

LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping

摘要

我们提出了一个通过 SAM 实现紧耦合激光雷达视觉惯性里程计的框架：LVI-SAM，该框架可实现高精度和鲁棒性的实时状态估计和地图构建。LVI-SAM 建立在因子图之上，由两个子系统组成：视觉惯性系统 (VIS) 和激光雷达惯性系统 (LIS)。这两个子系统以紧密耦合的方式设计，其中 VIS 利用 LIS 估计来促进初始化。通过使用激光雷达测量提取视觉特征的深度信息，提高了 VIS 的准确性。反过来，LIS 利用 VIS 估计进行初始猜测以支持扫描匹配。闭环首先由 VIS 识别，然后由 LIS 进一步细化。LVI-SAM 还可以在两个子系统之一出现故障时运行，从而提高其在无纹理和无特征环境中的稳健性。LVI-SAM 在从多个平台收集的各种规模和环境的数据集上进行了广泛评估。

1、 介绍

同步定位和地图构建 (SLAM) 是许多移动机器人导航任务所需的基础功能。在过去的二十年中，使用 SLAM 在具有挑战性的环境中使用单个感知传感器（例如激光雷达或相机）进行实时状态估计和映射取得了巨大成功。基于激光雷达的方法可以远距离捕捉环境的

细节。然而，当在无结构环境（例如长走廊或开阔地）中操作时，此类方法通常会失败。虽然基于视觉的方法特别适用于位置识别并且在纹理丰富的环境中表现良好，但它们的性能对光照变化、快速运动和初始化很敏感。因此，基于激光雷达和基于视觉的方法通常都与惯性测量单元 (IMU) 结合使用，以提高各自的稳健性和准确性。激光雷达惯性系统可以帮助纠正点云失真并解决短时间内缺乏特征的问题。可以通过 IMU 测量恢复度量尺度和姿态，以辅助视觉惯性系统。为了进一步提高系统性能，激光雷达、相机和 IMU 的融合越来越受到关注。

我们的工作与视觉惯性里程计 (VIO)、激光雷达惯性里程计 (LIO) 和激光雷达视觉惯性里程计 (LVIO) 关系最密切。我们注意到我们在本文中不考虑非惯性系统，尽管我们知道有成功的非惯性激光雷达视觉系统，例如 [1]、[2]。视觉惯性里程计 (VIO) 可以分为两大类：**基于滤波的方法和基于优化的方法**。基于滤波器的方法通常使用扩展卡尔曼滤波器 (EKF) 来使用来自相机和 IMU 的测量值来传播 (propagate) 系统状态。**基于优化的方法维护滑动窗口估计器并最小化视觉重投影误差以及 IMU 测量误差**。在我们的工作中，我们将我们的考虑限制在**单目相机**上。在最流行的公开 VIO 中，MSCKF [3]、ROVIO [4] 和 Open-VINS [5] 是基于过滤器的，而 OKVIS [6] 和 VINS-Mono [7] 是基于优化的。尽管 OKVIS 使用立体相机显示出卓越的性能，但它并未针对单目相机进行优化。VINS-Mono 在滑动窗口设置中执行非线性优化，并使用单目相机实

现最先进的精度 [8]。

基于他们的设计方案，激光雷达惯性里程计也可以分为两大类：**松耦合方法和紧耦合方法**。 LOAM [9] 和 LeGO-LOAM [10] 是松耦合系统，因为在优化步骤中不使用 IMU 测量。 紧耦合系统通常提供更高的准确性和鲁棒性，目前是正在进行的研究的主要焦点 [11]。 在公开可用的紧耦合系统中，LIO-mapping [12] 采用了 [7] 的优化流程并最小化了 IMU 和激光雷达测量的残差。由于 LIO 建图旨在优化所有测量，因此无法实现实时性能。 LIO-SAM [13] 通过引入激光雷达关键帧的滑动窗口来限制计算复杂性，利用因子图进行联合 IMU 和激光雷达约束优化。 专门为地面车辆设计的 LINS [14] 使用误差状态卡尔曼滤波器递归地校正机器人的状态。

