### 基于通用优化的多传感器局部测距估计框架

摘要：如今，机器人上配备了越来越多的传感器，以提高稳健性和自主能力。我们已经看到各种传感器套件配备在不同平台上，例如地面车辆上的立体摄像机，带有IMU（惯性测量单元）的单眼摄像机，以及带有空中机器人IMU的立体摄像机。尽管过去已经提出了许多用于状态估计的算法，但是它们通常应用于单个传感器或特定传感器套件。其中很少可以用于多种传感器选择。在本文中，我们提出了一种基于通用优化的测距估计框架，它支持多个传感器组。在我们的框架中，每个传感器都被视为一个通用因素。将共享共同状态变量的因子相加在一起以构建优化问题。我们进一步展示了视觉和惯性传感器的通用性，它们形成三个传感器套件（立体相机，带IMU的单目相机和带IMU的立体相机）。我们验证了我们的系统在公共数据集上的性能，以及通过多个传感器的实际实验。将结果与其他最先进的算法进行比较。我们强调我们的系统是一个通用框架，可以在姿势图优化中轻松融合各种传感器。我们的实现是开源的。

### 导言

实时6-DoF（自由度）状态估计是机器人技术的基础技术。准确的状态估计在各种智能应用中起着重要作用，例如机器人探索，自动驾驶，VR（虚拟现实）和AR（增强现实）。我们在这些应用中使用的最常见的传感器是相机。在过去的几十年中，已经提出了大量令人印象深刻的基于视觉的姿态估计算法，例如[1] - [5]。除了相机，IMU是另一种流行的状态估计选择。 IMU可以测量高频下的加速度和角速度，这对于实时应用中的低延迟姿态反馈是必需的。因此，有许多研究工作将视觉和IMU融合在一起，例如[6] - 12]。用于状态估计的另一种流行的传感器是LiDAR。基于LiDAR的方法[13]在局限环境中实现精确的姿态估计。尽管过去已经提出了许多算法，但它们通常应用于单个输入传感器或特定传感器套件。

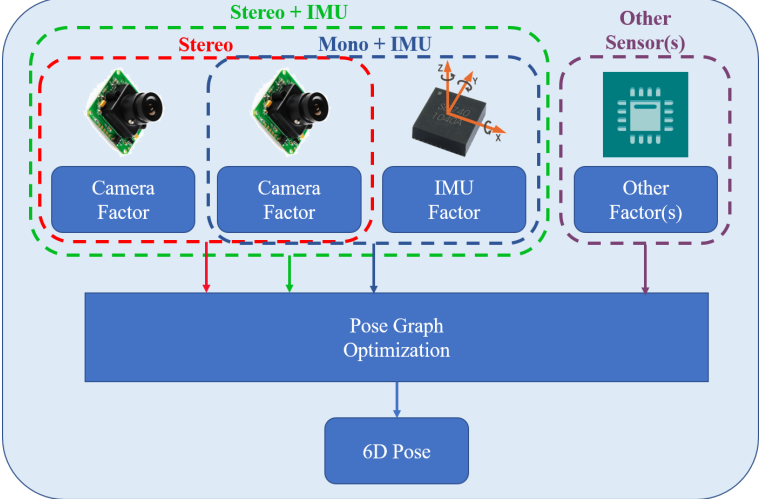


图1 所提出的状态估计框架的示例，其支持多种传感器选择，例如立体相机，具有IMU的单眼相机和具有IMU的立体相机。每个传感器都被视为一般因素。将共享共同状态变量的因子相加在一起以构建优化问题。

最近，我们已经看到配备各种传感器组的平台，例如地面车辆上的立体相机，带有移动电话上的IMU的单眼相机，带有空中机器人上的IMU的立体相机。 但是，由于大多数传统算法是针对单个传感器或特定传感器组设计的，因此无法将它们移植到不同的平台。 即使对于一个平台，我们也需要在不同的场景中选择不同的传感器组合。因此，需要支持不同传感器套件的通用算法。另一个实际要求是，在传感器发生故障的情况下，应移除非活动传感器，并应在系统中快速添加替代传感器。 因此，需要与多个传感器兼容的通用算法。

