# VINS-Mono：一种强大而通用的单眼视觉惯性状态估计器

**摘要**-一个摄像头和一个低成本惯性测量单元（IMU）构成一个单目视觉惯性系统（VINS），它是公制六自由度的最小传感器套件（尺寸，重量和功率）（DOF）状态估计。在这篇文章中，我们介绍了VINS-Mono：一种强大而通用的单目视觉惯性状态估计器。我们的方法从用于估计器初始化的稳健过程开始。基于紧密耦合的非线性优化方法用于通过融合预集成的IMU测量值和特征观察来获得高度精确的视觉惯性测距。环路检测模块与我们的紧密耦合配方相结合，可实现最小计算的重新定位。我们另外执行4DOF姿势图优化以强制实现全局一致性。此外，所提出的系统可以通过以有效的方式保存和加载地图来重用地图。可以通过全局姿势图优化将当前和先前的映射合并在一起。我们在公共数据集和现实世界的实验中验证了我们系统的性能，并与其他最先进的算法进行了比较。我们还在微型车辆平台上执行机载闭环自主飞行，并将算法移植到基于iOS的演示中。我们强调，所提出的工作是一种可靠，完整且通用的系统，适用于需要高精度定位的不同应用。我们开源我们的两台PC（https://github.com/HKUST-Aerial-Robotics/VINS?Mono）和iOS移动设备（https://github.com/HKUST-Aerial?Robotics/VINS-Mobile）的实现。

**索引术语**-单目视觉惯性系统（VINS），状态估计，传感器融合，同时定位和映射。

## 一、导言

状态估计无疑是广泛应用的最基本模块，例如机器人导航，自动驾驶，虚拟现实和增强现实（AR）。仅使用单目相机的方法由于尺寸小，成本低且易于硬件设置而在该领域引起了极大的兴趣[1]-[5]。然而，单眼视觉系统无法恢复公制尺度，因此限制了它们在现实世界机器人应用中的使用。最近，我们看到了通过低成本惯性测量单元（IMU）协助单眼视觉系统的增长趋势。这种单目视觉惯性系统（VINS）的主要优点是观察公制尺度，以及滚转角和俯仰角。这样可以启用需要度量标准状态估计的导航任务。此外，IMU测量的集成可以通过弥合由于照明变化，无纹理区域或运动模糊导致的视觉轨迹损失之间的差距而显着改善运动跟踪性能。单眼VINS不仅可以在地面机器人和无人机上广泛使用，而且在移动设备上也是可行的。它在尺寸，重量和功耗方面具有很大的优势，可以自我和环境感知。

但是，有几个问题会影响单眼VINS的使用。第一个是严格的初始化。由于缺乏直接距离测量，很难直接融合单个微观视觉结构和惯性测量。同时认识到VINS是高度非线性的这一事实，我们在估算器初始化方面看到了重大挑战。在大多数情况下，系统应该从一个已知的静止位置启动，并在开始时缓慢而小心地移动，这限制了它在实践中的使用。另一个问题是视觉惯性测距（VIO）的长期漂移是不可避免的。为了消除漂移，必须开发环路检测，重定位和全局优化。除了这些关键因素之外，对地图保存和重用的需求也在增长。

为了解决所有这些问题，我们提出了VINS-Mono，一种强大而通用的单目视觉惯性状态估计器，它是我们之前三个作品的组合和扩展[6]-[8]。 VINS-Mono包含以下功能：

1. 健壮的初始化过程，能够从未知的初始状态引导系统;
2. 紧密耦合，基于优化的单眼VIO，具有相机-IMU外部校准和IMU偏差校正;
3. 在线重定位和四自由度（DOF）全局姿态图优化;
4. 姿势图重用，可以保存，加载和合并多个局部姿势图。

在这些功能中，强大的初始化，重定位和姿势图重用是我们的技术贡献，来自我们以前的工作[6]-[8]。工程贡献包括开源系统集成，无人机导航的实时恶魔和移动应用。该整个系统已成功应用于小规模AR场景，中型无人机导航和大规模状态估计任务，如图1所示。

本文的其余部分结构如下。在第二节中，我们讨论相关文献。我们在第III节中概述了完整的系统管道。视觉和预集成IMU测量的预处理步骤在第IV节中介绍。在第五节中，我们讨论了估计器初始化过程。第VI节介绍了一种紧耦合，自校准，基于非线性优化的单眼VIO。第VII部分介绍了紧密耦合的重新定位。第VIII节讨论了全局姿态图优化和重用。实验结果显示在第IX节中。最后，本文结束了第十节的讨论和未来可能的研究方向。

