VINS-Mono翻译

2018年07月11日 22:48:51 Pancheng1 阅读数：732 更多

Abstract

由摄像机和低成本惯性测量单元(IMU)组成的单目视觉惯性系统(VINS)构成了用于度量六自由度状态估计的最小传感器套件。然而，由于缺乏直接距离测量，在IMU处理、估计器初始化、外部标定和非线性优化等方面提出了重大挑战。在本文中，我们提出了一种鲁棒的、通用的单目视觉惯性状态估计器VINSMono.我们的方法从一个稳健的程序开始，用于估计器初始化和故障恢复。采用一种基于紧耦合、非线性优化的方法，通过融合预计分的IMU测量数据和特征观测数据，获得高精度的视觉惯性里程计。结合我们紧耦合的公式，一个循环检测模块能够以最小的计算开销重新定位.此外，我们还执行四自由度姿态图优化，以加强全局一致性.我们验证了我们的系统在公共数据集和真实世界实验上的性能，并与其他最先进的算法进行了比较。我们还在MAV平台上执行星载闭环自主飞行，并将算法移植到基于IOS的演示中。我们强调，所提议的工作是一个可靠、完整和多功能的系统，适用于需要高精度定位的不同应用程序。我们为个人电脑和iOS移动设备开放了我们的实现方法。

I. INTRODUCTION

状态估计无疑是机器人导航、自主驾驶、虚拟现实(VR)和增强现实(AR)等广泛应用中最基本的模块。仅使用单目摄像机的方法由于其体积小、成本低和硬件设置简单而获得了社区的极大兴趣。然而，单目视觉系统无法恢复公制尺度，因此限制了它们在实际机器人应用中的应用。近年来，我们看到了一种发展趋势，即用低成本惯性测量单元(IMU)辅助单目视觉系统。这种单目视觉-惯性系统(VINS)的主要优点是具有公制尺度，以及滚动角和俯仰角，所有这些都是可观测的。这让需要度量状态估计的导航任务成为可能。此外，积分IMU测量可以显着地改善运动跟踪性能，弥补由于光照变化，纹理无区域，或运动模糊的视觉轨迹损失之间的差距。事实上，单目VINS不仅广泛应用于移动机器人、无人机和移动设备上，而且是满足充分自我感知和环境感知的最小传感器设置。

然而，所有这些优势都是有代价的。对于单目VIN，众所周知，加速度激励需要测量到尺度。这意味着单目VIN估计器不能从静止状态开始，而是从未知的移动状态发射。也认识到视觉惯性系统高度非线性的事实，在初始化估计方面还有重大挑战。两个传感器的存在也使得摄像机-IMU的外部校准至关重要。最后，为了消除在可接受的处理窗口内的长期漂移，提出了一个完整的系统，包括视觉惯性里程计，回环检测、重定位和全局优化。

为了解决所有这些问题，我们提出了VINS-MONO，一个强大的和多功能的单目视觉惯性状态估计。我们的解决方案开始于即时估计初始化。这个初始化模块也用于故障恢复。我们的解决方案的核心是一个鲁棒的基于紧耦合滑动窗非线性优化的单目视觉惯性里程计（VIO）。单目VIO模块不仅提供精确的局部姿态、速度和方位估计，而且还以在线方式执行摄像机IMU外部校准和IMU偏置校正。使用DBoW进行回环检测。重新定位是在对单目VIO进行特征级别融合的紧耦合设置中完成。这使得重新定位具有鲁棒性和精确性且有最小的计算开销。最后，几何验证回路被添加到姿态图中，并且由于来自单目VIO的可观察的滚转和俯仰角，生成四自由度（DOF）姿态图以确保全局一致性。

VINS-Mono结合并改进了我们以前在单目视觉-惯性融合方面的工作。它建立在我们的紧耦合，基于优化的单目VIO的公式之上，并结合了[9]中引入的改进的初始化过程。第一次尝试移植到移动设备是在[10]。与我们以前的工作相比，VINS-Mono的进一步改进包括改进的含偏置校正的IMU预积分、紧耦合重定位、全局姿态图优化、广泛的实验评估以及健壮和通用的开源实现。

整个系统完整且易于使用。它已经被成功应用于小规模AR场景、中型无人机导航和大规模状态估计任务。已经针对现有技术方法的其它状态示出了优异的性能。为此，我们总结了我们的贡献如下所示：

**—一个健壮的初始化过程，它能够从未知的初始状态引导系统。

*—一个紧耦合，优化的单目视觉惯性里程计与相机-IMU外部校准和IMU偏置估计。

*在线回环检测与紧耦合重定位。

*四自由度全局姿态图优化。

*用于无人机导航、大规模本地化和移动AR应用的实时性能演示。

*与ros完全积分的pc版本以及在iphone 6或更高版本上运行的IOS版本的开源版本。

论文的其余部分如下：在第二节中，我们讨论了相关的文献。我们在第三节中对完整的系统框架进行了概述。在第四节中给出了视觉和预处理IMU测量的预处理步骤。在第五节中，我们讨论了估计器的初始化过程。在第六节中提出了一种紧耦合、自标定、非线性优化的单目VIO。第七节和第八节分别给出了紧耦合重定位和全局姿态图优化。实施细节和实验结果见第九节。最后，第十节本文对研究方向进行了探讨和展望。

II. RELATED WORK

关于基于单目视觉的状态估计/里程测量/SLAM的学术著作非常广泛。值得注意的方法包括PTAM、SVO、LSD-SLAM、DSO和ORB-SLAM。显然，任何进行全面回顾都是不完整的。然而，在这一节中，我们跳过了关于只使用视觉的方法的讨论，而只专注于关于单目视觉惯性状态估计的最相关的结果。

处理视觉和惯性测量的最简单的方法是**松耦合**传感器融合，其中IMU被视为一个独立的模块，用于辅助从运动中获得的视觉结构的视觉姿态估计。融合通常由扩展卡尔曼滤波(EKF)完成，**其中IMU用于状态传播，而视觉姿态用于更新**。进一步说，**紧耦合视觉惯性算法要么基于EKF，要么基于图优化**，其中摄像机和IMU测量是从原始测量水平联合优化的。一种流行的基于EKF的VIO方法是MSCKF。MSCKF在状态向量中维护以前的几个摄像机姿态，并使用多个摄像机视图中相同特征的视觉测量来形成多约束更新。SR-ISWF是MSCKF的扩展。它采用SQuareroot格式实现单精度表示，避免了差的数值性质。该方法采用逆滤波器进行迭代再线性化，使其与基于优化的算法相当.批量图优化或捆绑调整（BA）技术维护和优化所有的测量，以获得最优状态估计。为了达到恒定的处理时间，流行的基于图的VIO方法通常通过边缘化有过去的状态和测量的有界滑动窗口来优化最近状态。由于对非线性系统迭代求解的计算要求很高，很少有基于图的非线性系统能够在资源受限的平台(如手机)上实现实时性能。

对于视觉测量处理，根据**视觉残差模型的定义，算法可分为直接法和间接法。**直接法最小光度误差，而间接法最小几何位移。直接方法由于其吸引区域小，需要很好的初始估计，而间接方法在提取和匹配特征时需要额外的计算资源。**间接方法由于其成熟性和鲁棒性，在实际工程部署中得到了广泛的应用。**然而，直接方法更容易扩展到稠密建图，因为它们是直接在像素级别上操作的。

在实践中，IMU通常以比摄像机更高的速度获取数据。不同的方法被提出来处理高速率的IMU测量。最简单的方法是在基于EKF的方法中使用IMU进行状态传播。在图优化公式中，为了避免重复的IMU重复积分，提出了一种有效的方法，即IMU预积分，这种方法是在[22]中首次提出的，它用欧拉角来参数化旋转误差。在我们先前的工作中，我们提出了一种流形上的IMU-preIntegration旋转公式，该文利用连续IMU误差状态动力学推导了协方差传播。然而，IMU偏置被忽略了。文[23]通过增加后验IMU偏置校正，进一步改进了预积分理论。

精确的初始值对于引导任何单目VINS是至关重要的。在[8]，[24]中提出了一种利用短期IMU预积分相对旋转的线性估计器初始化方法。但是，该方法不对陀螺仪偏置进行建模，无法在原始投影方程中对传感器噪声进行建模。在实际应用中，当视觉特性远离传感器套件时，这会导致不可靠的初始化。文[25]给出了单目视觉惯性初始化问题的一种封闭解.随后，文[26]提出了对这种封闭形式的解决方案的扩展，增加了陀螺仪的偏置校准。这些方法依赖于长时间内IMU测量的双重积分，无法模拟惯性积分的不确定性。在[27]中，提出了一种基于SVO的重初始化和故障恢复算法.这是一种基于松散耦合融合框架的实用方法。然而，需要额外的朝下的距离传感器来恢复公制尺度.在[17]中引入了一种建立在流行的ORB-SLAM之上的初始化算法.给出了一组ORB-SLAM的关键帧，计算了视觉惯性全BA的尺度、重力方向、速度和IMU偏置的初步估计。然而，据报道，规模收敛所需的时间可能超过10秒。这可能会给需要在一开始就进行规模评估的机器人导航任务带来问题。

VIO方法，不管它们所依赖的基本数学公式，在全局的平移和旋转中长期受到漂移的影响。为此，回环检测在长期操作中起着重要的作用.ORBSLAM能够关闭循环并重新使用地图，它利用了词袋模型。一个7自由度（位置、方向和尺度)的姿态图优化遵循回环检测。相对于单目Vins，由于IMU的加入，漂移只发生在4自由度，即三维平移，并围绕重力方向(偏航角)旋转。因此，本文选择在最小四自由度设定下，优化具有回路约束的姿态图。

III.OVERVIEW

提出的单目视觉惯性状态估计器的结构如图2所示.该系统从测量预处理(SectIV)开始，在其中提取和跟踪特征，对两个连续帧间的IMU测量进行预处理。初始化过程(SectV)提供了所有必要的值，包括姿态、速度、重力向量、陀螺仪偏置和三维特征位置，用于引导随后的基于非线性优化的VIO。VIO(SectVI)与重新定位(SectVII)模块紧密地融合了预先积分的IMU测量、特征观测和回环重新检测到的特征。最后，位姿图优化模块(SectVIII)接受几何验证的重定位结果，并进行全局优化以消除漂移。
VIO、重新定位和姿态图优化模块在多线程设置中同时运行.每个模块有不同的运行速度和实时保证，以确保在任何时候的可靠运行。

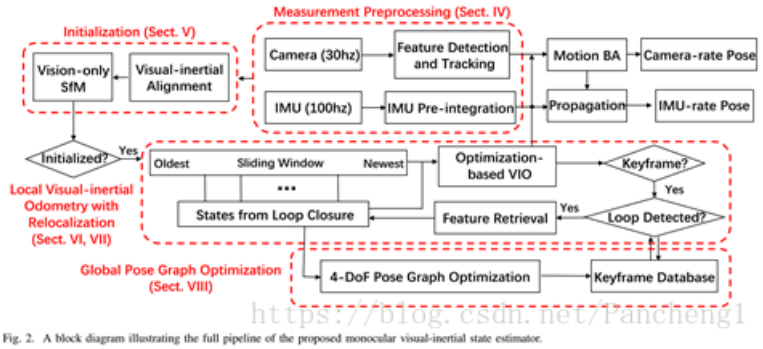


Fig. 2. A block diagram illustrating the full pipeline of the proposed monocular visual-inertial state estimator.

