

# VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator

一种稳健且多功能的单目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) 形成一个单目视觉惯性系统 (VINS)，这是用于度量六自由度状态估计的最小传感器套件(在尺寸、重量和功率上)。在本文中,我们提出了 VINS-Mono: 一种强大且通用的单目视觉惯性状态估计器。我们的方法从估计器初始化的稳健过程开始。通过融合预集成的 IMU 测量和特征观察,使用一种紧耦合的、基于非线性优化的方法来获得高精度的视觉惯性里程计。闭环检测模块与我们的紧耦合公式相结合,能够以最少的计算实现重定位。我们另外执行 4-DOF 位姿图优化以加强全局一致性。此外,所提出的系统可以通过有效的方式保存和加载地图来重用地图。当前和以前的地图可以通过全局位姿图优化合并在一起。我们在公共数据集和真实世界的实验上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。我们还在微型飞行器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,拟议的工作是一个可靠、完整且通用的系统,适用于需要高精度定位的不同应用。

**关键词:** 单目视觉惯性系统 (VINS)、状态估计、传感器融合、同时定位和建图

## 1、 介绍

状态估计无疑是机器人导航、自动驾驶、虚拟现实和增强现实 (AR) 等广泛应用的最基础模块。仅使用单目相机的方法因其体积小、成本低和易于硬件设置而在该领域获得了极大的兴趣 [1]-[5]。然而, 仅单目视觉系统无法恢复现实中的尺度, 因此限制了它们在现实世界机器人应用中的使用。最近, 我们看到了用低成本惯性测量单元 (IMU) 辅助单目视觉系统的增长趋势。这种单目视觉惯性系统 (VINS) 的主要优点是可以观察现实中的尺度以及横滚角和俯仰角。这使得需要度量状态估计的导航任务成为可能。此外, IMU 测量的集成可以弥合由于照明变化、无纹理区域或运动模糊造成的视觉轨迹损失之间的差距, 从而显著提高运动跟踪性能。单目 VINS 不仅可广泛用于地面机器人和无人机, 而且可用于移动设备。对于自我和环境感知, 它在尺寸、重量和功耗方面都具有很大的优势。

然而, 有几个问题影响单目 VINS 的使用。第一个是严格的初始化。由于缺乏直接的距离测量, 很难将单目视觉结构与惯性测量直接融合。还认识到 VINS 是高度非线性的事实, 我们在估计器初始化方面看到了重大挑战。在大多数情况下, 系统应该从已知的固定位置发射, 并在开始时缓慢小心地移动, 这限制了它在实践中的使用。另一个问题是视觉惯性里程计 (VIO) 无法避免长期漂移。为了消除漂移, 必须开发闭环检测、重定位和全局优化。除了这些关键问题外, 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, 一种强大且

通用的单目视觉惯性状态估计器，它是我们之前三项工作 [6]-[8] 的组合和扩展。VINS-Mono 包含以下功能：

- (1) 强大的初始化程序，能够从未知的初始状态引导系统；
- (2) 紧密耦合、基于优化的单目 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) 全局位姿图优化；
- (4) 姿态图重用，可以保存、加载和合并多个局部姿态图。

在这些特性中，鲁棒初始化、重定位和位姿图重用是我们的技术贡献，它们来自我们之前的作品[6]-[8]。工程贡献包括开源系统集成、无人机导航实时演示和移动应用程序。

整个系统已成功应用于小型 AR 场景、中型无人机导航、大型状态估计任务，如图 1 所示。

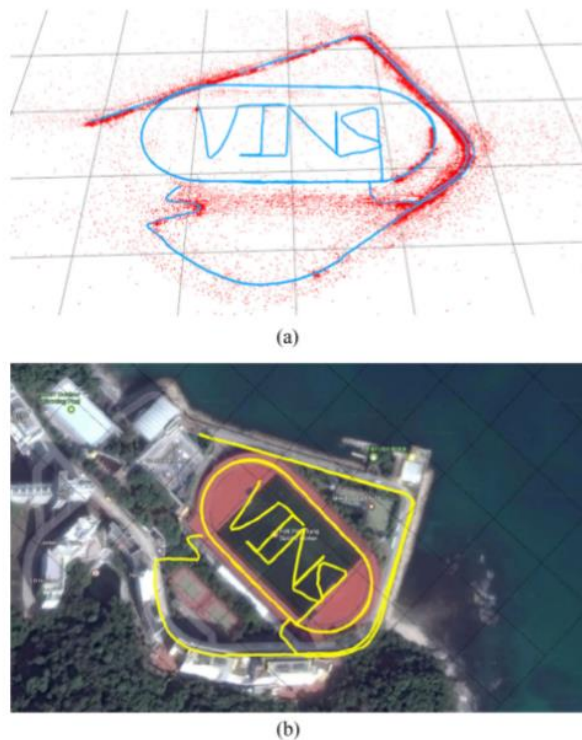


图 1. 所提出的单目视觉惯性状态估计器的户外实验结果。 在正常

步行条件下,通过手持单目相机-IMU 设置收集数据。它包括场地内的两个完整圆圈和附近车道上的两个半圆圈。总轨迹长度为 2.5 公里。实验视频可在多媒体附件中找到。(a) 轨迹(蓝色)和特征位置(红色)。(b) 与谷歌地图叠加的轨迹用于视觉比较。

本文的其余部分结构如下。在第二节中,我们讨论了相关文献。我们在第三节中概述了完整的系统流程。视觉和预集成 IMU 测量的预处理步骤在第四节中介绍。在第五节中,我们讨论了估计器的初始化过程。第六节介绍了一种紧密耦合、自校准、基于非线性优化的单目 VIO。第七节介绍了紧耦合的重定位。第八节讨论了全局位姿图优化和重用。实验结果显示在第九节中。最后,本文以第十节中的讨论和可能的未来研究方向作为结论。

## 2、 相关工作

关于基于单目视觉的状态估计/里程计/SLAM 的学术著作非常广泛。值得注意的方法包括 PTAM [1]、SVO [2]、LSD-SLAM [3]、DSO [5] 和 ORB-SLAM [4]。很明显,任何提供完整相关评论的尝试都是不完整的。然而,在本节中,我们跳过对纯视觉方法的讨论,只关注单目视觉惯性状态估计的最相关结果。

处理视觉和惯性测量的最简单方法是松耦合传感器融合 [9]、[10],其中 IMU 被视为一个独立模块来辅助视觉结构。融合通常由扩展卡尔曼滤波器 (EKF) 完成,其中 IMU 用于状态传播,而 visiononly pose 用于更新。此外,紧耦合的视觉惯性算法基于 EKF [11][13] 或图形优化 [14]-[19],其中相机和 IMU 测量从原始测量级

别联合优化。一种流行的基于 EKF 的 VIO 方法是 MSCKF [11], [12]。MSCKF 在状态向量中保留几个先前的相机姿势, 并使用跨多个相机视图的相同特征的视觉测量来形成多约束更新。SR-ISWF [20]、[21] 是 MSCKF 的扩展。它使用平方根形式 [14] 来实现单精度表示并避免不良的数值特性。这种方法采用逆滤波器进行迭代重新线性化, 使其等同于基于优化的算法。批图优化或捆绑调整技术维护和优化所有测量以获得最佳状态估计。为了实现恒定的处理时间, 基于图的 VIO 方法 [15]、[17]、[18] 通常通过边缘化过去的状态和测量来优化最近状态的有界大小的滑动窗口。由于非线性系统迭代求解的高计算需求, 很少有基于图的方法可以在资源受限的平台 (例如手机) 上实现实时性能。

对于视觉测量处理, 根据残差模型的定义, 算法可以分为直接法和间接法。直接方法 [2]、[3]、[22] 最小化光度误差, 而间接方法 [12]、[15]、[17] 最小化几何位移。由于吸引区域较小, 直接方法需要良好的初始猜测, 而间接方法在提取和匹配特征时会消耗额外的计算资源。由于其成熟性和稳健性, 间接方法在现实世界的工程部署中更常见。然而, 直接方法更容易扩展到密集建图, 因为它们直接在像素级别上操作。

IMU 通常以比相机高得多的速率获取数据。已经提出了不同的方法来处理高速 IMU 测量。最直接的方法是在基于 EKF 的方法 [9]、[11] 中使用 IMU 进行状态传播。在图形优化公式中, 开发了一种称为 IMU 预积分的有效技术, 以避免重复的 IMU 重新积分。

