

# 融合视觉信息的激光定位与建图

Laser Slam Based on Fusion of Visual Information

骆燕燕 陈 龙 (河北工业大学电气工程学院, 天津 300130)

**摘要:**针对大型和几何特征不够明显的环境,提出了激光与单目视觉融合的方法,进行同步定位与地图构建。结合视觉传感器获取信息丰富的特点,设计了激光与单目视觉融合的 SLAM 系统,设计了视觉回环检测环节,其中包括图像特征提取算法的选取、视觉关键帧的提取、视觉单词字典的构建和回环帧相似度的计算。进行了构建地图的实验,并与传统激光 SLAM 进行了对比。实验证明,该方法能够及时发现回环帧,有效抑制了误差的累积,起到较好的建图效果。

**关键词:**同步定位与建图,多传感器融合,回环检测,视觉词袋

**Abstract:**A new approach is designed by employing the visual information to aid the laser SLAM, and visual loop-closure detection is designed, which includes the selection of image feature extraction algorithm, the visual key frames extraction, the construction of bag of words and the calculation of the similarity of loop key frames. experiment is carried out to build maps.

**Keywords:**slam, multi-sensor fusion, loop-closure detection, bag of words

同步定位与地图构建<sup>[1]</sup>(SLAM)问题是目前移动机器人研究的热点,其解决方式是移动机器人实现自主导航的关键所在。

早期的基于扩展卡尔曼滤波(EKF)的 SLAM 算法,将非线性系统局部线性化,有效解决了对于非线性非高斯系统的 SLAM 问题。但对强非线性系统,线性化处理过程中难免会引入更多截断误差,甚至导致系统发散<sup>[2]</sup>。基于粒子滤波的 Fast SLAM 算法用一组样本(或称粒子)来近似表示系统的后验概率密度,避免了将非线性系统线性化,同时降低了计算复杂度。但在创建地图时多采用增量式地图的创建方法,会造成误差的逐渐积累,进而导致全局地图的不一致性,很难适用于大规模场景<sup>[3-4]</sup>。

基于图优化的方法主要以视觉 SLAM 为主。Raul Mur-Artal 等人提出的一种基于视觉关键帧提取的 ORB-SLAM 算法,在回环检测环节,该方法通过 ORB 特征构建词袋,配合词袋,提高了回环检测准确率和速度<sup>[5]</sup>。随后出现了很多基于 RGB-D 数据的 SLAM 算法。该类算法简化了三维重建的过程,但占用内存过多,实时性较差,很难应用于大规模场景<sup>[6-7]</sup>。本文拟在传统的基于图优化的激光 SLAM 系统中加入视觉回环检测环节,提高回环检测的可靠性。

## 1 改进的基于图优化的激光 SLAM

融合了视觉信息的激光 SLAM 系统由 4 部分组成,分别为

激光数据采集与扫描匹配、视觉数据处理与回环检测、后端图优化和地图构建。其流程图如图 1 所示。

### 1.1 激光扫描匹配

扫描匹配是指当前里程计位姿的观测信息与局部地图进行匹配,从而得到机器人的矫正位姿。主要有以下几个步骤:

1)进行扫描匹配时,首先要找到初始估计位姿。考虑到里程计位姿与估计位姿之间的误差,设在  $k-1$  时刻,机器人里程计位姿为  $X_{odo}^{k-1} = \begin{pmatrix} x_{odo}(k-1) \\ y_{odo}(k-1) \\ \theta_{odo}(k-1) \end{pmatrix}$ ,  $k-1$  时刻经过扫描匹配后的估计位姿

为  $X_{scan}^{k-1} = \begin{pmatrix} x_{scan}(k-1) \\ y_{scan}(k-1) \\ \theta_{scan}(k-1) \end{pmatrix}$ ;  $k$  时刻的里程计位姿为  $X_{odo}^k = \begin{pmatrix} x_{odo}(k) \\ y_{odo}(k) \\ \theta_{odo}(k) \end{pmatrix}$ , 扫描匹配之前的初始估计位姿为  $X_{est}^k = \begin{pmatrix} x_{est}(k) \\ y_{est}(k) \\ \theta_{est}(k) \end{pmatrix}$ , 初始位姿估计关系