## 相关工作

## 概述

所提出的单目视觉惯性状态估计器的结构如图2所示。系统从测量预处理开始（见第IV节），其中提取和跟踪特征，并且预先集成两个连续帧之间的IMU测量。初始化过程（参见第V节）提供了所有必要的值，包括姿态，速度，重力矢量，陀螺仪偏差和三维（3-D）特征位置，用于引导随后的基于非线性优化的VIO。具有重新定位功能的VIO（参见第VI节）（参见第VII节）模块紧密融合了预集成的IMU测量，功能观察结果。最后，姿势图优化模块（参见第VIII节）采用几何验证的重定位结果，并执行全局优化以消除漂移。它还实现了姿势图重用。VIO和姿势图优化模块在分离的线程中并发运行。

与OKVIS [15]相比，OKVIS [15]是一种适用于立体摄像机的最先进的VIO算法，我们的算法专为单目摄像机而设计。因此，我们特别提出初始化过程，关键帧选择标准，以及使用和处理大视场（FOV）摄像机以获得更好的跟踪性能。此外，我们的算法提出了一个完整的系统，其中包含循环闭包和姿势图重用模块。

我们现在定义我们在本文中使用的符号和框架定义。我们将视为世界坐标系。重力方向与世界坐标系的z轴对齐。是载体坐标系，我们将其定义为与IMU坐标系相同。是相机框架。我们使用旋转矩阵R和Hamilton四元数q来表示旋转。我们主要在状态向量中使用四元数，但是旋转矩阵也用于方便地旋转3-D向量。和是从载体坐标系到世界框架的旋转和平移。是拍摄第k个图像时的载体坐标系。是拍摄第k张图像时的相机坐坐标系。⊗表示两个四元数之间的乘法运算。是世界坐标系中的重力矢量。最后，我们将表示为某一数量的噪声测量或估计。

## 四、测量预处理

本节介绍惯性和单眼视觉测量的预处理步骤。对于视觉测量，我们跟踪连续帧之间的功能并检测最新帧中的新功能。对于IMU测量，我们在两个连续帧之间预先设置它们。

1. 视觉处理前端

对于每个新图像，由KLT稀疏光流算法[31]对现有特征进行跟踪。同时，检测新的角特征[32]以维持每个图像中的最小数量（100-300）的特征。检测器通过设置两个相邻特征之间的像素的最小间隔来实施统一的特征分布。二维（2-D）特征首先不失真，然后在通过异常值拒绝后投射到单位球面。使用RANSAC和基本矩阵模型进行异常值拒绝[33]。

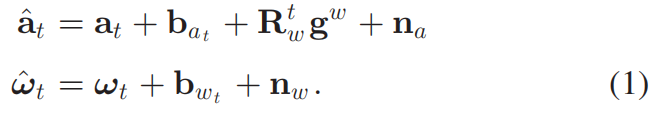
在此步骤中也选择了关键帧。我们有两个关键帧选择标准。第一个是除前一个关键帧之外的平均视差。如果跟踪特征的平均视差在当前帧和最新关键帧之间超出某个阈值，我们将帧视为新的关键帧。请注意，不仅平移而且旋转都会导致视差。但是，在仅旋转运动中不能对特征进行三角测量。为避免这种情况，我们在计算视差时使用陀螺仪测量的短期积分来补偿旋转。请注意，此旋转补偿仅用于关键帧选择，不参与VINS公式中的旋转计算。为此，即使陀螺仪包含大噪声或偏置，也只会导致关键帧选择结果不理想，并且不会直接影响估计质量。另一个标准是跟踪质量。如果跟踪要素的数量低于某个阈值，我们会将此帧视为新的关键帧。该标准是为了避免完全丢失特征轨迹。

1. IMU预整合

我们遵循以前基于连续时间四元数的IMU预整合推导[16]，并将IMU偏差的处理包括在内[19]和[24]。我们注意到，我们目前的IMU预整合程序与[19]和[24]几乎相同，但使用不同的推导。

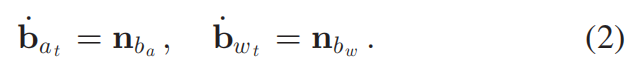
所以，我们这里只做一个简短的介绍。有关基于四元数的推导的详细信息，请参见附录A.

