低成本激光和视觉相结合的同步定位与建图研究

尹 磊^{1,2} 彭建盛^{1,3} 江国来^{2,4} 欧勇盛²

¹(广西科技大学电气与信息工程学院 柳州 545006)
 ²(中国科学院深圳先进技术研究院 深圳 518055)
 ³(河池学院物理与机电工程学院 宜州 546300)
 ⁴(中国科学院大学深圳先进技术学院 深圳 518055)

摘 要激光雷达和视觉传感是目前两种主要的服务机器人定位与导航技术,但现有的低成本激光雷 达定位精度较低且无法实现大范围闭环检测,而单独采用视觉手段构建的特征地图又不适用于导航应 用。因此,该文以配备低成本激光雷达与视觉传感器的室内机器人为研究对象,提出了一种激光和视 觉相结合的定位与导航建图方法:通过融合激光点云数据与图像特征点数据,采用基于稀疏姿态调整 的优化方法,对机器人位姿进行优化。同时,采用基于视觉特征的词袋模型进行闭环检测,并进一步 优化基于激光点云的栅格地图。真实场景下的实验结果表明,相比于单一的激光或视觉定位建图方 法,基于多传感器数据融合的方法定位精度更高,并有效地解决了闭环检测问题。

关键词 机器人; 非线性优化; 同步定位与地图构建; 稀疏姿态调整 中图分类号 TP 249 文献标志码 A doi: 10.12146/j.issn.2095-3135.20180717001

Research on Simultaneous Localization and Mapping Fusion with

Low-Cost Laser Sensors and Vision

YIN Lei^{1,2} PENG Jiansheng^{1,3} JIANG Guolai^{2,4} OU Yongsheng²

¹(School of Electrical and Information Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China)

²(Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China)

³(School of Physics and Electrical Engineering, Hechi University, Yizhou 546300, China)

⁴(Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Shenzhen 518055, China)

Abstract Laser lidar and vision sensors are the two mainstream three-dimensional sensing techniques in the applications of robot location and navigation. However, existing low-cost laser lidar usually has low location accuracy and cannot achieve loop closure detection in large areas. In this paper, an indoor robot equipped with low-cost laser lidar and camera was used for experiment. And a novel localization and mapping method was introduced by combing both lidar and image information. An optimization method based on sparse pose adjustment was used to optimize the robot pose by fusing laser points cloud and image feature points as constraints. At the same time, the bag of words model based on visual features was used for loop closure

收稿日期: 2018-07-17 修回日期: 2018-09-13

基金项目:国家自然科学基金项目(U1613210);深圳市基础研究计划项目(JCYJ20170413165528221)

作者简介: 尹磊,硕士研究生,研究方向为机器人导航; 彭建盛,副教授,研究方向为智能控制、智能机器人; 江国来,博士研究生,研究方向 为机器视觉、机器人导航; 欧勇盛(通讯作者),研究方向为智能机器人、智能控制和机器学习, E-mail: ys.ou@siat.ac.cn。

detection. The grid map was optimized by loop closure constraints. Real experimental results show that, the proposed method has better localization accuracy than either laser lidar or vision sensors, and loop closure detection also can be realized.

Keywords robot; nonlinear optimization; simultaneous localization and mapping; sparse pose adjustment

1 引 言

同步定位与地图构建(Simultaneous Localization and Mapping, SLAM)是机器人领域的技术 热点与难点,其解决的定位和建图问题被认为是 机器人实现自主导航的关键。该技术主要原理是 通过机器人配备的多种传感器来感知周围环境, 并计算出自己在当前环境下的位置。SLAM 被提 出后,先后经历了多个研究阶段。由于激光雷达 精度高、范围广,早期的 SLAM 研究往往以激光 雷达为主要的传感器。另外,早期的 SLAM 采 用扩展卡尔曼滤波方法估计机器人的位姿^[1],但 效果不好,对于某些强非线性系统,该方法会带 来更多的截断误差,从而导致无法精准地实现定 位与建图。而后基于粒子滤波的 SLAM 逐渐成 为主流, 粒子采样方法可有效避免非线性问题, 但同样引出了随着粒子数增多,计算量增大的问 题。一直到 2007 年, Grisetti 等^[2]提出了基于改 进粒子滤波的 SLAM 方法(Gmapping),该方法 通过改进的建议分布及自适应重采样技术有效地 提高了定位精度并降低了计算量,是激光 SLAM 的一个里程碑。Konolige 等^[3]在 2010 年提出了 Karto SLAM, 该方法采用图优化代替粒子滤波, 并采用稀疏点调整来解决非线性优化中矩阵直接 求解难的问题。Kohlbrecher 等^[4]在 2011 年提出 了 Hector SLAM, 该方法不需要里程计信息, 通 过利用高斯牛顿方法来解决扫描匹配问题,但对 传感器要求较高,需要高精度的激光雷达才能运 行。由谷歌于 2016 年提出的 Cartographer^[5]可以 说是最新的激光 SLAM 方案。该方法将获得的每 一帧激光数据,利用扫描匹配在最佳估计位置处 插入子图(Submap)中,且扫描匹配只与当前子 图有关。在生成一个子图后,会进行一次局部的 回环(Loop Close),而在所有子图完成后,利用 分支定位和预先计算的网格进行全局的回环。相 比于 Gmapping 和 Hector,该方案具有累积误差 较低,且不需要高成本设备的优点。