如图 2 所示。

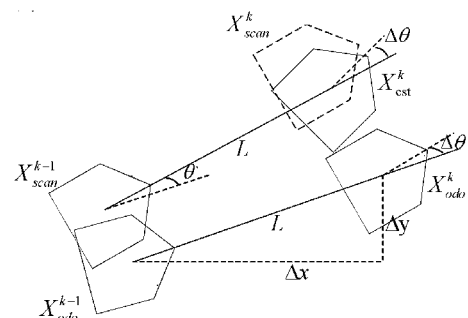


图 2 初始位姿估计关系图

图 2 中  $\Delta x$ 、 $\Delta y$ 、 $\Delta \theta$  表示里程计的位姿变化,  $\theta'$  表示在  $k-1$  时刻里程计位姿和扫描匹配位姿的朝向角偏差, 上述变量分别由(1)(2)式表达。

$$\begin{cases} \Delta x = x_{odo}(k) - x_{odo}(k-1) \\ \Delta y = y_{odo}(k) - y_{odo}(k-1) \\ \Delta \theta = \theta_{odo}(k) - \theta_{odo}(k-1) \end{cases} \quad (1)$$

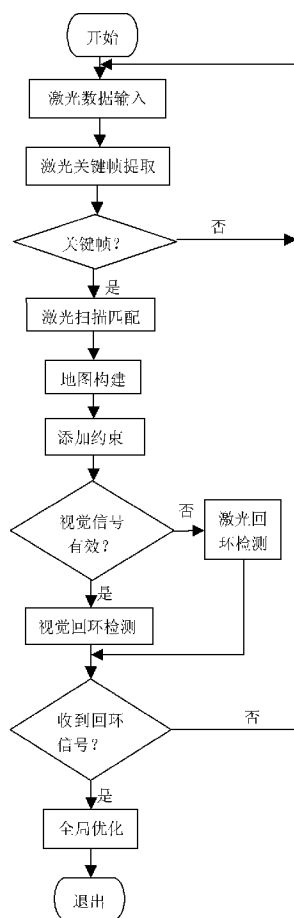


图 1 融合单目视觉的激光 SLAM 流程图

$$\theta' = \theta_{odo}(k-1) - \theta_{scan}(k-1) \quad (2)$$

则初始估计位姿  $X_{est}^k$  可以由式(3)计算得到。

$$\begin{cases} x_{est}(k) = x_{scan}(k-1) + \Delta x \cos(\theta') - \Delta y \sin(\theta') \\ y_{est}(k) = y_{scan}(k-1) + \Delta x \sin(\theta') + \Delta y \cos(\theta') \\ \theta_{est}(k) = \theta_{scan}(k-1) + \Delta \theta \end{cases} \quad (3)$$

2)在局部地图上,以初始估计位姿  $X_{est}^k$  为中心确定一个矩形区域,用以表示最终估计位姿  $X_{scan}^k$  的可能范围,并确定平移步长  $t_s$  和旋转步长  $r_s$  以及平移步数  $n_t$  和旋转步数  $n_r$ 。

3)将当前扫描数据通过坐标系转换投影到局部地图坐标系上。假设每次匹配共有  $m$  个点被击中,由于局部地图高斯模糊

的作用,击中点会有不同的分值,由公式  $M_{score} = \frac{\sum_{i=0}^m hit_i}{hitsum}$  计算,式中  $hit_i$  为第  $i$  个点的击中值,  $hitsum$  为每次匹配的总击中值。

4)选取拥有大于一定阈值的匹配值  $M_{score}$  的位姿,求和后取其均值作为扫描匹配最终估计位姿  $X_{scan}^k$ ,并计算此次扫描匹配协方差,用于后端优化。