我们现在定义在整个论文中使用的符号和框架定义。我们认为$(.)^w$是世界框架。重力的方向与世界帧的z轴对齐。$(\cdot)^b$是物体框架，我们把它定义为与IMU框架相同。$(\cdot)^c$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，但三维向量的旋转用旋转矩阵来方便表示。q_w^b,P_W^b是从B帧到W帧的旋转和平移。b_k是当取k个图像时的B帧。c_k是在获取k个图像时的相机帧。⊗表示两个四元数之间的乘法运算。$gw=[0,0,g]^T$是世界范围内的重力向量。最后，我们将(^.)表示为某一量的噪声测量值或估计值。

IV.MEASUREMENT PREPROCESSING

本节介绍VIO的预处理步骤。对于视觉测量，我们跟踪连续帧之间的特征，并检测最新帧中的新特征。对于IMU测量，我们将它们预先积分在两个连续的帧之间。请注意，我们使用的低成本IMU的测量值受到偏置和噪声的影响。因此，我们在IMU的预积分过程中特别考虑到偏置。

A.视觉处理前端

对于每一幅新图像，KLT稀疏光流算法对现有特征进行跟踪。同时，检测新的角点特征以保持每个图像中特征的最小数目(100-300)。该检测器通过设置两个相邻特征之间像素的最小间隔来执行均匀的特征分布。二维特征首先是不失真的，然后在通过异常点剔除后投射到一个单位球面上。利用基本矩阵模型的RANSAC进行外点剔除。（外点、异常点）

在此步骤中还选择了关键帧。我们有两个关键帧选择的标准。第一个是与前一个关键帧不同的平均视差。如果跟踪特征的平均视差介于当前帧和最新的关键帧之间，则将帧视为新的关键帧。请注意，不仅平移，旋转也会产生视差。然而，特征不能在旋转-纯运动中三角化.为了避免这种情况，在计算视差时，我们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，而不涉及VINS公式中的旋转计算。为此，即使陀螺仪含有较大的噪声或存在偏置，也只会导致次优的关键帧选择结果，不会直接影响估计质量。另一个标准是跟踪质量。如果跟踪的特征数量低于某一阈值，我们将此帧视为新的关键帧。这个标准是为了避免完全丢失特征轨迹。

B.IMU预积分

IMU预积分是在[22]中首次提出的，它将欧拉角的旋转误差参数化.在我们先前的工作中，我们提出了一个流形上的IMU预积分旋转公式，该文利用连续时间的IMU误差状态动力学推导了协方差传播，但忽略了IMU偏置.文[23]通过增加后验IMU偏置校正，进一步改进了预积分理论。本文通过引入IMU偏置校正，扩展了我们在前面工作[7]中提出的IMU预积分。

IMU的的原始陀螺仪和加速度计测量结果w^和a^如下：

$$\begin{aligned}\hat{\mathbf{a}}_t &= \mathbf{a}_t + \mathbf{b}_{a_t} + \mathbf{R}_w^t \mathbf{g}^w + \mathbf{n}_a \\ \hat{\boldsymbol{\omega}}_t &= \boldsymbol{\omega}_t + \mathbf{b}_{w_t} + \mathbf{n}_w.\end{aligned} \tag{1}$$

IMU测量值是在B帧中测量的，它结合了对抗重力和平台动力学的力量，并受到加速度偏置b_a、陀螺仪偏置b_w和附加噪声的影响。假设加速度计和陀螺仪测量中的附加噪声为高斯噪声，$n_a \sim N(0,\sigma_a^2)$，$n_w \sim N(0,\sigma_w^2)$。加速度计偏置和陀螺仪偏置被建模为随机游走，其导数为高斯性的，$n_{ba} \sim N(0,\sigma_{ba}^2)$，$n_{bw} \sim N(0,\sigma_{bw}^2)$。

$$\dot{\mathbf{b}}_{a_t} = \mathbf{n}_{b_a}, \quad \dot{\mathbf{b}}_{w_t} = \mathbf{n}_{b_w}. \tag{2}$$

给定对应于图像帧的两个时刻b_k和b_{k+1}，位置、速度和方向状态可以在时间间隔期$[t_k,t_{k+1}]$间在W帧内通过惯性测量传播：

$$\begin{aligned}\mathbf{p}_{b_{k+1}}^w &= \mathbf{p}_{b_k}^w + \mathbf{v}_{b_k}^w \Delta t_k \\ &\quad + \iint_{t\in[t_k,t_{k+1}]} (\mathbf{R}_t^w(\hat{\mathbf{a}}_t - \mathbf{b}_{a_t} - \mathbf{n}_a) - \mathbf{g}^w)\, dt^2 \\ \mathbf{v}_{b_{k+1}}^w &= \mathbf{v}_{b_k}^w + \int_{t\in[t_k,t_{k+1}]} (\mathbf{R}_t^w(\hat{\mathbf{a}}_t - \mathbf{b}_{a_t} - \mathbf{n}_a) - \mathbf{g}^w)\, dt \\ \mathbf{q}_{b_{k+1}}^w &= \mathbf{q}_{b_k}^w \otimes \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}_t - \mathbf{b}_{w_t} - \mathbf{n}_w)\mathbf{q}_t^{b_k}\, dt,\end{aligned} \tag{3}$$

where

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -\lfloor\boldsymbol{\omega}\rfloor_\times & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix}, \lfloor\boldsymbol{\omega}\rfloor_\times = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{4}$$

Δt_k是时间间隔$[t_k,t_{k+1}]$之间的持续时间。

可见，IMU状态传播需要帧b_k的旋转、位置和速度。当这些起始状态改变时，我们需要重新传播IMU测量值.特别是在基于优化的算法中，每次调整姿态时，都需要在它们之间重传IMU测量值。这种传播策略在计算上要求很高。为了避免再传播，我们采用了预积分算法。

在将参考帧从世界帧改为局部帧bk后，我们能对只与线性的加速度a^和角速度相关的部分进行预积分，如下所示：

$$\begin{aligned}\mathbf{R}_w^{b_k}\mathbf{p}_{b_{k+1}}^w &= \mathbf{R}_w^{b_k}(\mathbf{p}_{b_k}^w + \mathbf{v}_{b_k}^w \Delta t_k - \frac{1}{2}\mathbf{g}^w \Delta t_k^2) + \boldsymbol{\alpha}_{b_{k+1}}^{b_k} \\ \mathbf{R}_w^{b_k}\mathbf{v}_{b_{k+1}}^w &= \mathbf{R}_w^{b_k}(\mathbf{v}_{b_k}^w - \mathbf{g}^w \Delta t_k) + \boldsymbol{\beta}_{b_{k+1}}^{b_k} \\ \mathbf{q}_w^{b_k} \otimes \mathbf{q}_{b_{k+1}}^w &= \boldsymbol{\gamma}_{b_{k+1}}^{b_k},\end{aligned} \tag{5}$$

where,

$$\begin{aligned}\boldsymbol{\alpha}_{b_{k+1}}^{b_k} &= \iint_{t\in[t_k,t_{k+1}]} \mathbf{R}_t^{b_k}(\hat{\mathbf{a}}_t - \mathbf{b}_{a_t} - \mathbf{n}_a)dt^2 \\ \boldsymbol{\beta}_{b_{k+1}}^{b_k} &= \int_{t\in[t_k,t_{k+1}]} \mathbf{R}_t^{b_k}(\hat{\mathbf{a}}_t - \mathbf{b}_{a_t} - \mathbf{n}_a)dt \\ \boldsymbol{\gamma}_{b_{k+1}}^{b_k} &= \int_{t\in[t_k,t_{k+1}]} \frac{1}{2}\boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}_t - \mathbf{b}_{w_t} - \mathbf{n}_w)\boldsymbol{\gamma}_t^{b_k}\, dt.\end{aligned} \tag{6}$$

可以看出预积分项能通过IMU测量值单独得到，其中将b_k视为参考帧。$\boldsymbol{\alpha}_{b_{k+1}}^{b_k}, \boldsymbol{\beta}_{b_{k+1}}^{b_k}, \boldsymbol{\gamma}_{b_{k+1}}^{b_k}$只与$b_k$和$b_{k+1}$中的IMU偏置有关与其他状态无关。当偏置估计发生变化时，当偏置变化很小时，我们将$\boldsymbol{\alpha}_{b_{k+1}}^{b_k}, \boldsymbol{\beta}_{b_{k+1}}^{b_k}, \boldsymbol{\gamma}_{b_{k+1}}^{b_k}$按其对偏置的一阶近似来调整，否则就进行重传。这种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传播IMU测量。

对于离散时间的实现，可以采用不同的数值积分方法，如欧拉积分、中点积分、RK4积分等.这里选择了Euler积分来演示易于理解的过程(我们在实现代码中使用了中点积分)。

在开始时，$\boldsymbol{\alpha}_{b_k}^{b_k}, \boldsymbol{\beta}_{b_k}^{b_k}$是0，$\boldsymbol{\gamma}_{b_k}^{b_k}$是恒等四元数。α,β,γ在(6)中的平均值是按以下步骤逐步传播的。注意，加性噪声项n_a,n_w是未知的，在实现中被视为零。这导致了预积分的估计值，标记为(^·)：

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_{i+1}^{b_k} &= \hat{\boldsymbol{\alpha}}_i^{b_k} + \hat{\boldsymbol{\beta}}_i^{b_k}\delta t + \frac{1}{2}\mathbf{R}(\hat{\boldsymbol{\gamma}}_i^{b_k})(\hat{\mathbf{a}}_i - \mathbf{b}_{a_i})\delta t^2 \\ \hat{\boldsymbol{\beta}}_{i+1}^{b_k} &= \hat{\boldsymbol{\beta}}_i^{b_k} + \mathbf{R}(\hat{\boldsymbol{\gamma}}_i^{b_k})(\hat{\mathbf{a}}_i - \mathbf{b}_{a_i})\delta t \\ \hat{\boldsymbol{\gamma}}_{i+1}^{b_k} &= \hat{\boldsymbol{\gamma}}_i^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2}(\hat{\boldsymbol{\omega}}_i - \mathbf{b}_{w_i})\delta t \end{bmatrix}\end{aligned} \quad . \quad (7)$$

i是与[t_k,t_{k+1}]内的IMU测量相对应的离散时刻,δt是两个IMU测量i和i+1之间的时间间隔。

然后讨论协方差传播问题。由于四维旋转四元数γ_t^{bk}被过参数化，我们将其误差项定义为围绕其平均值的扰动：

$$\boldsymbol{\gamma}_t^{b_k} \approx \hat{\boldsymbol{\gamma}}_t^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2}\delta\boldsymbol{\theta}_t^{b_k} \end{bmatrix}, \quad (8)$$

