 direct SLAM 方法。下面就这两种方法对单目视觉 SLAM 与 RGB_D SLAM 进行简要的介绍,并介绍视觉 SLAM 的标志性成果以及国内外主要的研究单位。

1.1　基于特征的 SLAM 方法

基于特征的视觉 SLAM 方法指的是对输入的图像进行特征点检测及提取,并基于 2-D 或 3-D 的特征匹配计算相机位姿及对环境进行建图。如果对整幅图像进行处理,则计算复杂度太高,由于特征在保存图像重要信息的同时有效减少了计算量,从而被广泛使用。

早期的单目视觉 SLAM 的实现是借助于滤波器而实现的[1-4]。利用扩展卡尔曼滤波器(extended Kalman filter,EKF)来实现同时定位与地图创建,其主要思想是使用状态向量来存储相机位姿及地图点的三维坐标,利用概率密度函数来表示不确定性,从观测模型和递归的计算,最终获得更新的状态向量的均值和方差。但是由于 EKF 的引进,SLAM 算法会有计算复杂度及由于线性化而带来的不确定性问题。为了弥补 EKF 的线性化对结果带来的影响,在文献[5-7]里将无迹卡尔曼滤波器(Unscented Kalman filter, UKF)或改进的 UKF 引入到单目视觉 SLAM 中。该方法虽然对不确定性有所改善,但同时也增加了计算复杂度。此外,文献[8-9]利用 Rao-Blackwellized 粒子滤波(Particle filter)实现了单目视觉 SLAM。该方法避免了线性化,且对相机的快速运动有一定的弹力,但是为了保证定位精度,则需要使用较多的粒子,从而大大提高了计算复杂度。

之后基于关键帧的单目视觉 SLAM[10-13]逐渐发展起来,其中最具代表性的是 parallel tracking and mapping(PTAM)[10],该论文提出了一个简单、有效的提取关键帧的方法,且将定位和创建地图分为两个独立的任务,并在两个线程上进行。文献[13]是在关键帧的基础上提出的一个单目视觉 SLAM 系统,将整个 SLAM 过程分为定位、创建地图、闭环 3 个线程,且对这 3 个任务使用相同的 ORB 特征,且引进本质图的概念来加速闭环校正过程。

微软公司推出的 Kinect 相机,能够同时获得图像信息及深度信息,从而简化了三维重建的过程,且由于价格便宜,基于 RGB_D 数据的 SLAM 得到了迅速的发展[14-21]。文献[18]是最早提出的使用 RGBD 相机对室内环境进行三维重建的方法,在彩色图像中提取 SIFT 特征并在深度图像上查找相应的深度信息。然后使用 RANSAC 方法对 3-D 特征点进行匹配并计算出相应的刚体运动变换,再以此作为 ICP(iterative closest point)的初始值来求出更精确的位姿。RGBD SLAM 通常使用 ICP 算法来进行位姿估计,图 1 给出了 ICP 算法的基本流程。文献[20]与文献[18]类似,不同点在于对彩色图像进行的是 SURF 特征提取,并用 ICP 对运动变换进行优化后,最后使用 Hogman 位姿图优化方法求出全局最优位姿。

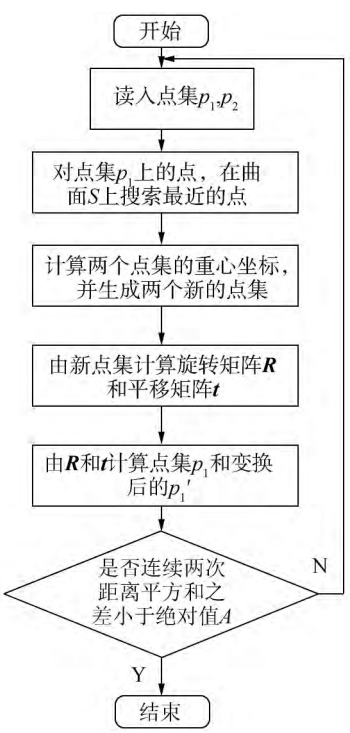


图 1　ICP 算法流程图

Fig.1　Flow diagram of ICP algorithm

1.2　直接的 SLAM 方法

直接的 SLAM 方法指的是直接对像素点的强度进行操作,避免了特征点的提取,该方法能够使用图像的所有信息。此外,提供更多的环境几何信息,有助于对地图的后续使用。且对特征较少的环境有更高的准确性和鲁棒性。

