

VIO综述

1.VIO可以解决什么问题？为什么要多传感融合？有何优势？

VIO背后本身是一种多传感融合技术，这是大趋势。因为我们现有的传感器无法一招独步天下，会出现单一传感器无法胜任的场景。比如视觉传感器在纹理特征丰富的场景下能很好地工作，在白墙，玻璃等弱纹理条件下表现不好，甚至不能工作。2D Lidar在走廊环境没法很好地进行匹配。IMU长时间使用会出现非常大的累计误差。差分GPS在天气比较好的情况下可以很好地定位，但是在高楼林立的地方或者天气恶劣时定位效果会大打折扣。因此利用各种传感器之间互补的特性进行多传感融合，是保证SLAM系统稳定工作的有效途径。

IMU传感器：能够测量传感器本体的加速度和角速度，与相机或者雷达有比较好的互补特性。

1. IMU本身高频测量得到角速度和加速度，具有明显的漂移，使得积分两次的位姿数据非常不准确。但是对于短时间的运动，这些高频的imu数据能提供一个较好的估计。这是相机不具备的优势，特别是对于卷帘门相机更是如此。当机器人运动过快，图像就会变得模糊，以致于无法很好提取特征，或者两帧相机运动视差较小，则无法很好地估计相机的运动。所以纯视觉slam无法工作的情况下，IMU可以提供一个较好的估计，保证系统有效。

2. 相比于IMU数据，相机的测量基本不会有太大的变动，所以相机的测量数据可以很好地修正IMU的估计

3. 当图像发生变化时，本质上我们没法知道是相机自身发生了运动，还是外界条件发生了变化，所以纯视觉SLAM难以处理动态的障碍物。而IMU能够感受到自己的运动信息，从某种程度上减轻动态物体的影响。

2. 主流方案

从数据融合方面来讲，主要分为松耦合和紧耦合两种方式。

1. 松耦合

采用独立的惯性定位模块和视觉定位模块，往往以惯性定位为中心，视觉测量辅助修正偏差。视觉定位模块不需要imu信息做辅助。在视觉失效的情况下，没有引入imu的信息减少漂移。

2. 紧耦合

使用IMU的数据去完成视觉部分的运动估计和空间地图恢复。一般会将两帧相机间imu数据预积分，然后约束两帧相机之间的运动。然后对于单目而言，imu可以恢复出尺度。单目视觉则无法做到。

业内主流的方案大多是基于紧耦合的（能充分利用传感器互补的特性），松耦合的方案资料很少。紧耦合的方案也主要分两种，一种是基于滤波，一种是基于优化。

1. 基于滤波的紧耦合方案

a. 比较经典的有MSCKF

谷歌tango里面的算法。在传统的EKF-SLAM框架中，特征点的信息会加入到特征向量和协方差矩阵里，这种方法的缺点是特征点的信息会给一个初始深度和初始协方差，如果不正确的话，极易导致后面不收敛，出现inconsistent的情况。Msckf维护一个pose的FIFO，按照时间顺序排列，可以称为滑动窗口(slide window)，一个特征点在滑动窗口的几个位姿(帧)都被观察到的话，就会在这几个位姿间建立约束，从而进行KF的更新。

b. ROVIO

基于扩展卡尔曼滤波:imu测量用于滤波器的状态传递过程，视觉信息在滤波器更新使用。

2. 基于优化的紧耦合方案

a. 比较经典的有OKVIS

相对应于MSCKF的filter-based SLAM派系，OKVIS是keyframe-based SLAM派系做visual-inertial sensor fusion的代表。OKVIS是将image观测和imu观测显式formulate成优化问题，一起去优化求解pose和3D map point。OKVIS的优化目标函数包括一个reprojection error term(重投影误差)和一个imu integration error term(imu积分误差)，其中已知的观测数据是每两帧之间的feature matching(特征匹配)以及这两帧之间的所有imu采样数据的积分，注意imu采样频率一般高于视频frame rate，待求的是camera pose和3D map point，优化针对的是一个bounded window内的frames。

b. 港科大的VINS

前端用LK光流法做特征跟踪，相近和imu的预处理（预积分过程）、单目惯性联合初始化（在线的标定过程）、基于滑动窗口的BA联合优化（这里跟OKVIS的大体一致，但是采用了不同的误差模型）、全局的图优化和回环检测等。

3. VIO新进展

1. Visual-Inertial Mapping with Non-Linear Factor Recovery

通过重建非线因子图，将回环约束加入到因子图中，进行全局非线性优化。

2. VIL-VIO: Stereo Visual Inertial LiDAR Simultaneous Localization and Mapping 系统由紧耦合双目vio和LiDAR mapping，以及 LiDAR enhanced visual loop closure.解决穿越隧道失效的问题。

3. Unsupervised Deep Visual-Inertial Odometry with Online Error Correction for RGB-D Imagery 学会了在没有惯性测量单元（IMU）内在参数或IMU和摄像机之间的外部校准的情况下执行视觉惯性里程计。

4. Selective Sensor Fusion for Neural Visual-Inertial Odometry. 论文集中在如何学习多传感器融合策略上。提出了一种针对单目VIO的端到端的多传感器选择融合策略。具体是指提出了两种基于不同掩蔽策略(masking strategies)的融合模态：确定性软融合和随机硬融合，并与先前提出的直接融合baseline进行比较。在测试期间，网络能够选择性地处理可用传感器模态的特征并且产生确定尺度下的轨迹。

5. VINet: Visual-inertial odometry as a sequence-to-sequence learning problem 利用FlowNet和RNN来做结合，处理VIO的问题。

4. 为什么选择VINS?

a. 完备的V-SLAM方案：包含IMU预积分，鲁棒的初始化，在线估计外参数，重定位以及高效的全局优化

b. 优异的性能：跟目前开源的学界和工业的方案相比

c. 扩展性好：有双目，单目，以及融合GPS

d. 针对现有雷达方案，由于更好地利用了IMU信息和视觉信息，可以处理打滑，以及长走廊场景，进一步提高系统的鲁棒性和精度

5. 后续改进方向

a. 加速vio后端，不依赖其他优化库，更好的优化算法，多线程等

b. 解决一些退化情况，融入里程计等其他信息,应用在移动机器人

c. 前端加入点线面等特征增强鲁棒性，硬件加速和多线程加速前端，提高帧率