在本文中，我们提出了一种基于通用优化的姿态估计框架，它支持多种传感器组合。我们使用视觉和惯性传感器进一步演示它，它形成三个传感器套件（立体相机，带有IMU的单目相机和带有IMU的立体相机）。我们可以轻松切换不同的传感器组合。我们强调本文的贡献如下：

* 基于优化的通用状态估计框架，支持多个传感器。
* 用视觉和惯性传感器详细论证状态估计，它们形成不同的传感器套件（立体摄像机，单目摄像机+ IMU，立体摄像机+IMU）。
* 对公共数据集和实际实验的拟议系统进行评估。
* 社区的开源代码

### 相关工作

在过去的几十年中，状态估计一直是一个热门的研究课题。 大量算法专注于精确的6-DoF姿态估计。我们已经看到许多令人印象深刻的方法适用于一种传感器，例如基于视觉的方法[1] - 5]，基于LiDAR的方法[13]，基于RGB-D的方法[14]。和基于事件的方法[15]。 使用单目相机的方法难以实现6-DoF姿势估计，因为无法从单个相机重新覆盖绝对比例。 为了提高可观察性和鲁棒性，将具有互补特性的多个传感器融合在一起。

多传感器功能的方法有两种趋势。一种是基于过滤器的方法，另一种是基于优化的方法。基于滤波器的方法通常通过EKF（扩展卡尔曼滤波器）实现。视觉和惯性测量通常被一起过滤以进行6-DoF状态估计。高速惯性传感器用于状态传播，视觉测量用于[9,16]中的更新。MSCKF [6,7]是一种流行的基于EKF的VIO（视觉惯性测距），它保持了几个摄像机姿势并利用多个摄像机视图来形成多约束更新。基于滤波器的方法通常较早地线性化状态并且受到由不准确的线性点引起的误差的影响。为了克服线性化误差引起的不一致性，提出了可观察性约束的EKF [17]来提高准确性和一致性。在[18]中提出了UKF（Unscented卡尔曼滤波器）算法，其中视觉，Li DAR和GPS测量被融合在一起。英国皇家空军是没有分析雅可比人的EKF的扩张。基于过滤器的方法对时间同步很敏感。任何迟到的测量都会引起麻烦，因为状态不能在过滤程序中传播回来。因此，需要特殊的排序机制来确保来自多个传感器的所有测量都是有序的。

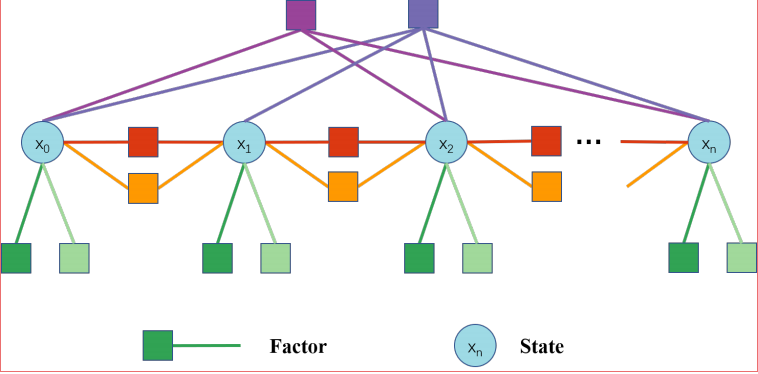


图2.姿势图的图示。 每个节点在一个时刻表示状态（位置，方向，速度等）。每条边代表一个因子，它是通过一次测量得出的。边缘约束一个状态，两个状态或多个状态。

基于优化的方法可以同时维护大量的测量和优化多个变量，这也称为捆绑调整（BA）。与基于滤波器的方法相比，基于优化的方法在时间同步方面具有优势。因为大束用作自然缓冲器，所以当多个传感器的测量结果无序时，它可以轻松处理。基于优化的算法在精度方面也优于基于滤波器的算法，但代价是计算复杂性。早期的优化求解器，如G2O [19]，利用Gauss-Newton和Levenberg-Marquardt方法来解决问题。尽管稀疏结构用于优化求解器，但复杂性随着状态和测量的数量呈二次方增长。为了实现实时性能，一些算法已经探索了增量求解器，而其他算法则限制了姿势图的大小。 iSAM2 [20]是一种有效的增量求解器，它可以重复使用先前的优化结果，以便在新测量结果出现时减少计算量。优化迭代仅更新了一小部分状态而不是整个姿势图。之后，在[21]中提出了一种加速求解器，它通过将密集结构重建为稀疏块来提高效率。保持固定大小的姿势图的方法称为滑动窗口方法。令人印象深刻的基于优化的VIO方法，例如[8,10,12]，在有界大小的滑动窗口上的优化变量。在[8,12]中，先前的状态被边缘化为先前因素而没有丢失信息。在本文中，我们采用基于滑动窗口优化的状态估计框架。

