VINS-Fusion论文精读：一种通用的基于优化的多传感器局部里程计估计框架

参考：

<https://zhuanlan.zhihu.com/p/627643416>https://zhuanlan.zhihu.com/p/627933712

# 摘要

在本文中，我们提出一种通用的基于优化的里程计估计框架，它支持多种传感器套件。每个传感器被作为我们框架中的一个通用因子。共享共同状态变量的因子被加到一起以构建优化问题。我们进一步展示视觉和惯性传感器的通用性，这形成三种传感器套件（双目相机、单目相机和IMU，以及双目相机和IMU）。

# 介绍

实时6自由度（DoF）状态估计是机器人学中的一种基础技术。IMU能够以高频测量加速度和角速度，这对于实时应用中低延迟位姿反馈是必要的。因此，很多研究工作将IMU和视觉融合到一起。另一个用于状态估计的主流传感器为激光雷达。基于激光雷达的方法在受限的局部环境中实现了精确的位姿估计。尽管在过去已经提出了很多算法，但是它们通常被应用于单个输入传感器或者特定的传感器套件。

在本文中，我们提出一种通用的基于优化的位姿估计框架，它支持多传感器组合。我们使用视觉和惯性传感器进一步进行演示，这形成了三种传感器套件（双目相机、单目相机和IMU，以及双目相机和IMU）。我们能够容易地在不同传感器组合之间进行切换。我们强调本文的贡献如下：

1）一种通用的基于优化的状态估计框架，它支持多传感器；

2）使用视觉和惯性传感器进行状态估计的详细推导，这些传感器形成不同的传感器套件（双目相机、单目相机+IMU以及双目相机+IMU）；

# 相关工作

在过去几十年中，状态估计一直是一个热门的研究课题。大量算法着重于精确的6自由度位姿估计。我们已经看到很多使用一种类型传感器工作的令人印象深刻的方法，例如基于视觉的方法、基于激光雷达的方法、基于RGB-D相机的方法和基于事件相机的方法。使用单目相机工作的方法难以实现6自由度位姿估计，因为无法从单个相机恢复绝对尺度。为了提高可观性和鲁棒性，将具有互补性质的多传感器融合到一起。

多传感器融合方法有两种趋势。一种是基于滤波的方法，另一种是基于优化的方法。

基于滤波的方法通常通过扩展卡尔曼滤波器（EKF）实现。视觉和惯性测量通常被一起滤波，以进行6自由度状态估计。高频的惯性传感器被用于状态传播，视觉测量被用于状态更新。

MSCKF是一种基于EKF的视觉惯性里程计（VIO），它维护若干相机位姿，并且利用多个相机视图来构成多约束更新。基于滤波的方法通常更早地线性化状态，并且受到不精确线性化点引起的误差的影响。

基于优化的方法维护大量的测量值，并且同时优化多个变量，这也称为捆集调整（BA）。

与基于滤波的方法相比，基于优化的方法在时间同步方面具有优势。

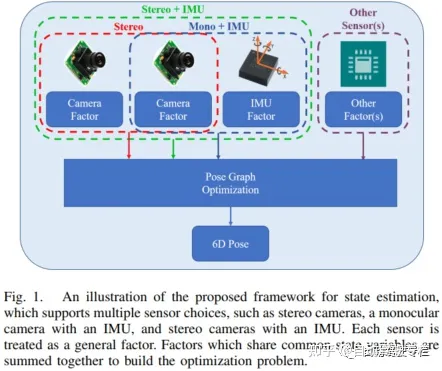
由于大的捆集作为自然缓冲区，因此当来自多个传感器的测量出现混乱时，它能够容易地处理这种情况。基于优化的算法在精度方面也优于基于滤波的算法，其代价为计算复杂度。

早期的优化求解器（例如G2O）利用Gauss-Newton和Levenberg-Marquardt方法来解决优化问题。尽管优化求解器中使用了稀疏的结构，但是复杂度随着状态和测量的数量呈平方增长。