1. IMU噪声和偏置：IMU测量，在载体坐标系中测量，结合了抵抗重力和平台动力学的力量，并受加速度偏差，陀螺仪偏差和加性噪声的影响。原始陀螺仪和加速度计测量值和由下式给出

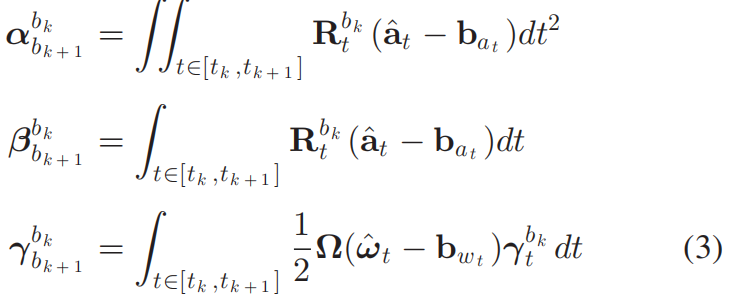


我们假设加速度和陀螺仪测量中的加性噪声是高斯白噪声，。加速度偏差和陀螺仪偏差被建模为随机游走，其导数是高斯白噪声

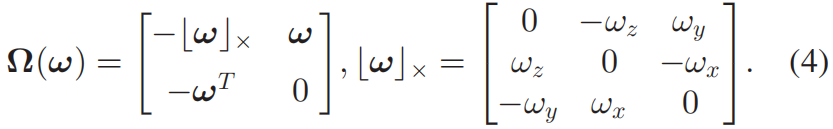




1. 预积分：对于两个连续帧和，在时间间隔中存在几个惯性测量。给定偏差估计，我们将它们集成在本地帧（拍摄第k个图像时的载体坐标系）中。

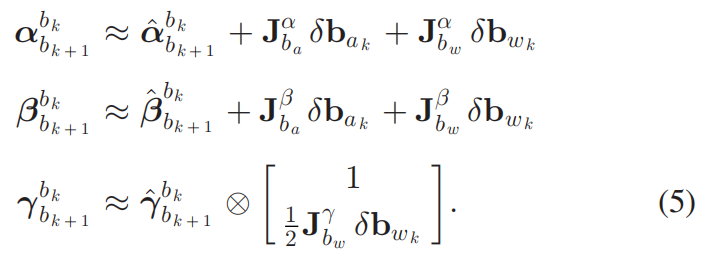


其中：



α，β和γ的协方差也相应地传播。可以看出，通过将作为给定偏差的参考帧，可以仅通过IMU测量获得预积分项（3）。

1. 偏差校正：如果偏差的估计微小变化，我们和调整为偏差的一阶近似值



否则，当偏差的估计发生显着变化时，我们会在新的偏差估计下进行重新传播。这个策略为基于优化的算法节省了大量的计算资源，因为我们不需要重复地进行IMU测量。

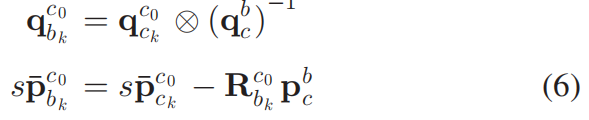
## 五、估计初始化

单目紧耦合VIO是一种高度非线性系统，需要在开始时进行准确的初始猜测。我们通过将IMU预整合与仅视觉结构松散地对齐来获得必要的初始值。

1. 滑动窗口中的仅视觉SfM

初始化过程从仅视觉SfM开始，以估计高达相机姿势和特征位置的图形。

我们在滑动窗口中维护几个帧以获得有限的计算复杂性。首先，我们检查最新帧和所有先前帧之间的特征对应关系。如果我们能够在最新帧和滑动窗口中的任何其他帧之间找到稳定的特征跟踪（超过30个跟踪特征）和足够的视差（超过20个像素）。我们使用五点算法恢复这两个帧之间的相对旋转和按比例转换[34]然后，我们直接设置比例并对这两个帧中观察到的所有特征进行三角测量。基于这些三角测量特征，执行透视点（PnP）方法[35]以估计窗口中所有其他帧的姿势。最后，应用全局完整束调整[36]以最小化所有特征观察的总重投影误差。由于我们还没有关于世界框架的任何知识，我们将第一个相机坐标系设置为SfM的参考坐标系。所有帧姿势和特征位置都相对于表示。给定相机和IMU之间的外部参数，我们可以将姿势从相机帧转换为主体（IMU）帧。



其中是未知缩放参数，将在下一个解决。

1. 视觉-惯性对准

视觉-惯性对准的图示如图3所示。基本思想是将符合规模的视觉结构与IMU预集成相匹配。

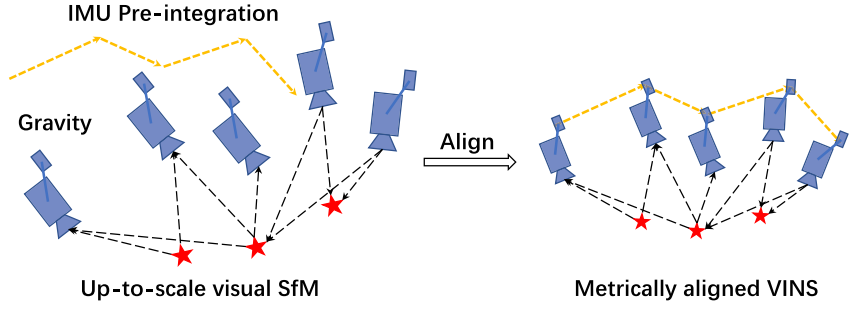
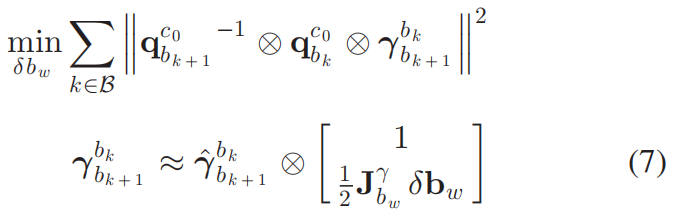


