

## （文献阅读）多传感器融合之 R2LIVE: A Robust, Real-time, LiDAR-Inertial-Visual tightly-coupled state Estimator and mapping

近期看了一个很震撼的 SLAM 视频，这一直是我想做但没做出的（尴尬）。所以就忍不住的看了全文。

文章地址: <https://arxiv.org/pdf/2102.12400.pdf>

视频展示: <https://www.bilibili.com/video/BV1xp4y187uz>

**摘要:** 在本文中,我们提出了一个鲁棒的、实时紧耦合的多传感器融合框架,该框架融合了来自 LiDAR、惯性传感器和视觉相机的测量结果,从而实现了鲁棒且准确的状态估计。我们提出的框架由两部分组成:基于滤波器里程计和因子图优化。为了保证实时性能,我们在误差状态迭代卡尔曼滤波器的框架内估计状态,并通过因子图优化进一步提高整体精度。利用所有传感器的测量优势,我们的算法对各种视觉故障、LiDAR 退化场景具有足够的鲁棒性,并且能够在机载计算平台上实时运行,如在室内进行的广泛测试所证明的那样,室外和不同规模的混合环境。此外,结果表明,我们提出的框架可以提高最新的 LiDAR 惯性或视觉惯性里程计的准确性。为了分享我们的发现并为社区做出贡献,我们在 Github 上开源了我们的代码 (<https://github.com/hku-mars/r2live>)。

### 文章细节

具有估计六自由度自我运动的能力,同时构建周围环境的密集和高精度地图,基于激光雷达 SLAM 已被广泛应用于自动驾驶车辆[1]、无人机[2,3]等领域。随着激光雷达技术的发展,低成本激光雷达(如激光雷达[4])的出现使激光雷达更为方便。遵循这一趋势,[5]-[8]的一些相关工作引起了学术圈对这一研究领域的关注。然而,基于激光雷达 SLAM 方法在那些几何特征很少的情况下,很容易失败(即退化),这对于那些具有小 FOV[9]的激光雷达来说更为重要。在这项工作中,为了解决基于激光雷达的测速法的退化问题,我们提出了一个激光-惯性-视觉融合框架来获得更高的鲁棒性和准确性的状态估计。我们工作的主要贡献是:

我们利用激光雷达、惯性传感器和相机传感器的测量结果，并以紧密耦合的方式融合它们。实验表明，我们的方法在各种具有挑战性的情况下，甚至在具有大量运动物体和小激光雷达视场的狭窄隧道状环境中，也具有足够的鲁棒性。

我们提出了一个基于高速率滤波器的测距法和低速率因子图优化的框架。基于滤波器的里程表将 LiDAR、惯性和相机传感器的测量结果融合在一个误差状态迭代的卡尔曼滤波器中，以实现实时性能。因子图优化局部地图中关键帧的位姿和视觉界标位置。

通过紧密融合不同类型的传感器，我们实现了高精度的状态估计。实验结果表明，该系统足够精确。

在本节中，我们回顾了与我们的工作密切相关的现有工作，包括仅限激光雷达的测速法和映射、激光雷达惯性融合和激光雷达-惯性-视觉方法。

## A 激光里程计与建图

zhang 人[10]首先提出了一个激光雷达测距和建图框架，该方法将[11]与点到平面和点到边缘的距离相结合。通过以不同的速度运行这两个模块，实现了良好的里程计和建图性能。为了使算法在计算有限的平台上实时运行，shan 等[12]提出了一种轻量级的地面优化 LOAM (Lego-loam)，消除了地面分割步骤中不可靠的特征。这些工作主要是基于多线激光雷达。我们之前的工作[9,13]通过考虑具有小 FOV 的固态激光雷达，开发了一个精确和鲁棒的算法。然而，这些方法仅基于激光雷达测量，非常容易受到无特征的环境或其他退化的情况。

## B 激光惯性里程计

现有的激光雷达-惯性融合的工作可以分为两类：松散耦合和紧密耦合。松散耦合方法分别处理两个传感器，以推断它们的运动约束，而紧耦合方法则通过联合优化直接融合激光雷达和惯性测量。与松散耦合方法相比，紧密耦合方法表现出较高的鲁棒性和准确性，因此近年来引起了越来越多的研究兴趣。例如，[14]的作者提出了一种基于激光雷达惯性里程计 LIOM 的先验和旋转约束优化方法的图优化。与前一种算法相比，LIO-SAM[15]在因子图中优化了关键帧姿态的滑动窗口，以获得更高的精度。同样，li 等人也一样，提出了基于滑动窗口优化的传统和固态激光器的 LiLi-OM 算法[16]。LINS[17]是第一个通过迭代卡尔曼滤波解决 6 个自我运动的紧密耦合的 LIO。为了降低计算卡尔曼增益时的高计算负载，