近几年,基于直接法的单目视觉里程计算法才

被提出[22-25]。文献[22]的相机定位方法依赖图像的每个像素点,即用稠密的图像对准来进行自身定位,并构建出稠密的 3-D 地图。文献[23]对当前图像构建半稠密 inverse 深度地图,并使用稠密图像配准(dense image alignment)法计算相机位姿。构建半稠密地图即估计图像中梯度较大的所有像素的深度值,该深度值被表示为高斯分布,且当新的图像到来时,该深度值被更新。文献[24]对每个像素点进行概率的深度测量,有效降低了位姿估计的不确定性。文献[25]提出了一种半直接的单目视觉里程计方法,该方法相比于直接法不是对整幅图像进行直接匹配从而获得相机位姿,而是通过在整幅图像中提取的图像块来进行位姿的获取,这样能够增强算法的鲁棒性。为了构建稠密的三维环境地图,Engel 等[26]提出了 LSD-SLAM 算法(large-scale direct SLAM),相比之前的直接的视觉里程计方法,该方法在估计高准确性的相机位姿的同时能够创建大规模的三维环境地图。

文献[27]提出了 Kinect 融合的方法,该方法通过 Kinect 获取的深度图像对每帧图像中的每个像素进行最小化距离测量而获得相机位姿,且融合所有深度图像,从而获得全局地图信息。文献[28]使用图像像素点的光度信息和几何信息来构造误差函数,通过最小化误差函数而获得相机位姿,且地图问题被处理为位姿图表示。文献[29]是较好的直接 RGB_D SLAM 方法,该方法结合像素点的强度误差与深度误差作为误差函数,通过最小化代价函数,从而求出最优相机位姿,该过程由 g2o 实现,并提出了基于熵的关键帧提取及闭环检方法,从而大大降低了路径的误差。

1.3 视觉 SLAM 主要标志性成果

视觉 SLAM 的标志性成果有 Andrew Davison 提出的 MonoSLAM[3],是第 1 个基于 EKF 方法的单目 SLAM,能够达到实时但是不能确定漂移多少,能够在概率框架下在线创建稀疏地图。DTAM[24]是 2011 年提出的基于直接法的单目 SLAM 算法,该方法通过帧率的整幅图像对准来获得相对于稠密地图的相机的 6 个自由度位姿,能够在 GPU 上达到实时的效果。PTAM[10]是由 Georg Klein 提出的第 1 个用多线程处理 SLAM 的算法,将跟踪和建图分为两个单独的任务并在两个平行的线程进行处理。KinectFusion[27]是第 1 个基于 Kinect 的能在 GPU 上实时构建稠密三维地图的算法,该方法仅使用 Kinect 相机获取的深度信息去计算传感器的位姿以及构建精确的环境 3-D 地图模型。2014 年提出的 LSD-SLAM[26]是直接的单目 SLAM 方法,即直接对图像的像素点进行处理,相比于之前的基于直接法的单目视觉里程计,不仅能计算出自身的位姿,还能构建出全局的半稠密且精确的环境地图。其中的追踪方法,直接在 sim3 上进行操作,从而能够准确地检测尺度漂移,可在 CPU 上实时运行。ORB_SLAM[13]是 2015 年出的比较完整的基于关键帧的单目 SLAM 算法,将整个系统分为追踪、地图创建、闭环控制 3 个线程进行处理,且特征的提取与匹配、稀疏地图的创建、位置识别都是基于 ORB 特征,其定位精确度很高,且可以实时运行。

1.4 SLAM 的主要研究实验室

SLAM 的主要研究实验室有:

1)苏黎世联邦理工学院的 Autonomous System Lab,该实验室在 tango 项目上与谷歌合作,负责视觉—惯导的里程计,基于视觉的定位和深度重建算法。

2)明尼苏达大学的 Multiple Autonomous Robotic Systems Laboratory,主要研究四轴飞行器导航,合作建图,基于地图的定位,半稠密地图创建等。

3)慕尼黑理工大学的 The Computer Vision Group,主要研究基于图像的 3-D 重建,机器人视觉,视觉 SLAM 等。

2 视觉 SLAM 关键性问题

2.1 特征检测与匹配

目前,点特征的使用最多,最常用的点特征有 SIFT(scale invariant feature transform)[30]特征,SURT(speeded up robust features)[31]特征和 ORB(oriented fast and rotated BRIEF)[32]特征。SIFT 特征已发展 10 多年,且获得了巨大的成功。SIFT 特征具有可辨别性,由于其描述符用高维向量(128 维)表示,且具有旋转不变性、尺度不变性、放射变换不变性,对噪声和光照变化也有鲁棒性。[33-36]在视觉 SLAM 里使用了 SIFT 特征,但是由于 SIFT 特征的向量维数太高,导致时间复杂度高。SURF 特征具有尺度不变性、旋转不变性,且相对于 SIFT 特征的算法速度提高了 3 到 7 倍。在文献[37-39]SURF 被作为视觉 SLAM 的特征提取方法,与 SIFT 特征相比,时间复杂度有所降低。对两幅图像的 SIFT 和 SURF 特征进行匹配时通常是计算两个特征向量之间的欧氏距离,并以此作为特征点的相似性判断度量。ORB 特征是 FAST[40]特征检测算子与 BRIEF[41]描述符的结合,并在其基础上做了一些改进。ORB 特征最大的优点是计算速度快,是 SIFT 特征的 100 倍,SURF 特征的 10 倍,其原因是 FAST 特征检测速度就很快,再加上 BRIEF 描述符是

