VINS-Mono：一种稳健而通用的单目视觉惯性状态估计器

Tong Qin, Peiliang Li, Shaojie Shen

摘要

由于传统方法存在的一些问题，本文作者提出了一个稳健而通用的单目视觉惯性状态估计器，并将成果集成为一个系统，包含的模块主要有：测量预处理模块、估计器初始化模块、单目 VIO 模块、重定位模块、位姿图优化和地图重用模块。本文主要贡献是单目 VIO 模块以及整个系统的集成，其余成果主要来源于作者其它文章的成果以及之前提出的一些方法

关键词：单目视觉惯性系统；状态估计；传感器融合；同步定位和绘图

1 引言

移动机器人的一个基本任务是在给定环境地图的条件下确定其所在的位置，然而环境地图并不是一开始就有的，当移动机器人进入未知的环境时，需要通过自身的传感器构建 3D 环境地图，并且同时确定自身在地图中的位置，这就是 SLAM(simultaneous localization and mapping) 问题。根据采用的视觉传感器不同视觉 SLAM 主要分为三类：单目视觉 SLAM、立体视觉 SLAM、RGB-D SLAM。

状态识别是机器人导航、自动驾驶、虚拟现实和增强现实（AR）等广泛应用的最基本模块。由于其体积小、成本低、硬件设置简单，仅使用单目摄像头的方法在该领域备受关注。然而，单目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，所以增加了低成本的惯性测量单元 (IMU）辅助单目视觉系统，即单目视觉惯性系统 (VINS)。但是有几个问题影响了单目 VINS 的使用，如严格的初始化、消除漂移、地图保存和重用需求等。

本次课程的论文复现实现一个稳健的、多功能的单目视觉-惯性状态估计器 (VINS-Mono)，是单目视觉惯性系统的实时 SLAM 框架。其具有稳健的初始化程序；紧密耦合的、基于优化的单目 VIO；在线重定位和四自由度（DOF）全局位姿图优化；位姿图重用等特点。

2 相关工作

学者们在基于单目视觉的状态估计、里程计、SLAM 方面的工作非常广泛。值得注意的方法包括 PTAM，SVO，LSD-SLAM，DSO，和 ORB-SLAM。但是在这篇文章中，跳过了关于纯视觉方法的讨论，而只关注与单目视觉-惯性状态估计最相关的结果。

2.1 视觉测量处理和 IMU

对于视觉测量处理，根据剩余模型的定义，算法可以分为直接方法和间接方法。直接方法 SVO、LSD-SLAM、使用立体摄像机的直接视觉惯性里程计等使光度误差最小，而间接方法基于 EKF 、使用非线性优化的基于关键帧、具有在线初始化的视觉惯性里程计等使几何位移最小。由于直接方法的吸引区域较小，因此需要一个良好的初始猜测，而间接方法在提取和匹配特征方面消耗了额外的计算资源。间接方法由于其成熟性和稳健性，在现实世界的工程部署中更经常被发现。然而，直接方法更容易扩展到密集绘图，因为它们是直接在像素层面上操作的。

IMU 通常以比相机高得多的速度获取数据。已经提出了不同的方法来处理高速 IMU 的测量。最直接的方法是在基于 EKF 的方法中使用 IMU 进行状态传播。在一个图形优化公式中，为了避免重复的 IMU 重新整合，开发了一种称为 IMU 预整合的有效技术。这种技术最早是在“视觉惯性辅助导航在无初始条件的建筑环境中实现高动态运动”中引入的，它使用欧拉角对旋转误差进行参数化。Shen 等人使用连续时间误差状态动力学推导出协方差传播。预积分理论通过增加后置 IMU 偏差校正而得到进一步改进。

2.2 传统的单目视觉惯性状态估计器

处理视觉和惯性测量的最简单方法是松散耦合的传感器融合，其中 IMU 被视为一个独立的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，仅视觉的位姿用于更新。更进一步说，紧密耦合的视觉惯性算法要么基于 EKF，要么基于图形优化，其中相机和 IMU 的测量值从原始测量水平上被联合优化。MSCKF 是一个流行的基于 EKF 的 VIO 方法。MSCKF 在状态向量中保留了之前的几个相机姿态，并使用多个相机视图中同一特征的视觉测量值来形成多约束更新。SR-ISWF 是 MSCKF 的一个扩展。它使用方根形式来实现单精度表示，避免了不良的数字特性。这种方法采用了反滤波的迭代重线性化，使其与基于优化的算法相当。批量图形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 方法通常通过边缘化掉过去的状态和测量值，对最近状态的有界尺寸的滑动窗口进行优化。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上实现实时性能，例如手机。