视觉 SLAM 由于传感器成本低、图像信息 丰富逐渐成为了 SLAM 的研究热点。但与激光 SLAM 相比,视觉 SLAM 更加复杂。Davison 等^[6] 在 2007 年首次提出的 MonoSLAM 被认为是众 多视觉 SLAM 的"发源地"。该方法以扩展卡 尔曼滤波为后端,追踪前端稀疏的特征点;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,最终获得后验概率分布的均值和方 差。Sim 等^[7]利用粒子滤波实现了视觉 SLAM, 该方法避开了线性化的问题并且精度高,但需要 使用大量的粒子,从而导致计算复杂度的提高。 随后,为了减少计算复杂度,提取关键帧变得极 为重要,其中最具代表性的是 2007 年 Klein 和 Murray 提出的 PTAM^[8]。该方法提出了简单有 效地提取关键帧的技术,并且该方法最为关键的 两点是: (1)实现了跟踪与建图的并行化,虽然 跟踪部分需要实时响应图像数据,但后端优化 却不需要实时计算。即后端优化可在后台慢慢运 行,在需要的时候将两个线程同步即可。这也是 首次提出了区分前后端的概念,引领了之后众多 SLAM 方法的架构设计。(2)第一次使用非线性 优化,而不是传统的滤波器。从 PTAM 提出后, 视觉 SLAM 研究逐渐转向了以非线性优化为主导 的后端。2016 年 Mur-Artal 和 Tardos^[9]提出了现 代 SLAM 中非常著名的第二代基于具有带方向性 的加速分段测试特征(FAST)关键点以及带旋转 不变性的二进制鲁棒独立基本特征(BRIEF)描述 子的特征点(Oriented FAST and Rotated BRIEF, ORB)^[10]的视觉 SLAM 系统 ORB-SLAM2, 是 现代 SLAM 中做得十分完善并且易用的系统之 一。该方法不仅支持单目、双目、深度摄像头三 种模式,并且将定位、地图创建、闭环分为三个 线程,且都使用 ORB 特征。ORB-SLAM2 的闭 环检测是一大亮点,该方法采用了词袋模型,有 效地防止了累积误差,并且能在丢失之后迅速找 回。但 ORB-SLAM2 的缺点也十分明显,图像特 征提取与匹配以及后端的优化都需要计算资源, 在嵌入式平台上实现实时运行存在一定的困难, 并且构建的稀疏特征点地图虽然可以很好地满足 定位,但无法提供导航、避障等功能。相比于提 取图像中的特征点,根据图像像素灰度信息来计 算相机运动的直接法从另一个方向实现了定位和 建图。Stühmer 等^[11]提出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,并构建出稠密的三维地图。Engel 等^[12] 对当前图像构建半稠密深度地图,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。Engel 等^[13]提出了 LSD-SLAM 算法,其核心是将直接 法应用到半稠密的单目 SLAM 中,这在之前的直 接法中很少见到。以往基于特征点只能构建稀疏 地图,而稠密地图又需要 RGB-D 这种可以提供 深度信息的相机。Forster 等^[14]于 2014 年提出了 半直接法单目视觉里程计(Semi-direct Monocular Visual Odometry, SVO), 一种被称为"稀疏直

接法"的方法,该方法将特征点与直接法混合使 用, 跟踪了一些关键点(如角点等), 然后按照直 接法根据关键点周围的信息估计相机运动及位 置。相比于其他方案, SVO 既不用消耗大量资 源去计算描述子,也不必处理过多的像素信息, 因此该方法可广泛用于无人机、手持增强现实 (Augmented Reality, AR)等设备上。Newcombe 等^[15]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。Gokhool 等^[16] 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。Kerl 等^[17]提出 了较好的直接 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 闭环检测方法,从而大大降低了路径的误差。

在多传感器融合方面,目前有视觉传感器、 激光雷达、惯性测量单元(Inertial Measurement Unit, IMU)及超声波传感器等多种传感器。目 前,主要融合方向为激光雷达结合视觉传感器 及 IMU 结合视觉传感器。Chen 等^[18]通过视觉传 感器结合 IMU 来进行机器人的精准位姿估计, 并在机器人上垂直安装二维激光雷达采集点云数 据实现三维建图。Houben 等^[19]采用三维激光数 据用于微型无人机定位,针对结构相似的环境激 光定位会出现模糊的问题,提出在环境中加入视 觉标记,提高定位精度,同时可用于微型无人机 的重定位。王消为等^[20]提出了一种双目视觉信 息和激光雷达数据融合的 SLAM 方法,该方法 基于改进的粒子滤波算法实现了在计算建议分布 时观测数据,其中同时包含视觉信息和激光雷达 数据。相对于里程计运动模型作为建议分布,该 方法有效地提高了定位和建图的精度。张杰和周