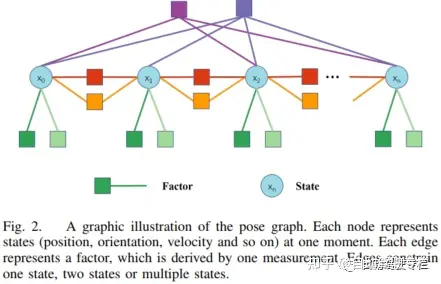
在本文中，我们采用了基于滑动窗口优化的状态估计框架。

# 系统概述

所提出框架的结构如图1所示，多种类型的传感器被自由地组合。



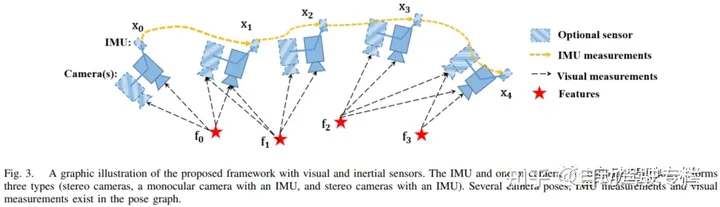
每个传感器的测量作为一个通用的因子。因子和它们相关的状态构成位姿图。位姿图的一种示例如图2所示。





每个节点表示一个时刻的状态（位置、姿态、速度等）。每条边表示一个因子，它通过一次测量获得。因子约束一个状态、两个状态或者多个状态。对于IMU因子，它通过连续的运动限制约束两个连续的状态。对于一个视觉路标，其因子约束了多个状态，因为它在多帧上被观测到。一旦构建了因子图，优化它等价于找到尽可能匹配所有边的节点配置。

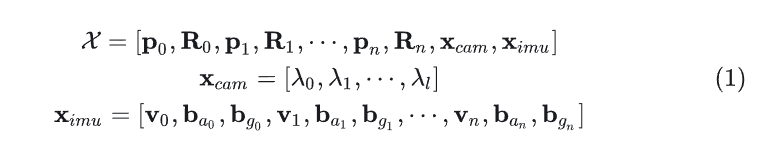
在本文中，我们具体演示了使用视觉和惯性传感器的系统。视觉和惯性传感器能够形成三种组合用于6自由度状态估计，即双目相机、单目相机和IMU、双目相机和IMU。图3显示了使用视觉和惯性传感器的所提出框架的图示。位姿图中存在若干相机位姿、IMU测量和视觉测量。IMU和其中一个相机是可选的。



# 方法

## A.问题定义

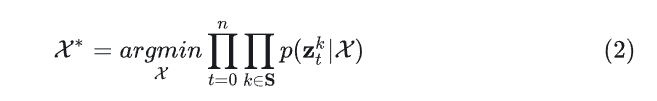
1）状态：我们需要估计的主要状态包括机器人中心的三维位置和姿态。除此之外，我们有其它可选的状态，它们与传感器有关。对于相机，需要估计视觉路标的深度或者三维位置。对于IMU，它产生另一个运动变量，即速度，还需要估计IMU的时变加速度计偏置和陀螺仪偏置。因此，对于视觉和惯性传感器，我们需要估计的整个状态定义如下：



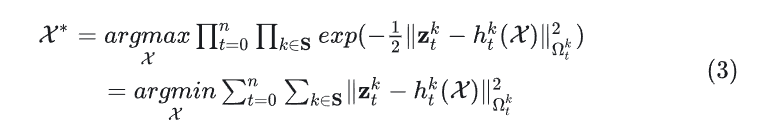
其中，P和R为基本的系统状态,，它们对应于世界坐标系下表示的机体位置和姿态。

为相机相关的状态，它包括第一帧中观测到的每个特征的深度。为IMU相关的变量，它由速度v、加速度计偏置和陀螺仪偏置组成。如果我们仅使用双目相机而不使用IMU，就能够忽略。传感器中心到机体中心的平移假定是已知的，它通过离线标定获得。为了简化符号，我们将IMU表示为机体中心（如果不使用IMU，我们将左相机表示为机体中心）。

