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

> ## LVI-SAM：通过平滑和映射紧密耦合的激光雷达-视觉-惯性里程计

Abstract— We propose a framework for tightly-coupled lidar- visual-inertial odometry via smoothing and mapping, LVI-SAM, that achieves real-time state estimation and map-building with high accuracy and robustness. LVI-SAM is built atop a factor graph and is composed of two sub-systems: a visual-inertial system (VIS) and a lidar-inertial system (LIS). The two sub- systems are designed in a tightly-coupled manner, in which the VIS leverages LIS estimation to facilitate initialization. The accuracy of the VIS is improved by extracting depth information for visual features using lidar measurements. In turn, the LIS utilizes VIS estimation for initial guesses to support scan-matching. Loop closures are first identified by the VIS and further refined by the LIS. LVI-SAM can also function when one of the two sub-systems fails, which increases its robustness in both texture-less and feature-less environments. LVI-SAM is extensively evaluated on datasets gathered from several platforms over a variety of scales and environments. Our implementation is available at https://git.io/lvi-sam**.**

> 摘要——我们提出了一个通过平滑和映射的紧密耦合的激光雷达-视觉-惯性测距框架，**LVI-SAM**，以高精度和鲁棒性实现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，可以提高 VIS 的精度。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 也可以在两个子系统之一出现故障时发挥作用，从而提高其在无纹理和无特征环境中的稳健性。 LVI-SAM 对从多个平台在各种规模和环境中收集的数据集进行了广泛评估。我们的实现可在 https://git.io/lvi-sam 获得**

1. INTRODUCTION 介绍

Simultaneous localization and mapping (SLAM) is a foun- dational capability required for many mobile robot navigation tasks. During the last two decades, great success has been achieved using SLAM for real-time state estimation and mapping in challenging settings with a single perceptual sensor, such as a lidar or camera. Lidar-based methods can capture the fine details of the environment at long range. However, such methods typically fail when operating in structure-less environments, such as a long corridor or a flat open field. Though vision-based methods are especially suitable for place recognition and perform well in texture-rich environments, their performance is sensitive to illumi- nation change, rapid motion, and initialization. Therefore, lidar-based and vision-based methods are each often coupled with an inertial measurement unit (IMU) to increase their re- spective robustness and accuracy. A lidar-inertial system can help correct point cloud distortion and account for a lack of features over short periods of time. Metric scale and attitudes can be recovered by IMU measurements to assist a visual- inertial system. To further improve system performance, the fusion of lidar, camera, and IMU measurements is attracting increased attention.

同时定位和建图 (SLAM) 是许多移动机器人导航任务所需的基础能力。在过去的二十年中，使用 SLAM 在具有挑战性的设置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精细细节。然而，当在无结构的环境中操作时，这些方法通常会失败，例如长走廊或平坦的开阔场地。尽管基于视觉的方法特别适用于地点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) 相结合，以提高它们各自的鲁棒性和准确性。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。公制尺度和姿态可以通过 IMU 测量恢复，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 测量的融合正引起越来越多的关注

Our work is most closely related to visual-inertial odom- etry (VIO), lidar-inertial odometry (LIO), and lidar-visual- inertial odometry (LVIO). We note that we do not consider non-inertial systems in this paper, though we are aware that there are successful non-inertial lidar-visual systems, such as [1], [2]. Visual-inertial odometry (VIO) can be grouped into two main categories: filter-based methods and optimization- based methods. Filter-based methods typically use an ex- tended Kalman Filter (EKF) to propagate system states using measurements from a camera and IMU. Optimization- based methods maintain a sliding window estimator and minimize the visual reprojection errors along with IMU measurement error. In our work, we limit our consideration to monocular cameras. Among the most popular publicly- available VIO pipelines, MSCKF [3], ROVIO [4] and Open- VINS [5] are filter-based, and OKVIS [6] and VINS-Mono [7] are optimization-based. Though OKVIS shows superior performance using a stereo camera, it is not optimized for a monocular camera. VINS-Mono performs non-linear optimization in a sliding window setting and achieves state- of-the-art accuracy with a monocular camera [8].

