

## Deep-Init:基于深度学习的视觉惯性里程计非联合初始化方法

史殿习, 高云琦, 宋林娜, 刘哲, 周晨磊, 陈莹

引用本文

史殿习, 高云琦, 宋林娜, 刘哲, 周晨磊, 陈莹. [Deep-Init:基于深度学习的视觉惯性里程计非联合初始化方法](#)[J]. 计算机科学, 2024, 51(7): 327-336.

SHI Dianxi, GAO Yunqi, SONG Linna, LIU Zhe, ZHOU Chenlei, CHEN Ying. [Deep-Init:Non Joint Initialization Method for Visual Inertial Odometry Based on Deep Learning](#) [J]. Computer Science, 2024, 51(7): 327-336.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

### Similar articles recommended (Please use Firefox or IE to view the article)

#### [自编码器端到端通信系统后门攻击方法](#)

Backdoor Attack Method in Autoencoder End-to-End Communication System  
计算机科学, 2024, 51(7): 413-421. <https://doi.org/10.11896/jsjcx.230400113>

#### [针对系统调用的基于语义特征的多方面信息融合的主机异常检测框架](#)

Host Anomaly Detection Framework Based on Multifaceted Information Fusion of Semantic Features for System Calls  
计算机科学, 2024, 51(7): 380-388. <https://doi.org/10.11896/jsjcx.230400023>

#### [基于彩色图像高频信息引导的深度图超分辨率重建算法研究](#)

Study on Algorithm of Depth Image Super-resolution Guided by High-frequency Information of Color Images  
计算机科学, 2024, 51(7): 197-205. <https://doi.org/10.11896/jsjcx.230400102>

#### [基于特征重要性的深度学习自动调度优化研究](#)

Study on Deep Learning Automatic Scheduling Optimization Based on Feature Importance  
计算机科学, 2024, 51(7): 22-28. <https://doi.org/10.11896/jsjcx.230500220>

#### [面向产线AI质检的少样本评测方法研究和验证](#)

Study and Verification on Few-shot Evaluation Methods for AI-based Quality Inspection in Production Lines  
计算机科学, 2024, 51(6A): 230700086-8. <https://doi.org/10.11896/jsjcx.230700086>

# Deep-Init: 基于深度学习的视觉惯性里程计非联合初始化方法

史殿习<sup>1,4</sup> 高云琦<sup>1,2</sup> 宋林娜<sup>3</sup> 刘哲<sup>3</sup> 周晨磊<sup>4</sup> 陈莹<sup>2</sup>

1 智能博弈与决策实验室 北京 100091

2 国防科技创新研究院 北京 100071

3 国防科技大学计算机学院 长沙 410073

4 天津(滨海)人工智能创新中心 天津 300457

(dxshi@nudt.edu.cn)

**摘要** 对于非线性的单目 VIO 系统来说,其初始化过程至关重要,初始化结果的好坏直接影响整个系统运行过程中状态估计的精度。为此,将深度学习引入单目 VIO 系统的初始化过程中,提出了一种高效的非联合初始化方法(简称 Deep-Init),其核心是使用深度神经网络对 IMU 中陀螺仪的偏置和噪声等随机误差项进行准确估计,得到初始化过程中的关键参数,即陀螺仪的 bias;同时,将 IMU 预积分与 SfM 进行松耦合,通过位置和旋转对齐,使用最小二乘法对绝对尺度、速度以及重力矢量等进行快速恢复,并将其作为初始值来引导非线性紧密耦合的优化框架。由于深度神经网络对陀螺仪数据进行补偿,从而大大提高了 IMU 中旋转估计量的准确性,有效提高了 IMU 数据的信噪比,同时减少了最小二乘方程失效的次数,因此进一步减少了计算量。使用去除误差项的陀螺仪数据的预积分量替换 SfM 中的旋转量,将 IMU 的旋转量作为真值,不仅避免了将不准确的 SfM 值作为真值进行初始化时所带来的误差,有效提升了系统状态估计的精度,而且能够有效地适应高速运动、光照变换剧烈和纹理重复等 SfM 估计效果差的场景。在 EuRoC 数据集上,对所提方法的有效性进行了实验验证,实验结果表明,所提出的初始化方法 Deep-Init 无论是精度还是耗时均取得了良好的效果。

**关键词:** 视觉惯性里程计;深度学习;初始化;惯性测量单元

**中图分类号** TP391.41

## Deep-Init: Non Joint Initialization Method for Visual Inertial Odometry Based on Deep Learning

SHI Dianxi<sup>1,4</sup>, GAO Yunqi<sup>1,2</sup>, SONG Linna<sup>3</sup>, LIU Zhe<sup>3</sup>, ZHOU Chenlei<sup>4</sup> and CHEN Ying<sup>2</sup>

1 Intelligent Game and Decision Lab(IGDL), Beijing 100091, China

2 National Innovation Institute of Defense Technology, Beijing 100071, China

3 College of Computer, National University of Defense Technology, Changsha 410073, China

4 Tianjin(Binhai) Artificial Intelligence Innovation Center, Tianjin 300450, China

**Abstract** For a non-linear monocular VIO system, its initialization process is crucial, and the initialization result directly affects the accuracy of the state estimation during the whole system operation. To this end, this paper introduces a deep learning method into the initialization process of the monocular VIO system and proposes an efficient non-joint initialization method(referred to as Deep-Init). The core of this method is to use a deep neural network to accurately estimate the random error terms such as bias and noise of the gyroscope in the IMU, to obtain the key parameter in the initialization process, i. e. the bias of the gyroscope. At the same time, we loosely couple the IMU pre-integration to the SfM. The absolute scale, velocity and gravity vector are quickly recovered by position and rotation alignment using least squares, which are used as initial values to guide the non-linear tightly coupled optimization framework. The accuracy of the rotation estimates in the IMU is greatly increased due to the compensation of the gyroscope data by the deep neural network, which effectively improves the signal-to-noise ratio of the IMU data. This also reduces the number of least squares equation failures, further reducing the computational effort. Using the pre-integrated amount of gyroscope data with the error term removed to replace the rotation amount in the SfM and using the IMU rotation amount as the true value, not only avoids the errors associated with initializing inaccurate SfM values as the true value but also effectively improves the accuracy of system state estimation. Moreover, it enables effective adaptation to scenarios where SfM estimation is

到稿日期:2023-05-06 返修日期:2023-10-15

基金项目:科技部科技创新 2030-重大项目(2020AAA0104802);自然科学基金(91948303)

This work was supported by the Science and Technology Innovation 2030 Major Project(2020AAA0104802) and National Natural Science Foundation of China(91948303).

通信作者:陈莹(selina.ychen@foxmail.com)

poor, such as high-speed motion, drastic lighting changes and texture repetition. The validity of the proposed method is verified on the EuRoC dataset, and the experimental results show that the proposed Deep-Init initialization method achieves good results in terms of both accuracy and time consumption.

**Keywords** Visual-inertial odometry, Deep learning, Initialization, Inertial measurement unit

## 1 引言

在智能机器人系统中,位姿估计是实现其自主运动能力的关键技术。精准的位姿估计具有极其广泛的应用场景,例如虚拟现实(Augmented Reality, AR)、文物重建以及自动驾驶等。常见的位姿估计方法包括基于相机的视觉里程计(Visual Odometry, VO)<sup>[1-3]</sup>以及基于相机和惯性测量单元(Inertial Measurement Unit, IMU)的视觉惯性里程计(Visual Inertial Odometry, VIO)<sup>[4-5]</sup>等。对于单目相机视觉里程计VO系统来说,由于单目相机的尺度不确定性,无法直接获得当前场景的尺度信息,因此难以将其直接应用到真实的场景中。对于单目视觉惯性里程计VIO系统来说,由于IMU长时间工作后积分可能导致位姿发散,从而造成不可忽略的累计误差,但可以通过IMU恢复尺度、roll和pitch角等信息,提升系统位姿估计的精度。而且,IMU还能够有效帮助VIO系统应对具有挑战的视觉场景,如快速运动、光照和视角变换剧烈以及运动模糊等。因此,在实际应用场景中,人们更倾向于使用VIO来进行位姿估计。