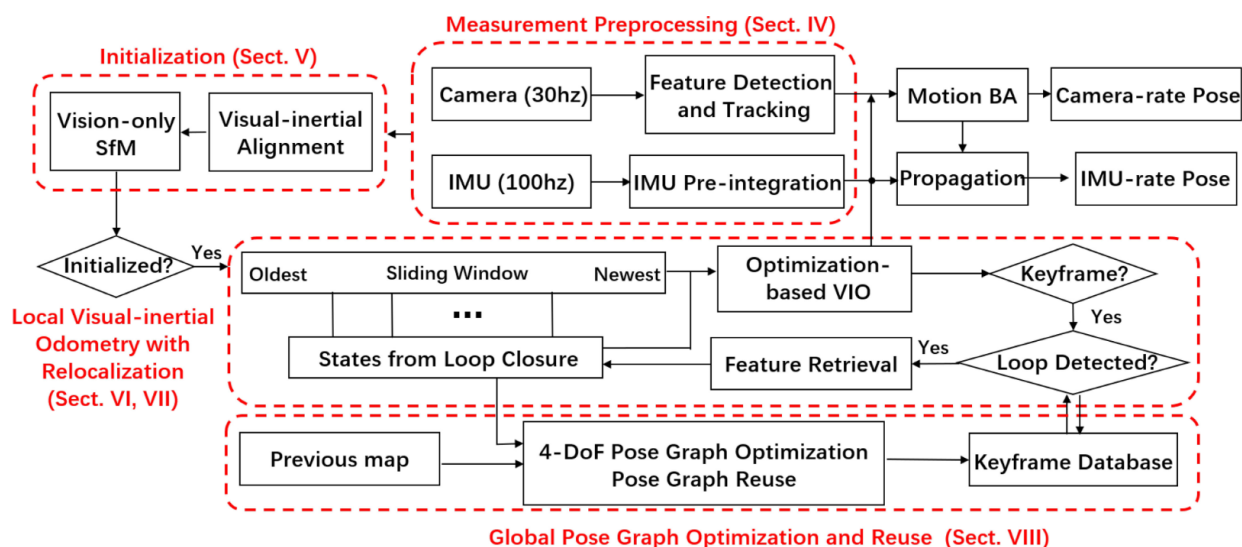
这种技术首先在 [23] 中引入, 它使用欧拉角参数化旋转误差。 [16] 使用连续时间误差状态动力学导出协方差传播。通过添加后验 IMU 偏差校正, 预积分理论在 [19] 和 [24] 中得到进一步改进。

准确的初始值对于引导任何单目 VINS 至关重要。 [17] 和 [25] 中提出了一种利用短期 IMU 预积分的相对旋转的线性估计器初始化方法。该方法无法对原始投影方程中的陀螺仪偏差和图像噪声进行建模。 [26] 中介绍了单眼视觉惯性初始化问题的封闭形式解决方案。后来, 在 [27] 中提出了通过添加陀螺仪偏置校准来扩展这种封闭形式的解决方案。这些方法无法对惯性积分中的不确定性进行建模, 因为它们依赖于 IMU 测量值在较长时间内的双重积分。在[28]中,提出了一种基于 SVO[2]的重新初始化和故障恢复算法。需要一个额外的向下距离传感器来恢复公制刻度。 [18] 中介绍了一种建立在流行的 ORB-SLAM [4] 之上的初始化算法。据报道, 尺度收敛所需的时间可以超过 10 s。这可能会给需要在一开始就进行比例估计的机器人导航任务带来问题。

里程计方法, 无论它们所依赖的基础数学公式如何, 都会在全局平移和方向上长期漂移。为此, 闭环在长期运行中发挥着重要作用。ORB-SLAM [4] 能够关闭循环并重用地图, 这利用了词袋 [29]。 7-DOF [30] (位置、方向和比例) 位姿图优化之后是闭环检测。



### 3、 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（参见第四节），其中提取和跟踪特征，并对两个连续帧之间的 IMU 测量值进行预积分。初始化过程（参见第五节）提供了所有必要的值，包括位姿、速度、重力矢量、陀螺仪偏差和三维 (3-D) 特征位置，用于引导后续基于非线性优化的 VIO。带有重定位（见第七节）模块的 VIO（见第六节）紧密融合了预集成的 IMU 测量、特征观察。最后，位姿图优化模块（参见第八节）采用经过几何验证的重定位结果，并执行全局优化以消除漂移。它还实现了位姿图的重用。VIO 和位姿图优化模块在单独的线程中同时运行。

与适用于立体相机的最先进的 VIO 算法 OKVIS [15] 相比，我们的算法专为单目相机设计。因此，我们特别提出了初始化程序、关键帧选择标准，并使用和处理大视场 (FOV) 相机以获得更好的跟踪性能。此外，我们的算法提供了一个完整的系统，带有闭环和位姿图重用模块。

我们现在定义在本文中使用的符号和框架。我们将  $(\cdot)^w$  视为世界坐标系。重力方向与世界坐标系的  $z$  轴对齐。 $(\cdot)^b$  为车体坐标系，且定义与 IMU frame 相同。 $(\cdot)^c$  是相机坐标系。我们使用旋转矩阵  $R$  和汉密尔顿四元数  $q$  来表示旋转。我们主要在状态向量中使用四元数，但旋转矩阵也用于 3-D 向量的 convenience rotation。 $q_b^w$  和  $p_b^w$  是从车体坐标系到世界坐标系的旋转和平移。 $b_k$  是拍摄第  $k$  张图像时的车体坐标系。 $c_k$  是拍摄第  $k$  张图像时的相机帧。 $\otimes$  表示两个四元数之间的乘法运算。 $g^w = [0, 0, g]^T$  是世界坐标系中的重力矢量。最后，我们将  $(\cdot)^\wedge$  表示为某个量的噪声测量或估计。

#### 4、 测量预处理

本节介绍惯性和单目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，我们将它们预先集成在两个连续帧之间。

##### A、 视觉处理前端

对于每个新图像，现有特征由 KLT 稀疏光流算法 [31] 跟踪。同时，检测到新的角点特征 [32] 以在每幅图像中保持最小数量 (100~300) 的特征。检测器通过设置两个相邻特征之间的最小像素间隔来强制执行均匀的特征分布。二维 (2-D) 特征首先不失真，然后在通过异常值拒绝后投影到单位球体。使用 RANSAC 和基本矩阵模型进行异常值剔除 [33]。

在此步骤中还选择了关键帧。我们有两个关键帧选择标准。第一个是与前一个关键帧的平均视差。如果跟踪特征的平均视差在当



前帧和最新关键帧之间超过一定阈值，我们将帧视为新关键帧。 请注意，不仅平移而且旋转都会导致视差。 但是，不能在仅旋转运动中对特征进行三角剖分。 为了避免这种情况，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，不涉及 VINS 公式中的旋转计算。 为此，即使陀螺仪含有较大噪声或存在偏差，也只会导致关键帧选择结果不理想，不会直接影响估计质量。 另一个标准是跟踪质量。 如果跟踪特征的数量低于某个阈值，我们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。

## B、 IMU 预积分

我们遵循之前基于连续时间四元数的 IMU 预积分推导 [16], 并将 IMU 偏差的处理包括在 [19] 和 [24] 中。 我们注意到当前的 IMU 预积分程序与 [19] 和 [24] 共享几乎相同的数值结果，但使用不同的推导。 所以，我们这里只做一个简单的介绍。 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**: 在车身框架中测量的 IMU 测量值结合了反重力和平台动力学，并受加速度偏置 $b_a$ 、陀螺仪偏置 $b_w$ 和附加噪声的影响。 原始陀螺仪和加速度计测量值 $\hat{w}$ 和 $\hat{a}$ 由下式给出：

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

我们假设加速度和陀螺仪测量中的附加的噪声是高斯白噪声， $n_a \sim \mathcal{N}(0, \sigma_a^2)$ ， $n_w \sim \mathcal{N}(0, \sigma_w^2)$ 。 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， $n_{b_a} \sim \mathcal{N}(0, \sigma_{b_a}^2)$ ， $n_{b_w} \sim \mathcal{N}(0, \sigma_{b_w}^2)$ ：

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

(2) **预积分**：对于两个时间连续的帧 $b_k$ 和 $b_{k+1}$ ，在时间区间 $[t_k, t_{k+1}]$ 内存在多个惯性测量值。给定偏差估计，我们将它们整合到局部坐标系 $\mathbf{b}_k$ 中为：

