VINS-Mono with Robust Initialization and Visual Weight Setting

Zewen Xu, Zeren Lv, Zhaolong Yang, Yidi Zhang, Hao Wei,

摘要

本系统是在 VINS-MONO [1]上改进的。通过观察，测试环境的场景比较固定，因此本系统采用多地图的模式来得到更高的精度轨迹。本系统相比原始的 VINS-MONO，主要修改了以下两个部分：(1)初始化采用极平面约束，利用旋转平移解耦的的方法估计陀螺仪的偏差 bg；(2)对视觉特征进行协方差估计，并设置了视觉观测的权重。

I. 鲁棒初始化

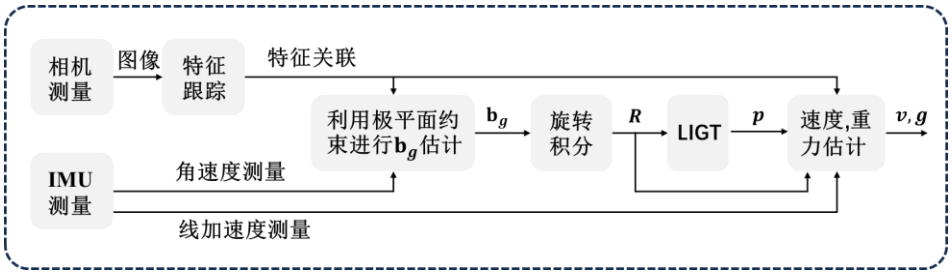

图 1 初始化流程图

为了得到更好的初始化结果，本系统使用 He 等人[2]提出的初始化方案替换了 VINS-MONO[1]中原有的初始化，流程如图 1 所示。我们利用 ceres 建立了极平面约束因子，实现对$\boldsymbol{b}_g$的估计，

$$\boldsymbol{b}_g^* = \underset{\boldsymbol{b}_g}{\operatorname{argmin}} \sum_{(i,j)\in\varepsilon} \lambda_{\boldsymbol{M}_{\mathrm{i,j}},min}, \tag{1}$$

其中ε为滑窗内的关键帧集合，$\lambda_{\boldsymbol{M}_{\mathrm{i,j}},min}$表示$\boldsymbol{M}_{\mathrm{i,j}}$矩阵的最小特征值，$\boldsymbol{M}_{\mathrm{i,j}}$为关键帧 i 和关键帧 j 之间的极平面约束矩阵,

$$M_{i,j} = \sum_{k=1}^{n} \boldsymbol{n}_k(\boldsymbol{b}_g)\, \boldsymbol{n}_k^T(\boldsymbol{b}_g), \tag{2}$$

其中$\boldsymbol{n}_k(\boldsymbol{b}_g)$表示关键帧 i，j 之间匹配的第 k 个特征构成的极平面的法向量，它是$\boldsymbol{b}_g$的函数，具体关系见参考文献[2]中的公式（6）。极平面约束的好处是无需重建三维特征，同时使得旋转和平移解耦，这样可以在提高初始化速度的同时得到更加精确的$\boldsymbol{b}_g$。之后利用 imu 测量和估计的$\boldsymbol{b}_g$积分得到关键帧的旋转$\boldsymbol{R}_{1i} \in SO(3), 1 < i \le n$,其中 n 为滑窗大小。进而利用多帧的线性全局平移约束(LiGT)[3]得到对全局平移$\boldsymbol{p}_{1i} \in \mathbb{R}^3, 1 < i \le n$的线性约束，

$$\begin{aligned} &\boldsymbol{L} \cdot \boldsymbol{P} = \boldsymbol{0} \\ &\boldsymbol{P} = (\boldsymbol{p}_{12}^T, \dots, \boldsymbol{p}_{1n}^T)^T, \\ &\boldsymbol{L} = \boldsymbol{A}^T \boldsymbol{A} \end{aligned} \tag{3}$$

需要注意的是$\boldsymbol{p}_{11}^T = (\boldsymbol{0}, \boldsymbol{0}, \boldsymbol{0})^T$,其中

$$
\boldsymbol{A}=\begin{bmatrix}
\boldsymbol{C}_{1,1} & \boldsymbol{D}_{1,1} & & & & & & & \\
\vdots & \vdots & & & & & & & \\
\boldsymbol{C}_{1,k_1} & \boldsymbol{D}_{1,k_1} & & & & & & & \\
\boldsymbol{B}_{2,1} & \boldsymbol{C}_{2,1} & \boldsymbol{D}_{2,1} & & & & & & \\
\vdots & \vdots & \vdots & & & & & & \\
\boldsymbol{B}_{2,k_2} & \boldsymbol{C}_{2,k_2} & \boldsymbol{D}_{2,k_2} & \ddots & & & & & \\
 & & & & \boldsymbol{B}_{n-3,1} & \boldsymbol{C}_{n-3,1} & \boldsymbol{D}_{n-3,1} & & \\
 & & & & \vdots & \vdots & \vdots & & \\
 & & & & \boldsymbol{B}_{n-3,k_{n-3}} & \boldsymbol{C}_{n-3,k_{n-3}} & \boldsymbol{D}_{n-3,k_{n-3}} & & \\
 & & & & & \boldsymbol{B}_{n-2,1} & \boldsymbol{C}_{n-2,1} & \boldsymbol{D}_{n-2,1} \\
 & & & & & \vdots & \vdots & \vdots \\
 & & & & & \boldsymbol{B}_{n-2,k_{n-2}} & \boldsymbol{C}_{n-2,k_{n-2}} & \boldsymbol{D}_{n-2,k_{n-2}}
\end{bmatrix}, \tag{4}
$$

其中k_i表示第 i 帧，i+1 帧和 i+2 帧共视特征的数量，矩阵$\boldsymbol{B},\boldsymbol{C},\boldsymbol{D}$的具体形式参考文献[2]中的公式（10）。利用估计出的旋转和平移替换掉 VINS-MONO 中视觉 imu 对齐函数的输入，得到重力方向和速度，完成初始化。

II. 视觉特征权重设置

为了更加合理的设置权重，本系统没有将所有特征在不同帧上的协方差都设定为一个常量，而是从残差数据中自适应的估计特征的协方差。具体来说，我们计算了残差的绝对中位差（MAD）。对满足正态分布的样本的 MAD 乘以一个比例因子 c = 1.4826，可以作为其分布标准偏差的一致估计量。因此，第 k 帧上的观测残差的标准偏差为，

$$
\hat{\sigma}\left(\{res_{k,i}\}\right)=c\,median\left(\left\{\left|res_{k,i}-median\left(\{res_{k,i}\}\right)\right|\right\}\right), \tag{5}
$$

信息矩阵为协方差矩阵的逆，所以第 k 帧的视觉观测的信息矩阵为，

$$
I=\begin{bmatrix} 1/\hat{\sigma}\left(\{res_{k,i}\}\right)^2 & 0 \\ 0 & 1/\hat{\sigma}\left(\{res_{k,i}\}\right)^2 \end{bmatrix}, \tag{6}
$$

由于观测不确定性的引入，在代码运行过程中，会自适应的调整 imu 和视觉测量的相对权重，也就是说在视觉误差较大的情况下，优化器会更加倾向于利用预积分项提供的信息。

III. 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] He Y, Xu B, Ouyang Z, et al. A Rotation-Translation-Decoupled Solution for Robust and Efficient Visual-Inertial Initialization[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 739-748.

[3] Cai Q, Zhang L, Wu Y, et al. A pose-only solution to visual reconstruction and navigation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 73-86.