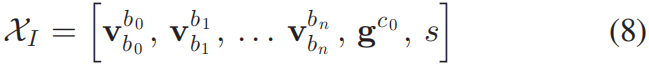
图3.估算器初始化的视觉惯性对齐过程的图示。基本思想是将符合规模的视觉结构与IMU预整合相匹配。

1）陀螺仪偏差校准：在窗口中考虑两个连续帧和，我们从视觉SfM得到旋转和，以及从IMU预积分得到相对约束。我们根据陀螺仪偏差线性化IMU预编程项，并最小化以下成本函数：



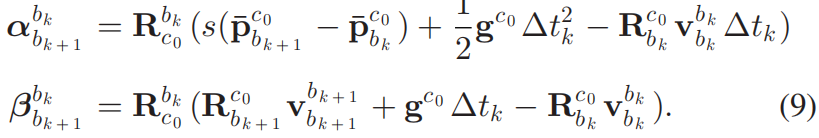
其中为索引窗口中的所有帧。通过这种方式，我们得到了陀螺仪偏置的初始校准。然后，我们使用新的陀螺仪偏差重新传播所有IMU预积分项和。

1. 速度，重力矢量和公制尺度初始化：陀螺仪偏差初始化后，我们继续初始化导航的其他基本状态，即速度，重力向量和度量标度。

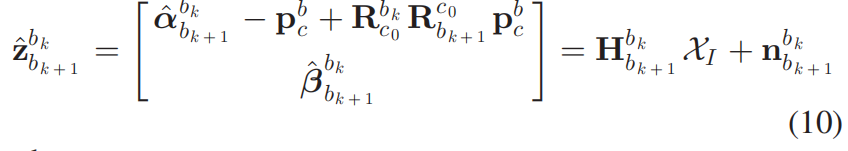


其中是在取第k个图像时载体坐标系中机器人的速度，是帧中的重力矢量，将单眼SfM缩放到公制单位。

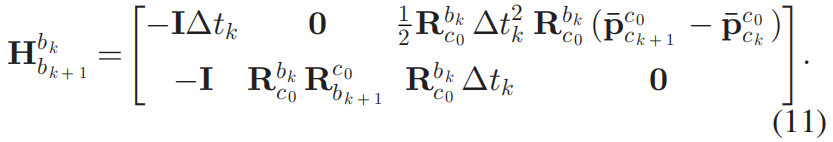
考虑窗口中的两个连续帧和，我们有以下等式：



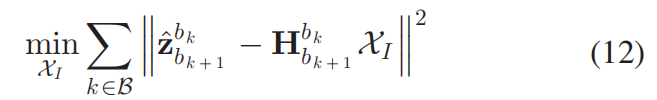
我们可以将（6）和（9）组合成以下线性测量模型：



其中，



可以看出，和是从符合规模的单眼视觉SfM获得的。是两个连续帧之间的时间间隔。通过解决这个线性最小二乘问题：



我们可以得到窗口中每一帧的载体帧速度，视觉参考系中的重力矢量，以及比例参数。

1. 重力细化：可以通过约束幅度来改进从先前线性初始化步骤获得的重力矢量。在大多数情况下，重力矢量的大小是已知的。这导致重力矢量仅剩余2-DOF。因此，我们在其切线空间上以两个变量扰动重力，这保留了2自由度。我们的引力向量受到扰动，，其中是已知的重力大小，是代表重力方向的单位向量。和是跨越切平面的两个正交基，如图4所示.和分别是朝向和的2-D扰动。我们可以任意地找到旋转切线空间的任何一组和。