我们之前的工作快速 Fast-LIO[5]提出了一个新的卡尔曼增益计算公式,结果的计算复杂度取决于状态维数而不是测量维数。这项工作在 uav 上的嵌入式计算机上运行时,实现了高达 50hz 的里程计和建图速率。

### C 激光-惯性-相机里程计

在激光雷达惯性方法的基础上,结合视觉传感器测量结果的激光雷达惯性视觉里程计具有较高的鲁棒性和精度。在[18]的工作中,激光雷达测量被用于为相机图像提供深度信息,形成一个类似于 rgb-d 相机的系统,可以利用现有的视觉 SLAM 工作,如 ORB-SLAM[19]。这是一种松散耦合的方法,因为它忽略了激光雷达测量所施加的对状态的直接约束。zuo 等人[20]提出了一种基于 MSCKF 框架的结合 IMU 测量、稀疏视觉特征、激光雷达平面和边缘特征与的在线时空校准的 LIC-Fusion 框架,并称该框架比现有最先进的方法更准确、稳健。他们的进一步工作称为 LIC-Fusion 2.0[21]改进了一种滑窗内多个激光雷达扫描的平面特征跟踪算法,使激光雷达扫描匹配更加鲁棒性。

据我们所知,我们的工作第一个开源的紧密耦合的激光雷达-惯性视觉融合系统。通过融合不同类型的传感器测量值,我们实现了更高的精度和鲁棒性的状态估计。结果表明,我们的系统比最先进的激光惯性和视觉惯性融合估计器(如快速激光[5]和静脉单[22])更准确和鲁棒。

我们的系统概述如图 2 所示,其中基于滤波器的里程计利用 LiDAR、相机和惯性传感器的测量值,在误差状态迭代卡尔曼滤波器框架内估计状态,如第 IV 节所述。为了进一步改善视觉测量,我们利用因子图优化来优化局部滑动窗口内的视觉界标,如第五部分中所述。

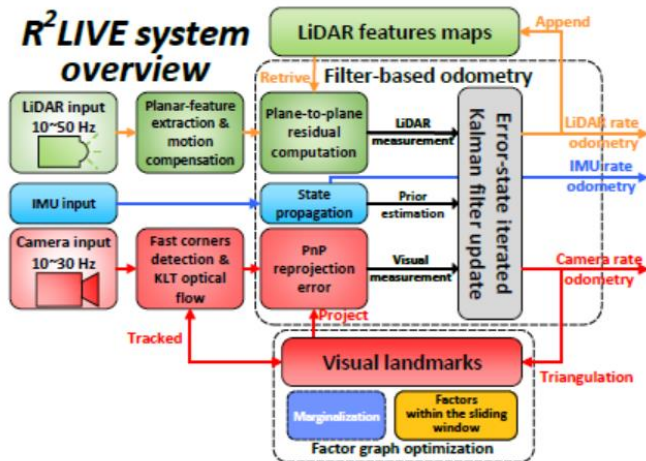


Fig. 2: The overview of our proposed method.

## 基础符号的定义

### IV. FILTER-BASED ODOMETRY

#### A. The boxplus “ $\boxplus$ ” and boxminus “ $\boxminus$ ” operator

In this paper, we make use of the “ $\boxplus$ ” and “ $\boxminus$ ” operations encapsulated on a manifold  $\mathcal{M}$  to simplify the notations and derivations. Let  $\mathcal{M}$  be the manifold on which the system state lies. Since a manifold is locally homeomorphic to  $\mathbb{R}^n$ , where  $n$  is the dimension of  $\mathcal{M}$ , we can use two operators, “ $\boxplus$ ” and “ $\boxminus$ ”, establishing a bijective map between the local neighborhood of  $\mathcal{M}$  and its tangent space  $\mathbb{R}^n$  [23]:

$$\boxplus: \mathcal{M} \times \mathbb{R}^n \rightarrow \mathcal{M}, \quad \boxminus: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^n \quad (1)$$

For the compound manifold  $\mathcal{M} = SO(3) \times \mathbb{R}^n$ , we have:

$$\begin{bmatrix} \mathbf{R} \\ \mathbf{a}_1 \end{bmatrix} \boxplus \begin{bmatrix} \mathbf{r} \\ \mathbf{a}_2 \end{bmatrix} \triangleq \begin{bmatrix} \mathbf{R} \cdot \text{Exp}(\mathbf{r}) \\ \mathbf{a}_1 + \mathbf{a}_2 \end{bmatrix}, \quad \begin{bmatrix} \mathbf{R}_1 \\ \mathbf{a}_1 \end{bmatrix} \boxminus \begin{bmatrix} \mathbf{R}_2 \\ \mathbf{a}_2 \end{bmatrix} \triangleq \begin{bmatrix} \text{Log}(\mathbf{R}_2^T \mathbf{R}_1) \\ \mathbf{a}_1 - \mathbf{a}_2 \end{bmatrix}$$

where  $\mathbf{r} \in \mathbb{R}^3$ ,  $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^n$ ,  $\text{Exp}(\cdot)$  and  $\text{Log}(\cdot)$  denote the Rodrigues' transformation between the rotation matrix and rotation vector<sup>2</sup>.