二进制串,大大缩减了匹配速度,而且具有旋转不变性,但不具备尺度不变性。文献[12-13,42-44]的 SLAM 算法中采用了 ORB 特征,大大加快了算法速度。ORB 特征匹配是以 BRIEF 二进制描述符的汉明距离为相似性度量的。

在大量包含直线和曲线的环境下,使用点特征时,环境中很多信息都将被遗弃,为了弥补这个缺陷,从而也提出了基于边特征的视觉 SLAM[45-46] 和基于区域特征的视觉 SLAM[47] 方法。

2.2　关键帧的选择

帧对帧的对准方法会造成大的累积漂浮,由于位姿估计过程中总会产生误差。为了减少帧对帧的对准方法带来的误差,基于关键帧的 SLAM[10-13,19,29] 方法被提出。

目前有几种选择关键帧的方法。文献[10,13]里当满足以下全部条件时该帧作为关键帧插入到地图里:从上一个关键帧经过了 n 个帧;当前帧至少能看到 n 个地图点,位姿估计准确性较高。文献[19]是当两幅图像看到的共同特征点数低于一定阈值时,创建一个新的关键帧。文献[29]提出了一种基于熵的相似性的选择关键帧的方法,由于简单的阈值不适用于不同的场景,对每一帧计算一个熵的相似性比,如果该值小于一个预先定义的阈值,则前一帧被选为新的关键帧,并插入地图里,该方法大大减少了位姿漂浮。

2.3　闭环检测(loop closing)方法

闭环检测及位置识别,判断当前位置是否是以前已访问过的环境区域。三维重建过程中必然会产生误差累积,实现闭环是消除的一种手段。在位置识别算法中,视觉是主要的传感器[3,48-50]。文献[51]对闭环检测方法进行了比较,且得出图像对图像[52-53]的匹配性能优于地图对地图[54],图像对地图[55]的匹配方法。

图像对图像的匹配方法中,词袋(bag of words)[56] 方法由于其有效性得到了广泛的应用[12-13,57-59]。词袋指的是使用视觉词典树(visual vocabulary tree)将一幅图像的内容转换为数字向量的技术。对训练图像集进行特征提取,并将其特征描述符空间通过 K 中心点聚类(k medians clustering)方法离散化为个簇,由此,词典树的第一节点层被创建。下面的层通过对每个簇重复执行这个操作而获得,直到共获得层。最终获得 W 个叶子节点,即视觉词汇。每层到每层的 K 中心聚类过程如图 2 所示[56]。

文献[60]对重定位和闭环检测提出了统一的方法,它们使用基于 16 维的 SIFT 特征的词典方法不断地搜索已访问过的位置。文献[61-62]使用基于 SURF 描述符的词典方法去进行闭环检测 SURF 特征,SURF 特征提取需要花费 400 ms 去进行。文献[63]使用 SIFT 特征执行全局定位,且用 KD 树来排列地图点。文献[59]提出了一种使用基于 FAST 特征检测与 BRIEF 二进制描述符词典,且添加了直接索引(direct index),直接索引的引入使得能够有效地获得图像之间的匹配点,从而加快闭环检测的几何验证。文献[12]用基于 ORB 特征的词典方法进行位置识别,由于 ORB 特征具有旋转不变性且能处理尺度变化,该方法能识别位置从不同的视角。文献[13]的位置识别方法建于文献[12]的主要思想上,即使用基于 ORB 特征的词典方法选出候选闭环,再通过相似性计算进行闭环的几何验证。

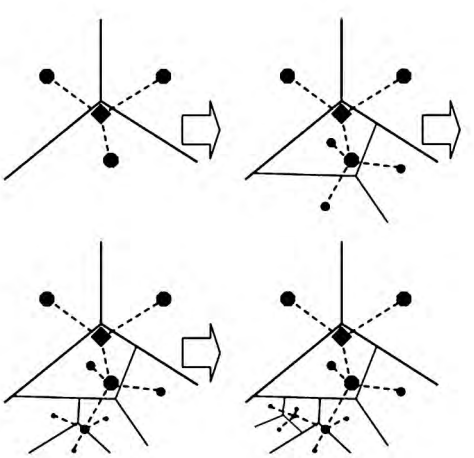


图 2　K 中心点聚类方法

Fig.2　Method of k medians clustering

2.4　地图优化

对于一个在复杂且动态的环境下工作的机器人,3-D 地图的快速生成是非常重要的,且创建的环境地图对之后的定位、路径规划及壁障的性能起到一个关键性的作用,从而精确的地图创建也是非常重要的。