最近，激光雷达-视觉-惯性系统因其在传感器**退化**任务中的鲁棒性而受到越来越多的关注 [15]。 [16] 提出了一个紧耦合的 LVIO 系统，它具有顺序处理流程，解决了从粗到细的状态估计问题。 粗略估计从 IMU 预测开始，然后由 VIO 和 LIO 进一步细化。 [16] 目前在 KITTI 基准 [17] 上达到了最先进的精度。 基于 MSCKF 的框架，[18] 具有在线时空多传感器校准功能。 [16] 和 [18] 的实现不公开。 我们的工作不同于上述工作，因为我们利用因子图进行全局优化，它可以通过闭环检测定期消除机器人产生的漂移。

在本文中，我们提出了一个通过 SAM 的紧耦合激光雷达-视觉-惯性里程计框架：LVI-SAM，用于实时状态估计和建图。 LVI-SAM

建立在因子图之上，由两个子系统组成，视觉惯性系统 (VIS) 和激光雷达惯性系统 (LIS)。当其中一个子系统检测到故障时，这两个子系统可以独立运行，或者当检测到足够多的特征时联合运行。VIS 执行视觉特征跟踪，并可选择使用激光雷达框架提取特征深度。通过优化视觉重投影和 IMU 测量的误差获得的视觉里程计用作激光雷达扫描匹配的初始猜测，并将约束引入因子图中。在使用 IMU 测量值消除点云偏移后，LIS 提取激光雷达边缘和平面特征，并将它们与滑动窗口中维护的特征图进行匹配。LIS 中估计的系统状态可以发送到 VIS 以方便其初始化。对于闭环，候选匹配首先由 VIS 识别并由 LIS 进一步优化。视觉里程计、激光雷达里程计、IMU 预积分和闭环的约束在因子图中联合优化。最后，优化后的 IMU bias 用于传播 (propagate) IMU 测量值，以 IMU 速率进行姿态估计。我们工作的主要贡献可以概括如下：

- 构建在因子图之上的紧耦合 LVIO 框架，可通过位置识别实现多传感器融合和全局优化。
- 我们的框架通过故障检测绕过故障子系统，使其对传感器退化具有鲁棒性。
- 我们的框架使用跨不同规模、平台和环境收集的数据进行了广泛验证。