由于单目视觉惯性里程计VIO是一个高度非线性的系统,如果要使VIO系统能够尽快收敛,则需要对一系列状态估计所需的参数(如重力、速度、Bias、噪声和尺度等信息)进行求解,上述求解过程被称为VIO系统的初始化。单目VIO系统<sup>[4-9]</sup>通常严重依赖于系统初始值,其初始化结果的好坏直接影响整个系统运行过程中状态估计的精度。因此,对于VIO系统来说,其初始化过程变得至关重要,错误的初始值将会导致错误的姿态估计,特别是在快速运动的设备以及AR/VR等多自由度估计的场景中,精确的初始化参数显得尤为重要。然而,由于仅使用单目相机并不能恢复尺度,因此对单目VIO系统进行初始化是十分困难的,并且在加入IMU之后,还需要一定的运动激励才能获得绝对尺度等信息,否则可能导致未知的初始姿态和速度,从而导致系统只能在静止或确定的运动模式下启动,极大地限制了VIO系统的使用场景。目前,现有的VIO系统初始化算法中<sup>[6,10,12]</sup>,通常将从运动恢复结构(Structure from Motion, SfM)<sup>[13]</sup>计算得到的姿态作为真值,但在高速运动造成的运动模糊、纹理缺失或者不明显的场景中,通过SfM求解得到的位姿是非常不准确的,从而难以精确地对系统状态进行估计。因此,如何在不使用任何先验信息的情况下,高效地对VIO系统进行初始化是一个极具挑战性的问题。同时,由于视觉算法对运动模糊和剧烈光照变化十分敏感,因此在VIO系统运行过程中,根据场景变化适时地进行重新初始化,使得系统快速适应不同的场景,提升系统的状态估计精度,也是一项十分重要和必要的工作。

针对上述问题,我们将深度学习方法引入单目VIO系统的初始化过程中,提出了一种高效的非联合初始化方法(简称

Deep\_Init),其核心是使用深度神经网络对IMU中陀螺仪的偏置和噪声等随机误差项进行准确估计,将IMU预积分<sup>[14]</sup>与SfM进行松耦合,对绝对尺度、速度、重力矢量等信息进行快速恢复,并将其作为初始值来引导非线性紧耦合的优化框架,减少计算量,简化了初始化过程。在EuRoC数据集上,对本文方法的有效性进行了实验验证。本文的贡献如下:1)将深度学习方法引入单目VIO系统初始化方法中,求解初始化中关键参数陀螺仪Bias和噪声等随机误差项,通过深度神经网络对陀螺仪数据进行补偿,有效提高了IMU数据的信噪比;进而使用去除误差项的陀螺仪数据预积分量与SfM进行松耦合(IMU的旋转量作为真值),对绝对尺度、速度、重力矢量等信息进行快速恢复,不仅避免了将不准确的SfM值作为真值进行初始化时所带来的误差,还有效提升了系统状态估计的精度,而且能够有效地适应高速运动、光照变换剧烈和纹理重复等SfM估计效果差的场景;2)将快速恢复位置的绝对尺度、速度、重力矢量等信息作为初始值来引导非线性紧密耦合的优化框架,在构建最小二乘方程时,有效降低了数据不可信造成方程失效的概率,进一步减少了计算量,从而有效提升初始化过程的效率;可以根据需要即时进行初始化,实时地对初始化参数进行校准;3)在EuRoC数据集上,对本文方法Deep\_Init的有效性进行了实验验证,实验结果表明,Deep\_Init无论是在精度还是耗时方面均取得了良好的效果。

## 2 相关工作

目前,根据初始化方式的不同,可以将单目VIO系统的初始化分为联合初始化、半联合初始化和非联合初始化3类方法<sup>[9]</sup>。下文将对这3类初始化方法的思想和相关工作分别进行描述。

### 2.1 联合初始化

联合初始化指通过初始化参数之间的约束关系,一步将系统中所有的初始化参数求解出来的方法。MK(Martinelli-Kaiser solution)方法<sup>[15]</sup>是最早的联合初始化方法,该方法假设已知系统的外部参数并且不考虑IMU Bias,通过IMU运动方程和特征点之间的空间约束,对尺度因子、重力、初始速度进行求解。由于MK方法不考虑IMU Bias,因此得到的结果不太准确。2016年Kaiser等<sup>[16]</sup>对MK方法进行了改进,加入了对IMU陀螺仪Bias的估计,并且实验证明是否估计加速度计Bias和更精确的重力向量模型对初始化的结果影响甚微,但是是否估计陀螺仪Bias则对初始化影响巨大。2018年Dominguez-Conti等<sup>[10]</sup>使用图像特征与IMU融合的方法,直接使用相机在二维平面上的测量信息与IMU的轨迹信息进行紧耦合,构造误差函数,从而直接求解初始化参数。2019年Campos等<sup>[7]</sup>基于MK方法,使用预积分处理IMU数据,并将预积分公式引入后续优化过程中,作为优化中的约束,通过连续的两次优化,分别求出陀螺仪bias、重力方向和

系统的初始速度、尺度等参数;而且,该方法还增加了两次初始化测试(可观测性测试和一致性测试),每通过一次测试,初始化则可以进一步进行 BA(Bundle Adjustment)优化,获得更精确的初始化参数,而当测试未通过时,则初始化停止。

## 2.2 非联合初始化

非联合初始化指分多步依次求解系统初始化参数的 VIO 初始化方法。VI-ORB-SLAM<sup>[11]</sup>中引入了一种非联合的初始化方法,该方法首先利用纯视觉信息,使用预积分算法中的旋转项,通过构造帧间视觉约束方程组,求解陀螺仪 bias;其次,忽略加速度计 bias,基于预积分算法中的速度项和位移项,构造帧间运动约束方程组,求解粗略的尺度和重力向量;最后,在第二步的基础上,加入加速度计 Bias 约束和细化的重力向量模型,求解加速度计 Bias、精细化重力向量和每一时刻的系统速度。

2017年 Qin 等<sup>[6]</sup>提出了一种非联合初始化方法,首先,仅进行视觉的三维重建,获取不考虑尺度信息时相机坐标系与 IMU 坐标系的关系;其次,进行 IMU 预积分和陀螺仪 Bias 校准,以及初始速度、重力矢量和尺度信息的求解;最后,进行重力细化。然而,该初始化方法并没有对加速度计 Bias 进行估计,因为他们认为加速度计 Bias 数值较低,又与陀螺仪 Bias 耦合严重,因此忽略了对加速度计 Bias 的估计。同时,通过实验表明,该方法只有在大于 30 度的旋转激励时,才能在初始化过程中有效地将加速度计 Bias 与陀螺仪 Bias 进行解耦,且一般情况下加速度计 Bias 对尺度估计的误差影响较小(10%以内)。2017年 Mustaniemi 等<sup>[12]</sup>提出了一种用于移动设备的视觉惯性初始化方法。首先,将相机与 IMU 数据在时间和空间上进行对齐;其次,通过对 SFM 得到的轨迹图进行微分,得到相机当前时刻的加速度和角速度值,再进行 RTS (Rauch-Tung-Striebel)平滑处理,并把这些值与经过低通滤波器的 IMU 数据一起构造残差;最后,使用最小二乘法进行尺度估计,得到尺度因子、重力方向和 IMU 偏置。2018年 Huang 等<sup>[17]</sup>在 VI-ORB-SLAM 的基础上增加了对系统外参数的估计,由于其初始化过程无法通过线性方程组直接求解,因此采用迭代法进行求解,同时通过优化的方法计算出加速度 Bias。

## 2.3 半联合初始化

半联合初始化是一种介于联合初始化与非联合初始化之间的方法,一般可以分为两步:第一步基于视觉信息和 IMU 陀螺仪数据,求出部分参数,这些参数可能是初始化参数,也可能是后续求解所需的参数;第二步在第一步的基础上,一步将所有剩余的初始化参数求解出来。半联合方法的初始化结果不仅比联合初始化精度高,而且还可以避免非联合初始化中因分布求解可能带来的误差问题。

2016年 Yang 等<sup>[8]</sup>提出的半联合初始化方法可以求解出尺度、重力方向、初始速度和相机与 IMU 之间的内外参数,但是忽略了对 IMU bias 的估计,该半联合初始化方法分为两步:第一步求解相机和 IMU 之间的旋转,也就是纯视觉初始化;第二步通过优化函数直接优化出重力方向、初始速度、相机与 IMU 之间的内外参和特征点深度。由于忽略了对 IMU

Bias 的估计,目前该方法的可用性并不高。2020年 Campos 等<sup>[18]</sup>提出的半联合方法可以求解尺度、重力方向、初始速度、IMU 陀螺仪和加速度计 Bias。同样,该方法也分为两步:第一步构造基于最大后验估计(MAP)的纯视觉最优估计方程,求解不考虑尺度的系统状态;第二步构造基于最大后验估计(MAP)的仅惯性的最优估计方程,直接优化出重力方向、初始速度、IMU 陀螺仪和加速度计 Bias。该方法的特点是将视觉部分和 IMU 部分解耦,可以适用于任何 VIO 系统的初始化估计,同时该方法在准确性上和时间性能上都取得了目前 VIO 初始化最好的结果。目前,该方法已经被整合进 ORB-SLAM 3 中,作为其中 VINS 系统的初始化方法使用。