其中$\delta\theta^{bk}_t$是三维小扰动。

我们可以导出连续时间线性化的误差项(6)的动力学：

$$\begin{aligned}\begin{bmatrix} \delta\dot{\boldsymbol{\alpha}}_t^{b_k} \\ \delta\dot{\boldsymbol{\beta}}_t^{b_k} \\ \delta\dot{\boldsymbol{\theta}}_t^{b_k} \\ \delta\dot{\mathbf{b}}_{a_t} \\ \delta\dot{\mathbf{b}}_{w_t} \end{bmatrix} &= \begin{bmatrix} 0 & \mathbf{I} & 0 & 0 & 0 \\ 0 & 0 & -\mathbf{R}_t^{b_k}[\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}]_\times & -\mathbf{R}_t^{b_k} & 0 \\ 0 & 0 & -[\hat{\boldsymbol{\omega}}_t - \mathbf{b}_{w_t}]_\times & 0 & -\mathbf{I} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta\boldsymbol{\alpha}_t^{b_k} \\ \delta\boldsymbol{\beta}_t^{b_k} \\ \delta\boldsymbol{\theta}_t^{b_k} \\ \delta\mathbf{b}_{a_t} \\ \delta\mathbf{b}_{w_t} \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 & 0 & 0 \\ -\mathbf{R}_t^{b_k} & 0 & 0 & 0 \\ 0 & -\mathbf{I} & 0 & 0 \\ 0 & 0 & \mathbf{I} & 0 \\ 0 & 0 & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{n}_a \\ \mathbf{n}_w \\ \mathbf{n}_{b_a} \\ \mathbf{n}_{b_w} \end{bmatrix} = \mathbf{F}_t\delta\mathbf{z}_t^{b_k} + \mathbf{G}_t\mathbf{n}_t,\end{aligned} \quad (9)$$

P_{bk}^{bk+1}可以通过初始协方差P_{bk}^{bk}=0的一阶离散时间协方差更新递归计算：

$$\begin{aligned}\mathbf{P}_{t+\delta t}^{b_k} = (\mathbf{I} + \mathbf{F}_t\delta t)\mathbf{P}_t^{b_k}(\mathbf{I} + \mathbf{F}_t\delta t)^T + (\mathbf{G}_t\delta t)\mathbf{Q}(\mathbf{G}_t\delta t)^T, \\ t \in [k, k+1].\end{aligned} \quad (10)$$

其中Q是噪声的对角线协方差矩阵(σ_a^2,σ_w^2,σ_{ba}^2,σ_{bw}^2)。

同时，δz_{bk}^{bk+1}的一阶Jacobian矩阵J_{bk+1}相对于δz_{bk}^{bk}也可以用初始Jacobian J_{bk}=I递归计算。

$$\mathbf{J}_{t+\delta t} = (\mathbf{I} + \mathbf{F}_t\delta t)\mathbf{J}_t, \quad t \in [k, k+1]. \quad (11)$$

利用这个递推公式，得到协方差矩阵p_{bk}^{bk+1}和Jacobian J_{bk+1}。 $\boldsymbol{\alpha}_{b_{k+1}}^{b_k}, \boldsymbol{\beta}_{b_{k+1}}^{b_k}, \boldsymbol{\gamma}_{b_{k+1}}^{b_k}$关于偏置的一阶近似可以写为:

$$\begin{aligned}\boldsymbol{\alpha}_{b_{k+1}}^{b_k} &\approx \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_a}^{\alpha}\delta\mathbf{b}_{a_k} + \mathbf{J}_{b_w}^{\alpha}\delta\mathbf{b}_{w_k} \\ \boldsymbol{\beta}_{b_{k+1}}^{b_k} &\approx \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_a}^{\beta}\delta\mathbf{b}_{a_k} + \mathbf{J}_{b_w}^{\beta}\delta\mathbf{b}_{w_k} \\ \boldsymbol{\gamma}_{b_{k+1}}^{b_k} &\approx \hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2}\mathbf{J}_{b_w}^{\gamma}\delta\mathbf{b}_{w_k} \end{bmatrix}\end{aligned} \quad (12)$$

其中J_{ba}^{α}是J_{bk+1}中的子块矩阵，其位置对应于 $\frac{\delta\boldsymbol{\alpha}_{b_{k+1}}^{b_k}}{\delta\mathbf{b}_{a_k}}$。 $\mathbf{J}_{b_w}^{\alpha}, \mathbf{J}_{b_a}^{\beta}, \mathbf{J}_{b_w}^{\beta}, \mathbf{J}_{b_w}^{\gamma}$也使用同样的含义。当偏置估计发生轻微变化时，我们使用(12)近似校正预积分结果，而不是重传。

现在我们可以写下IMU测量模型及其对应的协方差P_{bk}^{bk+1}：

$$\begin{bmatrix} \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} \\ \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \\ \hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_w^{b_k}(\mathbf{p}_{b_{k+1}}^w - \mathbf{p}_{b_k}^w + \frac{1}{2}\mathbf{g}^w\Delta t_k^2 - \mathbf{v}_{b_k}^w\Delta t_k) \\ \mathbf{R}_w^{b_k}(\mathbf{v}_{b_{k+1}}^w + \mathbf{g}^w\Delta t_k - \mathbf{v}_{b_k}^w) \\ \mathbf{q}_{b_k}^{w^{-1}} \otimes \mathbf{q}_{b_{k+1}}^w \\ \mathbf{b}_{ab_{k+1}} - \mathbf{b}_{ab_k} \\ \mathbf{b}_{wb_{k+1}} - \mathbf{b}_{wb_k} \end{bmatrix}. \quad (13)$$

V. ESTIMATOR INITIALIZATION

单目紧耦合VIO是一个高度非线性的系统.由于单目相机无法直接观测到尺度，因此，如果没有良好的初始值，很难直接将这两种测量结果融合在一起。人们可以假设一个固定的初始条件来启动单目vins估计器。然而，这种假设是不合适的，因为在实际应用中经常会遇到运动下的初始化.当IMU测量结果被大偏置破坏时，情况就变得更加复杂了。事实上，初始化通常是单目vins最脆弱的步骤。需要一个健壮的初始化过程来确保系统的适用性。

我们采用松耦合的传感器融合方法得到初始值.我们发现，只有视觉的SLAM，或SfM，具有良好的初始化性质。在大多数情况下，视觉系统可以通过从相对运动方法(如八点或五点算法或估计齐次矩阵)中导出初始值来引导自己。通过将度量IMU预积分与目视SfM结果相匹配，我们可以粗略地恢复尺度、重力、速度，甚至偏置。这足以引导非线性单目vins估计器，如图4所示。

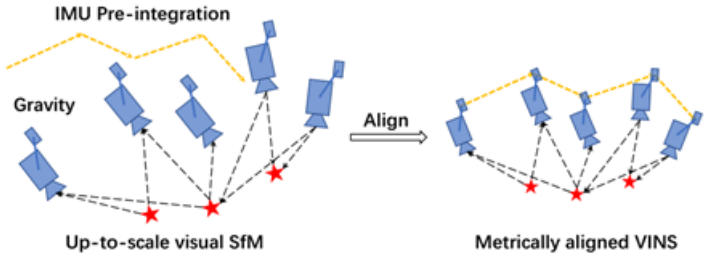


Fig. 4. An illustration of the visual-inertial alignment process for estimator initialization.

与在初始阶段同时估计陀螺仪和加速度计偏置的[17]相比，我们在初始阶段选择忽略加速度计偏置项。加速度计偏置与重力耦合，由于重力向量相对于平台动力学的大量级，以及初始阶段相对较短，这些偏置项很难被观测到。我们以前的工作对加速度计偏置标定进行了详细的分析。

A. Sliding Window Vision-Only SfM

初始化过程开始于一个只有视觉的SfM来估计一个高比例相机的姿势和特征位置的图表。

我们保持了一个帧的滑动窗口来限制计算复杂度。首先，我们检查了最新帧与之前所有帧之间的特征对应。如果我们能找到稳定的特征跟踪(超过30个跟踪特征)和足够的视差(20个以上的旋转补偿像素)在滑动窗口中的最新帧和任何其他帧之间，我们恢复相对旋转和这两个帧之间的上尺度平移使用五点算法。否则，我们会将最新的帧保存在窗口中，并等待新的帧。如果五点算法成功的话，我们可以任意设置标度，并对这两个帧中观察到的所有特征进行三角化。基于这些三角特征，采用透视n点(Pnp)方法来估计窗口中所有其他帧的姿态。最后，应用全局整束平差[36]将所有特征观测的总重投影误差降到最小。由于我们还没有任何世界帧的知识，我们设置了第一个相机帧$(\cdot)^{c0}$作为参考帧的SfM。所有帧的姿态(p_{ck}^{c0}，q_{ck}^{c0})和特征位置表示相对于$(\cdot)^{c0}$。假设摄像机和IMU之间有一个粗糙测度的外部参数(P_c^b,q_c^b)，我们可以将姿态从C帧转换到B帧(IMU)。

$$\begin{aligned}\mathbf{q}_{b_k}^{c_0} &= \mathbf{q}_{c_k}^{c_0} \otimes (\mathbf{q}_c^b)^{-1} \\ s\bar{\mathbf{p}}_{b_k}^{c_0} &= s\bar{\mathbf{p}}_{c_k}^{c_0} - \mathbf{R}_{b_k}^{c_0}\mathbf{p}_c^b,\end{aligned} \tag{14}$$

其中s是匹配视觉结构与米制尺度的尺度参数，解决这个缩放参数是实现成功初始化的关键。

B. Visual-Inertial Alignment

1)**陀螺仪偏置标定**：考虑窗口连续两帧b_k和b_{k+1}，我们从视觉sfM中得到旋转q_{bk}^{c0}和q_{bk+1}^{c0}，以及通过IMU预积分得到的相对约束γ_{bk+1}^{bk}。我们将IMU预积分项线性化，使陀螺仪偏置最小，并将下列成本函数降到最小：

$$\begin{aligned}&\min_{\delta b_w} \sum_{k\in\mathcal{B}} \left\| \mathbf{q}_{b_{k+1}}^{c_0}{}^{-1} \otimes \mathbf{q}_{b_k}^{c_0} \otimes \boldsymbol{\gamma}_{b_{k+1}}^{b_k} \right\|^2 \\ &\boldsymbol{\gamma}_{b_{k+1}}^{b_k} \approx \hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2}\mathbf{J}_{b_w}^{\gamma}\delta\mathbf{b}_w \end{bmatrix},\end{aligned} \tag{15}$$