#### B 连续时间运动学模型

所有这三个传感器，激光雷达、照相机和 IMU 都被预先校准了。此外，我们假设 LiDAR 和 IMU 之间的外参已知，但在线对相机-IMU 外参进行估计。此外，激光雷达通常按顺序扫描点，而激光雷达帧中的点可以在不同的姿势下被测量。利用 imu 进行运动补偿，因此假设激光雷达帧中的点可以同时被测量。通过这些，我们系统的输入数据序列可以简化为图 3。

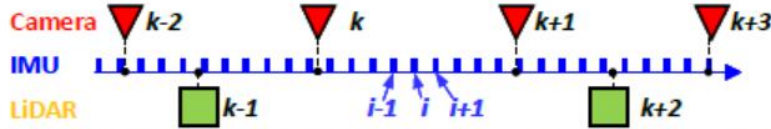


Fig. 3: Illustration of the input data sequences, where the frame rate of IMU, camera, and LiDAR is 200 Hz, 20 Hz and 10 Hz, respectively. The notation  $i$  denotes the index of IMU data while  $k$  denotes the index of LiDAR or camera measurements.

我们假设 imu、激光雷达和相机传感器之间的刚性连接在一起。并且已知 LIDAR 与 IMU 的外参（LIDA 系 w.r.t. IMU 系） ${}^I\mathbf{T}_L = ({}^I\mathbf{R}_L, {}^I\mathbf{p}_L)$  以及相机和 IMU 之间的外参(camera 系 w.r.t. IMU 系)  ${}^I\mathbf{T}_c = ({}^I\mathbf{R}_c, {}^I\mathbf{p}_c)$ 。为了方便起见，我们以 IMU 为载体坐标系，得到了以下连续的运动学模型：

$$\begin{aligned}
{}^G\dot{\mathbf{p}}_I &= {}^G\mathbf{V}_I \\
{}^G\dot{\mathbf{V}}_I &= {}^G\mathbf{R}_I(\mathbf{a}_m - \mathbf{b}_a - \mathbf{n}_a) + {}^G\mathbf{g} \\
{}^G\dot{\mathbf{R}}_I &= {}^G\mathbf{R}_I[\omega_m - \mathbf{b}_g - \mathbf{n}_g]_{\times} \\
\dot{\mathbf{b}}_g &= \mathbf{n}_{bg}, \dot{\mathbf{b}}_a = \mathbf{n}_{ba}
\end{aligned} \tag{2}$$

其中 ${}^G(\cdot)$ 表示在全局框架中表示的向量。第一个重力对齐的 imu 帧[22]),  $\mathbf{b}_a$  和  $\mathbf{b}_g$  是陀螺仪和加速度计的偏置, 它们被建模为由高斯噪声  $\mathbf{n}_{bg}$  和  $\mathbf{n}_{ba}$  的随机行走。

### C 离散模模型

我们以 imu 速率离散了连续模型 (2)。令  $\mathbf{x}_i$  作为第  $i$  次测量时的状态向量:

$\mathbf{x}_i = \begin{bmatrix} {}^G\mathbf{R}_{I_i}^T & {}^G\mathbf{p}_{I_i}^T & {}^G\mathbf{R}_{C_i}^T & {}^G\mathbf{p}_{C_i}^T & {}^G\mathbf{V}_i^T & \mathbf{b}_{gi}^T & \mathbf{b}_{ai}^T \end{bmatrix}$	$(3)$
---	-------

用零阶离散 (2) (即在一个采样时间周期  $\Delta t$  内的 imu 测量值是常数), 我们得到

$$\begin{aligned}
\mathbf{x}_{i+1} &= \mathbf{x}_i \boxplus (\Delta t \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{w}_i)) \tag{3} \\
\text{where} \\
\mathbf{u}_i &= [\omega_{m_i}^T \quad \mathbf{a}_{m_i}^T]^T, \quad \mathbf{w}_i = [\mathbf{n}_{gi}^T \quad \mathbf{n}_{ai}^T \quad \mathbf{n}_{bg_i}^T \quad \mathbf{n}_{ba_i}^T]^T \\
\mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{w}_i) &= \begin{bmatrix} \omega_{m_i} - \mathbf{b}_{gi} - \mathbf{n}_{gi} \\ {}^G\mathbf{V}_i \\ \mathbf{0}_{3 \times 1} \\ \mathbf{0}_{3 \times 1} \\ {}^G\mathbf{R}_{I_i}(\mathbf{a}_{m_i} - \mathbf{b}_{ai} - \mathbf{n}_{gi}) - {}^G\mathbf{g} \\ \mathbf{b}_{gi} \\ \mathbf{b}_{ai} \end{bmatrix}
\end{aligned}$$