2）代价函数：状态估计的本质为一个最大似然估计（MLE）问题。MLE由一段时间内机器人位姿的联合概率分布组成。在所有测量数据是独立的假设下，问题通常推导为：



其中，S为一组测量，它来自相机、IMU和其它传感器。我们假设测量的不确定性服从高斯分布，。因此，上述公式的负似然对数可以写成：

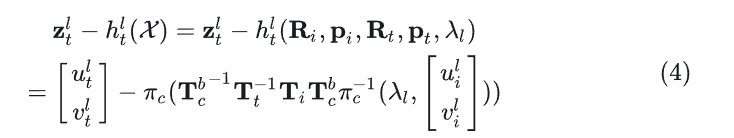


马氏范数被定义为为传感器模型，接着，状态估计被转换为一个非线性最小二乘问题，它也被称为捆集调整（BA）。

## B.传感器因子

1）相机因子：该框架支持单目和双目相机。每个相机的内参和相机之间的外参变换假设是已知的，它们能够容易地通过离线标定获取。对于每个相机帧，提取角点特征。KLT跟踪器在前一帧中跟踪这些特征。对于双目设置，跟踪器还在左图像和右图像之间匹配特征。根据特征关联，我们使用每一帧中的每个特征来构建相机因子。相机因子为重投影过程，其根据一个特征的首次观测将该特征投影到之后的帧中。

考虑图像i中首次观测到的特征l，其（l）在之后图像t中的观测残差定义为：

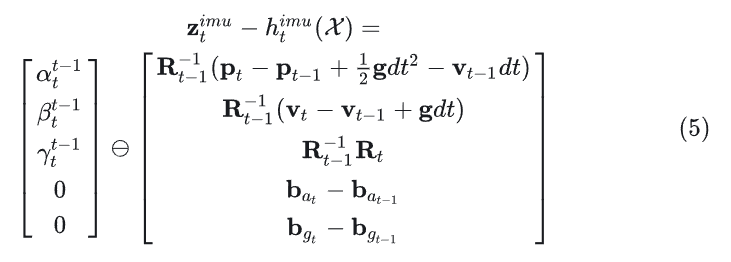


其中，为出现在图像i中的特征l的首次观测。为相同特征l在图像t中的观测。和为投影和反投影函数，它们依赖于相机模型（针孔、全景或者其它模型）。T为4x4的齐次变换矩阵：。为了简化表示，我们忽略了一些齐次项。为从机体中心到相机中心的外参变换，它通过离线标定获取。投影误差的协方差矩阵为像素坐标下的常值，它来自相机的内参标定结果。

该因子对于左相机和右相机都是通用的。我们能够在不同时间的空间中将左图像中的特征投影到另一左图像，同样，我们也能够在同一时间的空间中将左图像中的特征投影到右图像。对于不同的相机，应该使用不同的外参变换。

2）IMU因子：我们使用著名的预积分算法来构建IMU因子。我们假设，加速度计和陀螺仪测量中的附加噪声为高斯白噪声。时变的加速度计和陀螺仪偏置被建模为随机游走过程，其导数为高斯白噪声。由于IMU以比其它传感器更高的频率获取数据，因此两帧之间通常存在多个IMU测量数据。所以，我们在流形上预积分IMU测量，并且传递协方差矩阵。

在两个时刻t−1和t之间，预积分产生相对位置，相对速度和相对旋转。此外，预积分还传递相对位置、相对速度和相对旋转的协方差矩阵以及偏置的协方差矩阵。IMU残差能够被定义为：