军^[21]提出了一种激光雷达与视觉结合的 SLAM 方法,其中激光地图用于导航,视觉地图用于 复原目标场景;并提出一种改进的迭代最近点 法 (Iterative Closest Point, ICP)用于实现更快的 点云拼接,同时采用图优化的方法降低了累积误 差,保证地图精度。Shi 等^[22]在小型无人机上, 利用视觉里程计为二维激光的 ICP 提供初值,在 实时性和精确度方面都达到了较好的效果。Qin 等^[23]提出了视觉与 IMU 的紧耦合方案,将视觉 构造的残差项和 IMU 构造的残差项放在一起构 成一个联合优化问题。Li 等^[24]和 Lynen 等^[25]通 过扩展卡尔曼滤波器对视觉和 IMU 进行融合以 实时获取状态估计。

相对而言,激光 SLAM 的效果仍然要优于视 觉 SLAM,但激光 SLAM 由于其本身激光数据的 特性导致其无法有效地进行大范围的闭环检测。 对于低价格的激光雷达,由于激光点不够密集, 构建的地图常常会出现回到原来的位置地图,由 此造成定位不准确而出现偏差的情况,这是由累 积误差所导致的。同时在激光 SLAM 中,闭环检 测一直是一大难点:由于获取的激光数据是二维 点云数据,无明显特征且相互之间十分相似,所 以基于激光数据的闭环检测往往效果不好。由于 图像中包含了丰富的信息,故视觉 SLAM 在闭 环检测方面存在着天然优势。ORB-SLAM2 中提 出的词袋模型,采用了 ORB 特征配合词袋的方 法,具有很高的准确率和速度,是当前应用最广 的闭环检测手段。

针对低成本激光 SLAM 噪声大、精度低、难 以闭环的问题,本文提出激光结合视觉进行联合 优化的方法来提高定位建图精度,并通过视觉词 袋模型有效地解决激光闭环检测难的问题。

2 基于图优化的同步定位与地图构建框架

激光 SLAM 主要通过相邻帧激光的匹配来

计算相邻帧位姿变换,但由于激光数据并非完全 无噪声的数据, 尤其对于低成本激光雷达, 打出 去的激光点较为稀疏,从而导致计算得到的相邻 帧位姿变换存在误差,因此往往需要加入滤波或 优化的方法来使定位更加精准。而视觉 SLAM 是 通过特征点提取与匹配来反算位姿,但若其中一 帧出现问题就会导致误差累积并不断地增大。图 优化是 SLAM 后端优化的热门方法,通过构造节 点和约束边清晰地展示了一系列位姿和观测量的 关系,再通过非线性优化的手段求出最优变量从 而得到精准的位姿估计。该方法最早应用于视觉 SLAM 中,起初由于图像特征点数量过多,导致 矩阵维数过大, 求解困难, 所以该方法一直未能 成为主流方法。直到 2011 年海塞矩阵的稀疏性 被发现,极大地提高了运算速度,使得非线性优 化和图优化的思想成功应用于视觉 SLAM 和激光 SLAM 中。而根据位姿估计和优化的实时性要求 不同, SLAM 也被分为前端和后端两部分, 当前 主流的 SLAM 框架如图 1 所示。



Fig. 1 Simultaneous localization and mapping framework

前端主要通过传感器数据估计机器人的位 姿,但无论是图像还是激光,观测到的数据都含 有不同程度的噪声。相对而言,高精度激光雷达 噪声会小很多但成本过高,而通过低成本的激光 雷达及相机采集图像来进行位姿计算都会导致定 位与实际真值有累积误差,并且累积误差会随着 时间的增加而越来越大。后端优化的主要作用就 是通过滤波或优化的方式提高定位及构建的地图 精度,消除累积误差。本文采用图优化作为后 端,通过非线性优化寻找下降梯度的方式来迭代 实现误差最小化。

简单来说,图优化是以图的形式来描绘优化

问题。在 SLAM 中,图的节点表示位姿,边表示 位姿之间以及位姿与观测量的约束关系。机器人 在导航建图过程中,观测量为激光数据及通过摄 像头不停地捕捉到的外部环境信息,生成大量机 器人观测到的 ORB 特征点对应的三维空间点。 将所有数据放入图的框架中,如图 2 所示。



Fig. 2 Graph optimization framework

其中,X表示关键帧位姿;O表示观测量,包含 了特征点所对应的三维空间点坐标及二维激光数 据。视觉误差由重投影误差(图 3)来表示,重投 影误差的计算需要给定相邻帧所对应的两个相 机位姿、匹配好的特征点在两幅图像中的二维 坐标及所对应的三维空间点的三维坐标。纯视 觉 SLAM 往往通过特征点提取和匹配,再使用 EPnP 等方法来求取相邻帧位姿变换估计。但相 比于相邻帧的激光匹配,该方法误差较大,因 此本文使用激光扫描匹配得到的位姿估计作为 后端优化的初始值。而对于相邻帧图像中通过 特征点匹配得到的特征点对 p1、p2,本文采用的 深度摄像头可直接获取得到前一帧图像特征点 p_1 所对应的三维空间点 P 的坐标,将点 P 重新 投影到后一帧图像上形成图像中的特征点 \hat{p}_{γ} 。 由于位姿估计的误差及深度摄像头噪声的存 在, $\hat{p}_2 = p_2$ 并非完全重合, 两点之间的距离就