准确的初始值对于引导任何单目 VINS 至关重要。Shen 等人提出了一种线性估计器的初始化方法，利用短期 IMU 预集成的相对旋转。这种方法未能对原始投影方程中的陀螺仪偏差和图像噪声进行建模。Martinelli 提出了一个单目视觉-惯性初始化问题的闭合式解决方案。后来，Martinelli 又提出了通过添加陀螺仪偏差校准来扩展这一闭式解决方案。这些方法未能对惯性积分的不确定性进行建模，因为它们依赖于 IMU 测量值在很长一段时间内的双重积分。Faessler 等人提出了一种基于 SVO 的重新初始化和故障恢复算法。需要一个额外的朝下的距离传感器来恢复度量衡。Mur-Artal 等人介绍了一种建立在流行的 ORB-SLAM 之上的初始化算法。根据提出的内容，尺度收敛所需的时间可能超过 10 秒。这对于一开始就需要尺度估计的机器人导航任务来说，可能会带来问题。轨迹测量法，不管它们所依赖的基础数学公式是什么，都会受到全局平移和方向的长期漂移的影响。为此，循环闭合在长期运行中起着重要作用。ORB-SLAM 能够关闭循环并重用地图，它利用了词包的优势。一个 7 自由度（位置、方向和比例）位姿图的优化遵循循环检测。

3 本文方法

3.1 本文方法概述

为了解决传统方法存在的一些问题，如严格的初始化、消除漂移、地图保存和重用需求等，本文提出了 VINS-Mono，一个稳健的、多功能的单目视觉-惯性状态估计器[1]，它是本文作者以前三个作品的结合和扩展。VINS-Mono 包含以下特点：

1） 稳健的初始化程序，能够从未知的初始状态引导系统；

2） 紧密耦合的、基于优化的单目 VIO，具有相机-IMU 外在校准和 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） 位姿图重用，可以保存、加载和合并多个局部位姿图。