闭环检测成功后,往地图里添加闭环约束,执行闭环校正。闭环问题可以描述为大规模的光束平差法(bundle adjustment)[64-65] 问题,即对相机位姿及所有的地图点 3-D 坐标进行优化,但是该优化计算复杂度太高,从而很难实现实时。

一种可执行方法为通过位姿图优化(pose graph optimization)方法来对闭环进行优化,顶点为相机位姿,边表示位姿之间相对变换的图称为位姿图,位姿图优化即将闭环误差沿着图进行分配,即均匀分配到图上的所有位姿上。图优化通常由图优化框架

g2o(general graph optimization)[66] 里的 LM(levenberg-marquardt)算法实现。

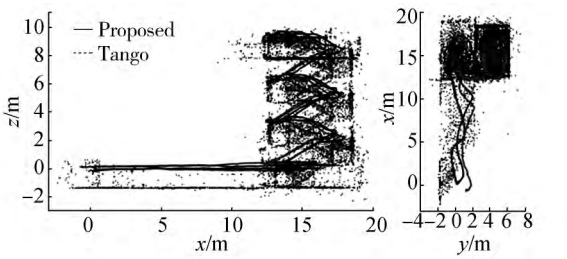
文献[29]提出的 RGB_D SLAM 算法的位姿图里每个边具有一个权重,从而在优化过程中,不确定性高的边比不确定性低的边需要变化更多去补偿误差,并在最后,对图里的每个顶点进行额外的闭环检测且重新优化整个图。文献[67]里,在闭环校正步骤使用了位姿图优化技术去实现旋转,平移及尺度漂浮的有效校正。文献[13]在闭环检测成功后构建了本质图,并对该图进行位姿图优化。本质图包含所有的关键帧,但相比于 covisibility 图[68],减少了关键帧之间的边约束。本质图包含生成树、闭环连接及 covisibility 图里权重较大的边。

3 视觉 SLAM 主要发展趋势及研究热点

3.1 多传感器融合

相机能够捕捉场景的丰富细节,而惯性测量单元(inertial measurement unit,IMU)有高的帧率且相对小的能够获得准确的短时间估计,这两个传感器能够相互互补,从而一起使用能够获得更好的结果。

最初的视觉与 IMU 结合的位姿估计是用滤波方法解决的,用 IMU 的测量值作为预测值,视觉的测量值用于更新。文献[69]提出了一种基于 EKF 的 IMU 与单目视觉的实时融合方法,提出一种测量模型能够表示一个静态特征被多个相机所观察时的几何约束,该测量模型是最优的且不需要在 EKF 的状态向量里包括特征的 3-D 坐标。文献[70]将融合问题分为两个线程进行处理,连续图像之间的惯性测量和特征跟踪被局部地在第 1 个线程进行处理,提供高频率的位置估计,第 2 个线程包含一个间歇工作的光束法平差的迭代估计,能够减少线性误差的影响。许多结果都已证明在准确性上基于优化的视觉 SLAM 优于基于滤波的 SLAM 方法。文献[71]将 IMU 的误差以全概率的形式融合到路标的重投影误差里,构成将被优化的联合非线性误差函数,其中通过关键帧来边缘化之前的状态去维持一个固定大小的优化窗口,保证实时操作。考虑到基于优化方法的视觉-惯导导航的计算复杂度问题,文献[72]通过预积分选出的关键帧之间的惯性测量来进行解决,预积分能够精确地概括数百个惯性测量到一个单独的相对运动约束,这个预积分的 IMU 模型能被完美地融合到视觉-惯性的因子图的框架下。该系统的实验结果表明该系统要比 Google 的 Tango 还要精确[72],如图 3 所示。



(a)正视图 (b)俯视图

图 3 与 Tango 的性能比较[72]

Fig.3 Performance comparison with tango[72]

共 160 m 的轨迹开始于(0,0,0)点,走到建筑物的第 3 层,然后再返回到初始点。其中,Google Tango 积累 1.4 m 的误差,而该提出的方法仅有 0.5 m的漂移。

即使在视觉与 IMU 融合的系统里,当机器人运动较为剧烈时,由于其不确定性增大,还是会导致定位失败,从而系统的鲁棒性有待进一步提高。并且为在现实生活中实现实时定位,其计算复杂度也需要进行改进。