从系统的角度来看，我们的工作是新颖的，代表了 VIO 和 LIO 中最新技术的独特集成，以实现 LVIO 系统，提供更高的鲁棒性和准

确性。 我们希望我们的系统可以作为一个坚实的基线，其他人可以轻松地为以此为基础来推进激光雷达-视觉-惯性里程计的最先进技术。

2、 通过 SAM 的激光雷达视觉惯性里程计

A、 系统概述

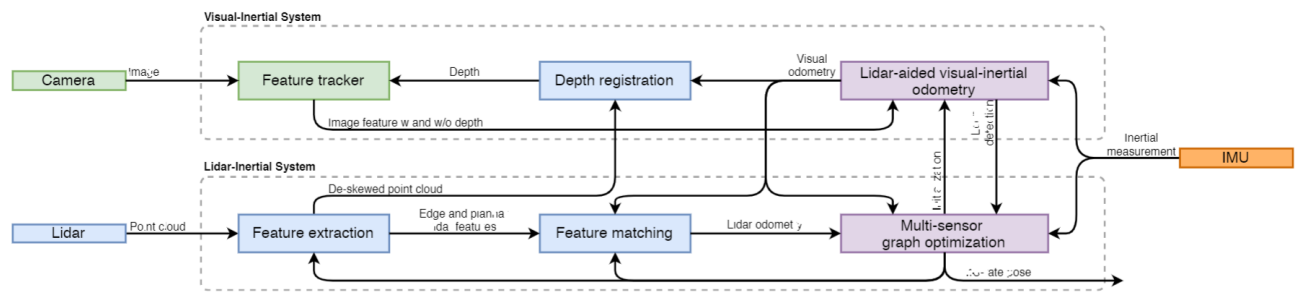


图 1: LVI-SAM 的系统结构。该系统接收来自 3D 激光雷达、相机和 IMU 的输入，可分为两个子系统：视觉惯性系统 (VIS) 和激光雷达惯性系统 (LIS)。VIS 和 LIS 可以独立运行，同时使用彼此的信息来提高系统的准确性和稳健性。系统以 IMU 速率输出姿态估计。

所提出的激光雷达-视觉-惯性系统的概述，该系统接收来自 3D 激光雷达、单目相机和 IMU 的输入，如图 1 所示。我们的框架由两个关键子系统组成：视觉惯性系统 (VIS) 和激光雷达惯性系统 (LIS)。VIS 处理图像和 IMU 测量，激光雷达测量是可选的。视觉里程计是通过最小化视觉和 IMU 测量的联合残差获得的。LIS 提取激光雷达特征并通过将提取的特征与特征图匹配来执行激光雷达里程计。特征图以滑动窗口的方式维护，以实现实时性能。最后，状态估计

问题可以表述为最大后验 (MAP) 问题, 通过使用 iSAM2 [19] 因子图联合优化 IMU 预积分约束、视觉里程计约束、激光雷达里程计约束和闭环约束来解决。 请注意, LIS 中采用的多传感器图形优化旨在减少数据交换并提高系统效率。

B、 视觉惯性系统

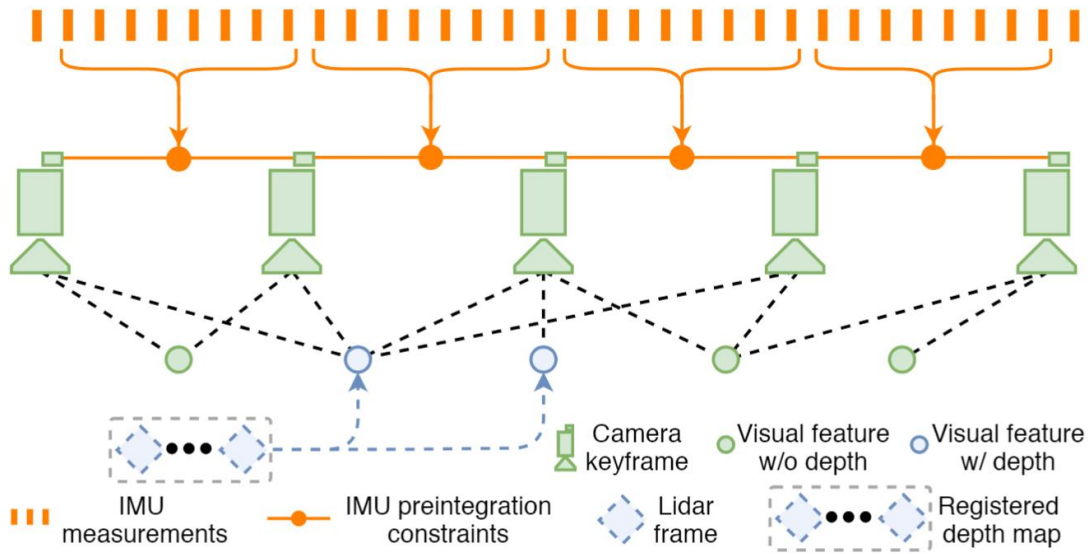


图 2: 我们的视觉惯性系统框架。该系统优化了 IMU 预积分、没有深度的视觉测量和有深度的视觉测量的残差。

我们为我们的 VIS 调整了 [7] 中的处理流程, 如图 2 所示。使用角检测器 [20] 检测视觉特征, 并通过 Kanade-Lucas-Tomasi 算法 [21] 进行跟踪。在 VIS 初始化后, 我们使用视觉里程计注册激光雷达帧, 并获得用于特征深度估计的稀疏深度图像。系统在滑动窗口设置中执行捆绑调整, 其中系统状态 $x \in X$ 可以写为:

$$x = [R, p, v, b]$$

$R \in SO(3)$ 是旋转矩阵, $p \in \mathbb{R}^3$ 是位置向量, v 是速度, $b = [b_a, b_w]$ 是 IMU bias。 b_a 和 b_w 分别是加速度和角速度的 bias 矢量。 平移变换 $T \in SE(3)$ 从传感器坐标系 \mathbf{B} 到世界坐标系 \mathbf{W} 表示为: $T = [R|p]$ 。 在接下来的部分中, 我们将给出改进 **VIS 初始化和特征深度估计** 的详细过程。 由于篇幅限制, 我们建议读者参考 [7] 了解更多细节, 例如残差的实现。

(1) 初始化

由于在初始化时解决了高度非线性问题, 基于优化的 VIO 通常会出现分歧。 初始化的质量在很大程度上取决于两个因素: 初始传感器运动和 IMU 参数的准确性。 在实践中, 我们发现当传感器以小速度或恒定速度行进时, [7] 经常无法初始化。 这是因为当加速度激励不够大时, 无法观察到**现实中的尺度**。 IMU 参数包括缓慢变化的 bias 和白噪声, 它们会影响原始加速度和角速度测量。 在初始化时很好地猜测这些参数有助于优化更快地收敛。

为了**提高 VIS 初始化的稳健性**, 我们利用了来自 LIS 的估计系统状态 x 和 IMU 偏差 b 。 因为深度可以从激光雷达直接观察到, 所以我们首先初始化 LIS 并获得 x 和 b 。 然后我们**根据图像时间戳将它们插值并关联到每个图像关键帧**。 请注意, 假定 IMU bias 在两个图像关键帧之间保持不变。 最后, 来自 LIS 的估计 x 和 b 被用作 VIS 初始化的初始猜测, 这显着提高了初始化速度和鲁棒性。 在补充视频 1 中可以找到使用和不使用 LIS 的 VIS 初始化的比较。

(2) 特征深度关联

在 VIS 初始化后,我们使用估计的视觉里程计将激光雷达帧注册到相机帧。由于 3D 激光雷达通常会产生稀疏扫描,因此我们堆叠多个激光雷达帧以获得密集的深度图。为了将特征与深度值相关联,我们首先将视觉特征和激光雷达深度点投影到以相机为中心的单位球体上。然后使用深度点的极坐标对深度点进行下采样并存储,以获得球体上的恒定密度。我们通过使用视觉特征的极坐标搜索二维 K-D 树,为视觉特征找到球体上最近的三个深度点。最后,特征深度是视觉特征与相机中心 O_c 形成的直线的长度,它与笛卡尔空间中三个深度点形成的平面相交。这个过程的可视化可以在图 3(a) 中找到,其中特征深度是虚直线的长度。