### 系统总览

提出的框架结构如图1所示。多种传感器可以自由组合。 每个传感器的测量被视为一般因子。 因子及其相关状态构成姿势图。 姿势图的图示如图2所示。每个节点在一个时刻表示状态（位置，方向，速度等）。 每条边代表一个因子，它是通过一次测量得出的。 因子限制一个状态，两个状态或多个状态。 对于IMU因子，它通过连续运动限制来约束两个连续状态。 对于视觉地标，其因子约束多个状态，因为它在多个帧上被观察到。构建图形后，优化它等于找到尽可能匹配所有边缘的节点配置。

在本文中，我们专门演示了具有视觉和惯性传感器的系统。视觉和惯性传感器可以形成用于6-DoF状态估计的三种组合，其是立体相机，具有IMU的单眼相机和具有IMU的立体相机。 所提出的具有视觉和惯性传感器的框架的图示如图3所示。姿势图中存在几种相机姿势，IMU测量和视觉测量。 IMU和其中一个摄像头是可选的。

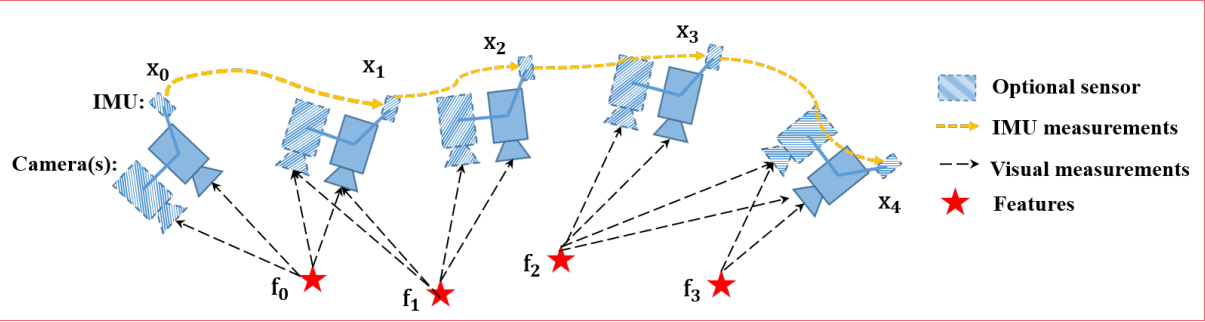


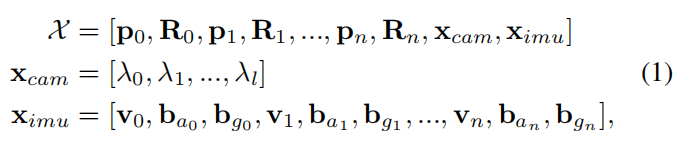
图3.使用视觉和惯性传感器的拟议框架的图示。 IMU和其中一个摄像头是可选的。因此，它形成三种类型（立体相机，具有IMU的单目相机和具有IMU的立体相机）。姿势图中存在多个相机姿势，IMU测量和视觉测量。

### 方法

#### 问题定义

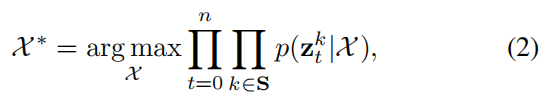
1. 状态：我们需要估计的主要状态包括机器人中心的3D位置和方向。 此外，我们还有其他可选状态，这些状态与传感器有关。对于相机，需要估计视觉地标的深度或3D位置。 对于IMU，它产生另一个运动变量，速度。此外，需要估计IMU的时变加速度偏差和陀螺仪偏差。因此，对于视觉和惯性传感器，我们需要估计的整个状态定义如下：