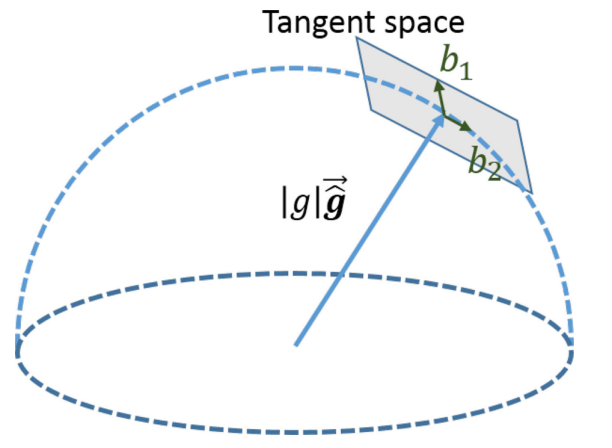


图4. 2-DOF重力扰动的图示。由于重力的大小是已知的，位于半径的球体上。重力在当前估计值附近被扰动为，其中和是跨越切线空间的两个正交基。

然后，我们用将代入（9），并将2D与其他状态变量一起求解。这个过程重复几次，直到g收敛。

4）完成初始化：在精炼重力矢量之后，我们可以通过将重力旋转到z轴来获得世界坐标系和相机坐标系之间的旋转。然后，我们将所有变量从参考帧旋转到世界帧。车身框架速度也将旋转到世界框架。视觉SfM中的平移组件将按比例缩放为公制单位。此时，初始化过程完成，所有这些度量值将被馈送到紧密耦合的单眼VIO。

## 六、紧密耦合的单声道VIO

在估计器初始化之后，我们继续使用基于滑动窗口的紧耦合单眼VIO，以实现高精度和稳健状态估计。滑动窗口配方的图示如图5所示

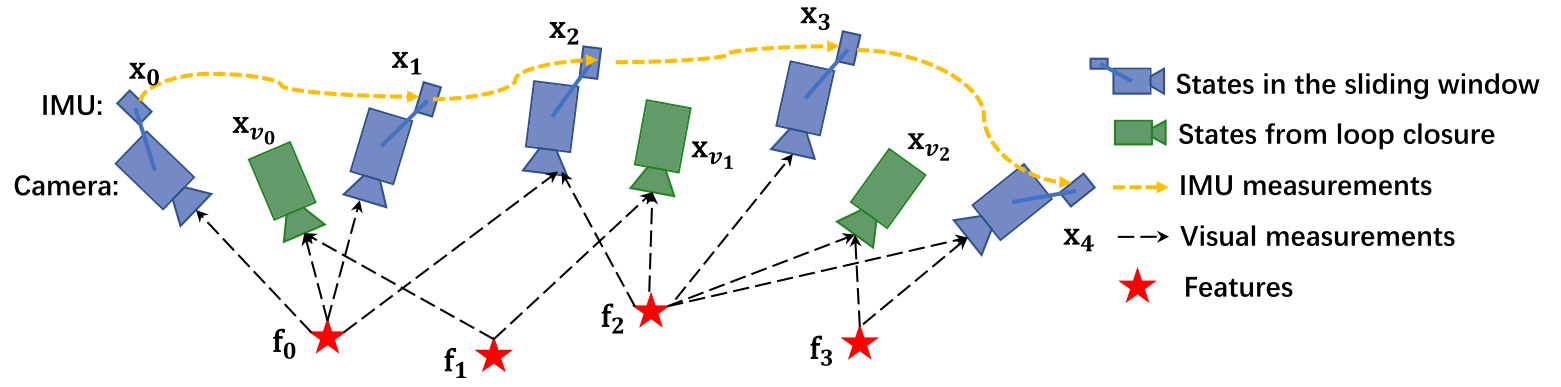
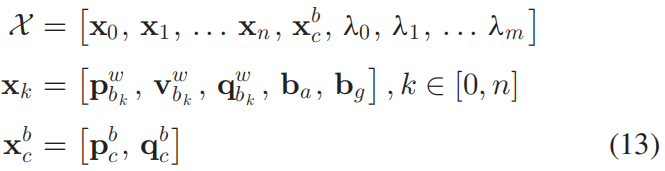


图5.具有重定位的滑动窗单眼VIO的图示。滑动窗口中存在多个相机姿势，IMU测量和视觉测量。它是一种紧密耦合的配方，具有IMU，视觉和环路测量。

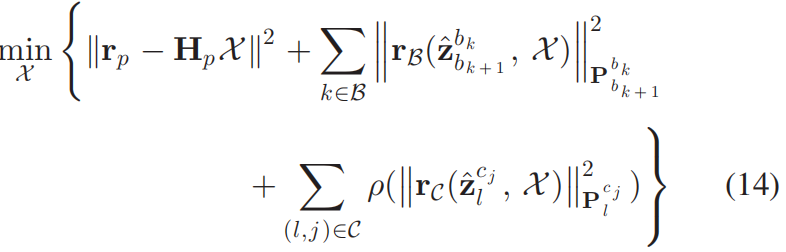
1. Formulation

滑动窗口中的完整状态向量定义为:

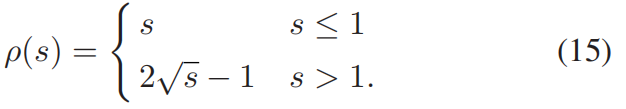


其中是捕获第k个图像时的IMU状态。它包含IMU在世界坐标系中的位置、速度和方向，以及IMU载体坐标系中的加速度偏差和陀螺仪范围偏差。 n是关键帧的总数，m是滑动窗口中的要素总数。是第一个特征与其第一个观测值的反距离。

我们使用视觉惯性束调整方法。我们将所有测量残差的先验和Mahalanobis范数之和最小化，以获得最大后验估计值。



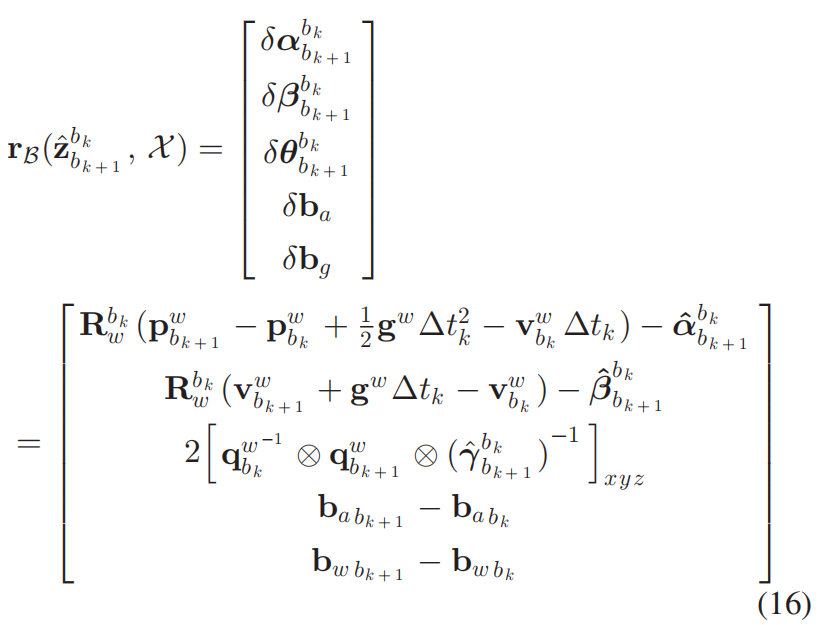
Huber规范[37]定义为



其中和分别是IMU和视觉测量的残差。剩余术语的详细定义将在第VI-B和VI-C节中介绍。B是所有IMU测量的集合，C是在当前滑动窗口中至少观察过两次的特征集。是边缘化的先验信息。 Ceres求解器[38]用于解决这个非线性问题。

1. IMU测量残差

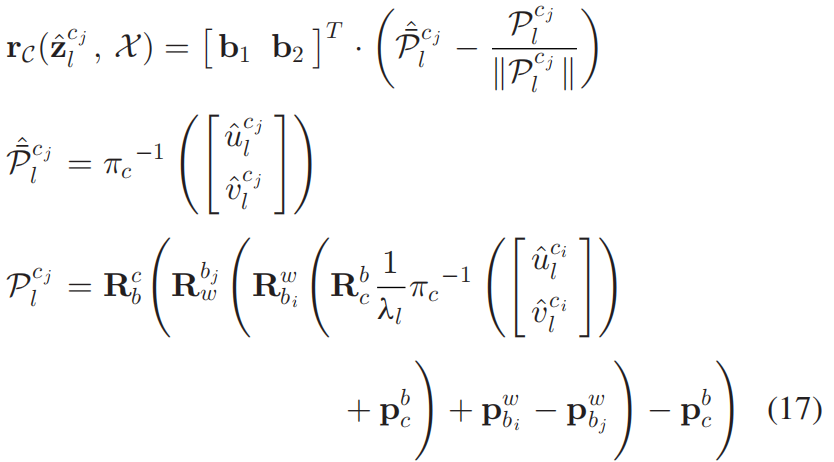
考虑滑动窗口中两个连续帧和内的IMU测量值，预集成IMU测量值的残差可定义为



其中提取四元数q的向量部分以用于错误状态表示。是四元数的3-D错误状态代表。是两个连续图像帧之间的预编辑IMU测量项。加速度计和陀螺仪偏差也包含在在线校正的剩余项中。

1. 视觉测量残差

与在广义图像平面上定义重投影误差的传统针孔相机模型相比，我们在单位球面上定义相机测量残差。几乎所有类型的摄像机（包括广角摄像机，鱼眼摄像机或全向摄像机）的光学系统都可以建模为单位光线，用于连接单位球体的表面。考虑在第i个图像中首次观察到的第l个特征，第j个图像中的特征观察的残差被定义为



其中是第i个图像中发现的个l特征的观察结果（像素坐标）。是对第j个图像中相同特征的观察。是反投影功能，其使用相机固有参数将像素位置转换为单位矢量。由于视觉残差的自由度是2，我们将残差矢量投影到切平面上。和是跨越的切平面的两个任意选择的正交基，如图6所示。如在（14）中使用的方差也从像素坐标传播到单位球上。