3.2 SLAM 与深度学习的结合

随着深度学习在计算机视觉领域的大成功,大家对深度学习在机器人领域的应用有很大的兴趣。SLAM 是一个大系统,里面有很多子模块,例如闭环检测,立体匹配等,都可通过深度学习的使用来获得更优的结果。

文献[73]提出了一种基于深度学习的立体匹配方法,用卷积神经网络来学习小图像块间的相似性,该卷积神经网络输出的结果用于线性化立体匹配代价。文献[74]通过整合局部敏感散列法和新的语义搜寻空间划分的优化技术,使用卷积神经网络和大的地图达到实时的位置识别。文献[75]使用卷积神经网络去学习视觉里程计的最佳的视觉特征和最优的估计器。文献[76]提出了一种重定位系统,使用贝叶斯卷积神经网络从单个彩色图像计算出六个自由度的相机位姿及其不确定性。

4 视觉 SLAM 的优缺点分析

4.1 单目视觉 SLAM 的优缺点

单目相机应用灵活、简单、价格低。但是,单目视觉 SLAM 在每个时刻只能获取一张图像,且只能依靠获得的图像数据计算环境物体的方向信息,无法直接获得可靠的深度信息,从而初始地图创建及特征点的深度恢复都比较困难。此外,尺度不确定性是单目 SLAM 的主要特点,它是主要的误差源之一,但是正是尺度不确定性才使得单目 SLAM 能够

在大尺度和小尺度环境之间进行自由转换。

4.2　双目视觉 SLAM 的优缺点

双目视觉 SLAM 利用外极线几何约束的原理去匹配左右两个相机的特征,从而能够在当前帧速率的条件下直接提取完整的特征数据,因而应用比较广泛,它直接解决了系统地图特征的初始化问题。但是系统设计比较复杂,系统成本比较高,且它的视角范围受到一定限制,不能够获取远处的场景,从而只能在一定的尺度范围内进行可靠的测量,从而缺乏灵活性。

4.3　RGBD SLAM 的优缺点

深度相机在获得彩色图像的同时获得深度图像,从而方便获得深度信息,且能够获得稠密的地图,但是成本高,体积大,有效探测距离太短,从而可应用环境很有限。

5　结束语

十几年来,视觉 SLAM 虽然取得了惊人的发展,但是仅用摄像机作为唯一外部传感器进行同时定位与三维地图重建还是一个很具挑战性的研究方向,想要实时进行自身定位且构建类似人眼看到的环境地图还有很长的科研路要走。为了弥补视觉信息的不足,视觉传感器可以与惯性传感器(IMU)、激光等传感器融合,通过传感器之间的互补获得更加理想的结果。此外,为了能在实际环境中进行应用,SLAM 的鲁棒性需要很高,从而足够在各种复杂环境下进行准确的处理,SLAM 的计算复杂度也不能太高,从而达到实时效果。

参考文献:

[1] DAVISON A J. SLAM with a single camera[C] //Proceedings of Workshop on Concurrent Mapping an Localization for Autonomous Mobile Robots in Conjunction with ICRA. Washington, DC, USA, 2002: 18-27.

[2] DAVISON A J. Real-time simultaneous localisation and mapping with a single camera [C]//Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC, USA, 2003: 1403-1410.

[3] 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.

[4] 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.

[5] MARTINEZ-CANTIN R, CASTELLANOS J A. Unscented SLAM for large-scale outdoor environments[C]//Proceedings of 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. Edmonton, Alberta, Canada, 2005: 3427-3432

[6] CHEKHLOV D, PUPILLI M, MAYOL-CUEVAS W, et al. Real-time and robust monocular SLAM using predictive multi-resolution descriptors[C]// Proceedings of the Second International Conference on Advances in Visual Computi. Lake Tahoe, USA, 2006: 276-285.

[7] HOLMES S, KLEIN G, MURRAY D W. A square root unscented kalman filter for visual monoSLAM[C]//Proceedings of 2008 International Conference on Robotics and Automation, ICRA. Pasadena, California, USA, 2008: 3710-3716.

[8] SIM R, ELINAS P, GRIFFIN M, et al. Vision-based SLAM using the Rao-Blackwellised particle filter[J]. IJCAI workshop on reasoning with uncertainty in robotics, 2005, 9(4): 500-509.

[9] LI Maohai, HONG Bingrong, CAI Zesu, et al. Novel Rao-Blackwellized particle filter for mobile robot SLAM using monocular vision [J]. International journal of intelligent technology, 2006, 1(1): 63-69.

[10] KLEIN G, MURRAY D. Parallel Tracking and Mapping for Small AR Workspaces[C]// IEEE and ACM International Symposium on Mixed and Augmented Reality. Nara, Japan, 2007: 225-234.