Fig. 3 Re-projection error

是误差。

重投影坐标计算公式及过程如下:

(1)通过世界坐标系到相机坐标系的变换关系(**R**, *t*),计算出世界坐标系下点 *P* 所对应的相机
坐标系下点 *P'*(*X'*, *Y'*, *Z'*)的三维坐标。

$$P' = \mathbf{R}P + \mathbf{t} = [X', Y', Z']^{\mathrm{T}}$$
(1)

其中, **R** 为旋转矩阵; **t** 为平移矩阵; **T** 表示转置。

(2) 将 P' 投至归一化平面,并将 Z' 归为 1 得到归一化坐标 P_c。

 $P_{c} = [u_{c}, v_{c}, 1]^{\mathrm{T}} = [X'/Z', Y'/Z', 1]^{\mathrm{T}}$ (2) 其中, (u_{c}, v_{c}) 为归一化平面上的二维坐标。

(3)根据相机内参模型,像素坐标系为归一 化平面横轴缩放了 f_x 倍,竖轴缩放了 f_y 倍,并且 原点平移了 c_x、c_y 个像素。因此,像素坐标 (u_s, v_s)计算公式如下:

$$u_s = f_x u_c + c_x$$

$$v_s = f_y v_c + c_y$$
(3)

则该点误差函数为 $e=p_2-p'_2$ 。其中, p_2 为直接通 过图像得到的像素坐标; p'_2 为根据公式(1)~(3) 计算得到的重投影坐标。

将上述误差函数扩展到相邻两帧之间,则最 小化代价函数为 $\frac{1}{2}\sum_{i=1}^{n} \|p_i - p'_i\|_2^2$ 。通过非线性优化 算法可以得到误差最小化时相应的位姿变换(\mathbf{R} , t) 和三维空间点坐标 P_i 。

相对于视觉误差,激光误差的获取更为简 单。激光 SLAM 往往需要通过扫描匹配来实现 相邻帧的位姿变换估计,而该估计值(**R**, *t*)无法 保证前一帧所有激光数据经过该位姿变换与后 一帧激光数据完全重合。因此,激光的误差定 义如下:

$$e = f(x) = P_2 - P_2' = P_2 - (\mathbf{R}P_1 + \mathbf{t})$$
(4)

之后,通过非线性优化来使误差函数最小 化,求得的位姿再返回到前端作为下一帧的参考 帧位姿。

3 后端优化与闭环检测

视觉 SLAM 构建的地图由特征点构成,特 征地图最大的问题在于无法用于导航,仅适合用 于定位。低成本激光雷达构建的栅格地图更适合 用于导航,但存在着激光稀疏、噪声大的问题。 因此,本文提出视觉激光融合的方法,在提高定 位精度的同时确保地图更加精准,还解决了激光 SLAM 构建的栅格地图难以回环的问题。而二 维栅格地图的构建主要依赖于一系列位姿和激 光测量数据,根据栅格被占用的概率来增量式 的建图。

由于同一时刻观测到的视觉信息和激光数据 并非完全相互独立,因此本文提出视觉激光联合 优化来充分利用数据之间的约束。加入视觉信息 的激光视觉联合优化的 SLAM 整体框架如图 4 所示。



图4 联合优化整体框架



3.1 误差函数

传统视觉相邻帧误差函数已在第2节中给出 具体形式,重投影坐标与三维空间点、上一帧图 像特征点以及位姿变换的关系如下:

$$Zp_1 = KP$$

$$Zp'_2 = K(RP + t)$$
(5)

其中, *K* 为相机内参; *Z* 为三维点的深度值。则 误差函数如下:

$$e = f(x) = p_2 - \frac{1}{Z} K(RP + t)$$
(6)

位姿变换(**R**, *t*)可写成对应的李代数形式 ξ[^], 李代数变换公式为:

$$\boldsymbol{R}\boldsymbol{P} + \boldsymbol{t} = \exp(\boldsymbol{\xi}^{\wedge})\boldsymbol{P} \tag{7}$$

将视觉相邻帧误差函数公式(6)以及激光误 差函数公式(4)改写成李代数形式并合并,则相 邻帧的整体误差函数为:

$$e = f(\xi) = \sum_{i=1}^{m} \left[p_i - \frac{1}{Z} K \exp(\xi^{\wedge}) P_i \right] + \sum_{i=1}^{n} \left[P_j - \exp(\xi^{\wedge}) P_{j-1} \right]$$
(8)

