视觉SLAM与构建地图方法概述

毕浩博,谭丹丹,兰崇嘉,高丽

(辽宁科技大学 计算机与软件工程学院,辽宁 鞍山 114051)

摘要:概述了基于视觉SLAM经典框架的五个组成部分,重点分析了特征点法的简单原理及特性、非线性系统和扩展卡尔曼滤波以及回环检测的词袋模型。最后展望了未来基于深度相机的SLAM研究和发展方向。

关键词:视觉 SLAM;视觉里程计;后端优化;回环检测

中图分类号:TP391 文献标识码:A

文章编号:1009-3044(2020)12-0210-02

开放科学(资源服务)标识码(OSID):



DOI:10.14004/j.cnki.ckt.2020.1323

Overview of Visual SLAM and Map Building Methods

BI Hao-bo, TAN Dan-dan, LAN Chong-jia, GAO Li

(Liaoning University of science and technology, College of computer and software engineering, Anshan 114051, China)

Abstract: This paper summarizes the five components of the visual slam framework, focusing on the simple principle and characteristics of the feature point method, the nonlinear system, the extended Kalman filter and the word bag model of loop detection. Finally, the future research and development of SLAM Based on depth camera are prospected.

Key words: vision SLAM; vision sensor; vision odometer; back end optimization; loop detection

1 背景

近年来,视觉同步定位和地图构建(SLAM)已逐渐稳步向学术方向发展全自动车辆的可靠性。简而言之,SLAM问题包括一组姿态下的车辆轨迹估计,同时建造环境地图。除了自我定位之外地图在避障、目标识别、任务规划等方面有着广泛的应用。通过读取多种传感器反馈的信息来估计本体运动信息和未知环境的特征信息,并融合反馈信息对本体位姿进行精确估计以及场景的空间建模是目前研究SLAM问题的常用方式。当前SLAM采用的传感器有激光和视觉等多种类型,视觉传感器其丰富的色彩和纹理等感知信息在提高帧间估计精度和闭环检测正确率方面较激光传感器单一的空间结构感知信息有着明显的优势和潜力。

常用的视觉传感器可分为:单目相机、双目相机、深度相机(RGB-D相机)。单目相机解决SLAM问题时,由于平移之后才能计算深度,以及无法确定真实尺度,会增加后端优化、回环检测时的计算量。而双目相机和RGB-D相机克服了单目相机无法知道距离的缺点,同时已知距离的情况下可以通过单张图像构建出场景的三维结构。双目相机和RGB-D相机的差异将在本文中进行讨论,此处略过。本次探究SLAM问题采用经典视觉SLAM框架,如图1所示。

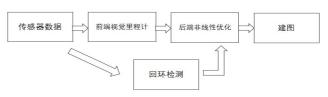


图1 整体视觉SLAM流程图

分为传感数据读取、视觉里程计(Visual Odometry, VO)、后端优化(Optimization)、回环检测(Loop Closing)、建图(Mapping) 五个部分,本文将着重讨论视觉里程计、后端优化、回环检测这三个部分。同时将工作环境尽量维持在静态、刚体,光照变化不明显、没有人为干扰的场景,以确保定位与建图的精准性。

本文主要对经典的视觉 SLAM 框架与地图构建问题的分析和讨论,系统地介绍和分析本次实验的传感器、视觉里程计、后端优化、回环检查,以及地图构建等五个部分。重点分析了特征点法的简单原理及特性、非线性系统和扩展卡尔曼滤波以及回环检测的词袋模型。最后展望了未来基于深度相机的SLAM 研究和发展方向。

2 视觉里程计

里程计在汽车领域有着广泛的应用,其目标是计量汽车行驶的距离,其计算方式在理想状态下简单来讲是将轮子周长与转动圈数的乘积。这种计算方式如果遇到打滑等特殊情况,就会出现偏差。当然在汽车领域这些误差微不足道,而对于机器

收稿日期:2020-02-28

基金项目:辽宁科技大学2019年校级大学生创新创业训练计划项目(项目编号:201910146320)

作者简介: 毕浩博(1997—), 男, 辽宁鞍山人, 学生, 软件工程专业; 谭丹丹(1981—), 女, 辽宁鞍山人, 通讯作者, 讲师, 硕士, 主要研究方向为人工智能、图形图像。

本栏目责任编辑: 唐一东

人和视觉领域来说差之分豪,谬之千里。在解决SLAM问题 中,我们不仅仅要考虑机器人位移的量,同时还要考虑机器人 自身的姿态,就将其统称为位姿。因此传统的里程计已经不能 满足该领域的需求。

可以确定机器人位资的方式很多,目前解决SLAM问题上 视觉里程计的表现更为优异,无论是在精准度上,还是在反馈 效率上都有着明显的优势。视觉里程计是根据拍摄图像来估 计相机的运动,由这个特点可以大概分为特征点法,直接法以 及Lucas-Kanade 光流法这三种方法。本章只介绍特征点法的 原理以及特性。