在这些功能中，稳健的初始化、重新定位和位姿图的重用是作者团队的技术贡献，这些贡献来自他们以前的工作[2][3][4]。

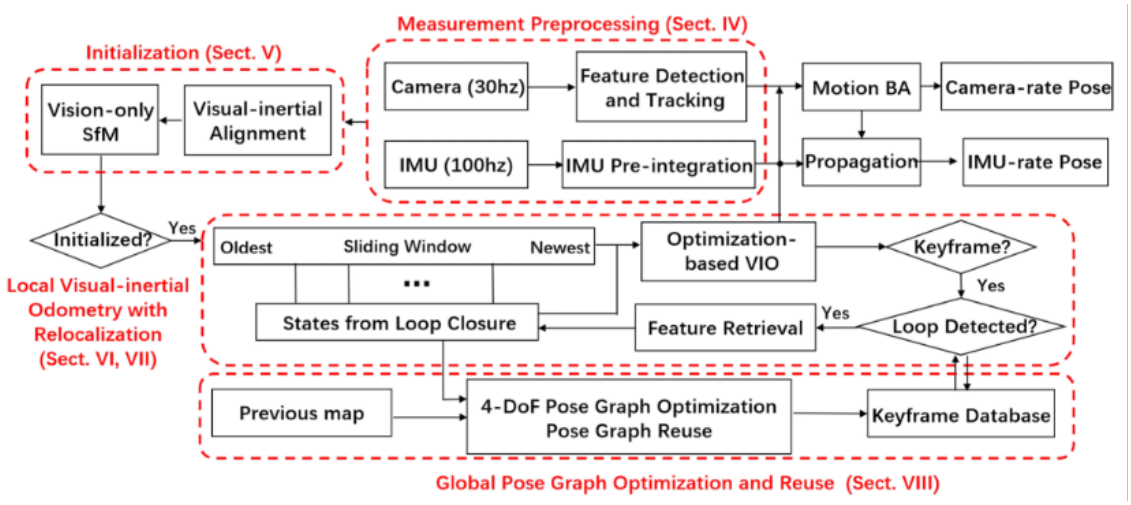

图 1: 框图为单目 VINS 的整体结构

单目视觉-惯性状态估计器的结构如图 1所示，系统从测量预处理开始，其中特征被提取和跟踪，两个连续帧之间的 IMU 测量被预整合。初始化程序提供所有必要的值，包括位姿、速度、重力矢量、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 与重定位模块紧密地融合了预先集成的 IMU 测量和特征观测。最后，位姿图优化模块采用经过几何学验证的重新定位结果，并进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不同的线程中同时运行。

3.2 测量预处理

对于视觉测量，我们在连续的帧之间跟踪特征，并在最新的帧中检测新的特征。对于 IMU 测量，我们在两个连续的帧之间对其进行预整合。

在视觉处理中，对于每张新的图像，KLT 稀疏光流算法都会跟踪现有特征。同时，检测新的角部特征，以保持每张图像中最小的特征数量，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统一的特征分布。二维（2-D）特征首先是不扭曲的，然后，在通过离群点剔除后投射到一个单位球体。离群点剔除是使用 RANSAC 与基本矩阵模型进行的。关键帧也在这一步骤中被选择。

对于 IMU 的预集成，沿用了之前基于连续时间四元数的 IMU 预积分推导方法，并且包括了对 IMU 偏差的处理。

3.3 估计器初始化

由于单目紧耦合 VIO 是一个高度非线性系统，所以在开始时需要一个准确的初始猜测。通过松散地对准 IMU 与纯视觉结构的预集成来获得必要的初始值。

首先，初始化程序从一个纯视觉的 SfM 开始，以估计一个按比例缩放的相机位姿和特征位置图。通过在一个滑动窗口中维护多个帧，以实现有限的计算复杂度。先检查最新帧和所有先前帧之间的特征对应关系。如果能在滑动窗口中找到稳定的特征跟踪（超过 30 个跟踪特征）和最新帧与任何其他帧之间的足够视差（超过 20 个像素）。使用五点算法恢复这两个帧之间的相对旋转和平移。然后，任

意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。基于这些三角特征，执行透视点（PnP）方法来估计窗口中所有其他帧的姿势。最后，应用全局全束调整来最小化所有特征观察的总重投影误差。由于我们还不知道关于世界坐标系的任何知识，我们将第一个相机坐标系设置为 SfM 的参考坐标系。

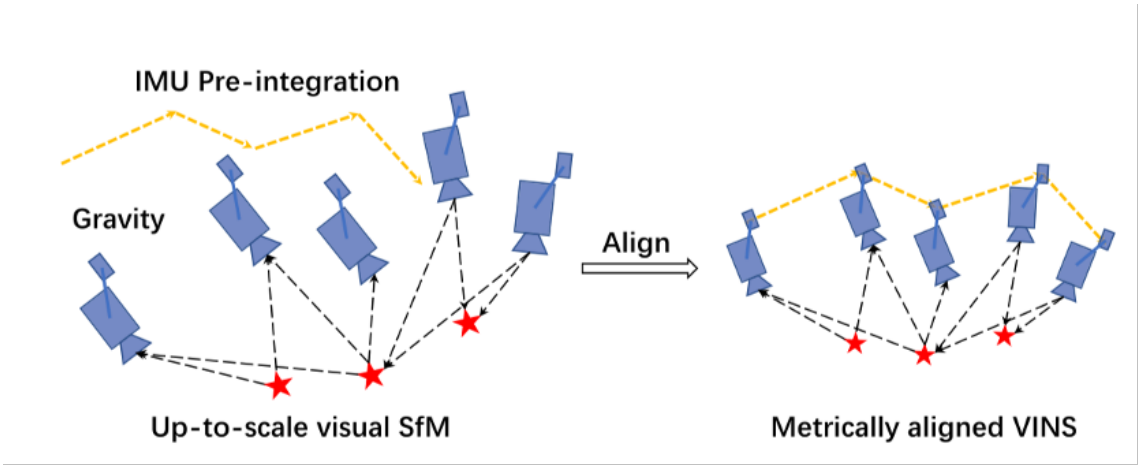

图 2: 估计器初始化的视觉惯性对准过程图解。其基本思想是将大规模视觉结构与 IMU 预集成相匹配。

视觉惯性对准的图示如图 2所示。其基本思想是将大规模视觉结构与 IMU 预集成相匹配。步骤：

1） 陀螺仪偏差校准：考虑窗口中的两个连续帧，从视觉 SfM 中获得旋转，以及 IMU 预积分中的相对约束。将 IMU 预积分项与陀螺仪偏差线性化，并最小化成本函数。