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &= \iint_{t \in [t_k, t_{k+1}]} \mathbf{R}_t^{b_k} (\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}) dt^2 \\ \beta_{b_{k+1}}^{b_k} &= \int_{t \in [t_k, t_{k+1}]} \mathbf{R}_t^{b_k} (\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}) dt \\ \gamma_{b_{k+1}}^{b_k} &= \int_{t \in [t_k, t_{k+1}]} \frac{1}{2} \boldsymbol{\Omega}(\hat{\boldsymbol{\omega}}_t - \mathbf{b}_{w_t}) \gamma_t^{b_k} dt \end{aligned} \quad (3)$$

where

$$\boldsymbol{\Omega}(\boldsymbol{\omega}) = \begin{bmatrix} -[\boldsymbol{\omega}]_{\times} & \boldsymbol{\omega} \\ -\boldsymbol{\omega}^T & 0 \end{bmatrix}, [\boldsymbol{\omega}]_{\times} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \quad (4)$$

$\alpha$ 、 $\beta$ 和 $\gamma$ 的协方差 $P_{b_{k+1}}^{b_k}$ 也相应地传播。可以看出，预积分项 (3)

可以通过将 $b_k$ 作为给定偏差的参考系，仅通过 IMU 测量获得。

### (3) **bias 校正**

如果偏差的估计变化很小，我们调整 $\alpha_{b_{k+1}}^{b_k}$ ， $\beta_{b_{k+1}}^{b_k}$ ，和 $\gamma_{b_{k+1}}^{b_k}$ 通过它们关于 bias 的一阶近似值作为

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &\approx \hat{\alpha}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_a}^{\alpha} \delta \mathbf{b}_{a_k} + \mathbf{J}_{b_w}^{\alpha} \delta \mathbf{b}_{w_k} \\ \beta_{b_{k+1}}^{b_k} &\approx \hat{\beta}_{b_{k+1}}^{b_k} + \mathbf{J}_{b_a}^{\beta} \delta \mathbf{b}_{a_k} + \mathbf{J}_{b_w}^{\beta} \delta \mathbf{b}_{w_k} \\ \gamma_{b_{k+1}}^{b_k} &\approx \hat{\gamma}_{b_{k+1}}^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} \mathbf{J}_{b_w}^{\gamma} \delta \mathbf{b}_{w_k} \end{bmatrix}. \end{aligned} \quad (5)$$

否则，当 bias 的估计值发生显著变化时，我们在新的 bias 估计

下进行重新传播。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。

## 5、 估计器初始化

单目紧耦合 VIO 是一个高度非线性的系统，需要在一开始就进行准确的初始猜测。我们通过将 IMU 预集成与仅视觉结构松散地对齐来获得必要的初始值。

### A、滑动窗口中的 Vision-Only SfM

初始化过程从仅视觉的 SfM 开始，以估计高比例相机位姿和特征位置的图形。

对于有限的计算复杂度，我们在一个滑动窗口中保持几个帧。首先，我们检查最新帧和所有先前帧之间的特征对应关系。如果我们能在滑动窗口中的最新帧和任何其他帧之间找到稳定的特征跟踪（超过 30 个跟踪的特征）和足够的视差（超过 20 个像素）。

我们使用五点算法[34]来恢复这两个帧之间的相对旋转和按比例平移。然后，我们任意设置尺度，并对在这两个帧中观察到的所有特征进行三角测量。基于这些三角化特征，执行透视点（PnP）方法[35]来估计窗口中所有其他帧的姿态。最后，应用全局全束平差[36]来最小化所有特征观测的总重投影误差。由于我们还没有任何关于世界帧的知识，我们将第一个相机坐标系 $(\cdot)^{c_0}$ 设置为 SfM 的参考帧。所有坐标系位姿 $(\bar{p}_{c_k}^{c_0}, q_{c_k}^{c_0})$ 和特征位置都相对于 $(\cdot)^{c_0}$ 表示。给定相机和 IMU 之间的外部参数 $(p_c^b, q_c^b)$ ，我们可以将位姿从相机坐标系转换为车体（IMU）坐标系。

$$\begin{aligned}
\mathbf{q}_{b_k}^{c_0} &= \mathbf{q}_{c_k}^{c_0} \otimes (\mathbf{q}_c^b)^{-1} \\
s\bar{\mathbf{p}}_{b_k}^{c_0} &= s\bar{\mathbf{p}}_{c_k}^{c_0} - \mathbf{R}_{b_k}^{c_0} \mathbf{p}_c^b
\end{aligned} \tag{6}$$

其中  $s$  是未知的缩放参数，将在接下来求解。

## B、视觉惯性对齐

视觉惯性对齐的图示如图 3 所示。基本思想是将 up-to-scale 视觉结构与 IMU 预积分相匹配。

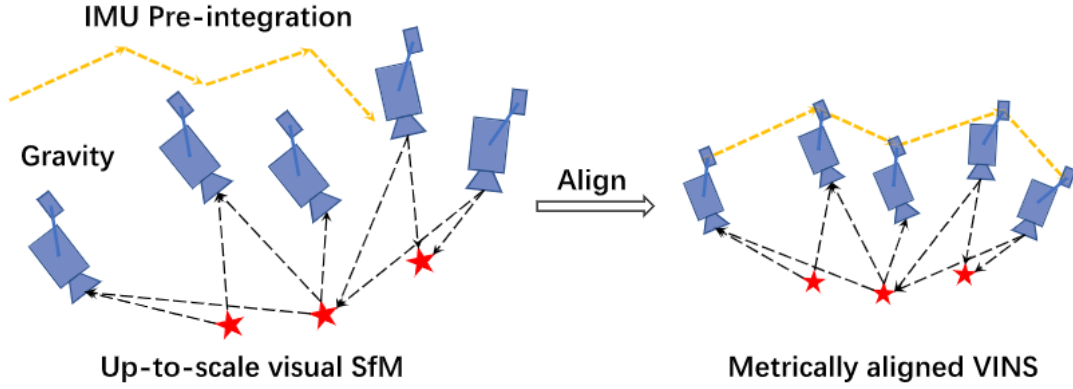


图 3、估计器初始化的视觉惯性对齐过程的图示。基本思想是将 up-to-scale 视觉结构与 IMU 预积分相匹配。

(1) 陀螺仪偏差校准：考虑窗口中的两个连续帧  $b_k$  和  $b_{k+1}$ ，我们从视觉 SfM 得到旋转  $\mathbf{q}_{b_k}^{c_0}$  和  $\mathbf{q}_{b_{k+1}}^{c_0}$ ，以及从 IMU 预积分得到相对约束  $\hat{\gamma}_{b_{k+1}}^{b_k}$ 。我们将 IMU 预积分项相对于陀螺仪偏差线性化，并最小化以下成本函数：

$$\begin{aligned}
\min_{\delta \mathbf{b}_w} \sum_{k \in \mathcal{B}} \left\| \mathbf{q}_{b_{k+1}}^{c_0}^{-1} \otimes \mathbf{q}_{b_k}^{c_0} \otimes \gamma_{b_{k+1}}^{b_k} \right\|^2 \\
\gamma_{b_{k+1}}^{b_k} \approx \hat{\gamma}_{b_{k+1}}^{b_k} \otimes \begin{bmatrix} 1 \\ \frac{1}{2} \mathbf{J}_{b_w}^\gamma \delta \mathbf{b}_w \end{bmatrix}
\end{aligned} \tag{7}$$

其中  $\mathcal{B}$  索引窗口中的所有坐标系。这样，我们得到了陀螺仪 bias  $b_w$  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  $\hat{\alpha}_{b_{k+1}}^{b_k}$ 、 $\hat{\beta}_{b_{k+1}}^{b_k}$  和  $\hat{\gamma}_{b_{k+1}}^{b_k}$ 。

(3) *Velocity, Gravity Vector, and Metric Scale Initialization*: 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，重力矢量和公制尺度。

$$\mathcal{X}_I = [\mathbf{v}_{b_0}^{b_0}, \mathbf{v}_{b_1}^{b_1}, \dots, \mathbf{v}_{b_n}^{b_n}, \mathbf{g}^{c_0}, s] \quad (8)$$

其中  $\mathbf{v}_{b_k}^{b_k}$  是拍摄第  $k$  张图像时身体坐标系中的速度， $\mathbf{g}^{c_0}$  是  $c_0$  坐标系中的重力矢量， $s$  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  $b_{k+1}$ ，我们有以下等式：