其中**B**代表窗口中的所有帧。利用第四部分导出的偏置雅可比，给出了γ_{bk+1}^{bk}对陀螺仪偏置的一阶近似。这样，我们得到了陀螺仪偏置b_w的初始校准。然后我们用新的陀螺仪偏置重新传播所有的imu预积分项$\alpha_{b_{k+1}}^{b_k}, \beta_{b_{k+1}}^{b_k}, \gamma_{b_{k+1}}^{b_k}$。

2)**速度、重力向量和米制尺度初始化**：在陀螺仪偏置初始化后，我们继续初始化导航的其他基本状态，即速度、重力向量和公制标度：

$$\mathcal{X}_I = \left[\mathbf{v}_{b_0}^{b_0}, \mathbf{v}_{b_1}^{b_1}, \cdots \mathbf{v}_{b_n}^{b_n}, \mathbf{g}^{c_0}, s\right], \tag{16}$$

其中，当取第k帧图像时，v_{bk}^{bk}是B帧中的速度，g^{c0}是c0帧中的重力向量，s将单目SfM缩放为公制单位。

考虑窗口中两个连续的帧b_k和b_{k+1}，那么(5)可以写成：

$$\begin{aligned}
\boldsymbol{\alpha}_{b_{k+1}}^{b_k} &= \mathbf{R}_{c_0}^{b_k}(s(\bar{\mathbf{p}}_{b_{k+1}}^{c_0} - \bar{\mathbf{p}}_{b_k}^{c_0}) + \frac{1}{2}\mathbf{g}^{c_0}\Delta t_k^2 - \mathbf{R}_{b_k}^{c_0}\mathbf{v}_{b_k}^{b_k}\Delta t_k) \\
\boldsymbol{\beta}_{b_{k+1}}^{b_k} &= \mathbf{R}_{c_0}^{b_k}(\mathbf{R}_{b_{k+1}}^{c_0}\mathbf{v}_{b_{k+1}}^{b_{k+1}} + \mathbf{g}^{c_0}\Delta t_k - \mathbf{R}_{b_k}^{c_0}\mathbf{v}_{b_k}^{b_k}).
\end{aligned} \tag{17}$$

我们可以将(14)和(17)合并成以下线性测量模型：

$$\hat{\mathbf{z}}_{b_{k+1}}^{b_k} = \begin{bmatrix} \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} - \mathbf{p}_c^b + \mathbf{R}_{c_0}^{b_k}\mathbf{R}_{b_{k+1}}^{c_0}\mathbf{p}_c^b \\ \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \end{bmatrix} = \mathbf{H}_{b_{k+1}}^{b_k}\mathcal{X}_I + \mathbf{n}_{b_{k+1}}^{b_k} \tag{18}$$

where,

$$\mathbf{H}_{b_{k+1}}^{b_k} = \begin{bmatrix} -\mathbf{I}\Delta t_k & \mathbf{0} & \frac{1}{2}\mathbf{R}_{c_0}^{b_k}\Delta t_k^2 & \mathbf{R}_{c_0}^{b_k}(\bar{\mathbf{p}}_{c_{k+1}}^{c_0} - \bar{\mathbf{p}}_{c_k}^{c_0}) \\ -\mathbf{I} & \mathbf{R}_{c_0}^{b_k}\mathbf{R}_{b_{k+1}}^{c_0} & \mathbf{R}_{c_0}^{b_k}\Delta t_k & \mathbf{0} \end{bmatrix} \tag{19}$$

可以看出，R_{bk}^{c0} ,R_{bk+1}^{c0},P_{ck}^{c0} ,P_{ck+1}^{c0}是从上尺度单目视觉Δt_k中得到的，Δt_k是两个连续帧之间的时间间隔。通过求解这个线性最小二乘问题：

$$\min_{\mathcal{X}_I} \sum_{k\in\mathcal{B}} \left\| \hat{\mathbf{z}}_{b_{k+1}}^{b_k} - \mathbf{H}_{b_{k+1}}^{b_k}\mathcal{X}_I \right\|^2, \tag{20}$$

我们可以得到窗口中每一帧的物体帧速度，视觉参照系$(\cdot)^{c0}$中的重力向量，以及尺度参数。

3)**重力精化**：通过约束量值，可以对原线性初始化步骤得到的重力向量进行细化。在大多数情况下，重力向量的大小是已知的。这导致重力向量只剩2自由度。因此，我们在其切线空间上用两个变量重新参数化重力。参数化将重力向量表示为$g \cdot \hat{\bar{\mathbf{g}}} + w_1\mathbf{b}_1 + w_2\mathbf{b}_2$，其中g是已知的重力大小， 是表示重力方向的单位向量，b1和b2是跨越切平面的两个正交基，如图5所示，w1和w2分别是b1和b2的对应位移。通过算法1的交叉乘积运算，可以找到一组b1，b2。然后用$g \cdot \hat{\bar{\mathbf{g}}} + w_1\mathbf{b}_1 + w_2\mathbf{b}_2$代替(17)中的g，并与其它状态变量一起求解w1和w2。此过程迭代到g^收敛为止。

Algorithm 1: Finding $\mathbf{b}_1$ and $\mathbf{b}_2$

```
if ḡ ≠ [1,0,0] then
    b1 ← normalize(ḡ × [1,0,0]);
else
    b1 ← normalize(ḡ × [0,0,1]);
end
b2 ← ḡ × b1;
```

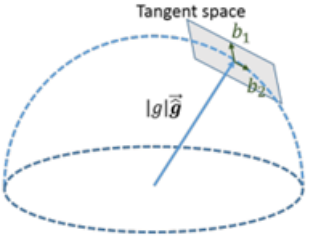



Fig. 5. Illustration of 2 DOF parameterization of gravity. Since the magnitude of gravity is known, $\mathbf{g}$ lies on a sphere with radius $g \approx 9.81m/s^2$. The gravity is parameterized around current estimate as $g \cdot \hat{\bar{\mathbf{g}}} + w_1\mathbf{b}_1 + w_2\mathbf{b}_2$, where $\mathbf{b}_1$ and $\mathbf{b}_2$ are two orthogonal basis spanning the tangent space.

4)**完成初始化**：经过对重力向量的细化，通过将重力旋转到z轴上，得到世界帧与摄像机帧c0之间的旋转q_{c0}^{W}。然后我们将所有变量从参考框架$(\cdot)^{c0}$旋转到世界框架$(\cdot)^{w}$。B帧的速度也将被旋转到世界框架。视觉SfM的转换组件将被缩放为公制单位。此时，初始化过程已经完成，所有这些度量值都将被输入到一个紧耦合的单目VIO中。

VI. TIGHTLY-COUPLED MONOCULAR VIO

在估计器初始化后，我们采用基于滑动窗口的紧耦合单目VIO进行高精度和鲁棒状态估计。图3显示了滑动窗口的图示。

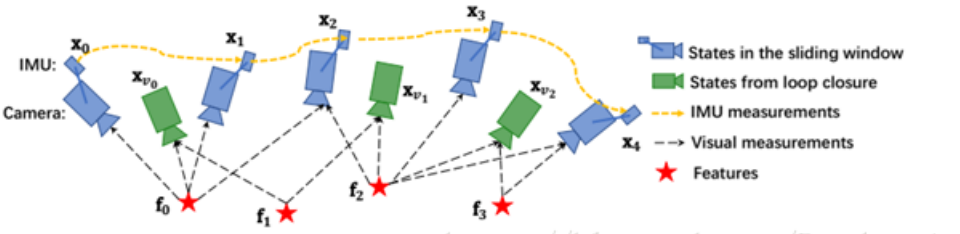


Fig. 3. An illustration of the sliding window monocular VIO with relocalization. It is a tightly-coupled formulation with IMU, visual, and loop measurements.

A. Formulation

滑动窗口中的完整状态向量定义为：

$$\begin{aligned}\mathcal{X} &= \left[\mathbf{x}_0, \mathbf{x}_1, \cdots \mathbf{x}_n, \mathbf{x}_c^b, \lambda_0, \lambda_1, \cdots \lambda_m\right] \\ \mathbf{x}_k &= \left[\mathbf{p}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{q}_{b_k}^w, \mathbf{b}_a, \mathbf{b}_g\right], k \in [0, n] \\ \mathbf{x}_c^b &= \left[\mathbf{p}_c^b, \mathbf{q}_c^b\right],\end{aligned} \tag{21}$$

其中x_K是捕获k^{th}图像时的IMU状态。它包含了IMU在世界帧中的位置、速度和方向，以及在IMU Body帧中的加速度计偏置和陀螺仪偏置，n是关键帧的总数，m是滑动窗口中的特征总数，λ_l是第一次观测到的lth特征的逆深度。

我们使用视觉惯性束调整公式（BA）.我们最小化所有测量残差的先验和Mahalanobis范数之和，得到最大后验估计：

$$\begin{aligned}\min_{\mathcal{X}} \Bigg\{ &\|\mathbf{r}_p - \mathbf{H}_p \mathcal{X}\|^2 + \sum_{k \in \mathcal{B}} \left\| \mathbf{r}_{\mathcal{B}}(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) \right\|_{\mathbf{P}_{b_{k+1}}^{b_k}}^2 + \\ &\sum_{(l,j) \in \mathcal{C}} \rho\left(\left\| \mathbf{r}_{\mathcal{C}}(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X}) \right\|_{\mathbf{P}_l^{c_j}}^2 \right) \Bigg\},\end{aligned} \tag{22}$$

where the Huber norm [37] is defined as:

$$\rho(s) = \begin{cases} 1 & s \geq 1, \\ 2\sqrt{s} - 1 & s < 1. \end{cases} \tag{23}$$

$r_B(z_{bk+1}^{bk},X)$和$r_C(z_l^{cj},X)$分别是IMU和视觉测量的残差。剩余条件的详细定义将在第六节的B和C中提出。B是所有IMU测量的集合，C是在当前滑动窗口中至少观察到两次的一组特征。$\{r_p,H_p\}$是来自边缘化的先验信息。CeresSolver被用来解决这个非线性问题。