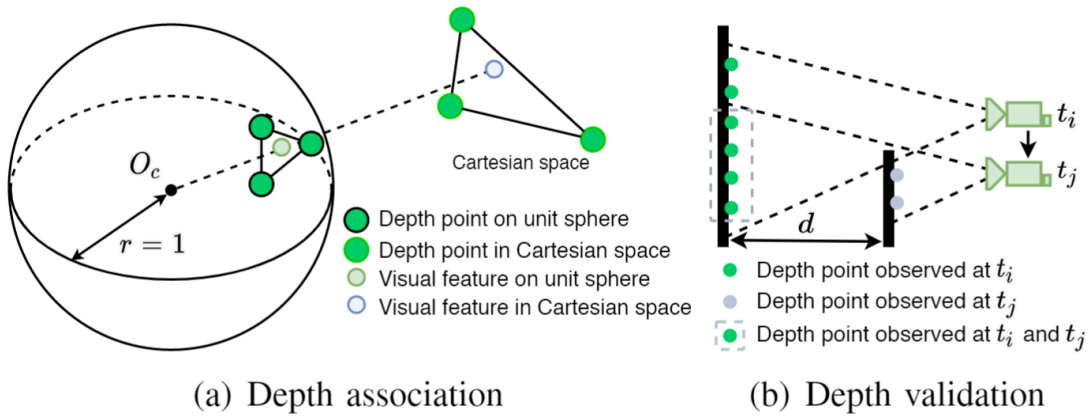


图 3: 视觉特征深度关联

我们通过检查三个最近的深度点之间的距离进一步验证相关的特征深度。这是因为堆叠来自不同时间戳的激光雷达帧可能会导致来自不同对象的深度模糊。图 3(b) 显示了这种情况的说明。在时

间 t_i 观察到的深度点用绿色表示。相机移动到 t_j 的新位置并观察新的灰色的深度点。然而，由于激光雷达帧堆叠， t_i 处的深度点（由灰色虚线圈出）在 t_j 处仍可观察到。使用来自不同对象的深度点关联特征深度会导致估计不准确。与[16]类似，我们通过检查特征的深度点之间的最大距离来拒绝此类估计。如果最大距离大于 2m，则该特征没有关联的深度。

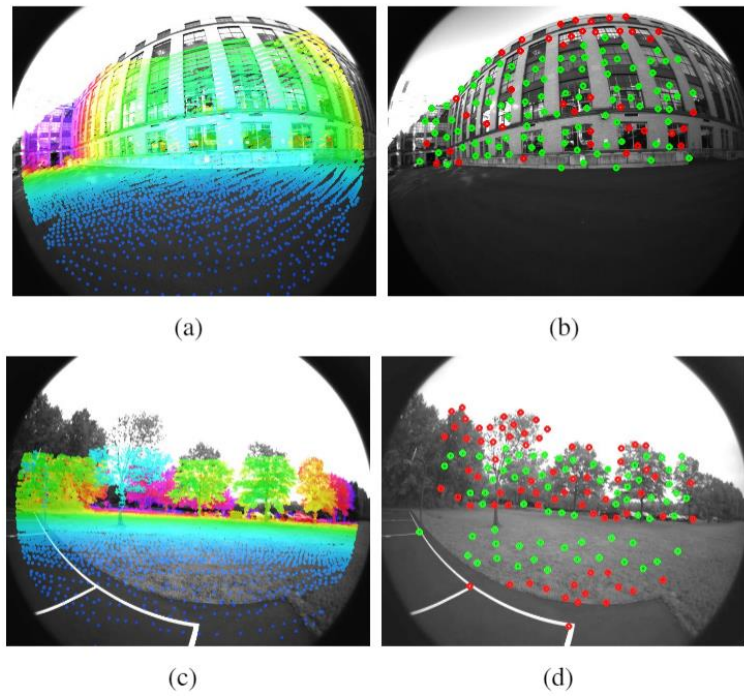


图 4：注册的深度图和视觉特征。在 (a) 和 (c) 中，深度图的颜色变化表示深度变化。在 (b) 和 (d) 中，绿点是与深度成功关联的视觉特征。深度关联过程失败的特征是红色的。

图 4 显示了注册深度图和视觉特征的演示。在图 4 (a) 和 (c) 中，使用视觉里程计注册的深度点被投影到相机图像上。在图 4 (b) 和 (d) 中，与深度成功关联的视觉特征为绿色。请注意，虽然深度