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &= \mathbf{R}_{c_0}^{b_k} (s(\bar{\mathbf{p}}_{b_{k+1}}^{c_0} - \bar{\mathbf{p}}_{b_k}^{c_0}) + \frac{1}{2} \mathbf{g}^{c_0} \Delta t_k^2 - \mathbf{R}_{b_k}^{c_0} \mathbf{v}_{b_k}^{b_k} \Delta t_k) \\ \beta_{b_{k+1}}^{b_k} &= \mathbf{R}_{c_0}^{b_k} (\mathbf{R}_{b_{k+1}}^{c_0} \mathbf{v}_{b_{k+1}}^{b_{k+1}} + \mathbf{g}^{c_0} \Delta t_k - \mathbf{R}_{b_k}^{c_0} \mathbf{v}_{b_k}^{b_k}). \end{aligned} \quad (9)$$

我们可以将 (6) 和 (9) 组合成如下线性测量模型：

$$\hat{\mathbf{z}}_{b_{k+1}}^{b_k} = \begin{bmatrix} \hat{\alpha}_{b_{k+1}}^{b_k} - \mathbf{p}_c^b + \mathbf{R}_{c_0}^{b_k} \mathbf{R}_{b_{k+1}}^{c_0} \mathbf{p}_c^b \\ \hat{\beta}_{b_{k+1}}^{b_k} \end{bmatrix} = \mathbf{H}_{b_{k+1}}^{b_k} \mathcal{X}_I + \mathbf{n}_{b_{k+1}}^{b_k} \quad (10)$$

where

$$\mathbf{H}_{b_{k+1}}^{b_k} = \begin{bmatrix} -\mathbf{I} \Delta t_k & \mathbf{0} & \frac{1}{2} \mathbf{R}_{c_0}^{b_k} \Delta t_k^2 \mathbf{R}_{c_0}^{b_k} (\bar{\mathbf{p}}_{c_{k+1}}^{c_0} - \bar{\mathbf{p}}_{c_k}^{c_0}) \\ -\mathbf{I} & \mathbf{R}_{c_0}^{b_k} \mathbf{R}_{b_{k+1}}^{c_0} & \mathbf{R}_{c_0}^{b_k} \Delta t_k & \mathbf{0} \end{bmatrix}. \quad (11)$$

可以看出， $\mathbf{R}_{b_k}^{c_0}$ ， $\mathbf{R}_{b_{k+1}}^{c_0}$ ， $\bar{\mathbf{p}}_{c_k}^{c_0}$ ， $\bar{\mathbf{p}}_{c_{k+1}}^{c_0}$  是从 up-to-scale 单目视觉 SfM 中得到的。 $\Delta t_k$  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

$$\min_{\mathcal{X}_I} \sum_{k \in \mathcal{B}} \left\| \hat{\mathbf{z}}_{b_{k+1}}^{b_k} - \mathbf{H}_{b_{k+1}}^{b_k} \mathcal{X}_I \right\|^2 \quad (12)$$

我们可以得到窗口中每一帧的体坐标系速度, 视觉参考坐标系中的重力矢量, 以及尺度参数 $(\cdot)^{c_0}$ 。

(4) Gravity Refinement: 从前面的线性初始化步骤中获得的重力向量可以通过约束大小来细化。在大多数情况下, 重力矢量的大小是已知的。这导致重力矢量只剩下 2-DOF。因此, 我们在其切线空间上用两个变量扰动引力, 它保留了 2-DOF。我们的重力矢量受到  $g(\bar{\hat{g}} + \delta g)$ ,  $\delta g = \omega_1 b_1 + \omega_2 b_2$  的扰动, 其中  $g$  为已知的重力大小,  $\bar{\hat{g}}$  为表示重力方向的单位向量。  $b_1$  和  $b_2$  是跨越切平面的两个正交基, 如图 4 所示。  $\omega_1$  和  $\omega_2$  分别是对  $b_1$  和  $b_2$  的二维扰动。我们可以任意找到任何一组  $b_1$  和  $b_2$  旋转切线空间。然后, 我们用  $g(\bar{\hat{g}} + \delta g)$  将  $g$  代入 (9), 并与其他状态变量一起求解  $2D \delta g$ 。这个过程重复几次直到  $\hat{g}$  收敛。

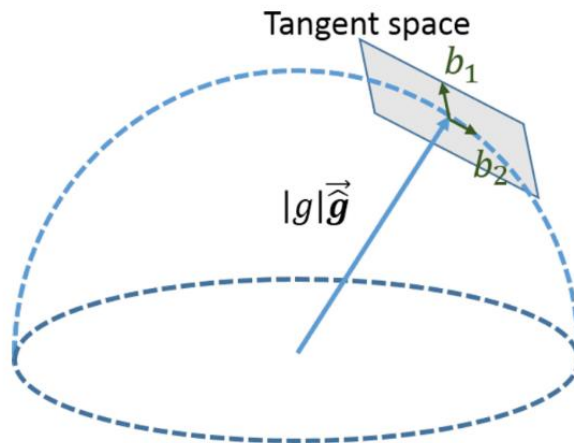


图 4、重力的 2-DOF 扰动示意图。由于重力大小已知,  $g$  位于半径为  $g \approx 9.81 \text{ m/s}^2$  的球体上。引力在当前估计周围受到扰动



$g(\bar{\hat{g}} + \delta g), \delta g = \omega_1 b_1 + \omega_2 b_2$ ，其中  $b_1$  和  $b_2$  是跨越切线空间的两个正交基。

(5) 完成初始化：对重力向量进行细化后，我们可以通过将重力旋转到  $z$  轴，得到世界坐标系和相机坐标系  $c_0$  之间的旋转  $q_{c_0}^w$ 。然后我们将所有变量从参考坐标系  $(\cdot)^{c_0}$  旋转到世界坐标系  $(\cdot)^w$ 。身体框架速度也将旋转到世界框架。来自视觉 SfM 的转换组件将被缩放为公制单位。此时，初始化过程完成，所有这些度量值将被馈送到紧耦合的单目 VIO。

## 6、紧耦合的单目 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目 VIO 来进行高精度和稳健的状态估计。滑动窗口公式的图示如图 5 所示。

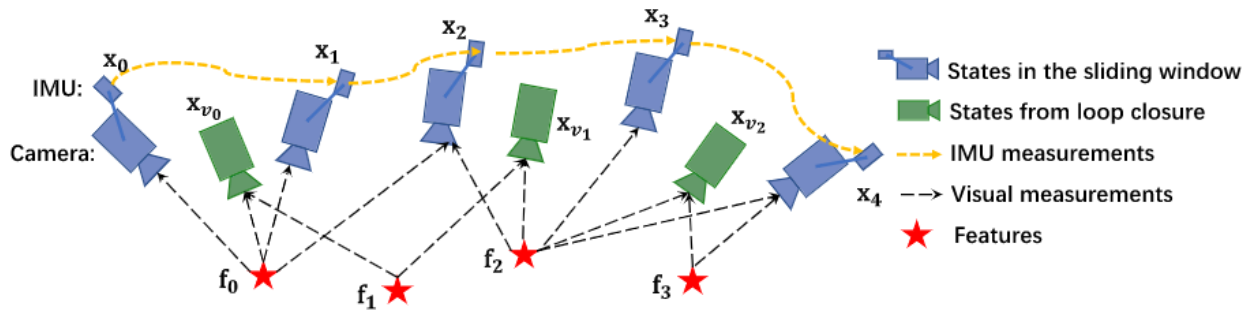


图 5、具有重定位的滑动窗口单目 VIO 的图示。滑动窗口中存在多个相机位姿、IMU 测量和视觉测量。它是与 IMU、视觉和闭环检测紧密耦合的公式。

### A、公式

滑动窗口中的全状态向量定义为：

$$\begin{aligned}
\mathcal{X} &= [\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{x}_c^b, \lambda_0, \lambda_1, \dots, \lambda_m] \\
\mathbf{x}_k &= [\mathbf{p}_{b_k}^w, \mathbf{v}_{b_k}^w, \mathbf{q}_{b_k}^w, \mathbf{b}_a, \mathbf{b}_g], k \in [0, n] \\
\mathbf{x}_c^b &= [\mathbf{p}_c^b, \mathbf{q}_c^b]
\end{aligned} \tag{13}$$

其中  $\mathbf{x}_k$  是拍摄第  $k$  张图像时的 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 主体坐标系中的加速度 bias 和陀螺仪 bias。  $n$  是关键帧总数， $m$  是滑动窗口中的特征总数。  $\lambda_l$  是第  $l$  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公式。我们最小化所有测量残差的先验和马氏范数之和以获得最大后验估计为：

$$\begin{aligned}
\min_{\mathcal{X}} \left\{ \|\mathbf{r}_p - \mathbf{H}_p \mathcal{X}\|^2 + \sum_{k \in \mathcal{B}} \left\| \mathbf{r}_B(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) \right\|_{\mathbf{P}_{b_{k+1}}^{b_k}}^2 \right. \\
\left. + \sum_{(l,j) \in \mathcal{C}} \rho(\|\mathbf{r}_C(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X})\|_{\mathbf{P}_l^{c_j}}^2) \right\} \tag{14}
\end{aligned}$$

where the Huber norm [37] is defined as

$$\rho(s) = \begin{cases} s & s \leq 1 \\ 2\sqrt{s} - 1 & s > 1. \end{cases} \tag{15}$$