其中,相邻帧共有 m 个匹配特征点, n 个激光数 据点。实际计算中可将 m、n 限制在一定数量内 从而减少计算复杂度。

单独的视觉 SLAM 有其自有的计算相邻帧位 姿变换算法,但由于图像数据的特点,计算得到 的(*R*, *t*)往往没有通过激光扫描匹配得到的位姿变 换精准。因此,采用激光扫描匹配得到的位姿变 换来估计误差函数中的位姿变换初始值。

3.2 稀疏姿态调整

根据观测模型可以很容易判断出误差函数不 是线性函数,所以本文采用非线性优化来求解 误差函数最小值所对应的变量。由于三维特征 点和激光数据量过多,且随着时间推移整体代 价函数所包含的多项式会变得越来越多,故本 文只优化误差函数中的位姿变量。同时,姿态图 的优化也会随着顶点约束的增加而变慢,所以本 文采用稀疏姿态调整法,利用矩阵的稀疏性来提 高优化速率。由于机器人位姿是由变换矩阵不 断计算得到且一一对应,即求得每两相邻帧之 $(\mathbf{0})$

间的位姿变换,亦即可获得机器人的当前位姿。 因此,将机器人位姿作为唯一变量并对其进行优 化,视觉观测量和激光点云数据作为位姿之间的 约束。设机器人位姿与位姿变换之间的关系为:

$$x=h(\xi)$$
 (9)
则误差函数可改写为关于位姿 x 的函数 $e=f(x)$ 。

$$x = [x_1, x_2, x_3, \cdots, x_k]^{\mathrm{T}}$$
 (10)

其中,k为待优化位姿个数。相应地, Δx 是对整 体自变量 x 的增量。因此, 当加入增量后, 目标 函数为:

其中, x 为位姿的集合, 即待优化变量为:

$$\frac{1}{2} \|f(x + \Delta x)\|^2 \approx \frac{1}{2} \sum_{t=1}^{K} \|e_t + J_t \Delta x_t\|^2$$
(11)

其中, J 表示雅克比矩阵, 为代价函数对自变量 的偏导数; K 为待优化位姿个数,相邻帧优化时 为2,全局优化时为当前帧到回环帧之间的位姿 个数。

位姿优化可看作是最小二乘问题,而解决最 小二乘问题的常用方法有梯度下降法、高斯牛顿 法及 Levenberg-Marquadt (L-M) 法。其中, L-M 法是对梯度下降法和高斯牛顿法的综合运用,效 果最佳,因此本文采用 L-M 法来求解上述最小 二乘问题。

通过加入拉格朗日乘子对误差函数进行改进:

$$\min_{\Delta x_k} \frac{1}{2} \| f(x) + \boldsymbol{J}(x) \Delta x \|^2 + \frac{\lambda}{2} \| \boldsymbol{D} \Delta x \|^2$$
(12)

其中, λ 为拉格朗日乘子;**D**为单位矩阵**I**。将 上式展开可得:

$$\frac{1}{2} \|f(x) + J(x)\Delta x\|^{2} + \frac{\lambda}{2} \|D\Delta x\|^{2}$$
$$= \frac{1}{2} [f(x) + J(x)\Delta x]^{T} [f(x) + J(x)\Delta x] + \frac{\lambda}{2} \Delta x^{T} D^{T} D\Delta x \qquad (13)$$

$$=\frac{1}{2}\left[\left\|f(x)\right\|^{2}+2f(x)^{\mathsf{T}}\boldsymbol{J}(x)\Delta x+\Delta x^{\mathsf{T}}\boldsymbol{J}(x)^{\mathsf{T}}\boldsymbol{J}(x)\Delta x+\lambda \Delta x^{\mathsf{T}}\boldsymbol{J}(x)^{\mathsf{T}}\boldsymbol{J}(x)\Delta x+\lambda \Delta x^{\mathsf{T}}\boldsymbol{J}(x)^{\mathsf{T}}\boldsymbol{J}(x)\Delta x+\lambda \Delta x^{\mathsf{T}}\boldsymbol{J}(x)^{\mathsf{T}}\boldsymbol{J}(x)^{\mathsf{T}}\boldsymbol{J}(x)^{\mathsf{T}}\boldsymbol{J}(x)\right]$$

对 Δx 求导并令其为 0, 可得:

$$\left[\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \boldsymbol{D}^{\mathrm{T}}\boldsymbol{D}\right] \Delta \boldsymbol{x} = -\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}} f(\boldsymbol{x}) \qquad (14)$$

定义公式(14) 左侧系数为 $H + \lambda D^{T} D$, 右侧 为g,则可得线性方程:

$$(\boldsymbol{H} + \lambda \boldsymbol{D}^{\mathrm{T}} \boldsymbol{D}) \Delta x = g \tag{15}$$

$$(\mathbf{H} + \lambda \mathbf{I}) \Delta x = g \tag{16}$$

根据公式(16)可以发现,当 λ 较小时,H矩 阵占主导地位,说明在该范围内二次近似效果更 好, 优化方法更近似于高斯牛顿法; 当 λ 较大 时,二次近似效果较差,优化方法更近似于梯度 下降法。 λ 的取值一般与自定义的参数 ρ 相关, 参数 ρ 可由公式(17)得到。

$$\rho = \frac{f(x + \Delta x) - f(x)}{J\Delta x} \tag{17}$$

其中,参数 ρ 表示近似模型与实际模型相似程 度, ρ 越接近1时, λ 越小,高斯牛顿法的近似 效果越好; ρ 越小时近似效果越差,优化方法更 近似于梯度下降法。