2.1 特征点法的简单原理及特性

视觉里程计的目标是根据拍摄图像来估计相机运动,而图 像用色彩和亮度的矩阵组成,如果用矩阵来估计相机运动,会 增加计算难度。因此目前采用图中较有特征性的点,通过改变 相机的角度,通过前后两张图片上的特征点来估计相机的位 资,所以我们也可以说特征是图像另一种的数字表达形式,因 此特征点法也成了解决SLAM问题广为流传的方法。

组成特征点的两个因素分别为关键点和描述子。描述特 征点在图像中的位置,大小以及方向等信息的量成为关键点。 而描述子是描述的是某一关键点周围像素的信息,通常描述子 是按照人为设定的方式去描述的,会选择图像中的角点、边缘 或者区块等一些具有代表性作为特征点。当然只通过角点这 些特征点是无法准确描述的,如今用 SIFT, SURF,ORB 等局部 图像特征来描述图像信息,更加稳定。

说到局部特征最为经典的就是SIFT,优点在于在提取局部 特征的同时考虑了图像变换过程中光强,尺度变换,旋转等可 能出现的变化,这也是最理想化的提取特征方式。但是随之而 来的就是需要消耗大量的计算资源,为了考虑到计算速度的问 题,适当地降低精准度和鲁棒性是比较可取的方式。目前最具 有代表性的实时图像特征就是ORB特征。它采用速度极快的 二进制描述子BRIEF,加快了提取特征的效率,并且保留了 SIFT特征的特性。因此ORB特征在实时SLAM里有相对更好 的表现力。为特征匹配提供了更精准的特征环境。

大体来讲,特征匹配解决了SLAM的数据关联的问题,通 过图像之间的描述子进行匹配,最后进行位资估计。由于图像 特征的局部特性,这样就会存在误匹配的现象,对于视觉SLAM 发展来说是一大瓶颈。但是仅通过局部特征来解决误匹配是 十分困难的事情。

3 后端优化

视觉里程计处理的图像信息是两个相邻时间点内的运动 轨迹,其存储的时间也十分短暂,这样会用过去的时间来更新 当前的运动状态,也就是运动状态会存在渐进性的问题,因此 后端优化的目标是解决整个运动时间内的状态估计问题,用过 去时间点内的信息和未来时间点的信息共同更新运动。后续 处理因系统的线性问题而方法各不相同,我们主要讨论非线性 系统和扩展卡尔曼滤波(EKF)。

3.1 非线性系统和扩展卡尔曼滤波

在SLAM问题中,运动方程和观测方程通常都不是线性函 数,例如相机模型需要调用相机内参和李代数表示位姿。这 并非是线性函数,如果用卡尔曼滤波器构成的线性系统处理该

函数,得到的分布列误差很大。其原因是卡尔曼滤波器假设所 有的状态和噪声都满足高斯分布,但是高斯分布经过非线性变 化,结果通常都不会是高斯分布。因此我们的目标是将卡尔曼 滤波的结果分布扩展到非线性系统中,这个过程被称为扩展卡 尔曼滤波器(简称EKF)。

简单描述其推导过程,先通过卡尔曼滤波器得到线性化 的无偏最优估计,后经过后验概率得到了单次线性近似下最大 后验估,从而解决了线性系统无法处理非线性问题的难题。扩 展卡尔曼滤波器也具有许多局限性。首先它考虑k时刻的状 态只通过k-1时刻相的状态,这样也就是一定程度上假设了 马尔可夫性。而非线性优化方法倾向于用使用所有的历史数 据,这样就增加了计算资源的使用。从程序实现上来说,EKF 需要存储状态量的均值和方差,并对它们进行维护和更新。如 果把路标也放进状态的话,由于视觉 SLAM 中路标数量很大, 这个存储量是相当可观的,且与状态量呈平方增长(因为要存 储协方差矩阵)。因此,EKF SLAM 普遍被认为不可适用于大 型场景。

4 回环检测

在SLAM经典框架中,视觉里程计提供特征点的提取和轨 迹、地图的初值,而后端负责对这所有的数据进行优化,这两个 模块的主要目的是估计相机运动。而只靠视觉里程计通过相 邻时间点上的数据,会将产生的误差累积到下一个时刻,使得 整个 SLAM 会出现累积误差,最终无法构建全局一致的轨迹和 地图。虽然后端能够估计最大后验误差,但是只有相邻关键帧 数据时,也无法解决累积误差问题。而回环检测模块的关键在 于其可以有效地检测出相机经过同一个地方,这样就可以反馈 给后端更多的有效信息,特别是得到一个全局一致的估计,也 就是估计的轨迹和地图在长时间下的正确性。因此对SLAM 的重要性不言而喻。

回环检测最简单的方法就是对任意两张图像都做一遍特 征匹配,根据正确匹配的数量来判断哪些图像存在关联。虽然 这种方法有很好的实效性,但这种想法存在一个强假设,就是 认定任意两个图像都可能存在回环。对于 N 个可能的回环,需