其中  $\mathbf{r}_B(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X})$  和  $\mathbf{r}_C(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X})$  分别是 IMU 和视觉测量的残差。剩余项的详细定义将在第 6-B 和 6-C 节中介绍。  $\mathcal{B}$  是所有 IMU 测量值的集合， $\mathcal{C}$  是在当前滑动窗口中至少观察到两次的特征集合。  $\{\mathbf{r}_p, \mathbf{H}_p\}$  是来自边缘化的先验信息。 Ceres 求解器 [38] 用于解决这个非线性问题。

## B、IMU 测量残差

考虑滑动窗口中两个连续帧  $b_k$  和  $b_{k+1}$  内的 IMU 测量值，预积分 IMU 测量值的残差可以定义为

$$\begin{aligned}
\mathbf{r}_B(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) &= \begin{bmatrix} \delta \alpha_{b_{k+1}}^{b_k} \\ \delta \beta_{b_{k+1}}^{b_k} \\ \delta \theta_{b_{k+1}}^{b_k} \\ \delta \mathbf{b}_a \\ \delta \mathbf{b}_g \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{R}_w^{b_k} (\mathbf{p}_{b_{k+1}}^w - \mathbf{p}_{b_k}^w + \frac{1}{2} \mathbf{g}^w \Delta t_k^2 - \mathbf{v}_{b_k}^w \Delta t_k) - \hat{\alpha}_{b_{k+1}}^{b_k} \\ \mathbf{R}_w^{b_k} (\mathbf{v}_{b_{k+1}}^w + \mathbf{g}^w \Delta t_k - \mathbf{v}_{b_k}^w) - \hat{\beta}_{b_{k+1}}^{b_k} \\ 2 \left[ \mathbf{q}_{b_k}^{w^{-1}} \otimes \mathbf{q}_{b_{k+1}}^w \otimes (\hat{\gamma}_{b_{k+1}}^{b_k})^{-1} \right]_{xyz} \\ \mathbf{b}_{a b_{k+1}} - \mathbf{b}_{a b_k} \\ \mathbf{b}_{w b_{k+1}} - \mathbf{b}_{w b_k} \end{bmatrix} \quad (16)
\end{aligned}$$

其中  $[\cdot]_{xyz}$  提取四元数  $\mathbf{q}$  的向量部分以表示误差状态。 $\delta \theta_{b_{k+1}}^{b_k}$  是四元数的 3-D 误差状态表示。 $[\hat{\alpha}_{b_{k+1}}^{b_k}, \hat{\beta}_{b_{k+1}}^{b_k}, \hat{\gamma}_{b_{k+1}}^{b_k}]$  是两个连续图像帧之间的预积分 IMU 测量项。加速度计和陀螺仪 bias 也包含在在线校正的残差项中。

### C、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，我们在单位球体上定义相机测量残差。几乎所有类型相机的光学器件，包括广角、鱼眼或全向相机，都可以建模为连接单位球体表面的单位光线。考虑在第  $i$  个图像中首先观察到的第  $l$  个特征，第  $j$  个图像中特征观察的残差定义为：

$$\begin{aligned}
\mathbf{r}_c(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X}) &= [\mathbf{b}_1 \ \mathbf{b}_2]^T \cdot \left( \hat{\mathcal{P}}_l^{c_j} - \frac{\mathcal{P}_l^{c_j}}{\|\mathcal{P}_l^{c_j}\|} \right) \\
\hat{\mathcal{P}}_l^{c_j} &= \pi_c^{-1} \left( \begin{bmatrix} \hat{u}_l^{c_j} \\ \hat{v}_l^{c_j} \end{bmatrix} \right) \\
\mathcal{P}_l^{c_j} &= \mathbf{R}_b^c \left( \mathbf{R}_w^{b_j} \left( \mathbf{R}_{b_i}^w \left( \mathbf{R}_c^b \frac{1}{\lambda_l} \pi_c^{-1} \left( \begin{bmatrix} \hat{u}_l^{c_i} \\ \hat{v}_l^{c_i} \end{bmatrix} \right) \right. \right. \right. \right. \\
&\quad \left. \left. \left. + \mathbf{p}_c^b \right) + \mathbf{p}_{b_i}^w - \mathbf{p}_{b_j}^w \right) - \mathbf{p}_c^b \right) \quad (17)
\end{aligned}$$

其中  $[\hat{u}_l^{c_i}, \hat{v}_l^{c_i}]$  是第  $i$  个图像中第  $l$  个特征的第一个观察值。  
 $[\hat{u}_l^{c_j}, \hat{v}_l^{c_j}]$  是对第  $j$  个图像中相同特征的观察。  $\pi_c^{-1}$  是反投影函数，它使用相机固有参数将像素位置转换为单位向量。 由于视觉残差的自由度为二，我们将残差向量投影到切平面上。  $\mathbf{b}_1$  和  $\mathbf{b}_2$  是两个任意选择的正交基，跨越  $\hat{\mathcal{P}}_l^{c_j}$  的切平面，如图 6 所示。方差  $\mathcal{P}_l^{c_j}$ ，如 (14) 中使用的，也从像素坐标传播到单位球体。

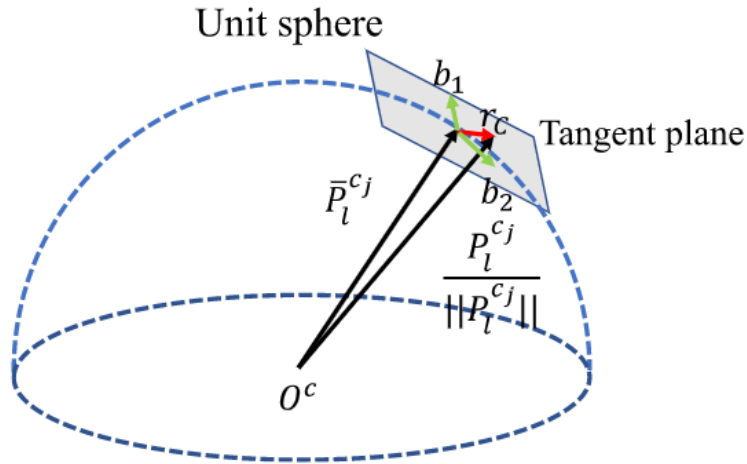


图 6、单位球体上视觉残差的图示。  $\hat{\mathcal{P}}_l^{c_j}$  是第  $j$  帧第  $l$  个特征观测的单位向量。  $\mathcal{P}_l^{c_j}$  是通过将其在第  $i$  帧中的第一个观测值转换到第  $j$  帧来预测单位球体上的特征测量。 残差定义在  $\hat{\mathcal{P}}_l^{c_j}$  的切平面上。

## D、边缘化

为了限制我们基于优化的 VIO 的计算复杂性，引入了边缘化。我们从滑动窗口中选择性地边缘化 IMU 状态  $x_k$  和特征  $\lambda_l$ ，同时将对应用于边缘化状态的测量值转换为先验值。

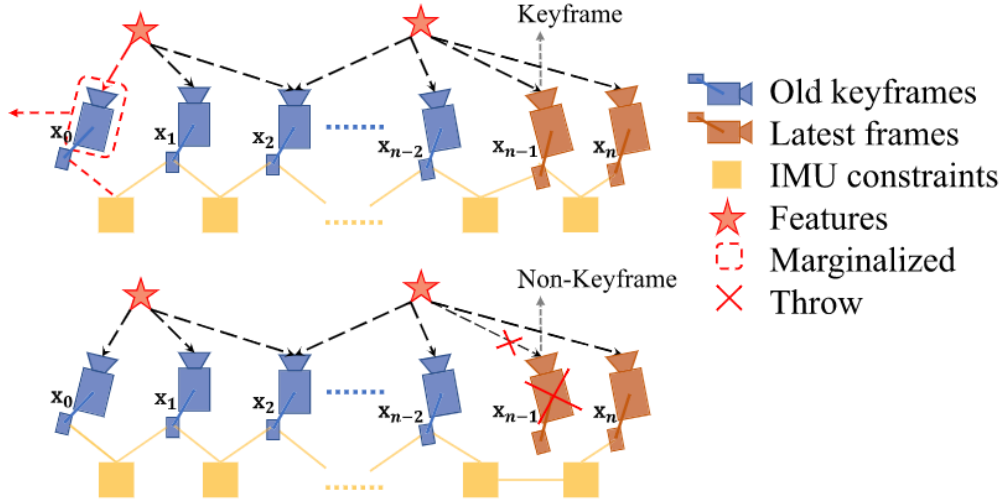


图 7、边缘化策略说明。如果第二个最新帧是关键帧，我们会将其保留在窗口中，并边缘化最旧的帧及其相应的视觉和惯性测量值。边缘化的测量变成了先验。如果第二个最新帧不是关键帧，我们将简单地删除该帧及其所有相应的视觉测量值。然而，预积分惯性测量保留用于非关键帧，并且预积分过程继续到下一帧。