B. IMU Measurement Residual

考虑滑动窗口中连续两个帧(b_k和b_{k+1})内的IMU测量，根据(13)中定义的IMU测量模型，预积分IMU测量的残差可以定义为：

$$
\begin{aligned}
\mathbf{r}_{\mathcal{B}}(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) &= \begin{bmatrix} \delta\boldsymbol{\alpha}_{b_{k+1}}^{b_k} \\ \delta\boldsymbol{\beta}_{b_{k+1}}^{b_k} \\ \delta\boldsymbol{\theta}_{b_{k+1}}^{b_k} \\ \delta\mathbf{b}_a \\ \delta\mathbf{b}_g \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}_w^{b_k}(\mathbf{p}_{b_{k+1}}^w - \mathbf{p}_{b_k}^w + \frac{1}{2}\mathbf{g}^w\Delta t_k^2 - \mathbf{v}_{b_k}^w\Delta t_k) - \hat{\boldsymbol{\alpha}}_{b_{k+1}}^{b_k} \\ \mathbf{R}_w^{b_k}(\mathbf{v}_{b_{k+1}}^w + \mathbf{g}^w\Delta t_k - \mathbf{v}_{b_k}^w) - \hat{\boldsymbol{\beta}}_{b_{k+1}}^{b_k} \\ 2\left[\mathbf{q}_{b_k}^{w^{-1}} \otimes \mathbf{q}_{b_{k+1}}^w \otimes (\hat{\boldsymbol{\gamma}}_{b_{k+1}}^{b_k})^{-1}\right]_{xyz} \\ \mathbf{b}_{ab_{k+1}} - \mathbf{b}_{ab_k} \\ \mathbf{b}_{wb_{k+1}} - \mathbf{b}_{wb_k} \end{bmatrix},
\end{aligned} \tag{24}
$$

其中，$[\cdot]_{xyz}$是提取四元数q的向量部分，以进行误差状态表示。$\delta\theta_{bk+1}^{bk}$是四元数的三维误差状态表示。$[\ ]^T$是在两个连续图像帧之间的间隔时间内仅使用噪声加速度计和陀螺仪测量值预积分的IMU测量项。加速度计和陀螺仪偏置也包括在线校正的剩余项中。

C. Visual Measurement Residual

与传统的在广义图像平面上定义再投影误差的针孔相机模型相比，我们定义了单位球面上摄像机的测量残差。几乎所有类型相机的光学，包括广角、鱼眼或全向相机，都可以模拟为连接单位球体表面的单位射线。考虑第一次在第i幅图像中观察到的第l个特征，第j幅图像中的特征观测的残差定义为：

$$
\begin{aligned}
\mathbf{r}_{\mathcal{C}}(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X}) &= \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix}^T \cdot (\hat{\mathcal{P}}_l^{c_j} - \frac{\mathcal{P}_l^{c_j}}{\|\mathcal{P}_l^{c_j}\|}) \\
\hat{\mathcal{P}}_l^{c_j} &= \pi_c^{-1}(\begin{bmatrix} \hat{u}_l^{c_j} \\ \hat{v}_l^{c_j} \end{bmatrix}) \\
\mathcal{P}_l^{c_j} &= \mathbf{R}_b^c(\mathbf{R}_w^{b_j}(\mathbf{R}_{b_i}^w(\mathbf{R}_c^b\frac{1}{\lambda_l}\pi_c^{-1}(\begin{bmatrix} u_l^{c_i} \\ v_l^{c_i} \end{bmatrix}) \\
&\qquad + \mathbf{p}_c^b) + \mathbf{p}_{b_i}^w - \mathbf{p}_{b_j}^w) - \mathbf{p}_c^b),
\end{aligned} \tag{25}
$$

其中$[u_l^{ci},v_l^{ci}]$是第一次观测第一次出现在i图像中的第l个特征。$[u\hat{}_l^{ci},v\hat{}_l^{ci}]$是对jth图像中相同特征的观察。$\pi_c^{-1}$是利用摄像机内参数将像素位置转换成单位向量的反投影函数。由于视觉残差的自由度是2，所以我们将残差向量投影到切平面 上，如图6所示，b1,b2是两个任意选择的正交基。我们可以很容易地找到一组b1,b2，如算法1所示。在(22)中使用的p_l^{cj}是正切空间中固定长度的标准协方差。

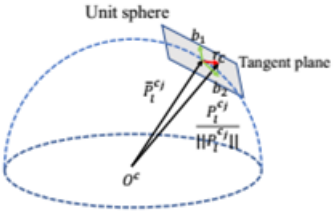


Fig. 6. An illustration of the visual residual on a unit sphere. $\hat{\mathcal{P}}_l^{c_j}$ is the unit vector for the observation of the l^{th} feature in the j^{th} frame. $\mathcal{P}_l^{c_j}$ is predicted feature measurement on the unit sphere by transforming its first observation in the i^{th} frame to the j^{th} frame. The residual is defined on the tangent plane of $\hat{\mathcal{P}}_l^{c_j}$.

D. Marginalization

为了限制基于优化的VIO的计算复杂度，本文引入了边缘化.我们有选择地将IMU状态X_K和特征λ1从滑动窗口边缘化，同时将对应于边缘状态的度量转换为先验。

如图7所示，当第二个最新的帧是关键帧时，它将停留在窗口中，而最老的帧与其相应的测量值被边缘化（边缘化的测量值作为先验信息）。否则，如果第二个最新的帧是非关键帧，我们丢掉视觉测量值和保留连接到这个非关键帧IMU测量值。为了保持系统的稀疏性，我们不丢弃非关键帧的所有测量值（留下IMU预积分值）。我们的边缘化方案的目的是在窗口中保持空间分离的关键帧。这确保了特征三角化有足够的视差，并且最大限度地提高了在大激励下获得加速度计测量值的概率。

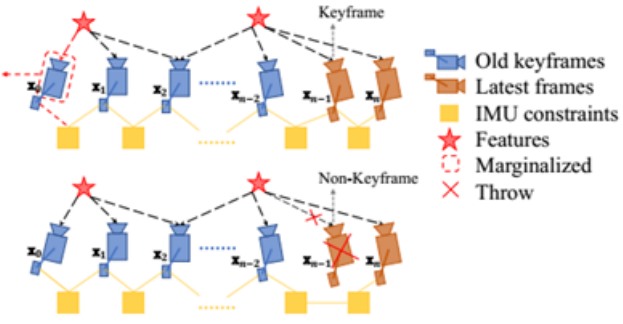


Fig. 7. An illustration of our marginalization strategy. If the second latest frame is a keyframe, we will keep it in the window, and marginalize the oldest frame and its corresponding visual and inertial measurements. Marginalized measurements are turned into a prior. If the second latest frame is not a keyframe, we will simply remove the frame and all its corresponding visual measurements. However, pre-integrated inertial measurements are kept for non-keyframes, and the pre-integration process is continued towards the next frame.

边缘化是利用Schur补充[39]进行的。我们基于与移除状态相关的所有边缘化度量来构造一个新的先验。新的先验项被添加到现有的先验项中。

我们确实注意到，边缘化导致了线性化点的早期固定，这可能导致次优估计结果。然而，由于小型漂移对于VIO来说是可以接受的，我们认为边缘化所造成的负面影响并不重要。

E. Motion-only Visual-Inertial Bundle Adjustment for Camera-RateState Estimation

对于计算能力较低的设备，如手机，由于对非线性优化的计算要求很高，使得紧耦合单目VIO无法实现摄像机速率输出。为此，我们采用了一种轻量级的只运动视觉惯性束调整（只运动的VI BA），以提高状态估计到相机的速率(30赫兹)。

单目视觉惯性集束调整（BA）的成本函数与(22)中的单目VIO的代价函数相同。然而，我们没有对滑动窗口中的所有状态进行优化，而是只对固定数量的最新IMU状态的姿态和速度进行了优化。我们将特征深度、外部参数、偏置和旧的IMU状态作为常量来处理，我们不希望优化这些状态。我们使用所有的视觉和惯性测量来进行仅运动的BA.这导致了比单帧PNP方法更平滑的状态估计。图8显示了提出方法的插图。与在最先进的嵌入式计算机上可能导致超过50 ms的完全紧耦合的单目VIO不同，这种只需运动的视觉惯性束调整只需大约5ms即可计算。这使得低延迟相机的姿态估计对无人机和AR应用特别有利.

F. IMU Forward Propagation for IMU-Rate State Estimation

IMU测量的速度比视觉测量高得多。虽然我们的VIO的频率受到图像捕获频率的限制，但是我们仍然可以通过最近的IMU测量来直接传播最新的VIO估计，以达到IMU速率的性能。高频状态估计可以作为回环检测的状态反馈。利用这种IMU速率状态估计进行的自主飞行实验在第九节的D中进行。

G. Failure Detection and Recovery

虽然我们紧耦合的单目视觉对各种具有挑战性的环境和运动是健壮的。由于强烈的光照变化或剧烈的运动，失败仍然是不可避免的。主动故障检测和恢复策略可以提高系统的实用性。故障检测是一个独立的模块，它检测估计器的异常输出。我们目前使用以下标准进行故障检测：

*在最新帧中跟踪的特征数小于某一阈值；

*最后两个估计器输出之间的位置或旋转有较大的不连续性；

*偏置或外部参数估计有较大的变化；

一旦检测到故障，系统将切换回初始化阶段。一旦单目VIO被成功初始化，将新建一个独立的位姿图段。

VII. RELOCALIZATION

我们的滑动窗口和边缘化方案限制了计算的复杂性，但也给系统带来了累积漂移。更确切地说，漂移发生在全局三维位置(x，y，z)和围绕重力方向的旋转(偏航)。为了消除漂移，提出了一种与单目VIO无缝集成的紧耦合重定位模块.重定位过程从一个循环检测模块开始，该模块标识已经访问过的地方。然后建立回环检测候选帧和当前帧之间的特征级连接。这些特征被对应紧密地集成到单目VIO模块中，从而得到无漂移状态估计，并且计算开销最小。多个特征的多观测直接用于重定位，从而提高了定位的精度和状态估计的平滑性。图9(a)示出了重新定位过程的图形说明。