2） 速度、重力矢量和度量尺度初始化：陀螺仪偏置初始化后，继续初始化导航的其他基本状态，即速度、重力向量和度量尺度。

3） 重力细化：可以通过约束大小来细化从上一个线性初始化步骤获得的重力向量。

4） 完成初始化：在细化重力向量之后，我们可以通过将重力旋转到 z 轴来获得世界帧和相机帧之间的旋转。然后，我们将所有变量从参考坐标系旋转到世界坐标系。身体坐标系的速度也将旋转至世界坐标系。视觉 SfM 的平移分量将缩放为公制单位。

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目 VIO

3.4 单目 VIO

在估计器初始化之后，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。

使用视觉-惯性束调整公式，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。然后考虑滑动窗口中两个连续帧内的 IMU 测量值，计算预积分 IMU 测量值的残差。然后计算视觉测量误差，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，本文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。

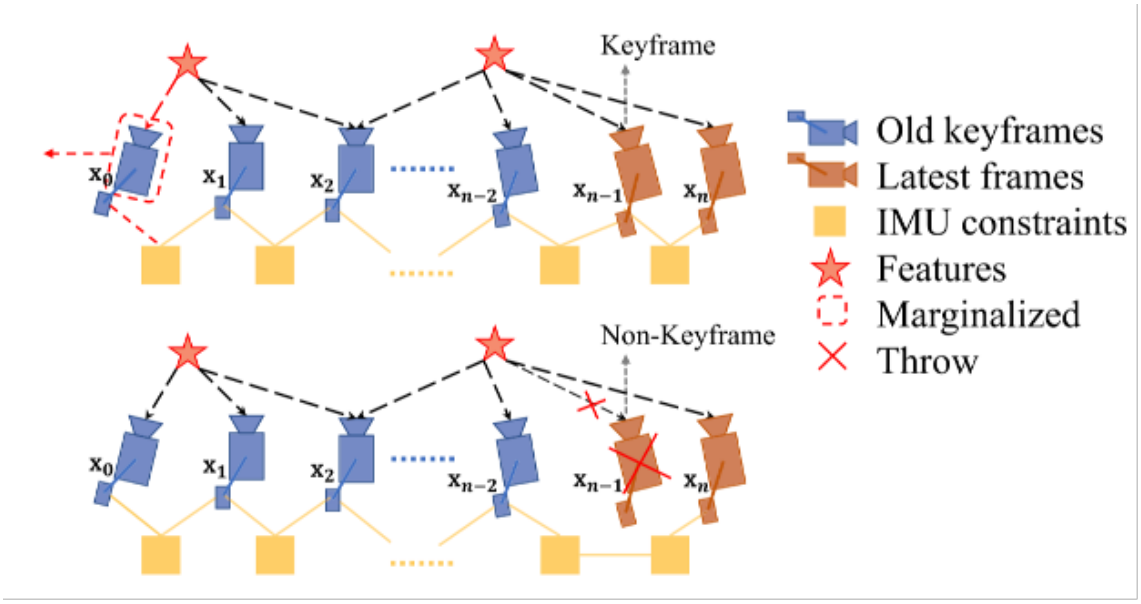

图 3: 边缘化策略

为了限制我们基于优化的 VIO 的计算复杂性，引入了边缘化。从滑动窗口中选择性地边缘化 IMU 状态和特征，同时将对应于边缘化状态的测量值转换为先验。如图 3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，则将抛出视觉测量并保留连接到该非关键帧的 IMU 测量。系统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。