如图 7 所示，当第二个最新帧是关键帧时，它将留在窗口中，而最旧的帧及其相应的测量值被边缘化。否则，如果第二个最新帧是非关键帧，我们将抛出视觉测量并保留连接到该非关键帧的 IMU 测量。我们不会边缘化非关键帧的所有测量以保持系统的稀疏性。我们的边缘化方案旨在保持窗口中空间分离的关键帧。这确保了特征三角测量有足够的视差，并最大限度地提高了在大激励下保持加速度计测量的可能性。

边缘化是使用 Schur 补码 [39] 进行的。我们基于与删除状态相关的所有边缘化测量构建一个新的先验。新的先验被添加到现有

的先验中。我们注意到边缘化导致线性化点的早期修复，这可能导致次优估计结果。然而，由于 VIO 可以接受小的漂移，我们认为边缘化造成的负面影响并不重要。

## E、 用于相机速率状态估计的纯运动视觉惯性优化

对于手机等计算能力较低的设备，由于非线性优化的大量计算需求，紧耦合的单目 VIO 无法实现相机速率输出。为此，除了全面优化外，我们还采用了轻量级的纯运动视觉惯性优化来将状态估计提高到相机速率 (30 Hz)。

仅运动视觉惯性优化的成本函数与 (14) 中单目 VIO 的成本函数相同。然而，我们不是优化滑动窗口中的所有状态，而是只优化固定数量的最新 IMU 状态的位姿和速度。我们将特征深度、外部参数、偏差和我们不想优化的旧 IMU 状态视为常量值。我们确实使用所有视觉和惯性测量来进行仅运动优化。这导致比单帧 PnP 方法更平滑的状态估计。所提出策略的图示如图 8 所示。与完全紧密耦合的单目 VIO 相比，在最先进的嵌入式计算机上可能会导致超过 50 毫秒，仅运动的视觉惯性优化仅 大约需要 5 毫秒来计算。这使得低延迟相机速率姿势估计成为可能，这对无人机和 AR 应用特别有益。



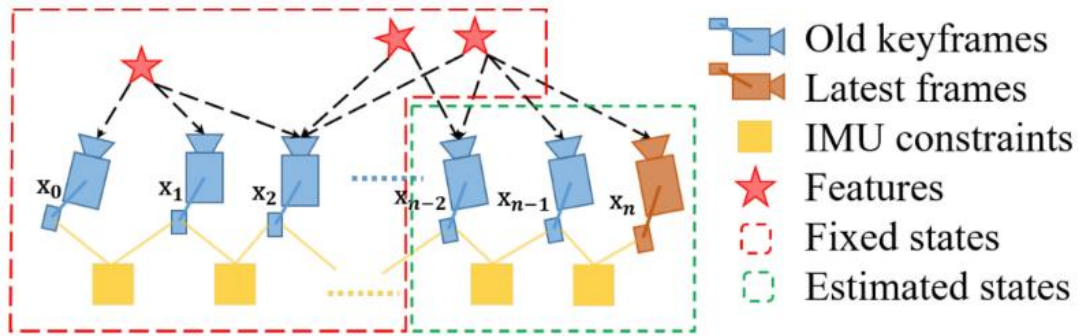


图 8、相机速率输出的纯运动优化图示

## F、 用于 IMU 速率状态估计的 IMU 前向传播

IMU 测量的速度比视觉测量高得多。尽管我们的 VIO 频率受到图像捕获频率的限制,但我们仍然可以使用最近的 IMU 测量值直接传播最新的 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。第 9-C 节介绍了利用此 IMU 速率状态估计的自主飞行实验。

## 7、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,但它也为系统引入了累积漂移。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,该模块识别已经访问过的地方。然后建立回环候选者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目 VIO 模块中,从而以最少的计算量实现无漂移状态估计。多个特征的多次观察直接用于重定位,从而获得更高的精度和更好的状态估计平滑度。重定位过程的图形说明如图 9 (a) 所示。

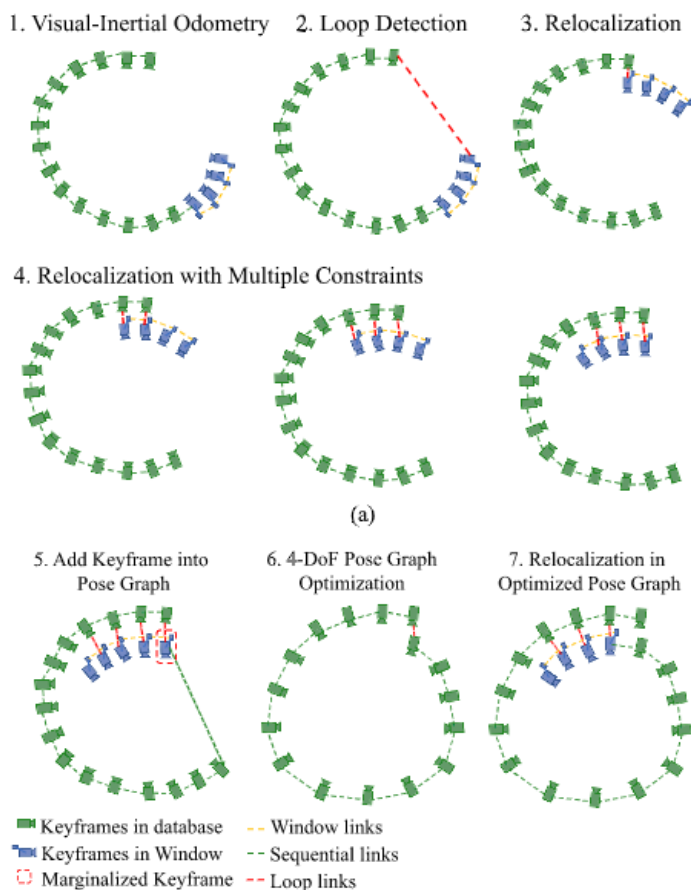


图 9. 说明重定位和位姿图优化过程的图表。(a) 重定位程序。它从仅 VIO 位姿估计 (蓝色) 开始。过去的状态被记录下来 (绿色)。如果检测到最新关键帧的循环 (参见第 7-A 节), 如第二个图中的红线所示, 则发生了重定位。请注意, 由于使用特征级对应进行重定位, 我们能够合并来自多个过去关键帧的闭环约束 (参见第 7-C 节), 如最后三个图所示。(b) 全局位姿图优化。当关键帧从滑动窗口边缘化时, 它被添加到姿势图中。如果此关键帧与任何其他过去的的关键帧之间存在循环, 则表示为 4-DOF 相对刚体变换的闭环约束也将添加到位姿图中。姿态图在单独的线程中使用所有相对姿态约束 (参见第 8-C 节) 进行优化, 并且重定位模块始终根据最新的姿态图配置运行。

## A、 闭环检测

我们利用 DBoW2 [29], 一种最先进的词袋位置识别方法, 用于闭环检测。除了用于单目 VIO 的角点特征外, BRIEF 描述符 [40] 还检测并描述了 500 多个角点。

额外的角点特征用于在循环检测中获得更好的召回率。描述符被视为查询视觉数据库的视觉词。DBoW2 在时间和几何一致性检查后返回闭环候选者。我们保留所有用于特征检索的 BRIEF 描述

符，但丢弃原始图像以减少内存消耗。

## B、 特征检索

当检测到循环时，通过检索特征对应关系建立局部滑动窗口和闭环候选之间的连接。通过 BRIEF 描述符匹配找到对应关系。描述符匹配可能会导致一些错误的匹配对。为此，我们使用两步几何异常值拒绝，如图 10 所示。