### D Propagation 过程

在我们的工作中, 我们利用流形上迭代误差状态卡尔曼滤波器[24]来估计状态向量  $\mathbf{x}_i$ , 其中状态估计误差  $\delta \hat{\mathbf{x}}_i$  在状态估计  $\hat{\mathbf{x}}_i$  的切空间中:

$$\begin{aligned}
\delta \hat{\mathbf{x}}_i &\triangleq \mathbf{x}_i \boxminus \hat{\mathbf{x}}_i \\
&= \begin{bmatrix} {}^G\delta \hat{\mathbf{r}}_{I_i}^T & {}^G\delta \hat{\mathbf{p}}_{I_i}^T & {}^I\delta \hat{\mathbf{r}}_{C_i}^T & {}^I\delta \hat{\mathbf{p}}_{C_i}^T & {}^G\delta \hat{\mathbf{v}}_i^T & \delta \hat{\mathbf{b}}_{gi}^T & \delta \hat{\mathbf{b}}_{ai}^T \end{bmatrix}^T \\
&\sim \mathcal{N}(\mathbf{0}_{21 \times 1}, \Sigma_{\delta \hat{\mathbf{x}}_i}) \tag{4}
\end{aligned}$$

Note that  $\delta \hat{\mathbf{x}}_i \in \mathbb{R}^{21}$  is in minimum dimension (the system dimension 21) and is a random vector with covariance  $\Sigma_{\delta \hat{\mathbf{x}}_i}$ .  ${}^G \delta \hat{\mathbf{r}}_{I_i}^T$  and  ${}^I \delta \hat{\mathbf{r}}_{C_i}^T$  are:

$${}^G \delta \hat{\mathbf{r}}_{I_i} = \text{Log}({}^G \hat{\mathbf{R}}_{I_i}^T {}^G \mathbf{R}_{I_i}^T), \quad {}^I \delta \hat{\mathbf{r}}_{C_i} = \text{Log}({}^I \hat{\mathbf{R}}_{C_i}^T {}^I \mathbf{R}_{C_i})$$

Once receiving a new IMU measurement, the state estimate is propagated by setting the process noise in (3) to zero:

$$\hat{\mathbf{x}}_{i+1} = \hat{\mathbf{x}}_i \boxplus (\Delta t \cdot \mathbf{f}(\hat{\mathbf{x}}_i, \mathbf{u}_i, \mathbf{0})). \quad (5)$$

The associated estimation error is propagated in the linearized error space as follows (see [24] for more details):

$$\begin{aligned} \delta \hat{\mathbf{x}}_{i+1} &= \mathbf{x}_{i+1} \boxminus \hat{\mathbf{x}}_{i+1} \\ &= (\mathbf{x}_i \boxplus (\Delta t \cdot \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, \mathbf{w}_i))) \boxminus (\hat{\mathbf{x}}_i \boxplus (\Delta t \cdot \mathbf{f}(\hat{\mathbf{x}}_i, \mathbf{u}_i, \mathbf{0}))) \\ &\sim \mathcal{N}(\mathbf{0}_{21 \times 1}, \Sigma_{\delta \hat{\mathbf{x}}_{i+1}}) \end{aligned} \quad (6)$$

where:

$$\begin{aligned} \Sigma_{\delta \hat{\mathbf{x}}_{i+1}} &= \mathbf{F}_{\delta \hat{\mathbf{x}}} \Sigma_{\delta \hat{\mathbf{x}}_i} \mathbf{F}_{\delta \hat{\mathbf{x}}}^T + \mathbf{F}_{\mathbf{w}} \mathbf{Q} \mathbf{F}_{\mathbf{w}}^T \\ \mathbf{F}_{\delta \hat{\mathbf{x}}} &= \left. \frac{\partial (\delta \hat{\mathbf{x}}_{i+1})}{\partial \delta \hat{\mathbf{x}}_i} \right|_{\delta \hat{\mathbf{x}}_i = \mathbf{0}, \mathbf{w}_i = \mathbf{0}}, \quad \mathbf{F}_{\mathbf{w}} = \left. \frac{\partial (\delta \hat{\mathbf{x}}_{i+1})}{\partial \mathbf{w}_i} \right|_{\delta \hat{\mathbf{x}}_i = \mathbf{0}, \mathbf{w}_i = \mathbf{0}} \end{aligned} \quad (7)$$

with their exact values computed in Appendix. A.