要检测 C_n^2 次,这是 $O(N^2)$ 的复杂度,随着轨迹变长增长太快, 在大多数实时系统当中是不实用的。尽管随机检测在有些实 现中有用,但设计回环检测的期望是可以找出可能出现回环的 点,这样就可以减少回环次数避免资源浪费。这样的方式大体 分为两种思路:基于里程计的几何关系,或基于外观。基于几 何关系我们可以知道,当前相机运动到了之前的某个位置附近 时,检测它们有没有回环关系,但是由于累积误差的存在,没法 正确地发现运动到了之前的某个位置附近。因此,回环检测也 无从谈起。另一种思路是基于外观的。它和前端后端的估计 都无关,仅根据两张图像的相似性确定回环检测关系。这种做 法摆脱了累计误差,使回环检测模块成为 SLAM 系统中一个相 对独立的模块,摆脱了基于几何关系的思路中累积误差对结果 的影响。

由于回环检测提供了当前数据与所有历史数据的关联,在 跟踪算法丢失之后,我们还可以利用回环检测进行重定位。因 此,回环检测对整个 SLAM 系统精度与鲁棒性的提升是非常明 显的。

(下转第216页)

的录入正常。



图7 音频设备和输入检测界面

3.4.2 电脑内音频的捕捉

如图 8 所示在 System audio 的右侧点击最左侧的按钮变成 黄色出现 Enable system audio 可进行电脑内音频的捕捉录制。



图8 捕捉电脑内音频的界面

3.4.3 软件控制设置

如图 9 所示 Control Setting 下方的两个复选框为软件控制设置,若最终录制的视频文件中不想出现鼠标点击动画,则 Mouse click animation前的复选框不勾选,否则勾选;若要在视频录制过程中使用 F10/F11 作为视频录制自动结束的快捷键则 Enable F10/F11 shortcut key 前的复选框勾选,否则不勾选,这里 F10 是一键直接停止视频录制的快捷键,F11 是出现停止录像的界面以便手动停止视频录制的快捷键。



图9 软件控制设置

4视频的制作

点击圆形红色按钮开始屏幕摄像,在倒计时三秒结束后,视频制作者在选定视频录制屏幕区域内的一切操作及输入的音频将被录制为视频文件,视频录制完毕后点击F10或者F11结束视频录制,视频制作完成。

5讨论和应用体会

Live Screen Capture 软件是教学中应用效果比较好的多媒体课件的一种,传统的多媒体视频课件的制作对制作者的摄像、视频编辑的水平要求较高,一般的医学教育工作者往往很难掌握,Live Screen Capture 软件操作方法简单易上手,是比较实用的多媒体课件制作方法之一。

参考文献:

- [1] 胡艳全, 余清波, 支平. 实例视频教学课件的制作方法[J]. 长春教育学院学报, 2014, 30(9): 88-92.
- [2] 徐振亚, 张雅茹. 谈音视频素材在医学多媒体课件中的应用 [J]. 中国医学教育技术, 2012, 26(4): 412-414.
- [3] 陈少锐, 陈健文, 皮荣标, 等. 网络视频课件在药理学实验教学的应用[J]. 药学教育, 2016, 32(2): 76-78.
- [4] 张红娜, 王秋阁. 应用 Authorware 制作多媒体教学课件[J]. 电脑知识与技术, 2015, 11(23): 132-133.
- [5] 蔡宏宇. 浅谈多媒体课件在基础医学教学中的应用[J]. 中国继续医学教育, 2018, 10(28): 35-37.

【通联编辑:王力】

(上接第211页)

5 视觉 SLAM 的发展

对于目前的视觉 SLAM来说,未来发展的方向大体来讲分为两大方向一是往轻量级、小型化方向发展,让 SLAM 能够在嵌入式或手机等小型设备上良好的运行,然后考虑以它为底层功能的应用。另一个方面,则是利用高性能计算设备,实现精密的三维重建、场景理解等功能。在这些应用中,我们的目的是完美地重建场景,而对于计算资源和设备的便携性则没有多大限制。由于可以利用 GPU,这个方向和深度学习亦有结合点,而如何将深度学习架构的应用贯穿于整个 SLAM 系统仍是一个巨大挑战。

参考文献:

- 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] 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.

- [3] Handa A, Bloesch M, Pătrăucean V, et al. Gvnn: neural network library for geometric computer vision[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 67–82.
- [4] 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.
- [5] Salas-Moreno, Renato F. Dense semantic SLAM[D]. London, UK: Imperial College, 2014.
- [6] Kitt B, Geiger A, Lategahn H. Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme[C]//2010 IEEE Intelligent Vehicles Symposium, June 21–24, 2010. La Jolla, CA, USA. IEEE, 2010: 486–492.
- [7] Jaderberg M, Simonyan K, Zisserman A. Spatial transformernetworks[C]//Advances in Neural Information Processing Systems. San Francisco, USA: Morgan Kaufmann, 2015: 2017– 2025.

【通联编辑:谢暖暖】