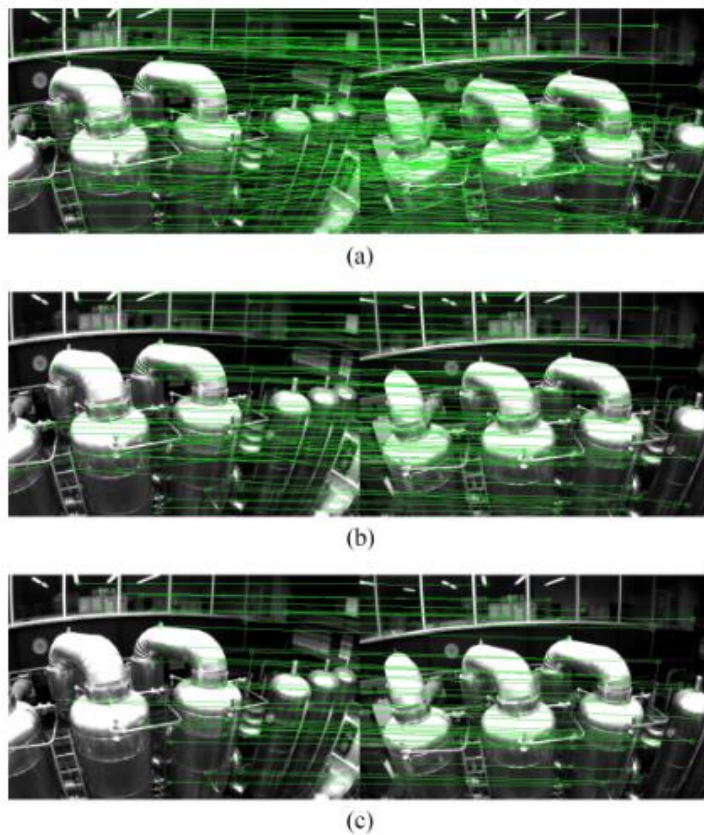


图 10、闭环期间用于特征检索的描述符匹配和异常值去除。(a) BRIEF 描述符匹配结果。(b) 第一步：2-D-2-D 异常值拒绝结果。(c) 第二步：3-D-2-D 异常值拒绝结果。

(1) 2-D-2-D：使用 RANSAC [33] 的基本矩阵测试。我们使用当前图像和闭环候选图像中检索到的特征的二维观察来执行基本矩阵测试。

(2) 3-D-2-D：使用 RANSAC [35] 的 PnP 测试。基于局部滑动窗口中特征的已知 3-D 位置和回环候选图像中的 2-D 观察，我们

执行 PnP 测试。

在异常值拒绝后,我们将这个候选者视为正确的闭环检测并执行重定位。

### C、 紧耦合重定位

重定位过程有效地将当前滑动窗口与过去的位姿对齐。 在重定位期间,我们将所有闭环帧的位姿视为常量。 我们使用所有 IMU 测量值、局部视觉测量值和检索到的特征对应关系联合优化滑动窗口。 我们可以轻松地将闭环坐标系  $v$  观察到的检索特征的视觉测量模型编写为与 VIO 中的视觉测量相同,如 (17)。 唯一的区别是闭环框架的位姿 $(\hat{\mathbf{q}}_v^w, \hat{\mathbf{p}}_v^w)$ ,取自位姿图(参见第 8 节),或直接取自过去的里程计输出(如果这是第一次重定位), 被视为常数。为此,我们可以用额外的循环项稍微修改 (14) 中的非线性成本函数作为:

$$\min_{\mathcal{X}} \left\{ \|\mathbf{r}_p - \mathbf{H}_p \mathcal{X}\|^2 + \sum_{k \in \mathcal{B}} \left\| \mathbf{r}_B(\hat{\mathbf{z}}_{b_{k+1}}^{b_k}, \mathcal{X}) \right\|_{\mathbf{P}_{b_{k+1}}^{b_k}}^2 \right. \\ \left. + \sum_{(l,j) \in \mathcal{C}} \rho(\|\mathbf{r}_C(\hat{\mathbf{z}}_l^{c_j}, \mathcal{X})\|_{\mathbf{P}_l^{c_j}}^2) \right. \\ \left. + \underbrace{\sum_{(l,v) \in \mathcal{L}} \rho(\|\mathbf{r}_C(\hat{\mathbf{z}}_l^v, \mathcal{X}, \hat{\mathbf{q}}_v^w, \hat{\mathbf{p}}_v^w)\|_{\mathbf{P}_l^{c_v}}^2)}_{\text{reprojection error in the loop-closure frame}} \right\} \quad (18)$$

其中  $\mathcal{L}$  是闭环帧中检索到的特征的观察集。  $(l, v)$  表示在闭环框架  $v$  中观察到的第  $l$  个特征。请注意,尽管成本函数与 (14) 略有不同,但要解决的状态维度保持不变,因为闭环框架的姿势被视为常数。 当使用当前滑动窗口建立多个闭环时,我们同时使用来自所有

帧的所有闭环特征对应进行优化。 这为重定位提供了多视图约束，从而获得更高的准确性和更好的平滑度。 重定位后保持一致性的全局优化将在第八节中讨论。

## 8、全局位姿图优化和地图重用

重新定位后，开发了额外的位姿图优化步骤，以确保将过去的位姿集注册到全局一致的配置中。

### A、 四个累积漂移方向

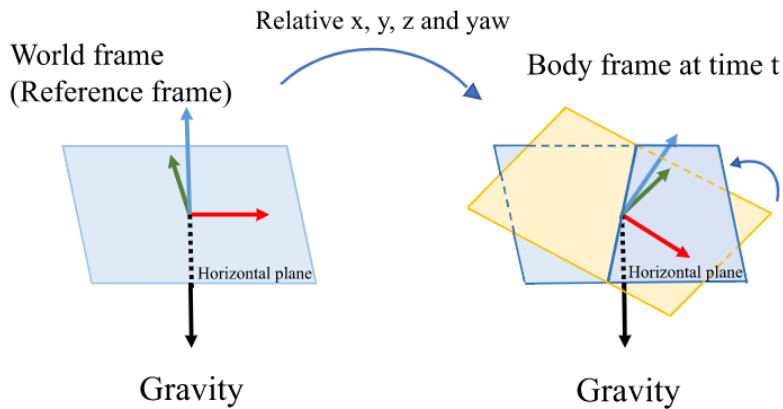


图 11、四个漂移方向的图示。 随着物体的移动， $x$ 、 $y$ 、 $z$  和偏航角相对于参考系发生相对变化。 绝对横滚角和俯仰角可以由重力矢量的水平面确定。

得益于重力的惯性测量，在 VINS 中可以完全观察到横滚角和俯仰角。 如图 11 所示，随着对象的移动，3-D 位置和旋转相对于参考系发生相对变化。 然而，我们可以通过重力矢量确定水平面，这意味着我们始终观察绝对横滚角和俯仰角。 因此，roll 和 pitch 是世界坐标系中的绝对状态，而  $x$ 、 $y$ 、 $z$  和 yaw 是相对于参考系的相对估计。 累积漂移仅发生在  $x$ 、 $y$ 、 $z$  和偏航角中。 为了充分利用有效信息并有效地纠正漂移，我们固定了无漂移的横滚和俯仰，并且仅在 4-DOF 中执行位姿图优化。



## B、 将关键帧添加到姿势图中

在 VIO 过程之后, 关键帧被添加到位姿图中。 每个关键帧作为位姿图中的一个顶点, 它通过两种类型的边与其他顶点相连, 如图 12 所示。

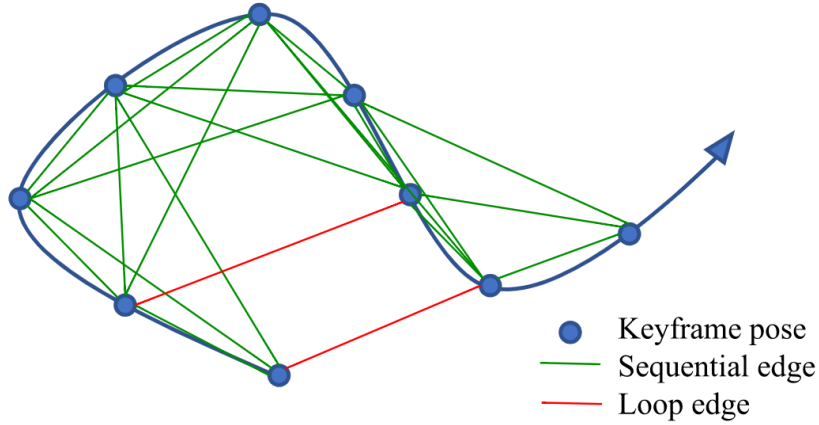


图 12、姿态图的插图。 关键帧充当位姿图中的一个顶点, 它通过连续边和闭环边连接其他顶点。 每条边代表相对平移和相对偏航。

(1) 连续边: 关键帧为其先前的关键帧建立多个连续边缘。 连续边表示两个关键帧之间的相对变换, 它直接取自 VIO。 考虑到关键帧  $i$  和它之前的关键帧之一  $j$ , 连续边仅包含相对位置  $\hat{\mathbf{p}}_{ij}^i$  和偏航角  $\hat{\psi}_{ij}$ 。

$$\begin{aligned}\hat{\mathbf{p}}_{ij}^i &= \hat{\mathbf{R}}_i^{w^{-1}} (\hat{\mathbf{p}}_j^w - \hat{\mathbf{p}}_i^w) \\ \hat{\psi}_{ij} &= \hat{\psi}_j - \hat{\psi}_i.\end{aligned}\tag{19}$$