## 2.4 小结

综上所述,由于目前现有的大部分的初始化方法均是传统的使用方程联立的方法,都是通过最小二乘的方法进行一步求解或者多步求解,因此一般均会受到数据质量的影响,存在关键参数的初始值缺失以及求解速度慢等问题。因此,通过充分利用深度学习方法的优点,提升数据质量并求解部分关键参数,将深度学习方法引入 VIO 系统的初始化过程中,目的是在简化初始化过程、实现快速初始化的基础上,进一步提升系统状态估计的精度。

## 3 基于深度学习的非联合初始化方法

为了便于描述和表达,我们定义了一些符号和坐标系。其中, $(\hat{\cdot})$ Gyro-Net 表示网络估计值, $(\cdot)_k^m$ 表示 body 坐标系下  $k$  时刻的 IMU 测量值, $(\cdot)_k^b$ 表示 body 坐标系下  $k$  时刻的 IMU 真值, $(\cdot)_c^b$ 表示由当前时刻相机坐标系到 body 坐标系的坐标变换, $(\bar{\cdot})$ 表示缺少单位的参数量。

### 3.1 方法框架

针对目前单目 VIO 系统初始化过程中存在的问题,为了实现单目 VIO 系统高效精确的初始化,本文提出了一种基于深度学习的非联合初始化方法。该方法的框架结构如图 1 所示,以 IMU 和相机数据为输入,主要由 Gyro-Net 网络求解陀螺仪 Bias 等随机误差项、基于补偿陀螺仪的 IMU 预积分、将 IMU 预积分与纯视觉 SfM 进行松耦合 3 部分来求解初始化参数。

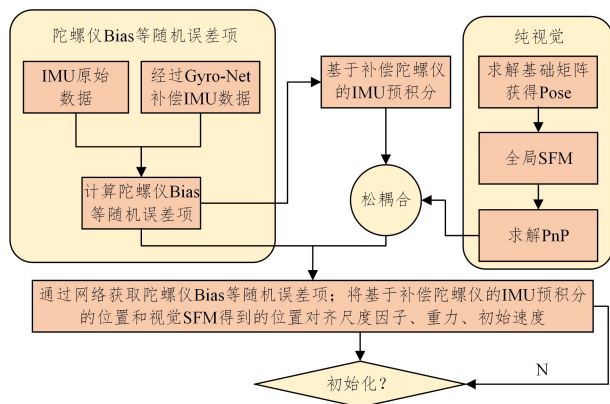


图 1 基于深度学习的非联合初始化方法框架

Fig. 1 Framework of non-joint initialization methods based on deep learning

该方法的核心思想如下:首先,将 IMU 作为网络输入,使用 Gyro-Net<sup>[19]</sup> 计算陀螺仪 Bias 和噪声等误差项,不仅有效简化了陀螺仪的偏置项  $b_o$  的求解过程,而且避免了现有算法<sup>[6]</sup> 中 SfM 在极端情况下旋转量估计不准确的问题。其次,对基于补偿陀螺仪数据的 IMU 预积分与纯视觉 SfM 的求解过程进行松耦合,将 Gyro-Net 旋转估计值的预积分量和 SfM 的位置估计值作为真值,进行初始化参数求解,通过平移约束对尺度因子  $s$ 、旋转外参数  $qbc$ 、重力向量和每帧的初始速度等参数进行求解。最后,由于本文采用的是经过网络补偿后的陀螺仪数据,其信噪比大大增加,数据可信度随之大大增强,因此,在基于上述数据建立最小二乘方程时,有效降低了数据不可信造成方程失效的概率,进一步减少了计算量,有效提升了初始化过程中的效率;而且,无需任何先验信息和激励运动,可以根据需要即时进行初始化,实时地对初始化参数进行校准。

### 3.2 陀螺仪 Bias 和噪声误差项估计

对于单目视觉惯性里程计 VIO 来说,由于 IMU 长时间工作后积分可能导致位姿发散,进而造成不可忽略的累计误差,尤其是陀螺仪的偏置及误差项对系统状态估计影响巨大。在 VIO 系统中,旋转量可以通过陀螺仪读数获取,也可以通过相机使用多视图几何的方法获取。在不考虑噪声等其他误差的情况下,两者的大小是一致的,但在实际场景中,两者之间肯定存在误差。陀螺仪测量值的误差包括测量噪声和陀螺仪 Bias 两个部分(测量噪声暂时可以忽略,因为其毕竟太小),而视觉观察值的误差则只有观测噪声(也可以忽略不记),因此两者之间差值的绝对值只有陀螺仪 Bias。现有方法大都采用图 2 中的 A 方法来构建残差,即把相机 SfM 计算当作真值与陀螺仪测量值来构造残差。本文方法采用 Gyro-Net 估计值与陀螺仪测量值构造残差方法 B。目前,现有的大部分传统视觉惯性初始化方法均是整个滑动窗口的所有的旋转做差,构成一个最小化误差模型,通常按照图 2 中的方法 A 构造残差,即把通过视觉计算的 SfM 当作真值与陀螺仪测量值构造残差,对陀螺仪 Bias 进行求解,即对陀螺仪偏置项  $b_o$  进行求解,其估计精度低,对后续系统运行状态估计精度造成严重影响。为了提升陀螺仪 Bias 的估计精度,本文采用将 Gyro-Net 估计值与陀螺仪测量值构造残差的方法(如图 2 中的方法 B),对陀螺仪 Bias 进行求解,即对偏置项  $b_o$  进行估计,以提升其精度。

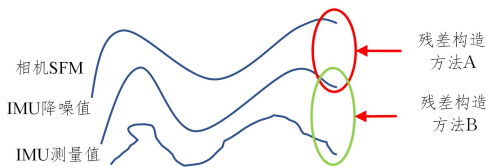


图 2 陀螺仪 Bias 求解示意图

Fig. 2 Schematic diagram of gyroscopic Bias solution

在上述的 VIO 系统初始化算法中,通常将从运动恢复结构(SfM)计算得到的姿态作为真值,但在高速运动造成的运动模糊、纹理缺失或者不明显的场景中,通过 SfM 求解得到的位姿是非常不准确的,从而导致难以精确地对系统状态进行估计。因此,在视觉估计不准确的场景下,将 SfM 作为

真实值进行残差构建缺少合理性,而且现有的 IMU 测量模型通常只考虑 Bias,而忽略了 IMU 数学模型中噪声等时变误差,这进一步造成了估计值不准确性的问题。针对这一问题,基于 Gyro-Net<sup>[19]</sup>,使用深度学习的方法对陀螺仪的 Bias 和噪声等误差项进行估计。本文使用的 Gyro-net 与文献<sup>[19]</sup> 中的方法的不同之处在于我们使用网络估计的是偏置、噪声和  $g$  灵敏度,而不是得到估计的旋转量。将 Gyro-Net 的估计结果替换 SfM 求解的旋转量,获得更准确的不考虑尺度的相机位姿,不仅可以避免使用视觉 SfM 估计的旋转量带来的估计不准确性问题,而且可以保证在视觉估计不准确的场景下也能进行正常的初始化,为此,我们将陀螺仪的测量模型建模为:

$$\omega_k^{bm} = S_g M_g \omega_k^b + S_{ga} a_k^b + \eta_g + b_g \quad (1)$$

其中,  $\eta$  是均值为 0 的高斯白噪声;  $b$  为偏置项;  $S_g$  是尺度因子; 轴间非对齐误差位为  $M_g$  (在 VIO 或者 SLAM 系统中通常将其设定为单位矩阵  $\mathbf{I}_3$ );  $S_{ga}$  为陀螺仪的  $g$ -灵敏度,表明加速度测量值对陀螺仪测量值的影响。

Gyro-Net 是一种基于深度学习的 IMU 陀螺仪随机误差补偿方法<sup>[19]</sup>,其核心是针对 IMU 中陀螺仪长期工作后朝向估计不准确以及数据可用性差的问题,采用一种半稠密的网络结构,使用 IMU 特征提取与选择模块(IFES Block)获得更加丰富的特征,通过跳跃连接和过渡层调整特征流向,实现特征提取、选择和压缩前后不同 Block 间的特征复用,进而实现对 IMU 信息的有效利用。同时,针对累计误差对整个航向估计的影响问题,采用绝对角度损失和相对角度损失相结合的方式,在减小 IMU 数据的局部误差的同时,进一步有效减少长时间的累计误差,进而有效提升系统运行的可靠性,使得该网络最终能够达到更加精准的角度估计值。