[11] KLEIN G, MURRAY D. Improving the agility of keyframe-based SLAM[C]// European Conference on Computer Vision. Marseille, France, 2008: 802-815.

[12] MUR-ARTAL R, TARDÓS J D. Fast relocalisation and loop closing in keyframe-based SLAM[C]// IEEE International Conference on Robotics and Automation. New Orleans, LA, 2014: 846-853.

[13] 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.

[14] KHOSHELHAM K, ELBERINK S O. Accuracy and resolution of Kinect depth data for indoor mapping applications [J]. Sensors, 2012, 12(2): 1437-1454.

[15] HOGMAN V. Building a 3-D Map from RGB-D sensors [D]. Stockholm, Sweden: Royal Institute of Technology, 2012.

[16] HENRY P, KRAININ M, HERBST E, et al. RGB-D mapping: Using depth cameras for dense 3-D modeling of indoor environments[C]//12th International Symposium on Experimental Robotics. Berlin, Germany, 2014: 477-491.

[17] DRYANOVSKI I, VALENTI R G, XIAO J Z. Fast visual odometry and mapping from RGB-D data[C]// IEEE In-

ternational Conference on Robotics and Automation. Piscataway, USA, 2013: 2305-2310.

[18] HENRY P, KRAININ M, HERBST E, et al. RGB-D mapping: using depth cameras for dense 3-D modeling of indoor environments[M]//KHATIB O, KUMAR V, PAPPAS G J. Experimental Robotics. Berlin Heidelberg: Springer, 2014: 647-663.

[19] HENRY P, KRAININ M, HERBST E, et al. RGB-D Mapping: Using Depth Cameras for Dense 3-D Modeling of Indoor Environments[C]// 12th International Symposium on Experimental Robotics. Berlin Germany, 2014: 477-491.

[20] HENRY P, KRAININ M, HERBST E, et al. RGB-D mapping: Using Kinect-style depth cameras for dense 3-D modeling of indoor environments[J]. International journal of robotics research, 2012, 31(5): 647-663.

[21] ENGELHARD N, ENDRES F, HESS J. Real-time 3-D visual SLAM with a hand-held RGB-D camera[C]// Proceeedings of the RGB-D workshop on 3-D Perception in Robotics at the European Robotics Forum. Västerås, Sweden, 2011.

[22] STÜHMER J, GUMHOLD S, CREMERS D. Real-time dense geometry from a handheld camera[C]// GOESELE M, ROTH S, KUIJPER A, et al. Pattern Recognition. Berlin Heidelberg: Springer, 2010: 11-20.

[23] ENGEL J, STURM J, CREMERS D. Semi-Dense Visual Odometry for a Monocular Camera[C]// International Conference on Computer Vision. Sydney, NSW, 2013: 1449-1456.

[24] NEWCOMBE R A, LOVEGROVE S J, DAVISON A J. DTAM: Dense tracking and mapping in real-time[C]// International Conference on Computer Vision. Barcelona, Spain, 2011: 2320-2327.

[25] FORSTER C, PIZZOLI M, SCARAMUZZA D. SVO: Fast semi-direct monocular visual odometry[C]// 2014 IEEE International Conference onRobotics and Automation. Hong Kong, China, 2014: 15-22.

[26] ENGEL J, SCHÖPS T, CREMERS D. LSD-SLAM: Large-Scale Direct Monocular SLAM[M]// FLEET D, PAJDLA T, SCHIELE B, et al, eds. Computer Vision-ECCV 2014. Switzerland: Springer International Publishing, 2014: 834-849.

[27] NEWCOMBE R A, IZADI S, HILLIGES O, et al. KinectFusion: Real-time dense surface mapping and tracking[C]// IEEE International Symposium on Mixed and Augmented Reality. Basel, Switzerland, 2011: 127-136.

[28] GOKHOOL T, MEILLAND M, RIVES P, et al. A dense map building approach from spherical RGBD images[C]// International Conference on Computer Vision Theory and Applications. Lisbon, Portugal, 2014: 1103-1114.

[29] KERL C, STURM J, CREMERS D. Dense visual SLAM for RGB-D cameras[C]// Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan, 2013: 2100-2106.

[30] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[31] BAY H, TUYTELAARS T, VAN GOOL L. SURF: speeded up robust features[M]// LEONARDIS A, BISCHOF H, PINZ A. Computer Vision-ECCV 2006. Berlin Heidelberg: Springer, 2006.

[32] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]// International Conference on Computer Vision. Barcelona, Spain, 2011: 2564-2571.