## E 先验

令式 (5)和(7) 中的在 (k+1) 激光雷达/相机测量中停止 (见图 4), 和传播的状态估计和协方差分别为  $\hat{\mathbf{x}}_{k+1}$  和  $\Sigma_{\delta \hat{\mathbf{x}}_{k+1}}$ 。在融合 k+1 测量之前,它们对状态  $\mathbf{x}_{k+1}$  施加了一个先验:

$$\mathbf{x}_{k+1} \boxminus \hat{\mathbf{x}}_{k+1} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\delta \hat{\mathbf{x}}_{k+1}}). \quad (8)$$

## F 初始化迭代后的更新

式 (8) 中的先验分布将与激光雷达或照相机的测量结果相融合, 以产生  $\mathbf{x}_{k+1}$  的最大后验估计值(表示为  $\tilde{\mathbf{x}}_{k+1}$ )。MAP 估计  $\tilde{\mathbf{x}}_{k+1}$  被初始化为先验估计  $\hat{\mathbf{x}}_{k+1}$ , 并由于问题的非线性性质进行了迭代改进。在每次迭代中, 真实状态  $\mathbf{x}_{k+1}$  和当前估计值  $\tilde{\mathbf{x}}_{k+1}$  之间的误差  $\delta \tilde{\mathbf{x}}_{k+1}$  定义为

$$\delta \tilde{\mathbf{x}}_{k+1} \triangleq \mathbf{x}_{k+1} \boxminus \tilde{\mathbf{x}}_{k+1}, \quad (9)$$

将通过最小化在 (8) 和激光雷达/视觉测量中考虑先验的后验分布来解决。因此, 用 (8) 表示的  $\mathbf{x}_{k+1}$  来表示的先验分布应该转换为用  $\delta \tilde{\mathbf{x}}_{k+1}$  表示的等价先验



$$\begin{aligned}\mathbf{x}_{k+1} \boxminus \hat{\mathbf{x}}_{k+1} &= (\tilde{\mathbf{x}}_{k+1} \boxplus \delta\tilde{\mathbf{x}}_{k+1}) \boxminus \hat{\mathbf{x}}_{k+1} \\ &\approx \tilde{\mathbf{x}}_{k+1} \boxminus \hat{\mathbf{x}}_{k+1} + \mathcal{H}\delta\tilde{\mathbf{x}}_{k+1} \\ &\sim \mathcal{N}(\mathbf{0}, \Sigma_{\delta\tilde{\mathbf{x}}_{k+1}}),\end{aligned}\quad (10)$$

where

$$\mathcal{H} = \left. \frac{(\tilde{\mathbf{x}}_{k+1} \boxplus \delta\tilde{\mathbf{x}}_{k+1}) \boxminus \hat{\mathbf{x}}_{k+1}}{\partial\delta\tilde{\mathbf{x}}_{k+1}} \right|_{\delta\tilde{\mathbf{x}}_{k+1}=\mathbf{0}}$$

is computed in detail in Appendix. B, (10) essentially imposes a prior distribution to  $\delta\tilde{\mathbf{x}}_{k+1}$  as below:

$$\delta\tilde{\mathbf{x}}_{k+1} \sim \mathcal{N}(-\mathcal{H}^{-1}(\tilde{\mathbf{x}}_{k+1} \boxminus \hat{\mathbf{x}}_{k+1}), \mathcal{H}^{-1}\Sigma_{\delta\tilde{\mathbf{x}}_{k+1}}\mathcal{H}^{-T}) \quad (11)$$

## G LiDAR 测量

如果 (k+1) -th 测量是激光雷达帧，我们利用文献[5]从原始三维点提取平面特征点，并如 Section IV-B 节补偿帧内运动。表示运动补偿后的特征点集  $L_{k+1}$ ，我们计算每个特征点  ${}^L\mathbf{p}_j \in L_{k+1}$  的残差，其中 j 是特征点的索引，上标 L 表示该点在激光雷达参考系中表示。

令  $\mathbf{x}_{k+1}$  的当前估计  $\tilde{\mathbf{x}}_{k+1}$ ，我们可将  ${}^L\mathbf{p}_j$  转换到世界坐标系下。正如之前的 LOAM [5,9]中所做的那样，我们搜索地图中最近的平面特征点，并使用它们来拟合一个具有正常  $\mathbf{u}_j$  的平面和一个平面内点  $\mathbf{q}_j$  的平面，测量残差为：

$$\mathbf{r}_l(\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j) = \mathbf{u}_j^T ({}^G\mathbf{p}_j - \mathbf{q}_j) \quad (12)$$

Let  $\mathbf{n}_j$  be the measurement noise of the point  ${}^L\mathbf{p}_j$ , we can obtain the true point location  ${}^L\mathbf{p}_j^{\text{gt}}$  by compensating the noise from  ${}^L\mathbf{p}_j$ :

$${}^L\mathbf{p}_j = {}^L\mathbf{p}_j^{\text{gt}} + \mathbf{n}_j, \mathbf{n}_j \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{n}_j}). \quad (13)$$