其中p和R是基本系统状态，对应于世界坐标系中表示的机器人的位置和方向。

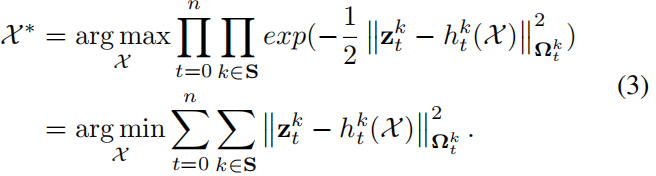


xcam是相机抖动状态，其包括在第一帧中观察到的每个特征的深度λ。ximu是IMU相关变量，由速度v，加速度偏差ba和陀螺仪偏差bg组成。如果我们只使用没有IMU的立体相机，则可以省略ximu。假设从传感器中心到身体中心的转换是已知的，其是离线校准的。为了简化表示法，我们将IMU表示为身体的中心（如果不使用IMU，我们将左侧摄像头表示为身体的中心）。

1. 成本函数：状态估计的性质是MLE（最大似然估计）问题。MLE由一段时间内机器人姿势的联合概率分布组成。在假设所有测量都是独立的情况下，问题通常来源于，



其中S是一组测量值，来自摄像机，IMU和其他传感器。我们假设测量的不确定性是高斯分布的。因此，上述方程的负对数似然性被写为，

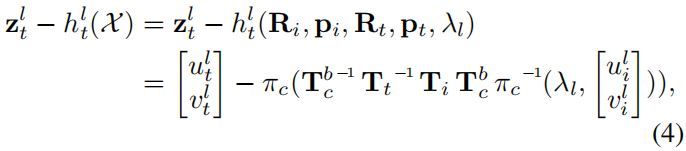


Mahalanobis范数定义为。是传感器模型，将在下一节中详述。然后将状态估计转换为非线性最小二乘问题，其也称为束调整（BA）。

#### 传感器因素

1. 摄像机因素：该框架支持单相和立体摄像机。应该知道每个摄像机的固有参数和摄像机之间的外在变换，这可以容易地离线校准。对于每个相机框架，检测角部特征[22]。KLT跟踪器[23]在前一帧中跟踪这些特征。对于立体声设置，跟踪器还匹配左图像和右图像之间的特征。根据特征关联，我们在每个帧中构造具有每个特征的相机因子。相机因子是重投影过程，它将第一次观察的特征投影到后续帧中。

考虑到在图像i中首次观察到的特征l，在下面的图像t中观察的残差被定义为：



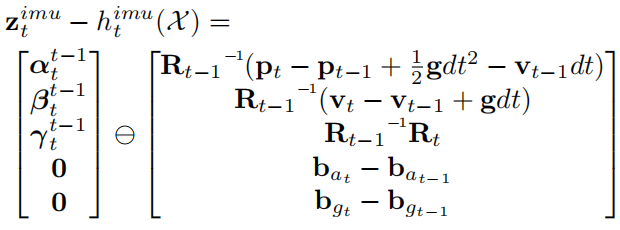
其中是对i图像中出现的l特征的第一次观察。是对t图像中相同特征的观察。和是投影和反投影功能，它们取决于相机模型（针孔，全向或其他模型）。 T是4x4齐次变换，是。我们省略了一些用于简洁表达的同类术语。 是从身体中心到摄像机中心的外在转换，它是离线校准的。重投影误差的协方差矩阵是像素坐标中的恒定值，其来自相机的固有校准结果。

这个因素对于左相机和右相机都是通用的。我们可以在时空空间中投影从左图像到左图像的特征，也可以在空间空间中将特征从左图像投影到右图像。对于不同的相机，应该使用不同的外在变换。

1. IMU因子：我们使用众所周知的IMU预处理算法[11,12]来构建IMU因子。我们假设加速度和陀螺仪测量中的加性噪声是高斯白噪声。时变加速度和陀螺仪偏差被建模为随机游走过程，其导数是高斯白噪声。

由于IMU以比其他传感器更高的频率获取数据，因此通常在两帧之间存在多个IMU测量。因此，我们使用协方差传播在流形上预先集成IMU测量。

详细的预整合可以在[12]找到。在两个时间点t-1和t内，预积分产生相对位置，速度和旋转。此外，预积分传播相对位置，速度和旋转的协方差，以及偏差的协方差。IMU残差可以定义为：