为了对 IMU 陀螺仪 Bias 和噪声等随机误差项进行精确估计,我们将 IMU 数据输入到训练好的网络模型中,采用式(2)预测初始状态下的误差项:

$$\omega^{\wedge} = (S_{ga} a_k^b + \eta_o + b_o)^{\wedge} = f(u_{k-N}^b, \dots, u_k^b) \quad (2)$$

其中,  $f(\cdot)$  表示由神经网络定义的函数,  $u$  表示加速度和角速度测量值,  $a$  表示加速度,  $\omega^{\wedge}$  表示陀螺仪误差项,  $\eta_o$  和  $b_o$  分别表示陀螺仪的噪声和 Bias。网络结构如图 3 所示,该网络由 IFES Block、IFC Block、跳跃连接和过渡层等部分组成。其中,  $T$  代表过渡层, IFC Block 用于压缩特征尺度, IFES Block 用于 IMU 特征提取与选择,  $T$  为过渡层,用于改变特征图的大小。

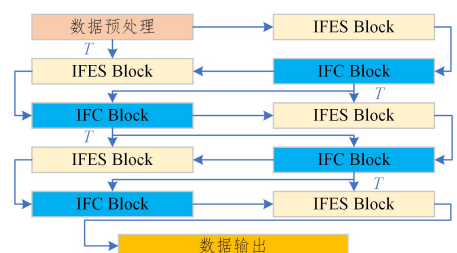


图 3 网络实现细节

Fig. 3 Network implementation details

在数据预处理阶段,需要完成数据增强和数据填充两个步骤。首先,文献<sup>[20]</sup>为 IMU 数据添加高斯噪声,进行数据

增强,避免训练过程中的过拟合,通过试错法将标准偏差为  $1 \times 10^{-3}$  的噪声添加到加速度分量中,将标准差为  $8 \times 10^{-5}$  的噪声添加到陀螺仪分量中;其次,由于 IMU 传感器工作时间越长,误差越大,因此使用首个采样时刻的值对 IMU 数据进行填充,学习误差较小的 IMU 的特征。

IFES Block 使用两组不同大小感受野卷积提取更丰富的 IMU 特征,对每个通道赋予不同的权重,并且动态调整两组不同感受野得到的特征。将文献[21]中提到的卷积层结构称为 IMU 特征压缩模块(IFC block)。为了防止特征在不同 block 传递时特征信息丢失的问题,使用跳跃连接和过渡层相结合的方式,对特征进行复用。

在现有的 VIO 系统中,由于噪声的时变性,通常 IMU 的测量模型并不考虑对噪声进行估计,而本文方法对陀螺仪测量模型进行了更精细化的建模,使得估计值更加准确,从而可以有效提升系统初始值的估计精度。

### 3.3 基于补偿陀螺仪的 IMU 预积分

由于 Gyro-Net 可以去除掉陀螺仪的 Bias 等随机误差,因此根据 Gyro-Net 的计算结果,对 IMU 预积分部分的计算方法进行改进,只对速度和位置的预积分量进行更新。接下来将详细介绍预积分的更新策略。

以第  $k$  帧相机坐标系为当前坐标系,对第  $k$  时刻和第  $k+1$  时刻内所有的角速度和线加速度进行积分<sup>[9,22]</sup>,可以得到位置、速度和旋转的积分量,如式(3)所示:

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

其中,  $(\cdot)_a^b$  表示从  $a$  坐标系转换到  $b$  坐标系,  $w$  表示世界坐标系,  $b_k$  表示第  $k$  帧的相机坐标系,  $b_{k+1}$  表示第  $k+1$  帧的相机坐标系,  $\mathbf{R}$  表示旋转矩阵,  $\mathbf{p}$  表示平移向量,  $\mathbf{v}$  表示速度,  $t$  表示时间,  $\mathbf{g}^w$  表示重力向量,  $\mathbf{q}$  表示旋转四元数。由于陀螺仪 bias 可以由网络进行估计,不需要在后续的过程进行求解,因此在计算旋转时去除了偏置项。

为了解决 IMU 在世界坐标系下难以求解的问题,使用坐标变换的方式进行解决,具体方法是需要把整个 IMU 积分的坐标系从世界坐标系变换到另一个相对坐标系(整个视觉系统或者整个滑动窗口中的第一帧的坐标系)。通过观察式(3)可知,IMU 的预积分需要依赖于前一帧的速度和旋转,每次进行非线性优化时,还要迭代更新前一帧的速度和旋转,这就导致每次迭代都需要重新进行积分。因此,考虑将优化变量在相邻相机帧中的 IMU 预积分项进行分离,通过对世界坐标系下的式(3)左右两边同时乘以  $\mathbf{R}_{b_k}^{b_{k+1}}$ ,可以得到:

$$\begin{aligned} \mathbf{R}_{b_k}^{b_{k+1}} \mathbf{P}_{b_{k+1}}^w &= \mathbf{R}_{b_k}^{b_{k+1}} \mathbf{P}_{b_k}^w + \mathbf{v}_{b_k}^w \Delta t_k - \frac{1}{2} \mathbf{g}^w \Delta t_k^2 + \alpha_{b_{k+1}}^{b_k} \\ \mathbf{R}_{b_k}^{b_{k+1}} \mathbf{v}_{b_{k+1}}^w &= \mathbf{R}_{b_k}^{b_{k+1}} \mathbf{v}_{b_k}^w + \beta_{b_{k+1}}^{b_k} \\ \mathbf{q}_{b_k}^w \otimes \mathbf{q}_{b_{k+1}}^w &= \gamma_{b_{k+1}}^{b_k} \end{aligned} \quad (4)$$

其中,  $\alpha$  表示位置的预积分量,  $\beta$  表示速度的预积分量,  $\gamma$  表示

旋转的预积分量。结合去除随机误差的陀螺仪数据,得到修正后的 IMU 预积分,如式(5)所示:

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &= \iint_{t \in [k, k+1]} \mathbf{R}_{b_k}^{b_{k+1}} (\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}) dt^2 \\ \beta_{b_{k+1}}^{b_k} &= \int_{t \in [k, k+1]} \mathbf{R}_{b_k}^{b_{k+1}} (\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}) dt \\ \gamma_{b_{k+1}}^{b_k} &= \int_{t \in [k, k+1]} \gamma_{b_k}^{b_{k+1}} \otimes \begin{bmatrix} 0 \\ \frac{1}{2} (\hat{\boldsymbol{\omega}}_t) \end{bmatrix} dt \end{aligned} \quad (5)$$

其中,  $\mathbf{b}_{a_t}$  表示加速度的偏置,  $\hat{\mathbf{a}}$  表示补偿过随机误差的陀螺仪数据,  $dt$  表示积分时间。这样就得到了连续时刻的 IMU 预积分公式,根据式(5)得到的 IMU 预积分只与不同时刻的  $\hat{\mathbf{a}}_t$  和  $\hat{\boldsymbol{\omega}}_t$  相关。式(5)可以直接从 IMU 数据的积分中得到,这样 IMU 的积分就不再需要依赖于全局的旋转,解决了在世界坐标系下 IMU 难以积分的问题。IMU 预积分技术本质是把参考坐标系进行转换,将 IMU 相关的所有全局旋转信息全部分离,然后就可以在不知道全局旋转的情况下进行积分。

在得到 IMU 的预积分基础上,我们可以将两帧图像之间的 IMU 进行积分。因为整个系统是基于视觉和惯导融合的解决方案,所以两种测量值之间需要进行加权平均,权重可以由 IMU 预积分得到的线性动力学方程得到。假设测量值都是基于高斯模型,则加权值为每一个测量值所对应的一个协方差矩阵。由于 IMU 在每一个时刻积分出来的值均具有误差,因此首先需要对误差进行分析,由 IMU 线性动力学方程<sup>[9]</sup>可以得到在  $t$  时刻误差项的导数,如式(6)所示:

$$\begin{bmatrix} \delta \dot{\alpha}_t^{b_k} \\ \delta \dot{\beta}_t^{b_k} \\ \delta \dot{\gamma}_t^{b_k} \\ \delta \dot{b}_{a_t} \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I} & 0 & 0 \\ 0 & 0 & -\mathbf{R}_t^{b_k} [\hat{\mathbf{a}}_t - \mathbf{b}_{a_t}] & -\mathbf{R}_t^{b_k} \\ 0 & 0 & -[\hat{\boldsymbol{\omega}}_t] & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \delta \alpha_t^{b_k} \\ \delta \beta_t^{b_k} \\ \delta \gamma_t^{b_k} \\ \delta b_{a_t} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ -\mathbf{R}_t^{b_k} & 0 & 0 \\ 0 & -\mathbf{I} & 0 \\ 0 & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} n_a \\ n_w \\ n_{b_a} \end{bmatrix} \quad (6)$$