图覆盖了图 4(a) 中的大部分图像，但由于验证检查失败，4(b) 中位于窗口角落的许多特征缺乏深度关联。

(3) 故障检测

由于剧烈运动、光照变化和无纹理环境，VIS 会出现故障。当机器人进行激进运动或进入无纹理环境时，跟踪特征的数量会大大减少。特征不足可能导致优化发散。我们还注意到，**当 VIS 失败时，会估计出较大的 IMU bias**。因此，当跟踪的特征数量低于阈值时，我们报告 VIS 失败，或者当估计的 IMU bias 超过阈值时。我们的系统需要主动故障检测，这样它的故障就不会破坏 LIS 的功能。一旦检测到故障，VIS 将重新初始化并通知 LIS。

(4) 闭环检测

我们利用 DBoW2 [22] 进行闭环检测。对于每个新的图像关键帧，我们提取 BRIEF 描述符 [23] 并将它们与之前提取的描述符相匹配。DBoW2 返回的闭环候选图像时间戳被发送到 LIS 以进行进一步验证。

C、 激光雷达惯性系统

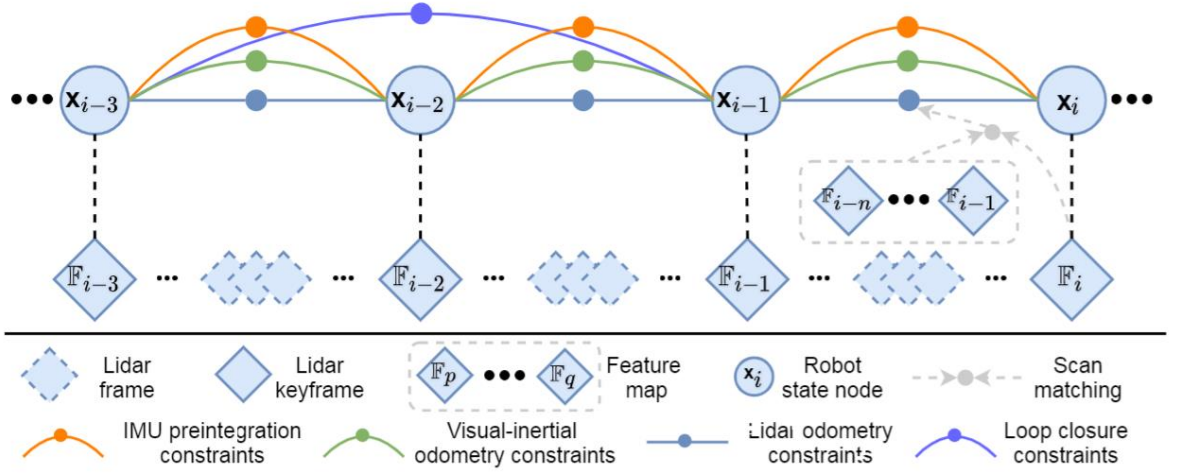


图 5：我们的激光雷达惯性系统的框架。 系统维护一个具有四种类型约束的因子图

如图 5 所示，所提出的激光雷达惯性系统改编自 [13]，维护全局姿态优化的因子图。 IMU 预积分约束、视觉里程计约束、激光雷达里程计约束和闭环约束四种类型的约束被添加到图中并联合优化。激光雷达里程计约束源自扫描匹配，我们将当前激光雷达关键帧与全局特征图进行匹配。闭环约束的候选者首先由 VIS 提供，然后通过扫描匹配进一步优化。我们为特征图维护一个激光雷达关键帧的滑动窗口，这保证了有限的计算复杂性。当机器人姿势的变化超过阈值时，将选择一个新的激光雷达关键帧。位于关键帧对之间的间歇性激光雷达帧被丢弃。选择新的激光雷达关键帧后，新的机器人状态 x 作为节点添加到因子图中。通过这种方式添加关键帧，不仅可以在内存消耗和地图密度之间取得平衡，还有助于保持一个相对稀疏