(2) 闭环边: 如果关键帧有闭环连接, 它通过位姿图中的闭环闭合边缘连接闭环闭合帧。 类似地, 闭环边仅包含定义与 (19) 相同的 4-DOF 相对位姿变换。 使用重定位的结果获得闭环边缘的值。

## C、 4-DOF 位姿图优化

我们将帧  $i$  和  $j$  之间的边缘残差定义为:



$$\mathbf{r}_{i,j}(\mathbf{p}_i^w, \psi_i, \mathbf{p}_j^w, \psi_j) = \begin{bmatrix} \mathbf{R}(\hat{\phi}_i, \hat{\theta}_i, \psi_i)^{-1}(\mathbf{p}_j^w - \mathbf{p}_i^w) - \hat{\mathbf{p}}_{ij}^i \\ \psi_j - \psi_i - \hat{\psi}_{ij} \end{bmatrix} \quad (20)$$

其中  $\hat{\phi}_i$  和  $\hat{\theta}_i$  是滚动角和俯仰角的固定估计值，它们是从单目 VIO 获得的。

通过最小化以下成本函数来优化连续边和闭环边的整个图：

$$\min_{\mathbf{p}, \psi} \left\{ \sum_{(i,j) \in \mathcal{S}} \|\mathbf{r}_{i,j}\|^2 + \sum_{(i,j) \in \mathcal{L}} \rho(\|\mathbf{r}_{i,j}\|^2) \right\} \quad (21)$$

其中  $\mathcal{S}$  是所有连续边的集合， $\mathcal{L}$  是所有闭环边的集合。尽管紧耦合的重定位已经有助于消除错误的闭环，但我们添加了另一个 Huber 范数  $\rho(\cdot)$  以进一步减少任何可能的错误循环的影响。相比之下，我们不对顺序边使用任何稳健的规范，因为这些边是从 VIO 中提取的，它已经包含足够的离群值拒绝机制。

位姿图优化和重定位（参见第 7-C 节）在两个单独的线程中异步运行。这使得在可用时立即使用最优化的位姿图进行重新定位。类似地，即使当前位姿图优化尚未完成，重新定位仍然可以使用现有位姿图配置进行。这个过程如图 9 (b) 所示。

#### D、位姿图合并

位姿图不仅可以优化当前地图，还可以将当前地图与先前构建的地图合并。如果我们加载了一个先前构建的地图并检测到两个地图之间的循环连接，我们可以将它们合并在一起。由于所有边都是相对约束，位姿图优化通过循环连接自动将两个图合并在一起。如图

13 所示，当前地图被循环边缘拉入先前地图。每个顶点和每条边都是相对变量，因此，我们只需要固定位姿图中的第一个顶点。

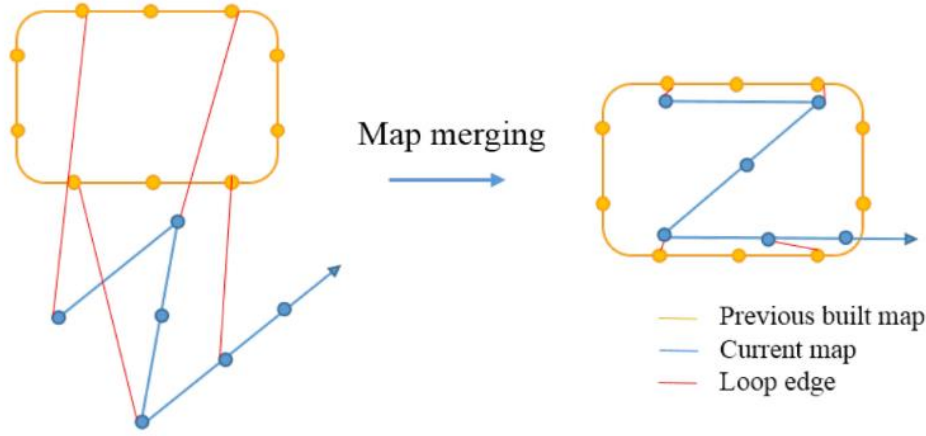


图 13、地图合并示意图。黄色图是之前建好的地图。蓝色图为当前地图。根据闭环连接合并两个地图。

## E、位姿图保存

我们的位姿图的结构非常简单。我们只需要保存顶点和边，以及每个关键帧（顶点）的描述符。丢弃原始图像以减少内存消耗。具体来说，我们为第  $i$  个关键帧保存的状态是：

$$[i, \hat{\mathbf{p}}_i^w, \hat{\mathbf{q}}_i^w, v, \hat{\mathbf{p}}_{iv}^i, \hat{\psi}_{iv}, \mathbf{D}(u, v, \text{des})] \quad (22)$$

其中  $i$  是帧索引， $\hat{\mathbf{p}}_i^w$  和  $\hat{\mathbf{q}}_i^w$  分别是来自 VIO 的位置和方向。如果这个帧有一个闭环帧， $v$  是闭环帧的索引。 $\hat{\mathbf{p}}_{iv}^i$  和  $\hat{\psi}_{iv}$  是这两个帧之间的相对位置和偏航角，它是从重定位中获得的。 $\mathbf{D}(u, v, \text{des})$  是特征集。每个特征都包含二维位置及其 BRIEF 描述符。

## F. 位姿图加载

我们使用相同的保存格式来加载关键帧。每个关键帧都是位姿图中的一个顶点。顶点的初始姿态是  $\hat{\mathbf{p}}_i^w$  和  $\hat{\mathbf{q}}_i^w$ 。闭环边由闭环信息  $\hat{\mathbf{p}}_{iv}^i$ ， $\hat{\psi}_{iv}$  直接建立。每个关键帧与其相邻关键帧建立几个连续边，如

(19) 所示。加载位姿图后,我们立即执行一次全局 4-DOF 位姿图。

位姿图保存和加载的速度与位姿图的大小呈线性相关。

## 9、实验结果

我们执行数据集和真实世界的实验以及两个应用程序来评估所提出的 VINS-Mono 系统。在第一个实验中,我们将所提出的算法与公共数据集上的另一种最先进的算法进行比较。我们执行数值分析以详细显示我们系统的准确性。然后,我们在室内环境中测试我们的系统,以评估重复场景中的性能。进行了大规模实验以说明其长期实用性。此外,我们将建议的系统应用于两个应用程序。对于空中机器人应用,我们使用 VINS-Mono 进行位置反馈以控制无人机遵循预定义的轨迹。然后,我们将我们的方法移植到 iOS 移动设备上。

## 10、结论和未来的工作

在本文中,我们提出了一种强大且通用的单目视觉惯性估计器。我们的方法具有 IMU 预积分、估计器初始化、在线外部校准、紧耦合 VIO、重定位和高效全局优化的最先进和新颖的解决方案。通过与其他最先进的开源实现进行比较,我们展示了卓越的性能。为了社区的利益,我们开源了 PC 和 iOS 实现。

尽管基于特征的 VINS 估计器已经达到实际部署的成熟度,但我们仍然看到许多未来研究的方向。取决于运动和环境,单目 VINS 可能会达到微弱可观察甚至退化的条件。我们对评估单目 VINS 的可观察性属性的在线方法感兴趣,以及在线生成运动计划以恢复可观

察性。另一个研究方向涉及在各种消费设备（例如 Android 手机）上大规模部署单目 VINS。该应用需要对几乎所有的传感器内在和外在参数进行在线校准，以及校准质量的在线识别。最后，我们有兴趣根据单目 VINS 的结果生成密集图。 [47] 中介绍了我们关于应用于无人机导航的单目视觉惯性密集映射的第一个结果。然而，仍然需要进行广泛的研究以进一步提高系统的准确性和鲁棒性。