This true point location together with the true state  $\mathbf{x}_{k+1}$  should lead to zero residual in (12), i.e.,

$$\mathbf{0} = \mathbf{r}_l(\mathbf{x}_{k+1}, {}^L\mathbf{p}_j^{\text{gt}}) = \mathbf{r}_l(\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j) + \mathbf{H}_j^l \delta\tilde{\mathbf{x}}_{k+1} + \alpha_j, \quad (14)$$

which constitutes a posteriori distribution for  $\delta\tilde{\mathbf{x}}_{k+1}$ . In (14),  $\mathbf{x}_{k+1}$  is parameterized by its error  $\delta\tilde{\mathbf{x}}_{k+1}$  defined in (9) and  $\alpha_j \sim \mathcal{N}(\mathbf{0}, \Sigma_{\alpha_j})$ :

$$\begin{aligned}\mathbf{H}_j^l &= \left. \frac{\partial \mathbf{r}_l(\tilde{\mathbf{x}}_{k+1} \boxplus \delta\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j)}{\partial \delta\tilde{\mathbf{x}}_{k+1}} \right|_{\delta\tilde{\mathbf{x}}_{k+1}=\mathbf{0}} \\ \Sigma_{\alpha_j} &= \mathbf{F}_{\mathbf{p}_j} \Sigma_{\mathbf{n}_j} \mathbf{F}_{\mathbf{p}_j}^T \\ \mathbf{F}_{\mathbf{p}_j} &= \left( \frac{\partial \mathbf{r}_l(\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j)}{\partial {}^L\mathbf{p}_j} \right) = {}^G\mathbf{R}_{I_{k+1}} {}^I\mathbf{R}_L\end{aligned}\quad (15)$$

The detailed computation of  $\mathbf{H}_j^l$  can be found in Appendix. C.

## H 视觉测量

如果 (k+1) -th 是相机图像，我们从未失真图像中提取快角特征点  $C_{k+1}$ ，并使用 KLT 光流跟踪当前滑动窗口中关键帧在  $C_{k+1}$  中看到的特征点（第五部分）。如果  $C_{k+1}$  中的一个特征点丢失或之前尚未被跟踪，我们将用最优估计的相机姿态对

三维空间（视觉地标）中的新特征点进行三角形分析。

在  $(k+1)$ -th 中，视觉地标与其跟踪的特征点之间的重投影误差用于更新当前状态估计  $\tilde{\mathbf{x}}_{k+1}$ 。对于提取的角点  ${}^C\mathbf{p}_s = [u_s \ v_s]^T \in C_{k+1}$ ，其中  $s$  是角点的索引，它在三维空间中的对应地标表示为  ${}^G\mathbf{p}_s$ ，那么  ${}^C\mathbf{p}_s$  的测量残差为

$$\begin{aligned} {}^C\mathbf{p}_s &= \left( {}^G\tilde{\mathbf{R}}_{I_{k+1}} {}^I\tilde{\mathbf{R}}_{C_{k+1}} \right)^T {}^G\mathbf{p}_s \\ &\quad - \left( {}^I\tilde{\mathbf{R}}_{C_{k+1}} \right)^T {}^G\tilde{\mathbf{p}}_{I_{k+1}} - {}^I\tilde{\mathbf{p}}_{C_{k+1}} \quad (16) \\ \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_s, {}^G\mathbf{p}_s) &= {}^C\mathbf{p}_s - \pi({}^C\mathbf{p}_s) \end{aligned}$$

where  $\pi(\cdot)$  is the pin-hole projection model.

Now considering the measurement noise, we have:

$${}^G\mathbf{p}_s = {}^G\mathbf{p}_s^{\text{gt}} + \mathbf{n}_{\mathbf{p}_s}, \quad \mathbf{n}_{\mathbf{p}_s} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{n}_{\mathbf{p}_s}}) \quad (17)$$

$${}^C\mathbf{p}_s = {}^C\mathbf{p}_s^{\text{gt}} + \mathbf{n}_{\mathbf{p}_s}, \quad \mathbf{n}_{\mathbf{p}_s} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{n}_{\mathbf{p}_s}}) \quad (18)$$

where  ${}^G\mathbf{p}_s^{\text{gt}}$  and  ${}^C\mathbf{p}_s^{\text{gt}}$  are the true value of  ${}^G\mathbf{p}_s$  and  ${}^C\mathbf{p}_s$ , respectively. With these, we obtain the first order Taylor expansion of the true zero residual  $\mathbf{r}_c(\mathbf{x}_{k+1}, {}^C\mathbf{p}_s^{\text{gt}})$  as:

$$\begin{aligned} \mathbf{0} &= \mathbf{r}_c(\mathbf{x}_{k+1}, {}^C\mathbf{p}_s^{\text{gt}}, {}^G\mathbf{p}_s^{\text{gt}}) \\ &\approx \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_s, {}^G\mathbf{p}_s) + \mathbf{H}_s^c \delta \tilde{\mathbf{x}}_{k+1} + \beta_s, \end{aligned} \quad (19)$$

which constitutes another posteriori distribution for  $\delta \tilde{\mathbf{x}}_{k+1}$ . In (19),  $\beta_s \sim \mathcal{N}(\mathbf{0}, \Sigma_{\beta_s})$  and:

$$\begin{aligned} \mathbf{H}_s^c &= \frac{\partial \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1} \boxplus \delta \tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_s, {}^G\mathbf{p}_s)}{\partial \delta \tilde{\mathbf{x}}_{k+1}} \Big|_{\delta \tilde{\mathbf{x}}_{k+1}=\mathbf{0}} \\ \Sigma_{\beta_s} &= \Sigma_{\mathbf{n}_{\mathbf{p}_s}} + \mathbf{F}_{\mathbf{p}_s} \Sigma_{\mathbf{p}_s} \mathbf{F}_{\mathbf{p}_s}^T \\ \mathbf{F}_{\mathbf{p}_s} &= \frac{\partial \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_s, {}^G\mathbf{p}_s)}{\partial {}^G\mathbf{p}_s} \end{aligned} \quad (20)$$

The detailed computation of  $\mathbf{H}_s^c$  and  $\mathbf{F}_{\mathbf{p}_s}$  is given in appendix.  $\square$

## I 误差状态迭代卡尔曼滤波器的更新

结合先验分布 (11)、激光雷达测量(14)引起的后验分布和视觉测量(19)引起的后验分布，得到了  $\delta \tilde{\mathbf{x}}_{k+1}$  的最大后验（MAP）估计：

$$\begin{aligned} \min_{\delta \tilde{\mathbf{x}}_{k+1}} & \left( \left\| \tilde{\mathbf{x}}_{k+1} \boxminus \hat{\mathbf{x}}_{k+1} + \mathcal{H} \delta \tilde{\mathbf{x}}_{k+1} \right\|_{\Sigma_{\delta \tilde{\mathbf{x}}_{k+1}}^{-1}}^2 \right. \\ & + \sum_{j=1}^{m_l} \left\| \mathbf{r}_l(\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_j) + \mathbf{H}_j^l \delta \tilde{\mathbf{x}}_{k+1} \right\|_{\Sigma_{\alpha_j}^{-1}}^2 \\ & \left. + \sum_{s=1}^{m_c} \left\| \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_s, {}^G\mathbf{p}_s) + \mathbf{H}_s^c \delta \tilde{\mathbf{x}}_{k+1} \right\|_{\Sigma_{\beta_s}^{-1}}^2 \right) \end{aligned}$$



$$\begin{aligned}
\mathbf{H}^T &= [\mathbf{H}_1^l, \dots, \mathbf{H}_{m_l}^l, \mathbf{H}_1^{cT}, \dots, \mathbf{H}_{m_c}^{cT}]^T \\
\mathbf{R} &= \text{diag}(\Sigma_{\alpha_1}, \dots, \Sigma_{\alpha_{m_l}}, \Sigma_{\beta_1}, \dots, \Sigma_{\beta_{m_c}}) \\
\tilde{\mathbf{z}}_{k+1}^T &= \left[ \mathbf{r}_l(\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_1), \dots, \mathbf{r}_l(\tilde{\mathbf{x}}_{k+1}, {}^L\mathbf{p}_{m_l}), \right. \\
&\quad \left. \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_1, {}^G\mathbf{p}_1), \dots, \mathbf{r}_c(\tilde{\mathbf{x}}_{k+1}, {}^C\mathbf{p}_{m_c}, {}^G\mathbf{p}_{m_c}) \right] \\
\mathbf{P} &= (\mathcal{H})^{-1} \Sigma_{\delta \tilde{\mathbf{x}}_{k+1}} (\mathcal{H})^{-T}
\end{aligned} \tag{21}$$

Following [5], we have the Kalman gain computed as:

$$\mathbf{K} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{P}^{-1})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \tag{22}$$

Then we can update the state estimate as:

$$\hat{\mathbf{x}}_{k+1} = \tilde{\mathbf{x}}_{k+1} \boxplus (-\mathbf{K} \tilde{\mathbf{z}}_{k+1} - (\mathbf{I} - \mathbf{K} \mathbf{H}) (\mathcal{H})^{-1} (\tilde{\mathbf{x}}_{k+1} \boxminus \hat{\mathbf{x}}_{k+1}))$$