[33] ALI A M, JAN NORDIN M. SIFT based monocular SLAM with multi-clouds features for indoor navigation[C]// 2010 IEEE Region 10 Conference TENCON. Fukuoka, 2010: 2326-2331.

[34] WU E Y, ZHAO L K, GUO Y P, et al. Monocular vision SLAM based on key feature points selection[C]// 2010 IEEE International Conference on Information and Automation (ICIA). Harbin, China, 2010: 1741-1745.

[35] CHEN C H, CHAN Y P. SIFT-based monocluar SLAM with inverse depth parameterization for robot localization[C]// IEEE Workshop on Advanced Robotics and Its Social Impacts, 2007. Hsinchu, China, 2007: 1-6.

[36] Zhu D X. Binocular Vision-SLAM Using Improved SIFT Algorithm[C]// 2010 2nd International Workshop on Intelligent Systems and Applications (ISA). Wuhan, China, 2010: 1-4.

[37] ZHANG Z Y, HUANG Y L, LI C, et al. Monocular vision simultaneous localization and mapping using SURF[C]// WCICA 2008. 7th World Congress on Intelligent Control and Automation. Chongqing, China, 2008: 1651-1656.

[38] YE Y. The research of SLAM monocular vision based on the improved surf feather[C]// International Conference on Computational Intelligence and Communication Networks. Hongkong, China, 2014: 344-348.

[39] WANG Y T, FENG Y C. Data association and map management for robot SLAM using local invariant features[C]//2013 IEEE International Conference on Mechatronics and Automation. Takamatsu, 2013.

[40] ROSTEN E, DRUMMOND T. Machine Learning for High-Speed Corner Detection[M]// LEONARDIS A, BISCHOF H, PINZ A, et al. European Conference on Computer Vision. Berlin Heidelberg: Springer, 2006: 430-443.

[41] CALONDER M, LEPETIT V, STRECHA C, et al.

BRIEF: Binary Robust Independent Elementary Features [C]// European Conference on Computer Vision. Crete, Greece, 2010: 778-792.

[42] FEN X, ZHEN W. An embedded visual SLAM algorithm based on Kinect and ORB features[C]// 2015 34th Chinese Control Conference. Hangzhou, China, 2015: 6026-6031.

[43] XIN G X, ZHANG X T, WANG X, et al. A RGBD SLAM algorithm combining ORB with PROSAC for indoor mobile robot[C]// 2015 4th International Conference on Computer Science and Network Technology (ICCSNT). Harbin, China, 2015: 71-74.

[44] LI J, PAN T S, TSENG K K, et al. Design of a monocular simultaneous localisation and mapping system with ORB feature[C]// International Conference on Multimedia and Expo (ICME), San Jose, California, USA, 2013: 1-4.

[45] EADE E, DRUMMOND T. Edge landmarks in monocular SLAM[J]. Image and vision computing, 2009, 27(5): 588-596.

[46] KLEIN G, MURRAY D. Improving the agility of keyframe-based SLAM[C]// European Conference on Computer Vision. Marseille, France, 2008: 802-815.

[47] CONCHA A, CIVERA J. Using superpixels in monocular SLAM[C]// IEEE International Conference on Robotics and Automation. New Orleans, LA, 2014: 365-372.

[48] SCHINDLER G, BROWN M, SZELISKI R. City-Scale Location Recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition. Ezhou, China, 2007: 1-7.

[49] ULRICH I, NOURBAKHSH I. Appearance-based place recognition for topological localization[C]// IEEE International Conference on Robotics and Automation. Anchorage, Alaska, 2010: 1023-1029.

[50] NEIRA J, RIBEIRO M I, TARDOS J D. Mobile robot localization and map building using monocular vision[C]// Proceedings of the 5th International Symposium on Intelligent Robotic Systems. Pisa, Italy, 1997: 275-284.

[51] WILLIAMS B, CUMMINS M, NEIRA J, et al. A comparison of loop closing techniques in monocular SLAM[J]. Robotics and autonomous systems, 2009, 57(12): 1188-1197.

[52] MUR-ARTAL R, TARDOS J D. ORB-SLAM: Tracking and mapping recognizable features[C]// IEEE International Conference on Robotics and Automation (ICRA). Berkeley, CA, USA, 2014.

[53] CUMMINS M, NEWMAN P. Accelerated appearance-only SLAM[C]// IEEE International Conference on Robotics and Automation. Pasadena, California, USA, 2008: 1828-1833.

[54] CLEMENTE L A, DAVISON A J, REID I D, et al. Mapping Large Loops with a Single Hand-Held Camera. [C]// Robotics: Science and Systems. Atlanta, GA, USA, 2007.

[55] CUMMINS M, NEWMAN P. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance[J]. International Journal of Robotics Research, 2008, 27(6): 647-665.