其中,  $\delta(\dot{\cdot})_{b_k}^{b_k}$  表示对位置预积分量进行求导,  $\delta \dot{b}_{a_t}$  表示对加速度偏置进行求导,  $\mathbf{I}$  表示单位矩阵,  $n$  表示噪声项。将式(6)进行简写得到式(7),可知下一个时刻的 IMU 测量误差与上一个时刻的测量误差成线性关系。

$$\delta z_t^{b_k} = \mathbf{F}_t \delta z_{t-1}^{b_k} + \mathbf{G}_t n_t \quad (7)$$

根据导数定义,由式(7)推导出下一个时刻的状态量,如式(8)所示:

$$\delta z_{t+\delta t}^{b_k} = (\mathbf{I} + \mathbf{F}_t \delta t) \delta z_t^{b_k} + (\mathbf{G}_t \delta t) n_t \quad (8)$$

则可以根据当前时刻的值,预测出下一个时刻的均值和协方差。协方差<sup>[13]</sup>的预测公式如式(9)所示:

$$\mathbf{P}_{t+\delta t}^{b_k} = (\mathbf{I} + \mathbf{F}_t \delta t) \mathbf{P}_t^{b_k} (\mathbf{I} + \mathbf{F}_t \delta t)^T + (\mathbf{G}_t \delta t) \mathbf{Q} (\mathbf{G}_t \delta t)^T \quad (9)$$

其中,  $\mathbf{P}$  表示协方差矩阵,  $\mathbf{Q}$  表示噪声的对角线协方差矩阵,矩阵雅可比行列式用于纠正偏差,在系统运行过程中进行修正,使用偏差来不断地更新状态。雅可比行列式的迭代公式如式(10)所示:

$$\begin{aligned} \mathbf{J}_{b_k} &= \mathbf{I} \\ J_{t+\delta t} &= (\mathbf{I} + F_t \delta t) J_t \end{aligned} \quad (10)$$

其中,  $\mathbf{J}_{b_k}$  表示第  $k$  帧的协方差矩阵,  $\mathbf{I} + F_t \delta t$  表示协方差传递的参数。由此可以得到, 对于偏差纠正的线性模型, 如式(11)所示:

$$\begin{aligned} \alpha_{b_{k+1}}^{b_k} &\approx \hat{\alpha}_{b_{k+1}}^{b_k} + \mathbf{J}_{\alpha}^{\alpha} \delta b_{\alpha_k} \\ \beta_{b_{k+1}}^{b_k} &\approx \hat{\beta}_{b_{k+1}}^{b_k} + \mathbf{J}_{\beta}^{\beta} \delta b_{\beta_k} \\ \gamma_{b_{k+1}}^{b_k} &\approx \hat{\gamma}_{b_{k+1}}^{b_k} \end{aligned} \quad (11)$$

其中,  $\mathbf{J}_{\alpha}^{\alpha}$  代表在状态更新时加速度偏置的雅可比矩阵,  $\mathbf{J}_{\beta}^{\beta}$  代表在状态更新时加速度偏置的雅可比矩阵。

根据中值积分可以得到离散形式的 PVQ 的增量方程, 如式(12)所示:

$$\begin{aligned} \begin{bmatrix} \delta \alpha_{k+1} \\ \delta \theta_{k+1} \\ \delta \beta_{k+1} \\ \delta b_{\alpha_{k+1}} \end{bmatrix} &= \begin{bmatrix} \mathbf{I} & f_{01} & \delta t & f_{03} \\ 0 & f_{11} & 0 & 0 \\ 0 & f_{21} & \mathbf{I} & f_{23} \\ 0 & 0 & 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta \alpha_k \\ \delta \theta_k \\ \delta \beta_k \\ \delta b_{\alpha_k} \end{bmatrix} + \\ &\begin{bmatrix} -\frac{1}{4} R_k \delta t^2 & -\frac{1}{4} R_{k+1} \delta t^2 & 0 \\ 0 & 0 & 0 \\ -\frac{R_k \delta t}{2} & -\frac{R_{k+1} \delta t}{2} & 0 \\ 0 & 0 & \delta t \end{bmatrix} \begin{bmatrix} n_{\alpha_k} \\ n_{\alpha_{k+1}} \\ n_{b_{\alpha_k}} \end{bmatrix} \end{aligned} \quad (12)$$

其中,  $f_{01} = \frac{\delta t}{2}$ ;  $f_{03} = -\frac{1}{4} (R_k + R_{k+1}) \delta t^2$ ;  $f_{11} = \mathbf{I} - \left(\frac{\hat{\omega}_k + \hat{\omega}_{k+1}}{2}\right)^\wedge \delta t$ ;  $f_{21} = -\frac{1}{2} R_k (\hat{a}_k - b_{\alpha_k})^\wedge \delta t - \frac{1}{2} R_{k+1} (\hat{a}_{k+1} - b_{\alpha_{k+1}})^\wedge \left[\mathbf{I} - \left(\frac{\hat{\omega}_k + \hat{\omega}_{k+1}}{2}\right)^\wedge \delta t\right] \delta t$ ;  $f_{23} = -\frac{1}{2} (R_k + R_{k+1}) \delta t$ ;  $n_{\alpha_k}$ ,  $n_{\alpha_{k+1}}$  和  $n_{b_{\alpha_k}}$  表示  $k$  时刻、 $k+1$  时刻以及加速度 Bias 的噪声项。

### 3.4 基于补偿陀螺仪的 IMU 与相机松耦合

融合注意力特征的无锚框视觉目标跟踪方法的网络离线训练过程如第 3.2 节所述, 其具体实现如图 3 所示。

基于单目相机与 IMU 紧耦合的 VIO 系统是一个高度非线性系统, 要想系统快速收敛, 则需要一个良好的系统初始值。而且, 由于单目相机无法直接获得尺度信息, 因此, 如果没有良好的初始值, 则很难直接将这两种测量结果有效地融合在一起。虽然可以假设一个静止的初始条件来启动单目 VIO 估计器, 但是这种假设并不合理, 因为在实际应用中, 经常会遇到运动状态下的初始化问题。而且, 当 IMU 测量结果被更大的偏置破坏时, 则情况会变得更加复杂。事实上, 对于基于单目相机的 VIO 系统来说, 初始化过程通常是其最脆弱的一个步骤, 若要保证系统的鲁棒性, 则需要一个鲁棒的初始化过程。

目前, 现有的 VIO 系统初始化过程中, 通常使用纯视觉 SLAM 或从运动中恢复结构 SFM 的方法, 对单目 VIO 系统进行初始化。在大多数情况下, 纯视觉系统可以通过从相对运动方法(如八点法或五点法或估计单应性矩阵)中导出初始

值来引导自己, 但是在初始化过程中, 对于高速运动、光照变换剧烈和纹理重复等 SFM 估计效果差的场景, 盲目地将 SFM 当作真值无法避免估计值不准确所带来的严重问题。

为了保障 VIO 系统的鲁棒性, 我们采用将 IMU 预积分与纯视觉 SFM 结果松耦合的方式, 分别计算 IMU 预积分结果以及纯视觉 SFM 结果, 再将两种结果的平移估计对齐, 采用松耦合的传感器融合方法得到系统的初始值(陀螺仪 bias、尺度因子、速度、重力向量), 粗略地快速恢复尺度、重力、速度等信息, 从而引导非线性单目 VINS 估计器, 在简化初始化过程的同时, 进一步减少计算量。使用松耦合的原因在于能够提升系统的运行频率, 可以更加快速地计算初始化过程, 满足初始化阶段系统快速高效的要求。

本文使用去除误差项的陀螺仪数据积分, 来获得旋转四元数与图像特征, 通过  $n$  点透视(Perspective- $n$ -Point, PnP)恢复得到旋转四元数, 根据图 4 构造旋转约束<sup>[23]</sup>来求解旋转外参数  $q_c^k$ 。图 4 中, 星号表示路标点,  $c_k$  表示  $k$  时刻的相机,  $b_k$  表示  $k$  时刻的 IMU。坐标系转换旋量为  $c \rightarrow b \rightarrow b_{k+1} \rightarrow b_k$  以及  $c_{k+1} \rightarrow c_k \rightarrow c \rightarrow b$ , 根据两种路径旋转量一致构造旋转约束。

$$q_{b_{k+1}}^{b_k} \otimes q_c^b = q_c^b \otimes q_{c_{k+1}}^{c_k} \quad (13)$$

$$([\cdot]_L^{b_{k+1}}]_L - [q_{c_{k+1}}^{c_k}]_R] q_c^b = Q_{k+1}^k \cdot q_c^b = 0$$