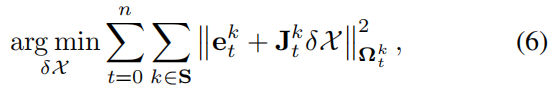


歧管上的专门用于非线性旋转的负操作在哪里。dt是两个时刻之间的时间间隔。 g是已知的重力矢量，其范数约为9.81。每两个相邻的帧在成本函数中构造一个IMU因子。

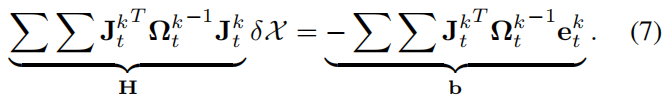
1. 其他因素：虽然我们只指定摄像机和IMU因素，但我们的系统不仅限于这两个传感器。其他传感器，如车轮速度计，LiDAR和雷达，可以毫不费力地添加到我们的系统中。关键是将这些测量值建模为一般残差因子，并将这些残差因子加入成本函数中。

#### 优化

在传统中，方程3的非线性最小二乘问题由Newton-Gaussian或Levenberg-Marquardt方法求解。成本函数相对于状态的初始猜测X线性化。然后，成本函数等于：



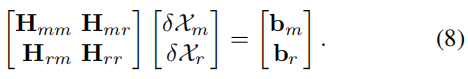
其中J是关于当前状态X的每个因子的雅可比矩阵。在线性化近似之后，该成本函数具有的闭合形式解。我们以牛顿-高斯为例，求解如下，



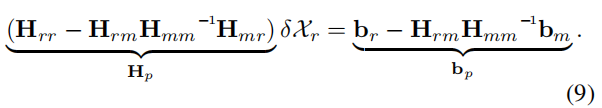
最后，当前状态X用更新，其中⊕是歧管旋转的加号运算。此过程迭代几次直到收敛。我们采用Ceres求解器[24]来解决这个问题，它利用先进的数学工具有效地获得稳定和最优的结果。

#### 边缘化

由于状态数随时间增加，计算复杂度将相应地以二次方式增加。为了限制计算复杂性，边缘化被合并而不会丢失有用的信息。边缘化程序将先前的测量转换为先前的术语，保留过去的信息。被边缘化的状态集合表示为Xm，剩余状态集合表示为Xr。通过对所有边际化因子（方程式7）求和，我们得到一个新的H和b。重新排列州的顺序后，我们得到以下关系：

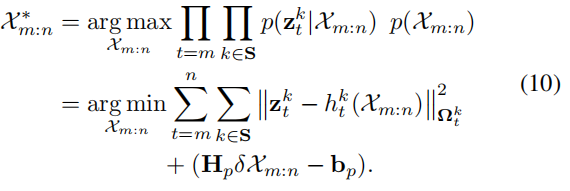


边缘化是使用Schur补充[25]进行的，如下所示：



我们为剩余的状态获得了新的先前。关于边缘化状态的信息被转换为先前条件而没有任何损失。具体而言，我们在系统中保留了十个空间相机框架。当一个新的关键帧到来时，我们将视觉和惯性因子边缘化，这与第一帧的状态有关。

在得到关于当前状态的先验信息之后，利用贝叶斯规则，我们可以将后验计算为似然和先验的乘积：。然后状态估计变为MAP（最大后验）问题。表示我们在滑动窗口中保持从瞬间m到瞬间n的状态。m之前的状态被边缘化并转换为前一个任期。因此，MAP问题写成：



与等式3相比，上述等式仅增加了先验项。它与Ceres求解器[24]的方程3相同。

#### 讨论

拟议的系统是一个总体框架。可以将各种传感器轻松添加到我们的系统中，只要它可以作为一般残留因子得出。由于我们的系统不是专门为某个传感器设计的，因此能够处理传感器故障情况。当传感器发生故障时，我们只需删除非活动传感器的因子，并添加其他替代传感器的新因子。