由于 IMU 测量的速度比视觉测量高得多，尽管 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。

3.5 重定位模块

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的漂移。为了消除漂移，作者使用了之前提出了一个与单目 VIO 无缝结合的紧密耦合的重新定位模块。重新定位的过程从一个识别已经被访问过的地方的环路检测模块开始。然后建立环路闭合候选者和当前帧之间的特征级连接。这些特征对应关系被紧密地整合到单目 VIO 模块中，从而以最小的计算量获得无漂移的状态估计。多个特征的多个观测值直接用于重新定位，从而获得更高的精度和更好的状态估计平滑度。

3.6 位姿图优化和地图重用模块

重新定位后，开发额外的姿势图优化步骤，以确保将过去的姿势集合注册到全局一致的配置中。该模块使用了作者团队之前提出方案：“单目视觉-惯性 SLAM 的重定位、全局优化和地图合并[4]”。

4 复现细节

4.1 与已有开源代码对比

本次复现的论文将作者团队已有论文的成果和引用他人的成果，集成为一个系统，形成一个单目视觉惯性系统的实时 SLAM 框架，主要用于自主无人机的状态估计和反馈控制，具有高效的 IMU 预集成与偏差校正、自动估计器初始化、在线外部校准、故障检测和恢复、循环检测和全局位姿图优化、

地图合并、位姿图重用。并参考 blog: https://blog.csdn.net/hltt3838/article/details/109739046 增加 GPS 融合代码。

4.2 实验环境搭建

1) 代码基于 ubuntu 16.04

2) 使用 ROS Kinetic，安装 ROS package

```
git clone -b 1.14.0 https://ceres-solver.googlesource.com/ceres-solver
sudo apt-get install cmake libgoogle-glog-dev libgflags-dev libatlas-base-dev
    libeigen3-dev libsuitesparse-dev
mkdir ceres-bin
cd ceres-bin
cmake ../ceres-solver
make -j4
make test
sudo make install
```

3) 安装 Ceres Solver

```
git clone -b 1.14.0 https://ceres-solver.googlesource.com/ceres-solver
sudo apt-get install cmake libgoogle-glog-dev libgflags-dev libatlas-base-dev
    libeigen3-dev libsuitesparse-dev
mkdir ceres-bin
cd ceres-bin
cmake ../ceres-solver
make -j4
make test
sudo make install
```

4) 构建 VINS-Mono

```
sudo chown -R freedom /home/freedom/catkin_ws/
cd ~/catkin_ws/src
# 保存代码到该目录下
cd ../
catkin_make
source ~/catkin_ws/devel/setup.bash
```

4.3 界面分析与使用说明

1） 打开三个 terminal，均使用以下命令：

```
source /opt/ros/kinetic/setup.bash
source ~/catkin_ws/devel/setup.bash
```

2） 在第一个终端使用`roslaunch vins_estimator euroc.launch`启动 vins_estimator

3） 在第二个终端使用`roslaunch vins_estimator vins_rviz.launch`启动 rviz 界面

4） 在第三个终端使用`rosbag play MH_01_easy.bag`读取公共数据集

4.4 创新点

本次复现的论文将多项成果转化为一个直接可以使用的集成的系统，使用最先进和新颖的解决方案，展示出卓越的性能。

5 实验结果分析

将 VINS-Mono 与 OKVIS 进行比较，OKVIS 是一种适用于单目和立体相机的 VIO，它是另一种基于优化的滑动窗口算法。本文的算法在许多细节上与 OKVIS 不同，例如 VINS-Mono 系统具有强大

的初始化和闭环功能。实验结果见实验演示视频。

6 总结与展望

在本文中，作者提出了一个稳健和通用的单目视觉惯性估计器。尽管基于特征的 VINS 估计器已经达到了现实世界部署的成熟度，但是仍然可以看到许多未来研究的方向，比如单目 VINS 的应用场景。此外，还可以将其扩展至双目视觉和立体视觉。同时可以增加建图能力，使其能更好的应用于无人机和机器人领域。

参考文献

[1] QIN T, LI P, SHEN S. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4): 1004-1020.

[2] QIN T, SHEN S. Robust initialization of monocular visual-inertial estimation on aerial robots[J]., 2017: 4225-4232.

[3] LI P, QIN T, HU B, et al. Monocular visual-inertial state estimation for mobile augmented reality[J]., 2017: 11-21.

[4] QIN T, LI P, SHEN S. Relocalization, global optimization and map merging for monocular visual-inertial SLAM[J]., 2018: 1197-1204.