其中,  $[\cdot]_L$  和  $[\cdot]_R$  分别代表四元数的左乘和右乘。将多个时刻公式累计, 同时为了增加  $q_c^k$  估计的鲁棒性, 加上 Huber 鲁棒核函数, 如式(14)所示:

$$\begin{aligned} \begin{bmatrix} \omega_1^0 \cdot Q_1^0 \\ \omega_{\frac{1}{2}} \cdot Q_{\frac{1}{2}} \\ \vdots \\ \omega_N^{N-1} \cdot Q_N^{N-1} \end{bmatrix} q_{bc} = Q_N \cdot q_{bc} = 0 \\ \omega_{k+1}^k = \begin{cases} 1, & r_{k+1}^k < \text{threshold} \\ \frac{\text{threshold}}{r_{k+1}^k}, & \text{otherwise} \end{cases} \end{aligned} \quad (14)$$

$$r_{k+1}^k = \text{acos}((\text{tr}(\hat{R}_{bc}^{-1} \mathbf{R}_{b_{k+1}}^{-1} \hat{R}_{bc} \mathbf{R}_{c_{k+1}})) - 1) / 2$$

其中,  $\omega_{k+1}^k$  为参数值,  $r_{k+1}^k$  为角度误差, 由于我们使用了补偿过误差项的陀螺仪数据, 因此可以有效增加  $Q_N$  的信噪比。

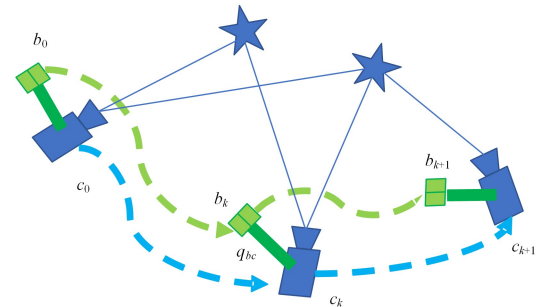


图 4 旋转外参数求解示意图

Fig. 4 Schematic diagram of the solution for outer rotation parameter

在陀螺仪偏置初始化后, 使用平移约束继续初始化其他基本状态量, 即速度、重力向量和尺度。最后, 需要估计的量为  $\gamma_i$ :

$$\chi_I = [\mathbf{v}_{0_n}^{b_n}, \mathbf{v}_{1_n}^{b_n}, \dots, \mathbf{v}_{n_n}^{b_n}, \mathbf{g}^{c_0}, s]^T \quad (15)$$

其中,  $\mathbf{v}_{n_n}^{b_n}$  是第  $n$  帧图像 body 坐标系的速度,  $\mathbf{g}^{c_0}$  是  $c_0$  坐标系中的重力向量,  $s$  是单目 SfM 到公制单位的尺度。根据视觉惯性对齐, 由平移约束可得:

$$T_{c_k}^{c_0} (T_k^b)^{-1} = T_{c_k}^{c_0} \quad (16)$$

$$s \bar{\mathbf{p}}_{c_k}^{c_0} = s \bar{\mathbf{p}}_{c_k}^{c_0} - \mathbf{R}_{c_k}^{c_0} \mathbf{p}_k^b$$

将式(12)的世界坐标系换成相机初始时刻坐标系  $c_0$ , 如式(17)所示:

$$\alpha_{b_{k+1}}^{b_k} = \mathbf{R}_{c_0}^{b_k} (s(\bar{\mathbf{p}}_{b_{k+1}}^{c_0} - \bar{\mathbf{p}}_{c_k}^{c_0}) + \frac{1}{2} \mathbf{g}^{c_0} \Delta t_k^2 - \mathbf{R}_{c_k}^{c_0} \mathbf{v}_k^{b_k} \Delta t_k) \quad (17)$$

$$\beta_{b_{k+1}}^{b_k} = \mathbf{R}_{c_0}^{b_k} (\mathbf{R}_{b_{k+1}}^{c_0} \mathbf{v}_{b_{k+1}}^{b_k} + \mathbf{g}^{c_0} \Delta t_k - \mathbf{R}_{c_k}^{c_0} \mathbf{v}_k^{b_k})$$

其中  $s$  代表尺度因子。

将式(16)代入式(17)得到:

$$\alpha_{b_{k+1}}^{b_k} = s \mathbf{R}_{c_0}^{b_k} (\bar{\mathbf{p}}_{b_{k+1}}^{c_0} - \bar{\mathbf{p}}_{c_k}^{c_0}) - \mathbf{R}_{c_0}^{b_k} \mathbf{R}_{b_{k+1}}^{c_0} \mathbf{p}_k^b + \mathbf{p}_k^b + \frac{1}{2} \mathbf{R}_{c_0}^{b_k} \mathbf{g}^{c_0} \Delta t_k^2 - \mathbf{v}_k^{b_k} \Delta t_k \quad (18)$$

构建待估计量  $\mathcal{Z}_I$  的优化方程, 如式(19)所示:

$$\hat{\mathcal{Z}}_{b_{k+1}}^{b_k} = \begin{bmatrix} \hat{\alpha}_{b_{k+1}}^{b_k} - \mathbf{p}_k^b + \mathbf{R}_{c_0}^{b_k} \mathbf{R}_{b_{k+1}}^{c_0} \mathbf{p}_k^b \\ \hat{\beta}_{b_{k+1}}^{b_k} \end{bmatrix} = \mathbf{H}_{b_{k+1}}^{b_k} \chi_I^k + \mathbf{n}_{b_{k+1}}^{b_k} \quad (19)$$

其中,  $\bar{\mathbf{p}}$  表示带尺度信息的位置数据,  $\mathbf{H}_{b_{k+1}}^{b_k} =$

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

的更新矩阵,  $\hat{\mathcal{Z}}_{b_{k+1}}^{b_k}$  表示状态增量,  $\mathbf{n}_{b_{k+1}}^{b_k}$  代表噪声项。

可以看出,  $\mathbf{R}_{c_0}^{b_k}$ ,  $\mathbf{R}_{b_{k+1}}^{c_0}$ ,  $\bar{\mathbf{p}}_{c_k}^{c_0}$ ,  $\bar{\mathbf{p}}_{b_{k+1}}^{c_0}$  是从带尺度的单目视觉中得到的,  $\Delta t_k$  是两个连续帧之间的时间间隔。将式(20)转换成线性最小二乘问题对状态量进行求解:

$$\min_{\chi_I} \sum_{k \in \mathbb{S}} \|\hat{\mathcal{Z}}_{b_{k+1}}^{b_k} - \mathbf{H}_{b_{k+1}}^{b_k} \chi_I^k\|^2 \quad (20)$$

基于上述计算, 则可以得到窗口中每一帧的惯性坐标系速度、视觉参照系 ( $\mathbf{g}$ )<sup>c0</sup> 的重力向量以及尺度等参数, 进而可以有效地对系统参数进行初始化。

## 4 实验及结果分析

本章通过一系列实验, 验证了本文提出的初始化方法的高效性和稳定性。在众多初始化参数中, 根据文献[16]可以知道陀螺仪 Bias 对位姿估计具有非常明显的影响。因此, 首先将本文方法得到的陀螺仪 Bias 与 VINS-Mono 得到的参数进行对比和分析; 其次, 将所有初始化参数进行求解, 并应用到 VINS-Mono 中, 对位姿估计结果进行分析; 最后, 为了达到快速初始化的目的, 对各个初始化方法的执行时间进行了测试和对比。

### 4.1 实现细节

为了寻找快速运动的序列, 在实验过程中使用 EuRoC 数据集进行测试。该数据集包含双目图像 (Aptina MT9V034 全局快门, WVGA 单色, 20 FPS 等)、同步 IMU 测量 (ADIS16448, 角速率和加速度, 200 Hz) 和地面真实状态 (VI-CON 和 Leica MS50)。EuRoC 数据集由无人机在两个场景

中采集的图像和 200Hz 的 IMU 数据 (未校准) 组成, 总共包含 11 个序列。训练集为 MH\_01\_easy, MH\_03\_medium, MH\_05\_difficult, V1\_02\_medium, V2\_01\_easy 和 V2\_03\_difficult; 测试集为 MH\_02\_easy, MH\_04\_difficult, V2\_02\_medium, V1\_03\_difficult 和 V1\_01\_easy。