### 实验结果

我们在数据集和实际实验中使用视觉和惯性传感器评估所提出的系统。在第一个实验中，我们将提出的算法与公共数据集上的另一个最先进的算法进行比较。然后，我们在大型户外环境中测试我们的系统。生成数值分析以详细显示我们系统的准确性。

#### 数据集

我们使用EuRoC MAV Visual-Inertial Datasets评估我们提出的系统[27]。这个数据集是在微型飞行器上收集的，其中包含立体图像（Aptina MT9V034全局快门，752x480单色，20 FPS），同步IMU测量（ADIS16448,200 Hz）。此外，地面真实状态由VICON和Leica MS50提供。我们使用三种不同的传感器组合运行数据集，这些传感器是立体相机，带有IMU的单目相机，带有IMU的立体相机。

在这个实验中，我们将我们的结果与OKVIS [8]进行了比较，OKVIS是一种最先进的VIO，可与立体相机和IMU配合使用。 OKVIS是另一种基于优化的滑动窗口算法。 OKVIS专为视觉惯性传感器而设计，而我们的系统是一个更通用的框架，支持多种传感器组合。我们使用EuRoC数据集中的所有序列测试了所提出的框架和OKVIS。我们通过RPE（相对姿势误差）和ATE（绝对轨迹误差）评估准确度。 RPE由[26]中提出的工具计算。两个序列，MH 05难和V2 02培养基的RPE（相对姿势误差）图分别显示在图4和图5中。

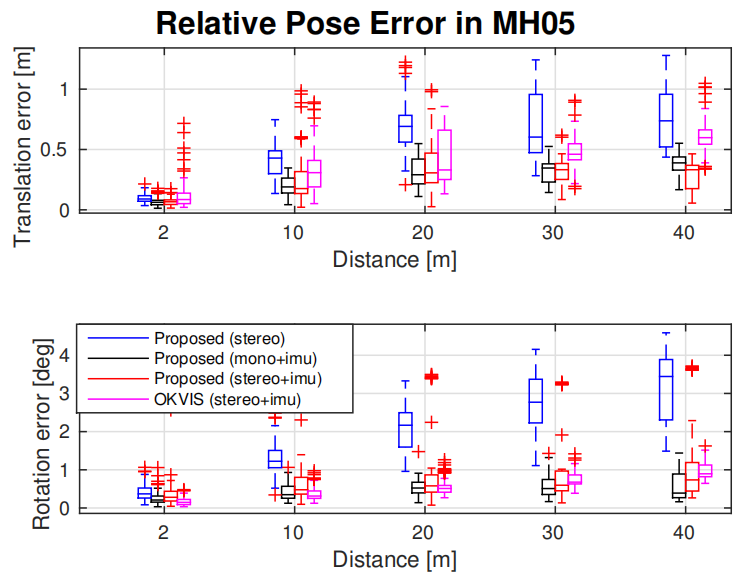


图4. MH 05中的相对姿势误差[26]很难。两个图分别是平移和旋转中的相对误差。

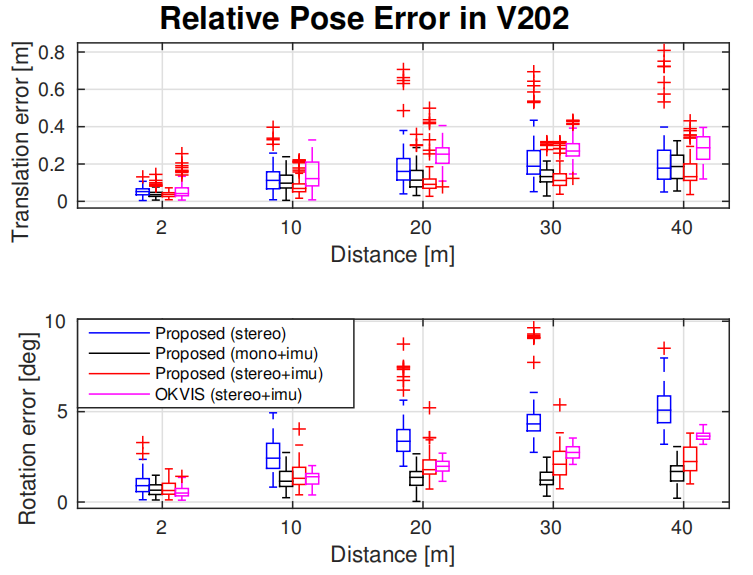
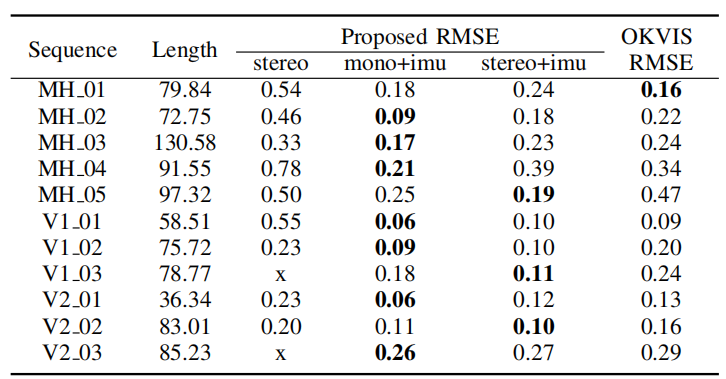