### 1.2 回环检测

在进行回环检测时,系统首先要判断视觉信号是否有效。机器人运行过程中可能出现图像模糊、机器人运动过快或当前图像缺少特征点的情况,此时无法检测到图像特征点,即使在回环点处,也无法进行回环检测。此时以激光回环检测信号作为标准,直到视觉信号恢复正常为止;当视觉信号有效时,以视觉回环检测为准,无论激光是否检测到回环,只要视觉检测到回环,都会发出回环信号。并将通过时间戳找到与视觉关键帧对应的激光关键帧,加入到优化列表中进行优化。

### 1.3 后端优化

由于本文只是优化位姿,不对路标点进行优化,节约了内存,能进行大规模建图。

优化变量为机器人在世界坐标系下的位姿,变量表达式设为  $c_i = [t_i, \theta_i]^T = [x_i, y_i, \theta_i]^T$ 。对于任意位姿  $c_i$  和  $c_j$ ,可设观测方程的表达式为  $g(c_i, c_j) = \begin{bmatrix} R_i(t_j - t_i) \\ \theta_j - \theta_i \end{bmatrix}$ ,式中  $R_i = \begin{bmatrix} \cos(\theta_i) & \sin(\theta_i) \\ -\sin(\theta_i) & \cos(\theta_i) \end{bmatrix}$  为旋转矩阵。用  $Z_{ij}$  表示两位姿之间的偏移,则误差函数可表示为  $e_{ij} = z_{ij} - g(c_i, c_j)$ 。可以设定待优化目标函数为  $F(c) = \sum_j e_{ij}^T \Lambda_{ij} e_{ij}$ ,式中  $i$  和  $j$  为建立约束的节点,  $\Lambda_{ij}$  为两约束节点间的信息矩阵(协方差矩阵的逆)。

选用易于扩展的 ceres 后端优化工具,并采用流行的 Levenberg-Marquardt(L-M)方法对上述目标函数进行优化。计算得到节点误差关于优化变量的 Jacobians 如式(4)所示:

$$\begin{aligned} \frac{\partial e_{ij}}{\partial t_i} &= \begin{bmatrix} -R_i^T \\ 0 \ 0 \end{bmatrix} & \frac{\partial e_{ij}}{\partial \theta_i} &= \begin{bmatrix} -\partial R_i^T(t_j - t_i) / \partial \theta_i \\ -1 \end{bmatrix} \\ \frac{\partial e_{ij}}{\partial t_j} &= \begin{bmatrix} R_i^T \\ 0 \ 0 \end{bmatrix} & \frac{\partial e_{ij}}{\partial \theta_j} &= [0 \ 0 \ 1]^T \end{aligned} \quad (4)$$

## 2 视觉回环检测

视觉回环检测通过视觉传感器采集信息,采集信息较激光传感器丰富,回环检测时容易发现回环帧。本文的视觉回环检测主要由图像特征点的提取、关键帧的选取与回环帧检测等环节组成,并通过常用的视觉词袋模型计算图像相似度。流程图如图3所示。

### 2.1 图像特征提取方法的选择

回环检测依赖图像特征点的提取,合适的特征提取算法能

提高特征提取速度和帧间匹配成功率。常用的图像特征点检测算法有 SIFT、SURF、FAST 和 ORB 等。

SIFT 特征点具有尺度不变性和旋转不变性,且不易受光照和噪声等因素的影响,但特征点提取和描述子计算时间较长,实时性差。SURF 算法是 SIFT 的改进版本,性能与 SIFT 相近。ORB 算法是 FAST 的改进版本,通常采用改进的 rBRIEF 作为描述子,特征点具有旋转不变性。

实时性能是选取特征提取方法的一个重要指标,本文对同一幅图像分别用不同方法进行了特征提取和描述子计算,并记录了不同算法的完成时间。表1为实验对比结果。

由表1可知,FAST 和 ORB 实时性能远优于 SIFT 和 SURF。考虑到 ORB 特征点具有旋转不变性,本文选用 ORB 特征提取算法,为回环检测环节提供图像的特征信息。