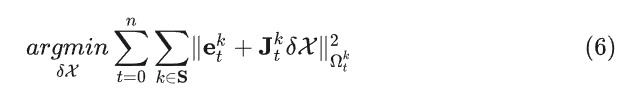


其中，⊖为流形上的减法运算，它专门用于非线性旋转。dt为两个时刻之间的时间间隔。g为已知的重力向量，其模长大约为9.81。每两个相邻帧构建代价函数中的一个IMU因子。

3）其它因子：尽管我们仅说明了相机和IMU因子，但是本文系统不限于这两种传感器。其它传感器（例如轮速计、激光雷达和毫米波雷达）也能够毫不费力地加入到我们的系统中。关键是将这些测量值建模为通用的残差因子，并且将这些残差因子加入到代价函数中（多传感器融合的必备条件）

## C.优化

传统上，公式3的非线性最小二乘问题通过Gauss-Newton或者Levenberg-Marquardt方法求解。代价函数相对于初始状态估计 线性化。接着，代价函数等价于：



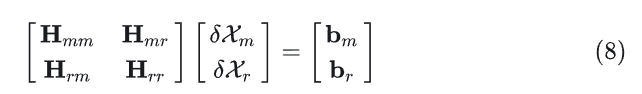
其中，J为每个因子相对于当前状态的雅可比矩阵。在线性化近似之后，该代价函数具有的闭式解。我们以Gauss-Newton为例，推导解如下：



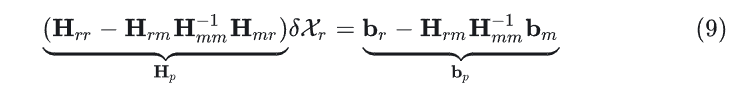
最后，当前状态由⊕更新，其中⊕为流形上旋转的加法运算。该过程迭代若干次，直到收敛。我们采用Ceres求解器来求解该问题，它利用先进的数学工具来高效地获取稳定且最优的结果。

## D.边缘化

由于状态的数量随着时间增加，因此计算复杂度将呈平方增加。为了限制计算复杂度，加入边缘化，而不损失有用的信息。边缘化过程将先前的测量转化为一个先验项，它保留了过去的信息。需要边缘化的一组状态被表示为，剩余的一组状态被表示为。通过累加所有的边缘化因子（公式7），我们获得一个新的H和b。在重新排列状态的顺序之后，我们获得如下关系：



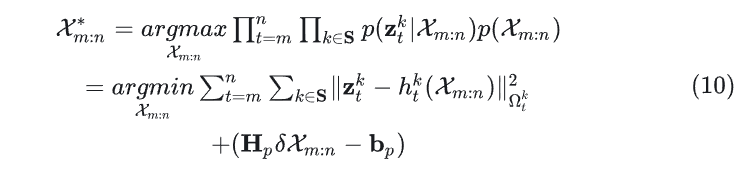
使用舒尔补来执行边缘化，如下：



我们获取剩余状态新的先验Hp、bp。有关边缘化状态的信息被转化为先验项，而没有任何损失。具体而言，我们在系统中保留十个空间相机帧。当传入新的关键帧时，我们边缘化视觉和惯性因子，它们与第一帧的状态有关。

在使用贝叶斯法则获得有关当前状态的先验信息后，计算后验为似然与先验乘积：

。接着，状态估计变成一个最大后验（MAP）问题。注意，我们在滑动窗口内保留从时刻m到时刻n的状态。在时刻m之前的状态被边缘化，并且转化为一个先验项。因此，MAP问题被写成：



与公式3相比，上述公式仅增加一个先验项。它与公式3相同，都是通过Ceres求解器来求解。

E.讨论

所提出系统是一个通用的框架。各种传感器能够被容易地加入到我们的系统中，只要它能够被导出为一个通用的残差因子。由于本文系统不是专门为特定传感器设计的，因此它能够处理传感器失效的情况。当传感器失效时，我们只删除非活跃传感器的因子，并且从其它可选的传感器中加入新的因子。