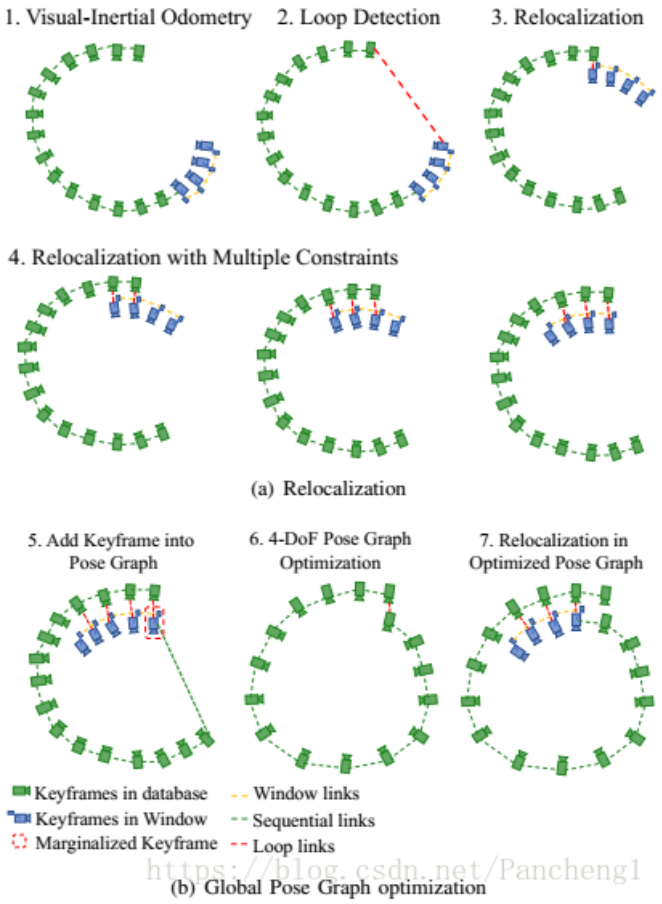


Fig. 9. A diagram illustrating the relocalization and pose graph optimization procedure. Fig. 9(a) shows the relocalization procedure. It starts with VIO-only pose estimates (blue). Past states are recorded (green). If a loop is detected for the newest keyframe (Sect. VII-A), as shown by the red line in the second plot, a relocalization occurred. Note that due to the use of feature-level correspondences for relocalization, we are able to incorporate loop closure constraints from multiple past keyframes (Sect. VII-C), as indicated in the last three plots. The pose graph optimization is illustrated in Fig. 9(b). A keyframe is added into the pose graph when it is marginalized out from the sliding window. If there is a loop between this keyframe and any other past keyframes, the loop closure constraints, formulated as 4-DOF relative rigid body transforms, will also be added to the pose graph. The pose graph is optimized using all relative pose constraints (Sect. VIII-B) in a separate thread, and the relocalization module always runs with respect to the newest pose graph configuration.

A. Loop Detection

我们利用DBoW 2，一种最先进的词袋位置识别方法来进行循环检测.除了用于单目VIO的角特征外，另外500个拐角被检测并由BRIEF描述符描述。额外的角落特征用于在回路检测中实现更好的召回率。描述符被视为用于查询可视化数据库的可视单词。DBoW 2在时间和几何一致性检查后返回回环检测候选项。我们保留所有用于特征检索的BRIEF描述符，丢弃原始图像以减少内存消耗。

我们注意到，我们的单目VIO能够使滚动和俯仰角可以被观察到。因此，我们不需要依赖旋转不变的特性，例如ORB SLAM中使用的ORB特性。

B. Feature Retrieval

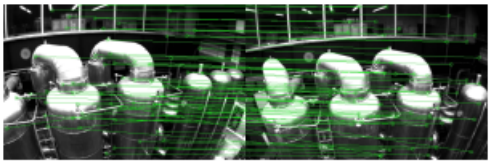
当检测到环路时，通过检索特征对应性建立本地滑动窗口与回环候选点之间的连接。通过BRIEF描述符匹配找到对应关系。直接描述符匹配可能会导致大量异常值。为此，我们使用两步几何离群点剔除，如图10所示。

*2D-2D：RANSAC的基本矩阵检验.我们利用当前图像中检索到的特征的二维观测和回环候选图像进行基本矩阵检验。

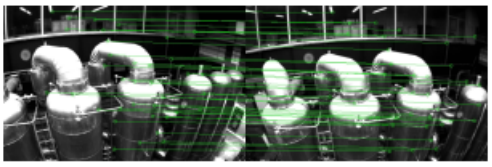
*3D-2D:RANSAC的PNP检验。基于已知的特征在局部滑动窗口中的三维位置，以及在环路闭合候选图像中的二维观测，进行PNP检验。

当超过一定阈值的不动点数时，我们将该候选点视为正确的循环检测并执行重新定位。

(a) BRIEF descriptor matching results



(b) First step: 2D-2D outlier rejection results



(c) Second step: 3D-2D outlier rejection results.

Fig. 10. Descriptor matching and outlier removal for feature retrieval during loop closure.

C. Tightly-Coupled Relocalization

重定位过程有效地使单目VIO保持的当前滑动窗口与过去的姿态图保持一致。在重新定位过程中，我们将所有闭环帧的姿态作为常量，利用所有IMU测量值、局部视觉测量和从回环中提取特征对应值，共同优化滑动窗口。我们可以轻松地为回环帧v所观察到的检索到的特征编写视觉测量模型，使其与VIO中的视觉测量相同，如(25)所示。唯一的区别是，从位姿图或直接从上一个里程计的输出(如果这是第一次重定位)获得的回环帧的姿态(q_v^w,p_v^w)被视为常数。为此，我们可以在(22)中稍微修改非线性代价函数，增加循环项：

$$\min_{\mathcal{X}} \Bigg\{ \|\mathbf{r}_p - \mathbf{H}_p \mathcal{X}\|^2 + \sum_{k\in\mathcal{B}} \left\| \mathbf{r}_{\mathcal{B}}(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) \right\|^2_{\mathbf{P}_{b_{k+1}}^{b_k}} + \sum_{(l,j)\in\mathcal{C}} \rho\left(\left\| \mathbf{r}_{\mathcal{C}}(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X}) \right\|^2_{\mathbf{P}_l^{c_j}} \right) + \sum_{(l,v)\in\mathcal{L}} \rho\left(\left\| \mathbf{r}_{\mathcal{C}}(\hat{\mathbf{z}}_l^v, \mathcal{X}, \hat{\mathbf{q}}_w^w, \hat{\mathbf{p}}_v^w) \right\|^2_{\mathbf{P}_l^{c_v}} \right) \Bigg\} \tag{26}$$

其中L是回环帧中检索到的特征的观测集。(l,v)是指在回环帧v中观察到的第l个特征。注意，虽然代价函数与(22)略有不同，但待解状态的维数保持不变，因为回环帧的构成被视为常数。当用当前滑动窗口建立多个回环时，我们使用来自所有帧的所有环路闭合特征对应进行优化。这就为重新定位提供了多视角的约束，从而提高了定位的精度和平滑性。请注意，过去的姿态和循环关闭帧的全局优化发生在重新定位之后，将在第八节中讨论。

VIII. GLOBAL POSE GRAPH OPTIMIZATION

重新定位后，局部滑动窗口移动并与过去的姿态保持一致。利用重定位结果，此附加姿态图优化步骤被开发以确保过去姿势的集合被登记为全局一致配置

由于我们的视觉惯性设置使滚动和俯仰角完全可观测，累积漂移只发生在四个自由度(x，y，z和偏航角)。为此，我们忽略了对无漂移滚转和俯仰状态的估计，只进行了四自由度姿态图的优化。

A. Adding Keyframes into the Pose Graph

当关键帧从滑动窗口中边缘化时，它将被添加到位姿图中。这个关键帧在位姿图中充当顶点，它通过两种类型的边与其他顶点连接：

1）序列边缘：关键帧将建立与其先前关键帧的若干顺序边。序列边缘表示局部滑动窗口中两个关键帧之间的相对转换，其值直接从VIO中获取。考虑到新边缘化的关键帧i及其先前的一个关键帧j，序列边缘只包含相对位置p_{ij}^{i}和偏航角 。

$$\begin{aligned}\hat{\mathbf{p}}_{ij}^{i} &= \hat{\mathbf{R}}_{i}^{w-1}(\hat{\mathbf{p}}_{j}^{w} - \hat{\mathbf{p}}_{i}^{w}) \\ \hat{\psi}_{ij} &= \hat{\psi}_{j} - \hat{\psi}_{i}.\end{aligned} \tag{27}$$

2）循环闭合边缘：如果新边缘化的关键帧有一个循环连接，它将与回环帧通过一个环闭合边在姿态图中连接。同样，闭环边缘只包含与(27)相同定义的四自由度相对位姿变换。利用重新定位的结果，得到了环路闭合边的值。

B.4-DOF Pose Graph Optimization

我们将帧i和j之间的边缘的残差定义为：

$$\mathbf{r}_{i,j}(\mathbf{p}_{i}^{w}, \psi_{i}, \mathbf{p}_{j}^{w}, \psi_{j}) = \begin{bmatrix} \mathbf{R}(\hat{\phi}_{i}, \hat{\theta}_{i}, \psi_{i})^{-1}(\mathbf{p}_{j}^{w} - \mathbf{p}_{i}^{w}) - \hat{\mathbf{p}}_{ij}^{i} \\ \psi_{j} - \psi_{i} - \hat{\psi}_{ij} \end{bmatrix}, \tag{28}$$

其中，φ^{i},θi是直接从单目VIO中得到的滚动角和俯仰角的估计。通过最小化以下代价函数，对顺序边和回环边的整个图进行了优化：

$$\min_{\mathbf{p},\psi} \left\{ \sum_{(i,j)\in\mathcal{S}} \|\mathbf{r}_{i,j}\|^{2} + \sum_{(i,j)\in\mathcal{L}} \rho(\|\mathbf{r}_{i,j}\|^{2}) \right\}, \tag{29}$$

其中S是所有序列边的集合，L是全环闭包边的集合。尽管紧耦合的重新定位已经有助于消除错误的闭环，但我们添加了另一个Huber规范ρ(·)，以进一步减少任何可能的错误循环的影响。相反，我们不对序列边缘使用任何鲁棒范数，因为这些边缘是从VIO中提取出来的，VIO已经包含了足够多的孤立点排除机制。

位姿图优化和重新定位异步运行在两个独立的线程中。以便在可用重定位时，能立即使用最优化的位姿图。同样，即使当前的姿态图优化尚未完成，仍然可以使用现有的姿态图配置进行重新定位。这一过程如图9(b)所示。

C. Pose Graph Management

随着行程距离的增加，姿态图的大小可能会无限增长，从而限制了系统的实时性。为此，我们实现了一个下采样过程，以将姿态图数据库保持在有限的大小。所有具有回环约束的关键帧都将被保留，而其他与其邻居方向过近或方向非常相似的关键帧可能会被删除。关键帧被移除的概率和其与邻居空间密度成正比。

IX. EXPERIMENTAL RESULTS

我们进行了三个实验和两个应用，以评估所提出的VINS-Mono系统.在第一个实验中，我们将所提出的算法与另一种最新的公共数据集算法进行了比较.通过数值分析，验证了系统的精度。然后在室内环境中测试我们的系统，以评估在重复场景中的性能。通过大量的实验验证了系统的长期实用性.此外，我们还将所提出的系统应用于两个应用程序。对于空中机器人的应用，我们使用vins-Mono作为位置反馈来控制无人机跟踪预定的轨迹。然后我们将我们的方法移植到iOS移动设备上，并与GoogleTango进行比较。

A. Dataset Comparison

我们使用欧洲MAV视觉-惯性数据集评估我们提出的VINS-Mono。这些数据集是在一架微型飞行器上收集的，它包含立体图像(Aptina MT9V034全球快门、WVGA单色、20 FPS)、同步IMU测量(ADIS 16448、200 Hz)和地面真实状态(Vicon和Leica MS 50)。我们只使用左边相机的图像。在这些数据集中观察到了较大的IMU偏置和光照变化。