The above process (Section IV-G to Section IV-I) is iterated until convergence (i.e., the update is smaller than a given threshold). The converged state estimate is then used to (1) project points in the the new LiDAR frame to the world frame and append them to the existing point cloud map; (2) triangulate new visual landmarks of the current frame if it is a keyframe; (3) serve as the starting point of the propagation in Section IV-D for the next cycle:

$$\hat{\mathbf{x}}_{k+1} = \tilde{\mathbf{x}}_{k+1}, \quad \hat{\Sigma}_{\delta \mathbf{x}_{k+1}} = (\mathbf{I} - \mathbf{K} \mathbf{H}) \tilde{\Sigma}_{\delta \mathbf{x}_{k+1}}$$

## 因子图优化

如第 iv-i 节所述，对新添加的关键帧中未跟踪的视觉标记进行三角化，以创建新的视觉地标。由于关键帧位姿估计误差，这种三角测量的精度通常较低。为了进一步提高视觉地标、关键帧姿态的质量，同时校准相机和 lidar-imu 系统之间的时间偏移，我们利用因子图优化来优化图像关键帧滑动窗口内的相机姿态和视觉地标。

我们的因子图优化类似于 VINS-Mono[22]，但进一步包含了由于激光雷达测量而造成的姿态约束，如图所示 5. 还包括了 imu 预积分的约束条件，以将激光雷达系数与相机因子连接起来。为了保持后端优化的轻量级，激光雷达在姿态图中的姿态是固定的，而激光雷达的原始点测量则不参与姿态图的优化。

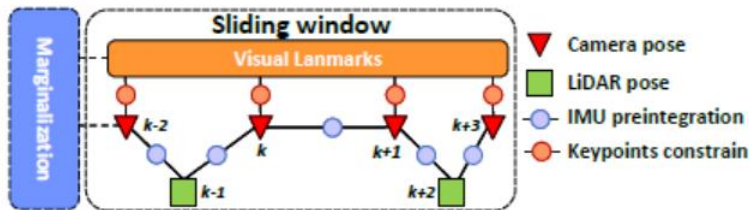


Fig. 5: Our factor graph optimization.

实验结果分析部分：



Fig. 7: We evaluate the robustness of our algorithm under scenarios with the aggressive motion (sequence in a~d) and sensor failure by intentionally blocking the camera (e) and LiDAR sensor (f).

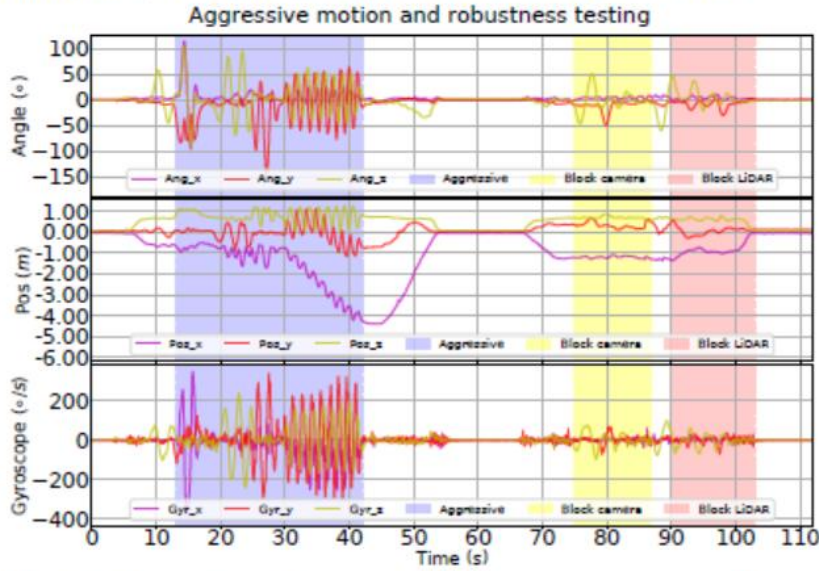


Fig. 8: Our estimated pose and the raw gyroscope reading of Experiment-1. The shaded area in blue, yellow and red represent different phases of aggressive motion, camera-failure and LiDAR-failure, respectively.



Fig. 9: We evaluate our algorithm in a Hong Kong MTR station consisting of cluttered lobby and very long narrow tunnels, as shown in (a). The tunnel is up to 190 meters long and is filled with moving pedestrians, making it extremely challenging for both LiDAR-based and camera-based SLAM methods. (b): the map built by our system is well aligned with the street map of MTR station. (c) Trajectory comparison among our system “R2LIVE”, the LiDAR-inertial system “Fast-LIO”, and visual-inertial system “VINS-Mono” and (our). The starting point is marked with  $\blacklozenge$  while the ending point of each trajectory is marked with  $\star$ . “VINS-Mono” stopped at middle way due to the failure of feature tracking.