表1 特征点提取算法性能对比

特征提取算法	提取角点个数/对	完成时间/ms
SIFT	987	897
SURF	1265	302
FAST	1521	18
ORB	785	19

### 2.2 关键帧提取和回环帧检测

通过视觉传感器采集的连续帧图像有很多冗余的信息,连续处理每帧图像会大大降低系统的运行效率,通过在连续帧中选取关键帧的方法可以有效提高系统的运行效率。关键帧选取必须满足一定条件。本文采取如下原则进行关键帧提取:

- 1)待插入关键帧的特征点数量不能少于 75 个,保证关键帧有足够的特征信息;
- 2)相邻关键帧之间至少间隔 25 帧连续帧,保证相邻关键帧的距离不至于太近,剔除冗余帧;
- 3)机器人在运行过程中可能停留在某个位置,为避免在同一地点连续采集关键帧,需要检测相邻关键帧间机器人的移动距离,只有移动距离大于一定阈值,才进行关键帧提取。

在进行回环帧检测的过程中,因为相邻关键帧间相似性较高,当发生回环检测时,可能连续检测到回环帧。为避免添加无用的约束,本文设定相邻回环关键帧至少间隔 15 帧,否则认为当前回环帧无效。

由于词袋的回环检测完全依赖于图像外观而未利用几何信息,且只能表达单词的有无,而不能体现单词的顺序,所以容易引发感知偏差。为此,本文加入了一个验证步骤,设立了回环的缓冲机制,认为单次检测的回环不足以确认是良好的回环,只有连续检测到的回环帧数大于 5 帧时,才认为是正确的回环。

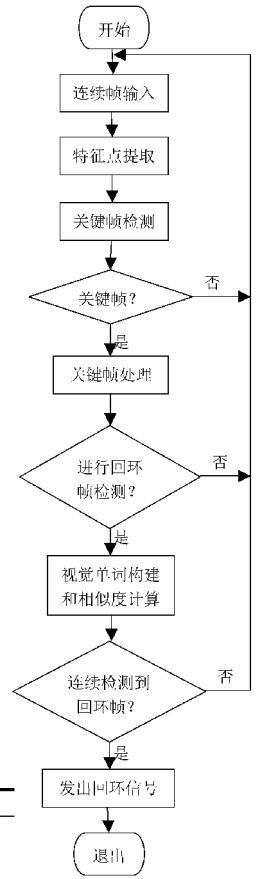


图3 视觉回环检测



图4 turtlebot 机器人

3 实验结果与分析

3.1 实验平台和环境

图 4 所示的实验平台为 TurtleBot 2.0 机器人,该机器人配备 Intel i5 处理器、内存 2.0 GB 的笔记本电脑,笔记本电脑安装 Ubuntu 14.04 操作系统和 Indigo 版本的 ROS。

本文选择搭载 Kinect V1 版本的 RGB-D 传感器,该传感器除了能采集图像数据,还可以测量图像的深度,通过 depthimage\_to\_laserscan 软件包能将图像深度数据转换成激光数据,模拟激光雷达构建地图。本文设计在路面平坦的大型走廊环境中进行试验,机器人运行速度为 0.25m/s,并记录了运行时间。

3.2 实验结果分析

本文分别用两种 SLAM 算法在同一环境下构建地图并进行对比,如图 5 所示。

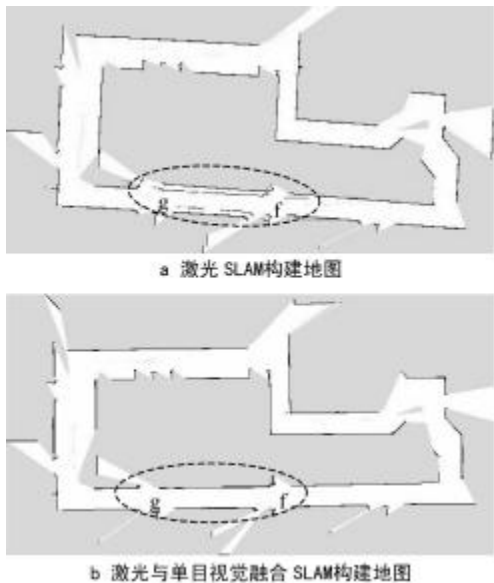


图 5 两种 SLAM 方法构建地图

图 5a 为激光 SLAM 所建地图,机器人从 f 点开始顺时针方向沿着走廊运动,运行一圈之后,再次经过 f 点时,由于获取的几何信息不够丰富,并没有及时发现回环点。如图 5a 中虚框所示,当机器人继续运行到 g 点时,依然没有发现回环点,此时可以明显发现地图的累积误差不断增大。