一般情况下, H 矩阵维数很大, 矩阵求逆的 复杂度为 $O(n^3)$ 。但由于H矩阵内部包含着各个 顶点之间的约束,而只有相邻顶点之间才具有直 接约束,从而导致 H 矩阵大部分元素为 0,具有 稀疏性。因此,利用*H*矩阵特有的稀疏性可以大 大提高运算速度。求解公式(16)可得 Δx ,沿着 梯度下降方向循环迭代 Δx ,最终获得目标函数 最小时所对应的自变量 x, 即机器人位姿。

3.3 闭环检测

闭环检测是 SLAM 中的一个核心问题,通 过对曾经走过的地点进行识别,能有效地减小累 积误差,提高定位精度。基于激光雷达的 SLAM 算法往往因为数据单一而无法有效地进行闭环检 测,而视觉图像丰富的纹理特征则刚好可以弥补 激光雷达的这一缺陷。本文采用视觉 SLAM 中 最常用的词袋模型(Bag-of-Words, BoW),通过 视觉特征来构建关键帧所对应的字典,在检测到 回环后通过回环帧与当前帧的匹配来计算当前位 姿,并将这一约束加入到后端中来进行回环帧与 当前帧之间的全局优化,提高定位精度的同时, 也防止了激光所构建的栅格地图常出现的无法闭 合的情况。

由于构建室内地图所采集到的图像数量过 多,且相邻图像之间具有很高的重复性,因此首 先需要进行关键帧的提取。本文关键帧选取机制 如下:

(1)距离上一次全局重定位已经过去了15帧;

(2)距离上一次插入关键帧已经过去了15帧;

(3)关键帧必须已经跟踪到了至少 50 个三维特 征点。

其中,(1)、(2)是其独特性的基础,因为短时间 视野内的特征不会发生明显变化;(3)保证了其 鲁棒性,过少的地图点会导致计算误差的不均 匀。相对而言,激光 SLAM 关键帧的选取要更为 简单且稳定,但采用激光作为关键帧选取参量容 易出现关键帧包含的图片特征点不够、图像不连 续等问题。因此,本文采用基于视觉的关键帧选 取机制,既保证相邻关键帧不会过于接近,又保 证有足够的信息进行关键帧之间的匹配。

常见的图像特征有尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)^[26]、加速稳健 特征(Speeded Up Robust Features, SURF)^[27]和 ORB等。其中,SIFT特征点具有旋转不变性和 尺度不变性,并且稳定性高,不易受到光照及噪 声的影响,是最理想的选择。然而,SIFT特征 提取速度慢,无法保证实时,故不适用于 SLAM 中。SURF效果与 SIFT相近,同样存在计算时间 过长的问题。因此,本文选择 ORB 来构建词袋 模型,ORB 在保持了旋转不变性和尺度不变性 的同时,速度要明显优于 SIFT和 SURF。

BoW 通过比对图像所包含的特征将当前帧 与每一个关键帧作相似度计算。首先当相似度表 明当前帧与某一关键帧足够相似时,则认为机器 人回到了该关键帧附近的位置,产生回环。然 后,在回环后对当前帧和该关键帧进行激光点云 的 ICP 匹配,从而以该关键帧的位姿为基准, 计算出当前机器人位姿,并将当前帧位姿、当前 帧与回环帧之间的图像特征点、激光点云作为约 束加入到图优化框架中,从而对检测到回环的关 键帧与当前关键帧之间的一系列位姿进行全局优 化。最后,根据优化后的位姿及每一帧位姿所携 带的激光数据重新构建地图,消除地图中无法闭 合的部分。

4 实 验

本文实验分为两部分:第一部分在小范围场 景下进行定点定位精度的对比实验,分别对传统 基于图优化的激光 SLAM 方法(即 Karto)和本文 提出的激光视觉相结合的方法进行定位数据采 集;第二部分为闭环实验,验证本文提出的方法 是否可以有效地解决激光 SLAM 可能出现的地图 不闭合情况。

4.1 实验平台和环境

本文实验在 Turtlebot2 上进行,配备笔记本电脑以及激光雷达和深度摄像头,电脑配置为 Intel Core i5 处理器、8 G 内存,运行 ubuntu14.04+ ROS Indigo 系统。二维激光雷达采用单线激光雷达 RPLIDAR A2,人为设置采样频率为 5~15 Hz, 测量半径为 8 m。深度相机采用奥比中光公司出品的 Astra 深度相机。其中,该相机深度测距的 有效范围为 0.6~8 m,精度达 3 mm;深度相机 视角可以达到水平 58°和垂直 45.5°。

