

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。

可以确定机器人位姿的方式很多,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姿,所以我们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,适当地降低精度和鲁棒性是比较可取的方式。目前最具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,最后进行位姿估计。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。这并非是非线性函数,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计,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实用,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。

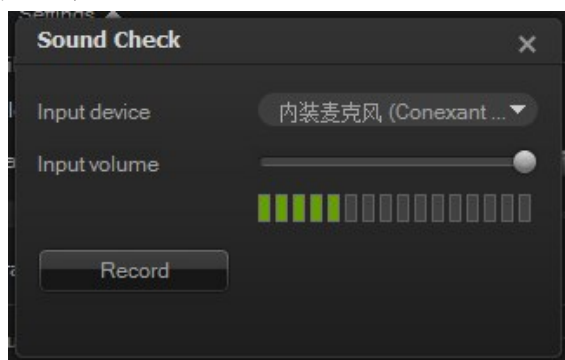


图 7 音频设备和输入检测界面

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。

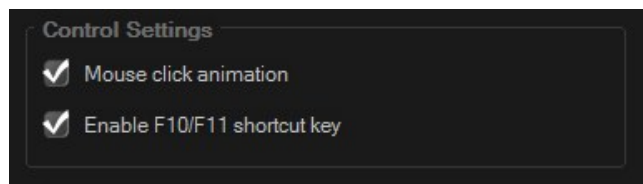


图 9 软件控制设置

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]. 药教教育, 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:

- [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] 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 transformer-networks[C]//Advances in Neural Information Processing Systems. San Francisco, USA: Morgan Kaufmann, 2015: 2017-2025.

【通联编辑:谢媛媛】