图 5b 为融合了视觉信息的激光 SLAM 所建地图,对于同

样的运行轨迹,由于视觉信息的融合,当机器人再次经过 f 点时,及时发现回了环点,并进行了全局优化,在运行到 g 点的过程中又进行了多次优化,及时消除了地图中的累积误差。

图 5a 中机器人第二次到达 g 点时,所用时间为 10 分 46 秒,图 5b 中第二次到达 g 点所用时间为 11 分 28 秒。系统在加入视觉回环检测环节后,会在一定程度上影响运行效率,但由于引入的视觉信息只是用于回环检测,不必对计算机器人进行跟踪、优化位姿和更新地图等,省却了很多耗时的操作。所以在同等运行速度和运行环境的前提下,所建地图 5b 与 5a 时间仅相差 42s,可知对系统的实时性能影响并不大。

实验证明,该方法能及时发现回环点,减少累计误差,起到较好的建图效果。

参考文献

[1]Smith R,Self M,Cheeseman P.Estimating Uncertain Spatial Relationships in Robotics [J].Machine Intelligence & Pattern Recognition, 2003,4(5):435-461

[2]王法胜,鲁明羽,赵清杰,等.粒子滤波算法[J].计算机学报,2014 (8):1679-1694

[3]Montemerlo M,Thrun S, Koller D, et al. Fast SLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem[C]//Proceedings of the National conference on Artificial Intelligence. 2002: 593-598

[4]Montemerlo M, Thrun S, Roller D, et al.FastSLAM 2.0:an improved particle filtering algorithm for simultaneous localization and mapping that provably converges [C]//International Joint Conference on Artificial Intelligence.Morgan Kaufmann Publishers Inc. 2003:1151-1156

[5]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

[6]Henry P,Krainin M,Herbst E,et al.Experimental Robotics [M]. Springer Berlin Heidelberg, 2014:647-663

[7]Henry P,Krainin M,Herbst E,et al.RGB-D mapping:Using Kinect-style depth cameras for dense 3D modeling of indoor environments[J].International Journal of Robotics Research, 2012, 31 (5):647-663

[收稿日期:2017.7.26]

(上接第 20 页)

表 1 人数统计测试结果

测试序号	实际人数	自动计数	误差率
1	80	78	97.5%
2	105	101	96.1%
3	76	74	97.3%
4	93	90	96.7%

测试结果在实验室环境下得出,人数统计正确率可达 96% 以上,可以满足实际应用需求,满足实际嵌入式应用中对实时性和检测精度的要求,该方法有较好的鲁棒性。

参考文献

[1]朱秋煜,唐利,郁铭,等.一种基于立体视觉的公交车客流计数方法[J].中国图象图形学报,2009,14(11):2391-2395

[2]Zhao Xi,Dellandr'ea E,Chen Liming. A People Counting System Based on Face Detection and Tracking in a Video[C]//

Proceedings of the 6th IEEE International Conference on Advanced Video and Signal-based Surveillance. Genoa, Italy: IEEE Press, 2009

[3]常庆龙,夏洪山,黎宁.一种归一化前景和角点信息的复杂场景人数统计方法[J].电子与信息学报,2014,36(2):312-317

[4]Wu S, Kan M, He Z, et al. Funnel-Structured Cascade for Multi-View Face Detection with Alignment-Awareness [J]. Neurocomputing, 2017,221(C):138-145

[5]Yan S,Shan S,Chen X, et al. Locally Assembled Binary (LAB) feature with feature-centric cascade for fast and accurate face detection[C]// Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. IEEE, 2008:1-7

[6]Henriques J F, Rui C, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters [J].IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(3):583-596

[收稿日期:2017.7.14]