1. 边缘化

为了限制基于优化的VIO的计算复杂性，结合了边缘化。我们选择性地从滑动窗口边缘化IMU状态xk和特征λl，同时将对应于边缘化状态的测量值转换为先验状态。

如图7所示，当第二个最新帧是关键帧时，它将停留在窗口中，并且最旧的帧被其边缘化以及其相应的测量值。其他方面，如果第二个最新帧是非关键帧，我们会抛出可视测量并保持连接到此非关键帧的IMU测量。为了保持系统的稀疏性，我们不会将非关键帧的所有度量标准边缘化。我们的边缘化计划旨在保持窗口中空间隔离的关键帧。这确保了特征三角测量的足够视差，并且最大化了主要维持加速度计测量的大激励的概率。

边缘化是使用Schur补充[39]进行的。我们基于与移除状态相关的所有边缘化测量来构建新的先验。新的先验被添加到现有的先前。

我们注意到边缘化会导致林化点的早期修复，这可能导致估计结果不理想。然而，由于小型漂移对VIO来说是可以接受的，我们认为边缘化造成的负面影响并不重要。

1. 用于摄像机速率状态估计的仅运动视觉惯性优化

对于具有低计算能力的设备，例如移动电话，由于非线性优化的大量计算需求，紧耦合的单眼VIO无法实现相机速率输出。为此，除了完全优化之外，我们采用轻量级运动视觉惯性优化来将状态估计提升到摄像机速率（30 Hz）。

仅运动视觉惯性优化的成本函数与（14）中单眼VIO的成本函数相同。然而，我们只是优化固定数量的最新IMU状态的姿势和速度，而不是优化滑动窗口中的所有状态。我们将特征深度，外部参数，偏差和旧IMU状态视为我们不希望优化为常数值。我们确实将所有视觉和惯性测量用于仅运动优化。这导致比单帧PnP方法更平滑的状态估计。图8显示了所提出策略的示意图。与完全紧密耦合的单眼VIO相比，它可以在最先进的嵌入式计算机上产生超过50毫秒的时间，仅运动的视觉惯性优化只需要大约5毫秒来计算。这使得低延迟相机速率姿态估计能够对无人机和AR应用特别有益。

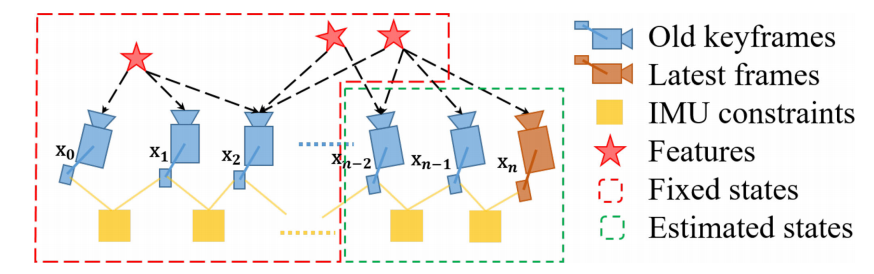


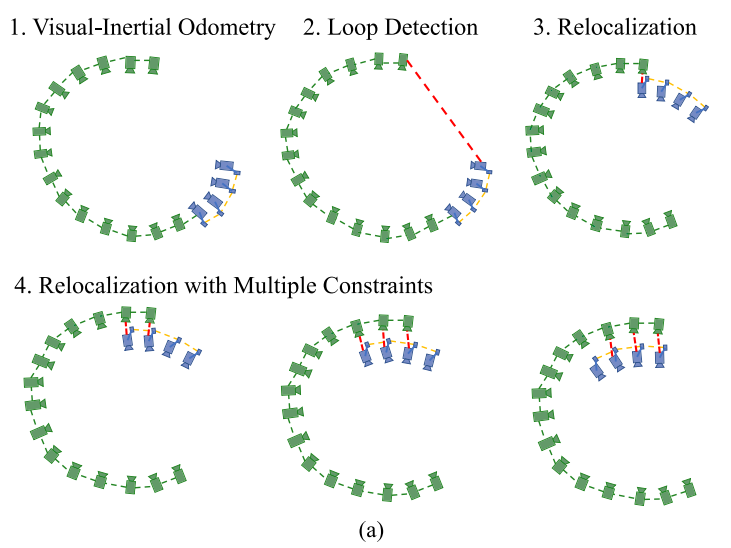
图8.相机速率输出的仅运动优化示意图。

1. 用于IMU速率状态估计的IMU前向传播

IMU测量的速度远远高于视觉测量。虽然我们的VIO频率受到图像捕获频率的限制，但我们仍然可以通过最近的IMU测量直接传播最佳VIO估计值，以实现IMU速率性能。高频状态估计可以用作闭环闭合的状态反馈。利用这种IMU速率状态估计的自主飞行试验见第IX-C节。