的因子图以进行实时优化。由于篇幅限制，我们建议读者参考 [13] 了解实施细节。在以下部分中，我们将重点关注提高系统稳健性的新过程。

(1) 初始猜测

我们发现初始猜测在扫描匹配的成功中起着至关重要的作用，尤其是当传感器进行剧烈运动时。LIS 初始化前后初始猜测的来源不同。

在 LIS 初始化之前，我们假设机器人从零速度的静止位置开始。然后我们整合原始 IMU 测量，假设 bias 和噪声为零值。两个激光雷达关键帧之间的综合平移和旋转变化产生扫描匹配的初始猜测。我们发现，当初始线速度小于 10 m/s，角速度小于 $180^\circ/\text{s}$ 时，这种方法可以成功地初始化系统。LIS 初始化后，我们在因子图中估计 IMU bias、机器人位姿和速度。然后将它们发送到 VIS 以帮助其初始化。

在 LIS 初始化之后，我们可以从两个来源获得初始猜测：具有校正偏差的集成 IMU 测量和 VIS。当它可用时，我们使用视觉惯性里程计作为初始猜测。如果 VIS 报告失败，我们将切换到 IMU 测量进行初始猜测。这些过程提高了纹理丰富和无纹理环境中的初始猜测准确性和鲁棒性。

(2) 故障检测

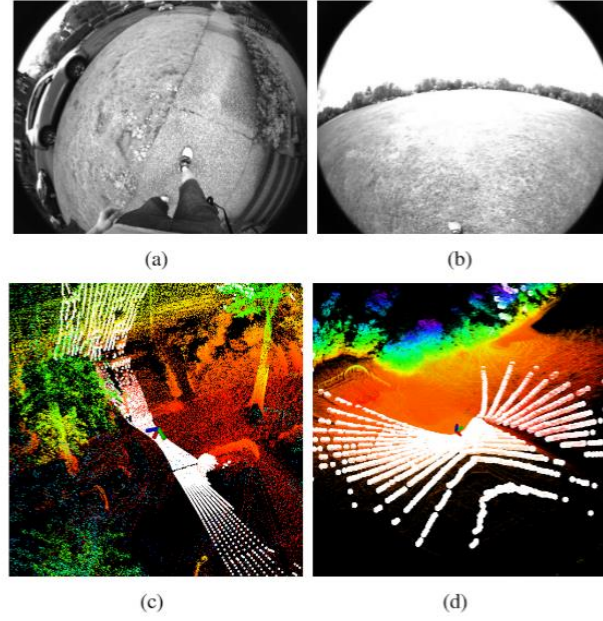


图 6: 扫描匹配受限的退化建图场景。在 (a) 和 (c) 中, 激光雷达面向地面放置。在 (b) 和 (d) 中, 激光雷达处于开放、无结构的环境中。白点表示激光雷达扫描的内容。颜色变化表示海拔变化。

尽管激光雷达可以远距离捕捉环境的细节, 但它仍然会遇到扫描匹配受限的降级场景。这种情况如图 6 所示。我们采用 [24] 中的方法进行 LIS 故障检测。扫描匹配中的非线性优化问题可以表述为迭代求解线性问题:

$$\min_T \|AT - b\|^2$$

其中 A 和 b 是从 T 处的线性化获得的。当 $A^T A$ 的最小特征值小于第一次优化迭代的阈值时, LIS 报告失败。发生故障时, 不会

将激光雷达里程计约束添加到因子图中。 我们建议读者参考 [24] 以了解这些假设所基于的详细分析。

3、 试验

4、 结论

我们提出了 LVI-SAM，这是一个通过 SAM 的紧耦合激光雷达-视觉-惯性里程计框架，用于在复杂环境中执行实时状态估计和建图。拟议的框架由两个子系统组成：视觉惯性系统和激光雷达惯性系统。这两个子系统旨在以紧耦合的方式进行交互，以提高系统的稳健性和准确性。通过对各种规模、平台和环境的数据集进行评估，我们的系统显示出与现有公开可用方法相当或更好的准确性。我们希望我们的系统将作为一个坚实的基线，其他人可以轻松地以此为基础来推进激光雷达-视觉-惯性里程计的最新技术水平。