本文实验在中国科学院深圳先进技术研究院 B 栋办公楼进行,以机器人起始位置建立世界坐 标系,在 B 栋 5 楼 B、C 区电梯之间选择 5 个标 志性位置,分别用 Karto 和本文提出的方法进行 位姿测量,实验过程如图 5 所示。从 0 点出发, 以 0 点为世界坐标系原点,0 到 1 方向为 *x* 轴方



(d) 位置 3

(e) 位置 4 图 5 实验过程截图 (f) 位置 5

19

Fig. 5 Screenshot of the experimental process

向,2到3方向为一y轴方向,依次沿1、2、3、 4、5运动,各点实际坐标如表1所示。

机器人实际位姿由时间和速度来控制。为保 证视觉特征提取的稳定性,机器人线速度保持匀 速0.2 m/s,角速度 30 (°)/s,即机器人只有0.2 m/s 匀速前进、30 (°)/s 匀速右转及停止三个状态。 其中,前进15 s 为前进3 m;右转3 s 为右转 90°。最终通过发布线速度和角速度指令并计时 来控制机器人准确到达5个标志点。 第二部分实验场景为中国科学院深圳先进技 术研究院 B 栋 4 楼 B 区工位及 C 区 3 楼。其中, 4 楼工位为小范围闭环, C 区 3 楼为大范围回环。

4.2 实验结果分析

定位结果及误差对比如表 1 所示。表 1 中数 据说明仅依靠激光进行 SLAM 时初始误差较小, 但随着距离变远,测量值与实际位姿之间误差逐 渐增大。而实验过程中使用的激光成本较低,光 束本身不够密集,导致无法保证有足够的可用数

表1 位姿估计对比

 Table 1
 Pose estimation contrast

编 号	实际	实际	激光 SLAM					激光+视觉			
	位置	姿态	估计位置	估计姿	位置误	姿态误	作	古计位置	估计姿	位置误	姿态误
	(m)	(°)	(m)	态(°)	差(cm)	差(°)		(m)	态(°)	差(cm)	差(°)
1	(3,0)	0	(2.995,-0.002)	0.032	0.54	0.032	(2.99	94,-0.003)	0.034	0.67	0.034
2	(6,0)	270	(5.987, -0.003)	271.842	1.33	1.842	(5.99	90,-0.003)	270.232	1.04	0.232
3	(6,-8)	270	(6.025, -8.032)	272.175	4.06	2.175	(6.0)	19,-8.021)	270.421	2.83	0.421
4	(6,-16)	180	(6.038,-15.954)	182.621	5.96	2.621	(6.02	28,-15.973)	181.074	3.88	1.074
5	(3,-16)	180	(2.969,-15.946)	183.024	6.23	3.024	(2.98	31,-15.965)	181.871	3.98	1.871



(a) Karto(工位)



(b) 激光+视觉 SLAM(工位)



(c) Karto (C 🗵)



(d) 激光+视觉 SLAM(C区)



```
图 6 不同方法所构建地图对比
```



据过滤掉测量过程中测量值本身的误差,从而产 生累积误差。本文提出的激光视觉联合优化位姿 的方法虽然同样存在累积误差,但加入了视觉信 息约束可有效地缩小累积误差,可获得更高的定 位精度。

图 6(a)~(d)分别是 Karto^[3]及本文提出的方

法所构建的栅格地图;图 6(e)是 ORB-SLAM2 所构建的 C 区 3 楼稀疏特征点地图。其中,机器 人位姿由一系列红色箭头表示。图 6(a)、(b)为 工位环境的小范围闭环,从中可以看到,图 6(a) 白线圈处有不重合的现象,但由于场景较小,效 果并不明显。图 6(c)、6(d)为 C 区大范围场景

515 5

所构建的地图,由于场景较大,随着时间推移产 生的累积误差会不断增大。图 6(c)中可以明显看 到白线圈出的部分地图并未重合,这是因为累积 误差影响了机器人的定位效果,而由于定位偏差 的加大,根据机器人位姿及激光数据所构建的地 图会出现不闭合的现象。相对而言,图 6(d)中机 器人移动一圈回到起始位置时检测到了闭环,根 据起始位置计算出了当前帧位姿,并将当前帧位 姿作为约束,进行全局优化,再通过优化后的所 有位姿及所携带的激光数据更新地图信息,消除 无法闭合的情况。实验结果表明,本文提出的方 法可以有效地检测回环,并优化两点之间的所有 位姿,位姿优化之后再根据每一帧位姿所携带的 激光数据重新构建地图,成功消除了无法闭合的 地图部分。但由于加入了视觉信息,建图过程中 对机器人速度限制变大,同时对动态障碍物的抗 干扰能力变小,只能在无人时进行建图,且需要 避免摄像头面向大面积白墙等无特征物体。