在这些实验中，我们将vins-mono和OKVIS进行了比较，这是一种最先进的VIO，可以用单目和立体相机工作。OKVIS是另一种基于优化的滑动窗口算法.我们的算法与OKVIS在许多细节上是不同的，如技术部分所示。该系统具有良好的初始化和闭环控制功能。我们使用MH-03和MH-05两种序列来证明该方法的性能。为了简化表示法，我们使用vins来表示我们只使用单目VIO的方法，而vins_loop表示含重新定位和姿态图优化的完全版本。我们分别用OKVISMono和OKVIS_stereo表示OKVIS的结果。为了进行公平的比较，我们丢弃前100个输出，并使用接下来的150个输出对齐地面真值，并比较其余的估计器输出。

MH-03序列轨迹如图11所示。我们只比较平移误差，因为旋转运动在这个序列中是可以忽略的。图12显示了x,y,z误差与时间的关系，以及平移误差与距离的关系。在误差图中，具有回环的vins-mono具有最小的平移误差.我们在MH 05上观察到类似的结果。该方法具有最小的平移误差。平移和旋转误差如图14所示。由于该序列运动平稳，偏角变化不大，只发生位置漂移。显然，循环闭合能力有效地约束了累积漂移。OKVIS在滚动和俯仰角度估计方面表现更好。一个可能的原因是VINSMono采用了预积分技术，即IMU传播的一阶近似，以节省计算资源。

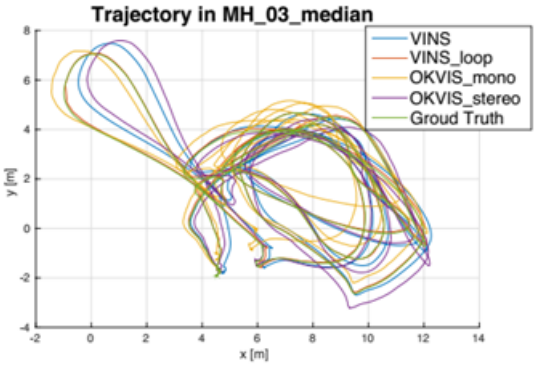


Fig. 11. Trajectory in MH_03_median, compared with OKVIS.

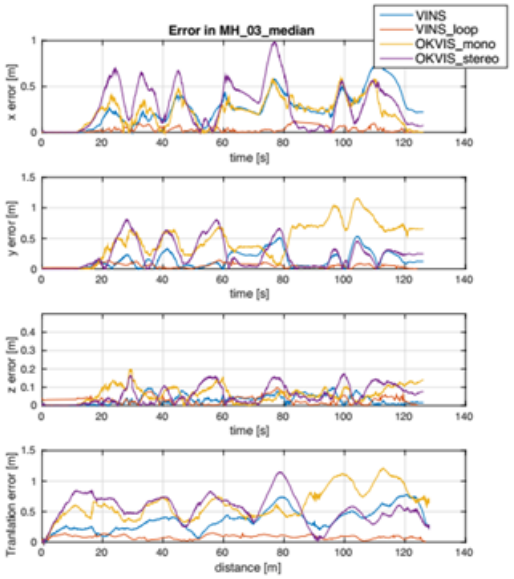


Fig. 12. Translation error plot in MH_03_median.

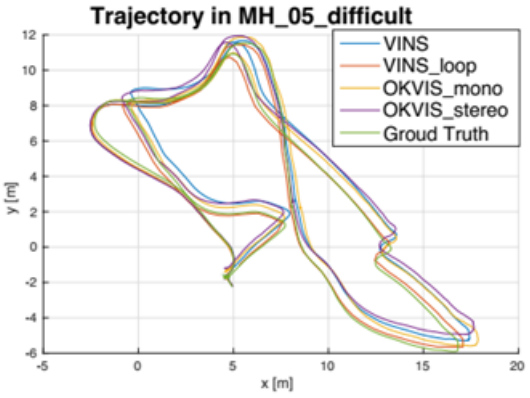


Fig. 13. Trajectory in MH_05_difficult, compared with OKVIS..

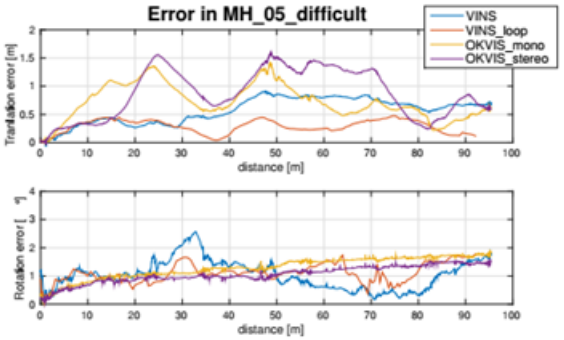


Fig. 14. Translation error and rotation error plot in MH_05_difficult.

vins-Mono在所有Euroc数据集中表现良好，即使在最具挑战性的序列中，V1-03包括剧烈性的运动、纹理减少的区域和显著的光照变化。由于采用了专用的初始化过程，该方法在V1-03中难以快速初始化。

对于纯VIO，vin-Mono和OKVIS具有相似的精度，很难区分哪个比较好。然而，vins-Mono在系统级别上优于OKVIS。它是一个完整的系统，具有鲁棒的初始化和闭环功能来辅助单目视觉。

B. Indoor Experiment

在室内实验中，我们选择实验室环境作为实验区域。我们使用的传感器套件如图15所示。它在DJIA3控制器中包含一个单目照相机(20Hz)和一个IMU(100 Hz)。我们手握传感器套件，在实验室以正常的速度行走。如图16所示，我们遇到行人，光线较弱，纹理较少，玻璃和反射。视频可以在多媒体附件中找到。

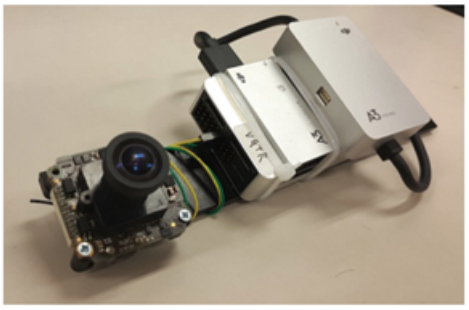
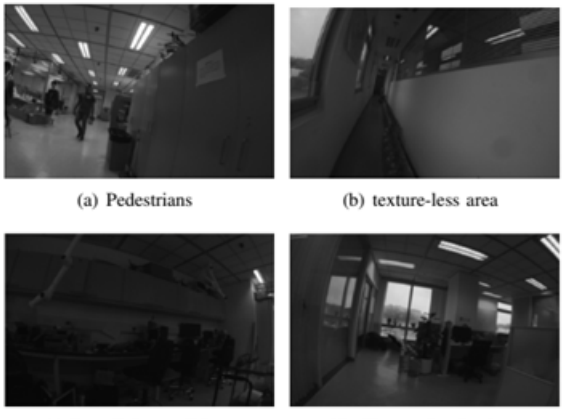


Fig. 15. The device we used for the indoor experiment. It contains one forward-looking global shutter camera (MatrixVision mvBlueFOX-MLC200w) with 752×480 resolution. We use the built-in IMU (ADXL278 and ADXRS290, 100Hz) for the DJI A3 flight controller.

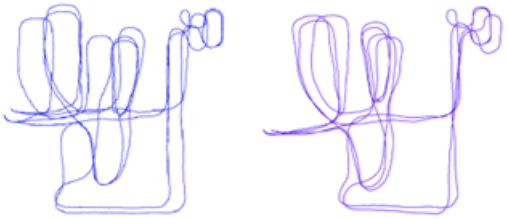
我们将我们的结果与OKVIS进行了比较，如图17所示。图17(a)是OKVIS的VIO输出。图17(b)是所提出的无回环方法的VIO结果。图17(c)是所提出的具有重新定位和环路闭合的方法的结果。当我们在室内循环时，会出现明显的漂移。OKVIS和只有VIO版本的vins-Mono在x,y,z和偏航角上积累了大量漂移.我们的重新定位和循环关闭模块有效地消除了所有这些漂移。



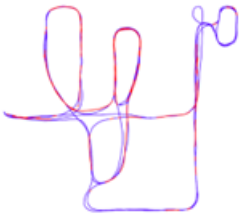
(a) Pedestrians (b) texture-less area

(c) Low light condition (d) Glass and reflection

Fig. 16. Sample images for the challenging indoor experiment.



(a) Trajectory of OKVIS. (b) Trajectory of VINS-Mono without loop closure.



(c) Trajectory of VINS-Mono with relocalization and loop closure. Red lines indicate loop detection.

Fig. 17. Results of the indoor experiment with comparison against OKVIS.

C. Large-scale Environment

1)走出实验室：我们在室内和室外混合的环境中测试vins-mono。传感器套件与图15所示的相同。我们从实验室的一个座位上开始，在室内空间里走来走去。然后我们下了楼梯，在大楼外的操场上走来走去。接下来，我们回到楼里上楼。最后，我们回到了实验室的同一个座位。整个轨道超过700米，持续约10分钟。在多媒体附件中可以找到实验的视频。

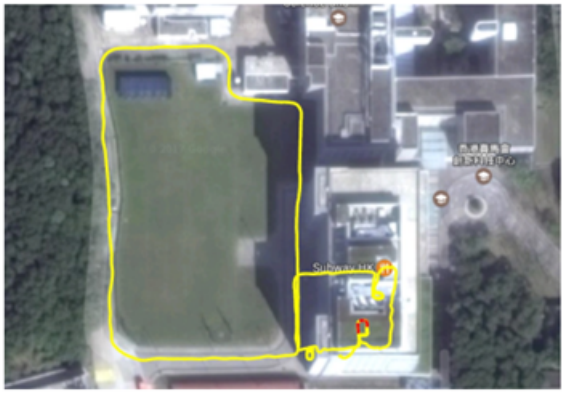
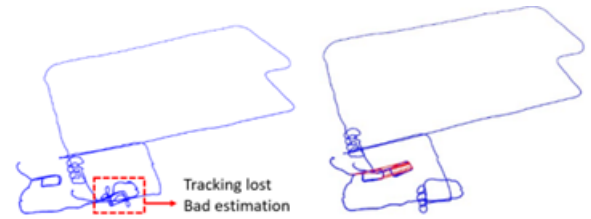
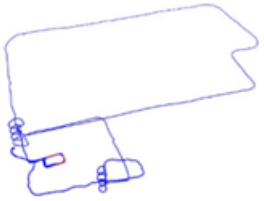


Fig. 18. The estimated trajectory of the mixed indoor and outdoor experiment aligned with Google Map. The yellow line is the final estimated trajectory from VINS-Mono after loop closure. Red lines indicate loop closure.