本文方法是基于 PyTorch 1.10.1 实现的, 选择 ADAM 优化器[22], 学习率设置为 0.01。设置 1800 个 epochs 进行训练。实验中使用的 GPU 是 GTX 3090, CPU 是 Intel i7-8700。每个训练序列需要 33s 左右, 显存消耗在 6GB 以内, 在测试序列上验证了本文的初始化方法。

### 4.2 陀螺仪 Bias

我们选取了初始化过程中的关键参数陀螺仪 Bias, 对不同时刻开始初始化的估计值进行计算。同时, 选取了明亮场景且具有清晰纹理的序列 MH\_02\_easy 以及黑暗环境快速运动的序列 MH\_04\_difficult, 在整个时间序列上计算陀螺仪 Bias, 保证恢复的准确性。图 5、图 6 分别给出了 MH\_04\_difficult 序列和 MH\_02\_easy 序列在初始化阶段陀螺仪 Bias 的恢复效果, 都包含了在 25 个不同起始时间的结果。

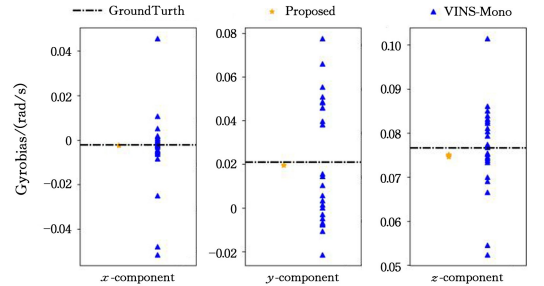


图 5 MH\_04\_difficult 序列在初始化阶段陀螺仪 Bias 的恢复效果  
Fig. 5 Recovery effects of gyroscopic Bias during the initialization phase of MH\_04\_difficult sequence

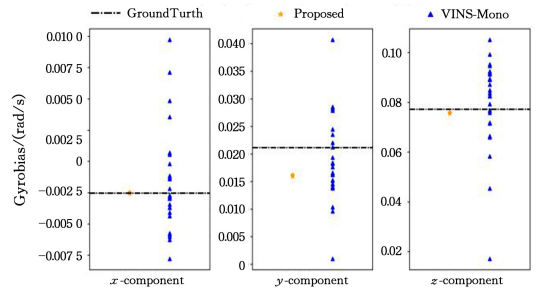


图 6 MH\_02\_easy 序列在初始化阶段陀螺仪 Bias 的恢复效果  
Fig. 6 Recovery effect of gyroscope Bias during the initialization phase of MH\_02\_easy sequence

根据图 5 可知, 在 MH\_04\_difficult 序列中, 由于 VINS-Mono 在求解  $b_w$  的过程中需要使用 SfM, 因此在黑暗和快速运动等视觉遇到挑战的场景下, 其估计值及其不稳定, 而本文方法在具有挑战性的环境下, 仍然可以进行准确的估计。在图 5 中可以看到, 本文方法的恢复值非常稳定, 同时估计的数据也非常稳定, 不会产生很大的差异。在明亮且纹理清晰的 MH\_02\_easy 序列中, 在图 6 中可以看到, 依赖 SfM 的初始

化方法恢复的 $b_{\omega}$ 在真值左右浮动,而本文方法则可以得到非常鲁棒的估计结果,并且十分接近真实值。

为了能够清楚地知道在整个序列上 $b_{\omega}$ 的恢复效果,我们

将上述两个序列在整个时间轴上对 $b_{\omega}$ 进行恢复。如图7和图8所示,两个序列 $b_{\omega}$ 的恢复值均在真值上下浮动,且浮动数值极小。

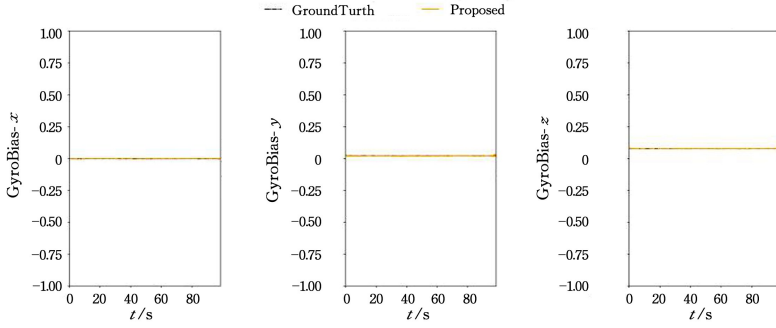


图7 MH\_04\_difficult 在整个序列上的 $b_{\omega}$ 估计值

Fig. 7  $b_{\omega}$  estimate of MH\_04\_difficult over the whole series

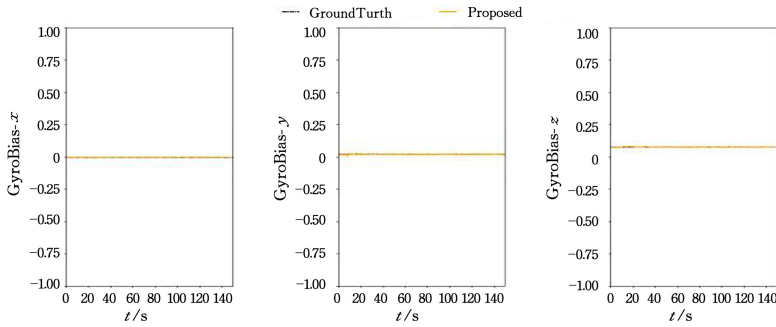


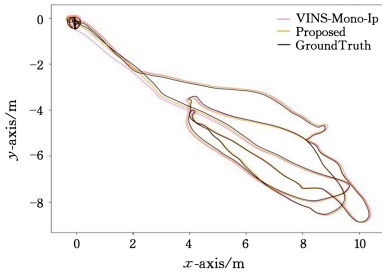
图8 MH\_02\_easy 在整个序列上的 $b_{\omega}$ 估计值

Fig. 8  $b_{\omega}$  estimate of MH\_02\_easy over the whole series

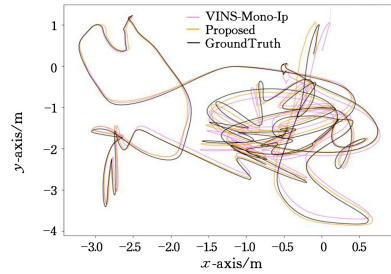
### 4.3 姿态估计效果

将恢复好的初始化参数应用到 VINS-Mono 系统中,测试位姿估计的精度。我们选择了 MH\_02\_easy 和 V1\_03\_difficult 两个序列,对整体性能表现进行测试。在实验中,

将本文方法与 VINS-Mono(loop closure)<sup>[4]</sup>进行对比。两个序列的轨迹估计结果如图9所示,从图中可以看出,本文方法的位姿估计精度更高,特别是在转角和快速运动等场景中。



(a) MH\_02\_easy



(b) V1\_03\_difficult

图9 MH\_02\_easy 和 V1\_03\_difficult 序列轨迹估计结果

Fig. 9 Estimation results of MH\_02\_easy and V1\_03\_difficult sequence trajectory

为了能清晰地看到在每个时刻两种算法的差异,我们计算两种算法的绝对位移误差(APE)。如图10所示,可以看到,几乎在所有时刻本文方法都具有更高的精度。由于 V1\_03\_difficult 序列具有快速运动的特点,为了更好地证明本文方法的有效性,我们对其旋转和平移两部分的误差分别进行了计算,结果如图11所示。

同时,为了定量分析,将本文方法在所有序列上进行了测试,结果如表1所列。从表1可以看到,本文方法的误差较小。

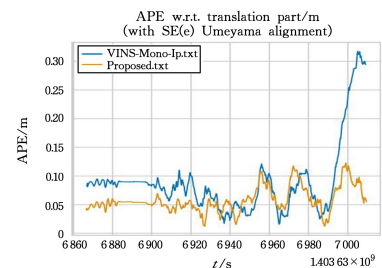


图10 MH\_02\_easy 序列在本文方法和 VINS-Mono 中的 APE  
Fig. 10 APE of MH\_02\_easy sequence in our method and VINS-Mono