图5. V2 02培养基中的相对姿势误差[26]。两个图分别是平移和旋转中的相对误差。

TABLE I

RMSE [M]在EUROC数据集中。



EuRoC数据集中所有序列的ATE的RMSE（均方根误差）如表所示。 I.通过霍恩的方法[28]，估计的轨迹与基本事实一致。只有立体声的情况在V1 03困难和V2 03困难序列中失败，其中运动过于激进以使视觉跟踪无法生存。涉及IMU在所有序列中成功工作的方法。这是一个很好的例子，表明IMU可以通过在视觉轨迹由于光照变化，纹理减少区域或运动模糊而失败时缩小间隙来显着改善运动跟踪性能。

从相对姿态误差和绝对轨迹误差，我们可以看出，仅立体视觉方法在大多数序列中表现最差。在Stereo vision情况下，位置和旋转漂移明显随距离而增大。换句话说，IMU在状态估计中显着地有利于视觉。由于IMU测量重力矢量，它可以有效地抑制侧倾角和俯仰角的漂移。具有IMU的立体摄像机并不总是表现最佳，因为它需要比具有IMU的单目摄像机更精确的校准。不准确的内在和外在校准将在系统中引入更多噪声。通常，多传感器融合增加了系统的稳健性。在大多数序列中，我们的结果优于OKVIS。

#### 真实世界的实验

在这个实验中，我们使用自行开发的传感器套件来演示我们的框架。传感器套件如图6所示。它包含立体声摄像机（mvBlueFOX-MLC200w，20Hz）和DJI A3控制器2，内置IMU（200Hz）和GPS接收器。 GPS位置被视为基本事实。我们手持传感器套件并在室外的地面上走路。我们使用三种不同的组合运行状态估计，这些组合是立体相机，具有IMU的单目相机和具有IMU的立体相机。

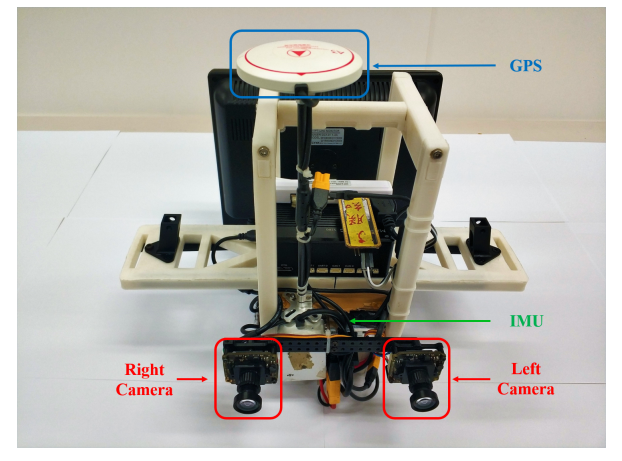


图6.自行开发的室外环境中使用的传感器套件。它包含立体声摄像机（mvBlueFOX-MLC200w，20Hz）和DJI A3控制器，包括内置IMU（200Hz）和GPS接收器。

为了进行准确性比较，我们在地面上走了两圈并将我们的估算与GPS进行了比较。轨迹如图7所示，RPE（相对姿态误差）如图8所示。与数据集实验相同，在仅立体场景中发生了明显的位置漂移。在IMU的帮助下，准确性提高了很多。表II中显示了更多室外实验的RMSE。涉及IMU的方法总是比仅立体声的情况更好。