轨迹如图19所示。图19(a)是OKVIS的轨迹。当我们上楼时，OKVIS显示出不稳定的特征跟踪，导致估计错误。我们看不到红色街区楼梯的形状。VINS-Mono的唯一结果如图19(b)所示.有闭环的轨迹如图19(c)所示。该方法的楼梯形状清晰。闭环轨迹与谷歌地图对齐，以验证其准确性，如图18所示。



(a) Estimated trajectory from OKVIS (b) Estimated trajectory from VINS-Mono with loop closure disabled



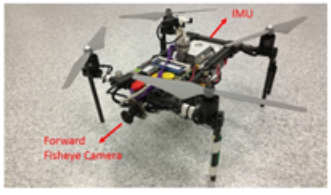
(c) Estimated trajectory from VINS-Mono with loop closure

Fig. 19. Estimated trajectories for the mixed indoor and outdoor experiment. In Fig. 19(a), results from OKVIS went bad during tracking lost in texture-less region (staircase). Figs. 19(b) 19(c) shows results from VINS-Mono without and with loop closures. Red lines indicate loop closures. The spiral-shaped blue line shows the trajectory when going up and down the stairs. We can see that VINS-Mono performs well (subject to acceptable drift) even without loop closure.

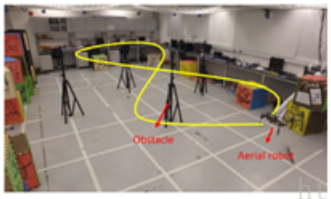
OKVIS x,y和z轴的最终漂移为[13.80,-5.26,7.23]米。VINS-Mono无环闭路的最终方向为[-5.47,2.76,-0.29]m，占总弹道长度的0.88%，小于OKVIS的2.36%。经回环修正，最终漂移有界于[-0.032,0.09,-0.07]m，与总弹道长度相比，这是微不足道的。虽然我们没有地面真值，但我们仍然可以直观地检查优化后的轨道是否平滑，并能精确地与卫星地图对齐。

2)环游校园：这张环绕整个科大校园的非常大规模的数据集是用一个手持的VI-Sensor 4记录下来的。该数据集覆盖的地面长度约为710米，宽度为240米，高度变化为60米。总路径长度为5.62km。数据包含25赫兹图像和200赫兹IMU，持续1小时34分钟。对VINS-Mono的稳定性和耐久性进行测试是一个非常有意义的实验。

Fig. 20. The estimated trajectory of the very large-scale environment aligned with Google map. The yellow line is the estimated trajectory from VINS-Mono. Red lines indicates loop closure.



(a) Aerial robot testbed



(b) Testing environment and desired figure eight pattern

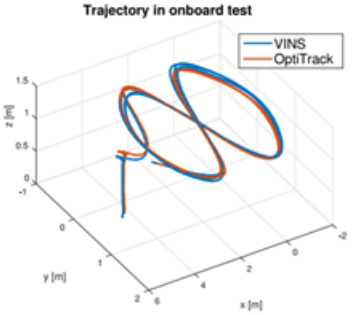


Fig. 22. The trajectory of loop closure-disabled VINS-Mono on the MAV platform and its comparison against the ground truth. The robot follows the trajectory four times. VINS-Mono estimates are used as real-time position feedback for the controller. Ground truth is obtained using OptiTrack. Total length is 61.97m. Finial drift is 0.18m.

在这个大规模的测试中，我们将关键帧数据库的大小设置为2000，以提供足够的循环信息并实现实时性能。我们运行此数据集时，英特尔i7-4790 CPU运行在3.60GHz。时间统计数据显示在表中。估计的轨迹与图20中的谷歌地图一致。与谷歌地图相

比，我们的结果在这个非常长时间的测试中几乎没有漂移。

TABLE I
TIMING STATISTICS

Tread	Modules	Time (ms)	Rate (Hz)
1	Feature detector	15	25
	KLT tracker	5	25
2	Window optimization	50	10
3	Loop detection	100	
	Pose graph optimization	130	

D. Application I: Feedback Control of an Aerial Robot

如图21(a)所示，我们将vins-Mono应用于航空机器人的自主反馈控制。我们使用了一个具有752×480分辨率的前瞻性全球快门相机(MatrixVisionMvBlueFOXMLC200w)，并配备了190度鱼眼镜头。DJIA3飞行控制器用于IMU测量和姿态稳定控制。星载计算资源是Intel i7-5500 U CPU，运行在3.00GHz。传统的针孔摄像机模型不适用于大视场摄像机。我们使用MEI模型对此相机进行校准，由[43]中引入的工具包进行校准。

在本实验中，我们使用VINS_Mono的状态估计来测试自主轨迹跟踪的性能。在这个实验中，回环检测被禁止。四转子被命令跟踪一个图8模式，每个圆圈半径为1.0米，如图21(b)所示。在弹道周围设置了四个障碍物，以验证VINS-Mono无闭环的准确性。在实验过程中，四转子连续四次跟踪这一轨迹。100 Hz星载状态估计支持对四转子的实时反馈控制.

地面真相是用OptiTrack 5获得的。总弹道长度为61.97 m。最终漂移为[0.08，0.09，0.13]m，结果为0.29%的位置漂移。平移和旋转的细节以及它们相应的误差如图23所示。

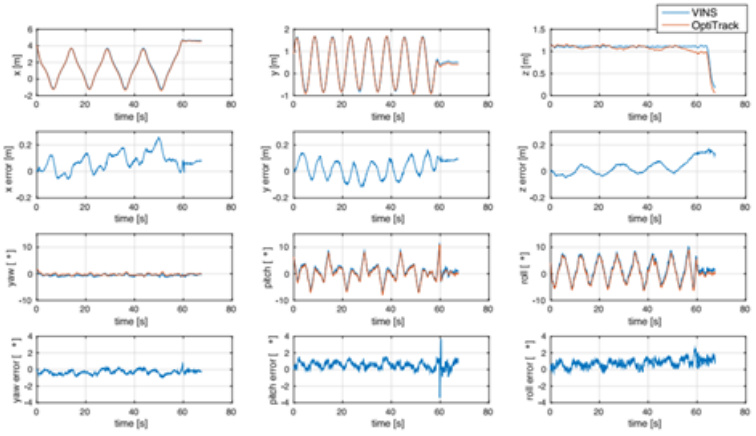


Fig. 23. Position, orientation and their corresponding errors of loop closure-disabled VINS-Mono compared with OptiTrack. A sudden in pitch error at the 60s is caused by aggressive breaking at the end of the designed trajectory, and possible time misalignment error between VINS-Mono and OptiTrack.

E. Application II: Mobile Device

我们将vins-Mono移植到移动设备上，并提供一个简单的AR应用程序来展示其准确性和健壮性。我们将我们的移动实现命名为vins-Mobile6，并将其与GoogleTangoDevice 7进行了比较，后者是移动平台上商业上最好的增强现实解决方案之一。

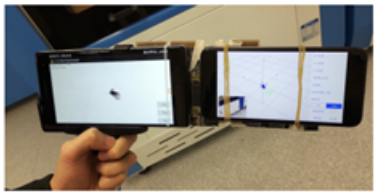
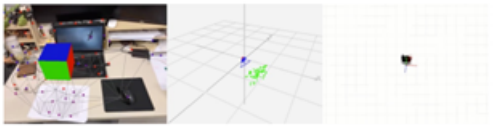


Fig. 24. A simple holder that we used to mount the Google Tango device (left) and the iPhone7 Plus (right) that runs our VINS-Mobile implementation.

VINS-Mono运行在iPhone7Plus上。我们使用iphone采集的30 Hz图像，分辨率640×480，以及内置的InvenSenseMP67B 6轴陀螺仪和加速度计获得的100 Hz IMU数据。如图24所示，我们将iphone与一个启用Tango功能的联想phab 2 Pro一起安装。探戈装置使用全局快门、鱼眼相机和同步IMU进行状态估计。首先，我们在平面上插入一个虚拟立方体，该虚拟立方体是从估计

的视觉特征中提取出来的，如图25(a)所示。然后，我们拿着这两个装置，以正常的速度在房间内外行走。当检测到回路时，我们使用四自由度姿态图优化.消除x,y,z和偏航漂移。



(a) Beginning

(b) Door opening

(c) End

Fig. 25. From left to right: AR image from VINS-Mobile, estimated trajectory from VINS-Mono, estimated trajectory from Tango Fig: 25(a): Both VINS-Mobile and Tango are initialized at the start location and a virtual box is inserted on the plane which extracted from estimated features. Fig. 25(b): A challenging case in which the camera is facing a moving door. The drift of Tango trajectory is highlighted. Fig. 25(c): The final trajectory of both VINS-Mobile and Tango. The total length is about 264m.

有趣的是，当我们打开一扇门时，探戈的偏航估计会跳到一个很大的角度，如图25(b)所示。其原因可能是由于不稳定的特征跟踪或主动故障检测和恢复而导致的估计器崩溃。然而，vins-Mono在这个具有挑战性的案例中仍然工作得很好。旅行了大约264米后，我们回到起点。最后的结果可以在图25(c)中看到，探戈的轨迹在最后一圈会漂移，而我们的vins会回到起点。通过对四自由度姿态图的优化，消除了总轨迹的漂移.这也证明了，与开始相比，立方体被标记到图像上的同一位置。

诚然，探戈比我们的实施更准确，尤其是对于局部状态的估计。实验结果表明，该方法可以在通用移动设备上运行，并且具有比较特殊工程设备的潜力。实验还证明了该方法的鲁棒性。视频可以在多媒体附件中找到。

X. CONCLUSION AND FUTURE WORK

本文提出了一种鲁棒、通用的单目视觉惯性估计器.我们的方法既具有先进的和新的解决方案的IMU预积分，估计器初始化和故障恢复，在线外部校准，紧耦合视觉惯性校正，重新定位，和有效的全局优化。我们通过与最先进的开源实现和高度优化的行业解决方案进行比较，显示出更好的性能。我们开放个人电脑和iOS的实现，以造福社会。

虽然基于特征的VINS估计器已经达到了现实部署的成熟程度，但我们仍然看到了未来研究的许多方向。单目VINS可能会根据运动和环境而达到弱可观测甚至退化的状态。我们最感兴趣的是在线方法来评估单目vins的可观测性，以及在线生成运动计划来恢复可观测性。另一个研究方向是在大量消费设备上大规模部署单目VINS，例如移动电话。这一应用要求在线校准几乎所有传感器的内在和外部参数，以及在线识别校准质量。最后，我们感兴趣的是制作由单目vins给出的稠密地图。我们在[44]中首次给出了用于无人机导航的单目视觉-惯性稠密地图的结果。然而，为了进一步提高系统的精度和鲁棒性，还需要进行广泛的研究。