## 七、重定位

我们的滑动窗口和边缘化方案限制了计算的复杂性，但它也引入了系统的累积漂移。为了消除漂移，提出了一种与单眼VIO无缝集成的紧密耦合的再生模块。重定位过程从一个循环检测模块开始，该模块识别已经访问过的地方。然后建立循环闭合候选日期和当前帧之间的特征级连接。这些特征对应关系紧密集成到单眼VIO模块中，从而以最小的计算量实现无漂移状态估计。多个特征的多个观察被直接用于重定位，从而导致更高的准确度和更好的状态估计平滑度。重定位过程的图解说明如图9（a）所示。



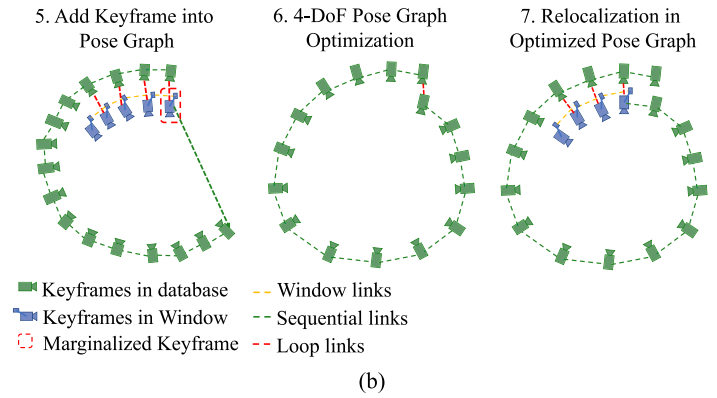


图9.说明重定位和姿势图优化程序的图。 （a）重新定位程序。它从仅VIO姿势估计（蓝色）开始。记录过去的州（绿色）。如果检测到最新关键帧的循环（参见第VII-A节），如第二个图中的红线所示，则发生重新定位。请注意，由于使用了重定位的特征级别对应，我们能够合并来自多个过去关键帧的循环闭包约束（参见第VII-C节），如最后三个图中所示。 （b）全局姿势图优化。当从滑动窗口边缘化时，关键帧被添加到姿势图中。如果在此关键帧和任何其他过去的关键帧之间存在循环，则闭环约束（公式化为4-DOF相对刚体变换）也将添加到姿势图中。在单独的线程中使用所有相对姿势约束（参见第VIII-C节）优化姿势图，并且重定位模块始终针对最新的姿势图配置运行。

1. 循环检测

我们利用DBoW2 [29]，这是一种最先进的词袋位置识别方法，用于循环检测。除了用于单眼VIO的角色功能之外，还有Jane描述符[40]检测和描述了500多个角。附加的角部特征用于在环路检测上实现更好的召回率。描述符被视为查询可视数据库的可视单词。 DBoW2返回循环闭包可以在时间和几何一致性检查后执行。我们保留所有简要描述符以进行特征检索，但丢弃原始图像以减少内存消耗。

1. 特征检索

当检测到循环时，通过检索特征对应关系建立本地滑动窗口和循环闭包候选之间的连接。通过BRIEF描述符匹配找到对应关系。描述符匹配可能导致一些错误的匹配对。为此，我们使用两步几何异常值抑制，如图10所示。

1）2-D-2-D：使用RANSAC进行基本矩阵测试[33]。

我们使用对图像中的检索特征的二维观察和闭环候选图像来执行基本矩阵测试。

2）3-D-2-D：使用RANSAC进行PnP测试[35]。基于局部滑动获取中的特征的已知3-D位置以及环闭合候选图像中的2-D观察，我们执行PnP测试。

在异常值拒绝后，我们将此候选者视为正确的循环检测并执行重定位。

1. 紧耦合重新定位

重定位过程有效地将当前的滑动窗口与过去的姿势对齐。在重定位期间，我们将所有循环闭包帧的姿势视为常量。我们使用所有IMU测量，局部视觉测量测量和检索到的特征对应来共同优化滑动窗口。我们可以轻松地为环路闭合框架v观察到的检索特征编写视觉测量模型，使其与VIO中的视觉测量结果相同，如（17）所示。唯一的区别是闭环框架的姿势（qw v，pw v）取自姿势图（参见第VIII节），或直接来自过去的odometry输出（如果这是第一次重新计算） ），被视为常数。为此，我们可以使用附加循环稍微修改（14）中的非线性代价函数，其中L是检索到的特征的集合。

闭环框架。 （l，v）表示观察到的第l个特征

循环闭包框架v。注意虽然是成本函数

与（14）略有不同，状态的维度为