[56] NISTER D, STEWENIUS H. Scalable Recognition with a Vocabulary Tree[C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). New York, NY, USA, 2006: 2161-2168.

[57] ANGELI A, FILLIAT D, DONCIEUX S, et al. A fast and incremental method for loop-closure detection using bags of visual words[J]. IEEE transactions on robotics, 2008, 24(5): 1027-1037.

[58] CUMMINS M, NEWMAN P. Highly scalable appearance-only SLAM-FAB-MAP 2.0[C]// Robotics: Science and Systems V, University of Washington. Seattle, USA, 2009.

[59] 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.

[60] EADE E D, DRUMMOND T W. Unified loop closing and recovery for real time monocular SLAM [C]// British Machine Vision Conference. Leeds, UK, 2008: 1-10.

[61] GÁLVEZ-LÓPEZ D, TARDOS J D. Real-time loop detection with bags of binary words[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, California, USA, 2011: 51-58.

[62] CUMMINS M, NEWMAN P. Appearance-only SLAM at large scale with FAB-MAP 2.0[J]. International journal of robotics research, 2011, 30(9): 1100-1123.

[63] LOWE D G. Distinctive Image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.

[64] TRIGGS B, MCLAUCHLAN P F, HARTLEY R I, et al. Bundle Adjustment-A Modern Synthesis [M]// TRIGGS B, ZISSERMAN A, SZELISKI R. Vision Algorithms: Theory and Practice. Berlin Heidelberg: Springer , 2000: 298-372.

[65] HARTLEY R, ZISSERMAN A. Multiple view geometry in computer vision[M]. 2nd ed. Cambridge U K: Cambridge University Press, 2003.

[66] KÜMMERLE R, GRISETTI G, STRASDAT H. G2o: A general framework for graph optimization[C]// IEEE International Conference on Robotics and Automation. Shanghai, China, 2011: 3607-3613.

[67] STRASDAT H, MONTIEL J M M, DAVISON A J. Scale drift-aware large scale monocular SLAM[C]// Proceedings of Robotics: Science and Systems. Zaragoza, Spain, 2010.

[68] STRASDAT H, DAVISON A J, MONTIEL J M M, et al. Double window optimisation for constant time visual SLAM [C]// International Conference on Computer Vision. Barcelona, Spain, 2011: 2352-2359.

[69] MOURIKIS A I, ROUMELIOTIS S I. A multi-state constraint Kalman filter for vision-aided inertial navigation [C]// Proceedings of the 2007 IEEE International Conference on Robotics and Automation (ICRA). Roma, Italy, 2007.

[70] MOURIKIS A I, ROUMELIOTIS S I. A dual-layer estimator architecture for long-term localization [C]// Proceedings of the 2008 Workshop on Visual Localization for Mobile Platforms at CVPR. Anchorage, Alaska, 2008.

[71] LEUTENEGGER S, FURGALE P, RABAUD V, et al. Keyframe-based visual-inertial slam using nonlinear optimization [C]// Proceedings of 2013 Robotics: Science and Systems (RSS). Berlin, Germany, 2013.

[72] Google. Project tango. URL https://www.google.com/atap/projecttango/.

[73] ŽBONTAR J, LE CUN Y. Stereo matching by training a convolutional neural network to compare image patches [J]. The journal of machine learning research, 2015, 17 (1): 2287-2318.

[74] SÜNDERHAUF N, SHIRAZI S, DAYOUB F. On the performance of ConvNet features for place recognition [C]// 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg, Germany, 2015: 4297-4304.

[75] COSTANTE G, MANCINI M, VALIGI P, et al. Exploring representation learning with CNNs for frame-to-frame ego-motion estimation [J]. IEEE robotics and automation letters, 2016, 1(1): 18-25.

[76] KENDALL A, CIPOLLA R. Modelling uncertainty in deep learning for camera relocalization [C]// 2016 IEEE International Conference on Robotics and Automation. Stockholm, Sweden, 2016: 4762-4769.

作者简介：

权美香，女，1992 年生，博士，主要研究方向为单目视觉 SLAM，VIN，移动机器人视觉导航。

朴松昊，男，1972 年生，教授，博士生导师，中国人工智能学会常务理事，机器人文化艺术专业委员会主任，主要研究方向为机器人环境感知与导航、机器人运动规划、多智能体机器人协作。主持或参加了国家自然科学基金、国家“863”计划重点及面上项目、机器人技术与系统国家重点实验室基金、教育部“985”项目、三星国际合作项目等多个项目。发表学术论文 60 余篇，其中被 SCI、EI、ISTP 检索 60 多篇，出版专著一部。

李国，男，1989 年生，博士，主要研究方向为 SLAM、机器学习。