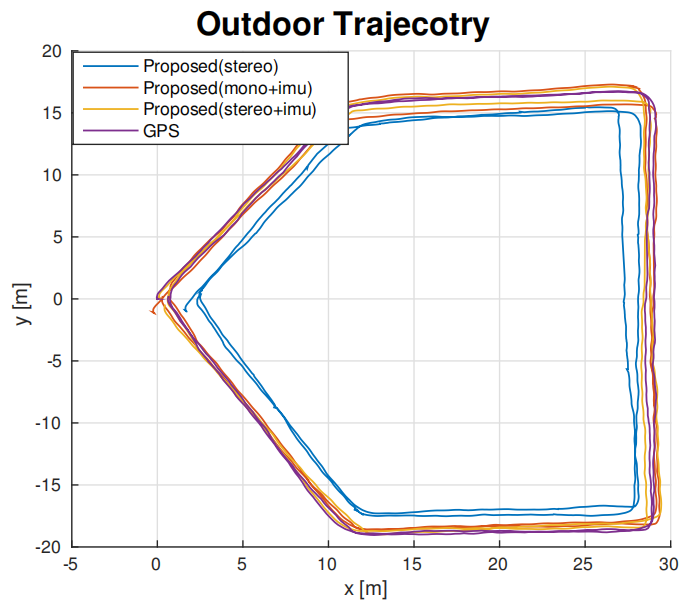
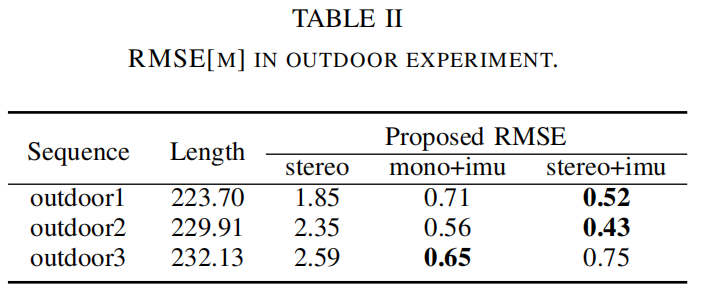


图7.室外实验的估计轨迹



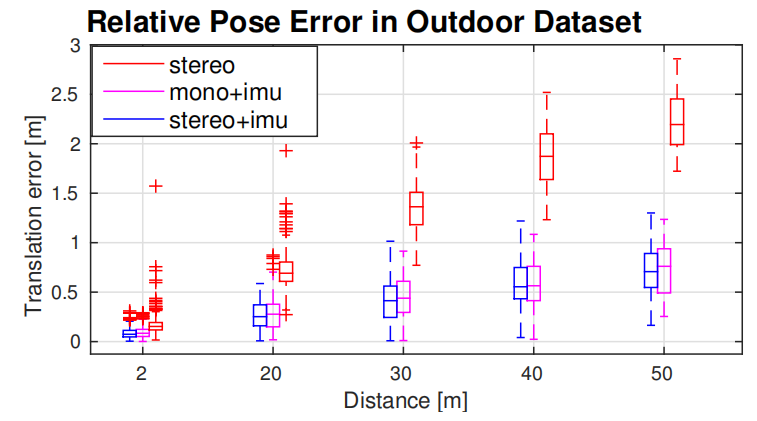


图8.室外实验中的相对姿态误差[26]。

### 结论

在本文中，我们提出了一种基于通用优化的局部姿态估计框架。所提出的框架可以支持多个传感器组合，这在鲁棒性和实用性方面是期望的。我们使用视觉和惯性传感器进一步演示它，它形成三个传感器套件（立体相机，带有IMU的单目相机和带有IMU的立体相机）。请注意，虽然我们只展示了相机和IMU的因子公式，但我们的框架也可以推广到其他传感器。我们在公共数据集和实际实验中使用多个传感器验证我们系统的性能。数值结果表明我们的框架能够融合不同设置的传感器数据。

在未来的工作中，我们将使用全球传感器（例如GPS）扩展我们的框架，以实现局部准确和全局感知的姿势估计。