5 总结与展望

本文研究了 SLAM 中多传感器融合的问题, 针对低成本激光雷达噪声大、视觉构建的稀疏特 征地图不适用于导航的问题,提出了激光视觉相 结合定位建图的方案。通过采用基于稀疏姿态调 整的方法,将激光数据和图像信息进行联合优 化,有效地提高了定位和建图的精度,并通过视 觉词袋模型实现闭环检测,解决了激光闭环检测 难实现的问题。但由于加入了视觉信息,系统鲁 棒性变差,且室内环境光照变化往往较小,同时 存在大量的动态障碍物,如摄像头捕捉到移动的 行人时会导致视觉约束的错误,从而影响整体的 定位建图效果。同时,当遇到白墙等物体时无法 提取足够的特征,视觉约束失效从而导致建图失 败。下一步工作将利用激光雷达和视觉各自的优 点,提高整体的鲁棒性,保证建图的效果。

参 考 文 献

- Smith R, Self M, Cheeseman P. Estimating uncertain spatial relationships in robotics [C] // IEEE International Conference on Robotics and Automation, 2003: 435-461.
- [2] 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.
- [3] Konolige K, Grisetti G, Kümmerle R, et al. Efficient sparse pose adjustment for 2D mapping
 [C] // IEEE/RSJ International Conference on Intelligent Robots and Systems, 2010: 22-29.
- [4] Kohlbrecher S, Stryk OV, Meyer J, et al. A flexible and scalable SLAM system with full 3D motion estimation [C] // IEEE International Symposium on Safety, Security, and Rescue Robotics, 2011: 155-160.
- [5] Hess W, Kohler D, Rapp H, et al. Real-time loop closure in 2D LIDAR SLAM [C] // IEEE International Conference on Robotics and Automation, 2016: 1271-1278.
- [6] Davison AJ, Reid ID, Molton ND, et al. MonoSLAM: real-time single camera SLAM
 [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6): 1052-1067.
- [7] Sim R, Elinas P, Griffin M. Vision-based SLAM using the rao-blackwellised particle filter [J].
 IJCAI Workshop on Reasoning with Uncertainty in Robotics, 2005, 9(4): 500-509.
- [8] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces [C] // IEEE/ACM International Symposium on Mixed and Augmented Reality, 2007: 1-10.
- [9] Mur-Artal R, Tardos JD. ORB-SLAM2: an opensource SLAM system for monocular, stereo, and RGB-D cameras [J]. IEEE Transactions on Robotics, 2016, 33(5): 1-8.
- [10] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C] // IEEE International Conference on Computer Vision, 2012: 2564-2571.

- [11] Stühmer J, Gumhold S, Cremers D. Real-time dense geometry from a handheld camera [C] // Proceedings of the 32th DAGM Conference on Pattern Recognition, 2010: 11-20.
- [12] Engel J, Sturm J, Cremers D. Semi-dense visual odometry for a monocular camera [C] // IEEE International Conference on Computer Vision, 2013: 1449-1456.
- [13] Engel J, Schöps T, Cremers D. LSD-SLAM: largescale direct monocular SLAM [C] // European Conference on Computer Vision, 2014: 834-849.
- [14] Forster C, Pizzoli M, Scaramuzza D. SVO: fast semi-direct monocular visual odometry [C] // IEEE International Conference on Robotics and Automation, 2014: 15-22.
- [15] Newcombe RA, Izadi S, Hilliges O, et al. KinectFusion: real-time dense surface mapping and tracking [C] // IEEE International Symposium on Mixed and Augmented Reality, 2011: 127-136.
- [16] 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, 2014: 656-663.
- [17] Kerl C, Sturm J, Cremers D. Dense visual SLAM for RGB-D cameras [C] // IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014: 2100-2106.
- [18] Chen M, Yang S, Yi X, et al. Real-time 3D mapping using a 2D laser scanner and IMU-aided visual SLAM [C] // IEEE International Conference on Real-Time Computing and Robotics, 2018.
- [19] Houben S, Droeschel D, Behnke S. Joint 3D

laser and visual fiducial marker based SLAM for a micro aerial vehicle [C] // IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, 2017.

- [20] 王消为, 贺利乐, 赵涛. 基于激光雷达与双目视 觉的移动机器人 SLAM 研究 [J]. 传感技术学报, 2018(3): 394-399.
- [21] 张杰,周军. 一种改进 ICP 算法的移动机器人激 光与视觉建图方法研究 [J]. 机电工程, 2017(12): 1480-1484.
- [22] Shi J, He B, Zhang L, et al. Vision-based real-time 3D mapping for UAV with laser sensor [C] // IEEE/ RSJ International Conference on Intelligent Robots and Systems, 2016: 4524-4529.
- [23] Qin T, Li P, Shen S. VINS-Mono: a robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2017, 3(4): 1-17.
- [24] Li M, Mourikis AI. 3-D motion estimation and online temporal calibration for camera-IMU systems [C] // IEEE International Conference on Robotics and Automation, 2013: 5709-5716.
- [25] Lynen S, Achtelik MW, Weiss S, et al. A robust and modular multi-sensor fusion approach applied to MAV navigation [C] // IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013: 3923-3929.
- [26] Lowe DG. Distinctive image features from scaleinvariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [27] Bay H, Tuytelaars T, Gool LV. SURF: speeded up robust features [C] // European Conference on Computer Vision, 2006: 404-417.