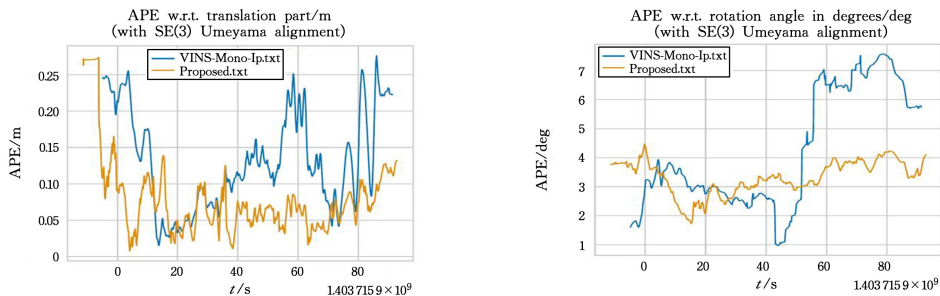


图 11 V1\_03\_difficult 序列在本文方法和 VINS-Mono 中的 APE 和 AOE

Fig. 11 APE and AOE of V1\_03\_difficult sequence in our method and VINS-Mono

表 1 本文方法在旋转和平移上的性能

Table 1 Performance of the proposed method in terms of rotations and translations

|                 | Proposed              |         |       |                     |       |       |              | VINS-Mono    |              |
|-----------------|-----------------------|---------|-------|---------------------|-------|-------|--------------|--------------|--------------|
|                 | Orientation Error/deg |         |       | Translation Error/m |       |       | RMSE/<br>deg | RMSE/m       | RMSE/m       |
|                 | E_roll                | E_pitch | E_yaw | E_x                 | E_y   | E_z   |              |              |              |
| MH_02_easy      | 0.285                 | 0.181   | 0.692 | 0.047               | 0.048 | 0.029 | 0.769        | <b>0.073</b> | 0.112        |
| MH_04_difficult | 0.232                 | 0.334   | 0.958 | 0.120               | 0.087 | 0.058 | 1.040        | <b>0.159</b> | 0.169        |
| V1_01_easy      | 0.223                 | 0.292   | 0.430 | 0.021               | 0.024 | 0.015 | 0.565        | <b>0.035</b> | 0.069        |
| V1_03_difficult | 0.701                 | 0.682   | 3.328 | 0.037               | 0.072 | 0.060 | 3.469        | <b>0.101</b> | 0.158        |
| V2_02_medium    | 0.830                 | 0.639   | 1.402 | 0.054               | 0.039 | 0.032 | 1.752        | 0.074        | <b>0.062</b> |
| Mean            | 0.454                 | 0.426   | 1.366 | 0.056               | 0.054 | 0.039 | 1.519        | <b>0.088</b> | 0.114        |

注:加粗数据为最优结果,斜体数据为次优结果。

#### 4.4 计算消耗

实时性是衡量初始化过程好坏的一个重要指标,将本文方法与 VINS-Mono 的初始化过程<sup>[4]</sup>进行对比,对比结果如表 2 所列。从表中可以看出,本文方法能够用更短的时间获得更加精确的初始化参数。

表 2 本文初始化方法和 VINS-Mono 的初始化方法耗时计算

Table 2 Time-consuming calculation of our initialization method and VINS-Mono's initialization method

|                 | (s)       |          |
|-----------------|-----------|----------|
|                 | VINS-Mono | Proposed |
| MH_02_easy      | 0.131341  | 0.110667 |
| MH_04_difficult | 0.126414  | 0.112763 |
| V1_01_easy      | 0.162555  | 0.082867 |
| V1_03_difficult | 0.185794  | 0.079762 |
| V2_02_medium    | 0.117559  | 0.097099 |
| Mean            | 0.144733  | 0.096632 |

**结束语** 本文提出了一种基于深度学习的视觉惯性里程计非联合初始化方法。本文通过将深度学习引入初始化方法中,求解陀螺仪误差项( $b_w$ 等),避免了直接使用 SFM 作为真值来进行初始化,不仅能够使得初始化不再受限于高速运动、光照变换剧烈和纹理重复等 SFM 估计效果不好的场景,而且还能够简化初始化过程。通过增加初始化阶段 IMU 数据的信噪比,使得在使用最小二乘时,能够使用更少的帧来进行求解,进一步减少计算量。在 EuRoC 数据集上,无论是精度还是耗时均取得了良好的效果。

#### 参考文献

[1] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: a versatile and accurate monocular SLAM system[J]. IEEE Transactions on Robotics, 2015, 31(5): 1147-1163.  
 [2] MUR-ARTAL R, TARDÓS J D. Orb-slam2: An open-source

slam system for monocular, stereo, and rgb-d cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.

[3] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(3): 611-625.  
 [4] 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.  
 [5] GENEVA P, ECKENHOFF K, LEE W, et al. OpenVINS: A Research Platform for Visual-Inertial Estimation[C]// Proceedings of the IEEE International Conference on Robotics and Automation. IEEE, 2020.  
 [6] QIN T, SHEN S. Robust initialization of monocular visual-inertial estimation on aerial robots[C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 4225-4232.  
 [7] CAMPOS C, MONTIEL J M M, TARDÓS J D. Fast and robust initialization for visual-inertial SLAM[C]// 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019: 1288-1294.  
 [8] YANG Z, SHEN S. Monocular visual-inertial state estimation with online initialization and camera-IMU extrinsic calibration [J]. IEEE Transactions on Automation Science and Engineering, 2016, 14(1): 39-51.  
 [9] CAMPOS C, ELVIRA R, RODRÍGUEZ J J G, et al. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual-Inertial, and Multimap SLAM[J]. IEEE Transactions on Robotics, 2021, 37(6): 1874-1890.  
 [10] DOMÍNGUEZ-CONTI J, YIN J, ALAMI Y, et al. Visual-inertial slam initialization: A general linear formulation and a gravity-observing non-linear optimization[C]// 2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR).

- IEEE, 2018: 37-45.
- [11] MUR-ARTAL R, TARDÓS J D. Visual-inertial monocular SLAM with map reuse[J]. IEEE Robotics and Automation Letters, 2017, 2(2): 796-803.
- [12] MUSTANIEMI J, KANNALA J, SÄRKKÄ S, et al. Inertial-based scale estimation for structure from motion on mobile devices[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems(IROS). IEEE, 2017: 4394-4401.
- [13] SNAVELY N, SEITZ S M, SZELISKI R. Modeling the world from internet photo collections [J]. International Journal of Computer Vision, 2008, 80(2): 189-210.
- [14] FORSTER C, CARLONE L, DELLAERT F, et al. On-manifold preintegration for real-time visual-inertial odometry[J]. IEEE Transactions on Robotics, 2016, 33(1): 1-21.
- [15] MARTINELLI A. Closed-form solution of visual-inertial structure from motion[J]. International Journal of Computer Vision, 2014, 106(2): 138-152.
- [16] KAISER J, MARTINELLI A, FONTANA F, et al. Simultaneous state initialization and gyroscope bias calibration in visual inertial aided navigation[J]. IEEE Robotics and Automation Letters, 2016, 2(1): 18-25.
- [17] HUANG W, LIU H. Online initialization and automatic camera-IMU extrinsic calibration for monocular visual-inertial SLAM [C]//2018 IEEE International Conference on Robotics and Automation(ICRA). IEEE, 2018: 5182-5189.
- [18] CAMPOS C, MONTIEL J M M, TARDÓS J D. Inertial-only optimization for visual-inertial initialization[C]//2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2020: 51-57.
- [19] GAO Y Q, SHI D X, LI R H, et al. Gyro-Net: IMU Gyroscopes Random Errors Compensation Method Based on Deep Learning [C]//IEEE Robotics and Automation Letters.
- [20] ESFAHANI M A, WANG H, WU K, et al. OriNet: Robust 3-D orientation estimation with a single particular IMU[J]. IEEE Robotics and Automation Letters, 2019, 5(2): 399-406.
- [21] BROSSARD M, BONNABEL S, BARRAU A. Denoising imu gyroscopes with deep learning for open-loop attitude estimation [J]. IEEE Robotics and Automation Letters, 2020, 5(3): 4796-4803.
- [22] LUPTON T, SUKKARIEH S. Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions[J]. IEEE Transactions on Robotics, 2011, 28(1): 61-76.
- [23] HORAUD R, DORNAIKA F. Hand-eye calibration[J]. The International Journal of Robotics Research, 1995, 14(3): 195-210.



**SHI Dianxi**, born in 1966, Ph.D, professor, Ph.D supervisor. His main research interests include artificial intelligence, robot operating system, distributed computing, and cloud computing.



**CHEN Ying**, born in 1985, Ph.D, assistant research fellow. Her main research interests include artificial intelligence algorithm and framework design and optimization.

(责任编辑:喻黎)