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) 密切相关。我们注意到在本文中我们没有考虑非惯性系统，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：基于过滤器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 的测量值来传播系统状态。基于优化的方法保持滑动窗口估计器并最小化视觉重投影误差和 IMU 测量误差。在我们的工作中，我们将考虑限制在单目相机上。在最流行的公开可用的 VIO 管道中，MSCKF [3]、ROVIO [4] 和 OpenVINS [5] 是基于过滤器的，而 OKVIS [6] 和 VINS-Mono [7] 是基于优化的。虽然 OKVIS 在使用立体相机时表现出卓越的性能，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，并通过单目相机实现最先进的精度

Based on their design scheme, lidar-inertial odometry can also be grouped into two main categories: loosely-coupled methods and tightly-coupled methods. LOAM [9] and LeGO- LOAM [10] are loosely-coupled systems, as the IMU mea- surements are not used in the optimization step. Tightly- coupled systems, which usually offer improved accuracy and robustness, are presently a major focus of ongoing research [11]. Among the publicly-available tightly-coupled systems, LIO-mapping [12] adapts the optimization pipeline of [7] and minimizes the residuals of IMU and lidar measurements. Because LIO-mapping is designed to optimize all measure- ments, real-time performance is not achieved. LIO-SAM [13], which bounds computational complexity by introducing a sliding window of lidar keyframes, utilizes a factor graph for joint IMU and lidar constraint optimization. Specifically designed for ground vehicles, LINS [14] uses an error-state Kalman filter to correct the state of the robot recursively.

根据他们的设计方案，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] 适应了 [7] 的优化管道并最小化了 IMU 和激光雷达测量的残差。由于 LIO 映射旨在优化所有测量，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

Recently, lidar-visual-inertial systems have attracted in- creased attention due to their robustness in sensor degraded tasks [15]. [16] proposes a tightly-coupled LVIO system that features a sequential processing pipeline, which solves the state estimation problem from coarse to fine. The coarse estimation starts with IMU prediction and is then further refined by VIO and LIO. [16] currently achieves the state- of-the-art accuracy on the KITTI benchmark [17]. Based on the framework of MSCKF, [18] features online spatial and temporal multi-sensor calibration. The implementations of [16] and [18] are not publicly available. Our work is different from the aforementioned work, as we leverage a factor graph for global optimization, which can regularly eliminate a robot's incurred drift via loop closure detection.
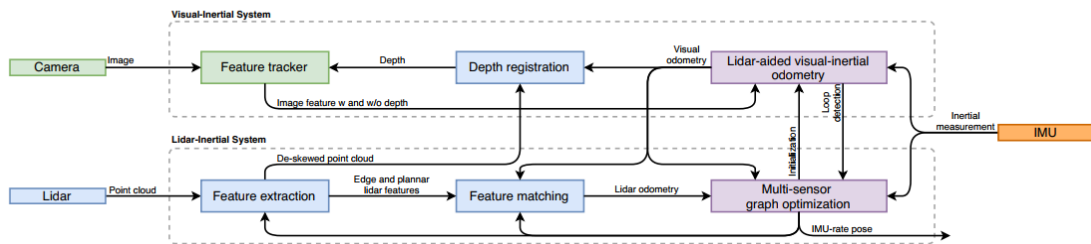
In this paper, we present a framework for tightly-coupled lidar-visual-inertial odometry via smoothing and mapping, LVI-SAM, for real-time state estimation and mapping. Built atop a factor graph, LVI-SAM is composed of two sub- systems, a visual-inertial system (VIS) and a lidar-inertial system (LIS). The two sub-systems can function indepen- dently when failure is detected in one of them, or jointly when enough features are detected. The VIS performs visual feature tracking and optionally extracts feature depth using lidar frames. The visual odometry, which is obtained by optimizing errors of visual reprojection and IMU measure- ments, serves as an initial guess for lidar scan-matching and introduces constraints into the factor graph. After point cloud de-skewing using the IMU measurements, the LIS extracts lidar edge and planar features and matches them to a feature map that is maintained in a sliding window. The estimated system

state in the LIS can be sent to the VIS to facilitate its initialization. For loop closure, candidate matches are first identified by the VIS and further optimized by the LIS. The constraints from visual odometry, lidar odometry, IMU preintegration, and loop closure are optimized jointly in the factor graph. Lastly, the optimized IMU bias terms are leveraged to propagate IMU measurements for pose estimation at the IMU rate. The main contributions of our work can be summarized as follows:

> 在本文中，我们通过平滑和映射 LVI-SAM 提出了一个紧密耦合的激光雷达-视觉-惯性测距框架，用于实时状态估计和映射。 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)。当在其中一个子系统中检测到故障时，这两个子系统可以独立运行，或者在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帧提取特征深度。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 测量进行点云去偏斜后，LIS 提取激光雷达边缘和平面特征，并将它们与在滑动窗口中维护的特征图匹配。 LIS 中估计的系统状态可以发送到 VIS 以促进其初始化。对于闭环，候选匹配首先由 VIS 识别，然后由 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，利用优化的 IMU 偏置项来传播 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总结如下

- A tightly-coupled LVIO framework built atop a factor graph, which achieves both multi-sensor fusion *and* global optimization aided by place recognition.

> 一个建立在因子图之上的紧密耦合的 LVIO 框架，它通过位置识别实现了多传感器融合和全局优化。

- Our framework bypasses failed sub-systems via failure detection, making it robust to sensor degradation.

> 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。

- Our framework is extensively validated with data gath- ered across varied scales, platforms, and environments.

> 我们的框架通过跨不同规模、平台和环境收集的数据进行了广泛验证

Our work is novel from a systems standpoint, representing a unique integration of the state-of-the-art in VIO and LIO to achieve an LVIO system offering improved robustness and accuracy. We hope that our system can serve as a solid baseline which others can easily build on to advance the state-of-the-art in lidar-visual-inertial odometry.

> 从系统的角度来看，我们的工作是新颖的，代表了 VIO 和 LIO 中最先进技术的独特集成，以实现具有更高稳健性和准确性的 LVIO 系统。我们希望我们的系统可以作为一个可靠的基线，其他人可以轻松地在此基础上推进激光雷达视觉惯性里程计的最新技术。

2. LIDAR VISUAL INERTIAL ODOMETRY VIA SMOOTHING AND MAPPING---通过平滑和映射的激光雷达视觉惯性里程计

2.1. *System Overview*

An overview of the proposed lidar-visual-inertial system, which receives inputs from a 3D lidar, a monocular camera, and an IMU, is shown in Figure 1. Our framework is composed of two key sub-systems: a visual-inertial system (VIS) and a lidar-inertial system (LIS). The VIS processes images and IMU measurements, with lidar measurements being optional. Visual odometry is obtained by minimizing the joint residuals of visual and IMU measurements. The LIS extracts lidar features and performs lidar odometry by matching the extracted features with a feature map. The feature map is maintained in a sliding-window manner for real-time performance. Lastly, the state estimation problem, which can be formulated as a maximum a posteriori (MAP) problem, is solved by jointly optimizing the contributions of IMU preintegration constraints, visual odometry constraints, lidar odometry constraints, and loop closure constraints in a factor graph using iSAM2 [19]. Note that the multi-sensor graph optimization employed in the LIS is intended to reduce data exchange and improve system efficiency.

> 拟议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、单目相机的输入和 IMU，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 测量的联合残差来获得的。 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测距。为了实时性能，特征图以滑动窗口的方式维护。最后，状态估计问题可以表述为最大后验 (MAP) 问题，通过联合优化因子图中 IMU 预积分约束、视觉测距约束、激光雷达测距约束和闭环约束的贡献来解决： iSAM2 [19]。请注意，LIS 中采用的多传感器图优化旨在减少数据交换并提高系统效率
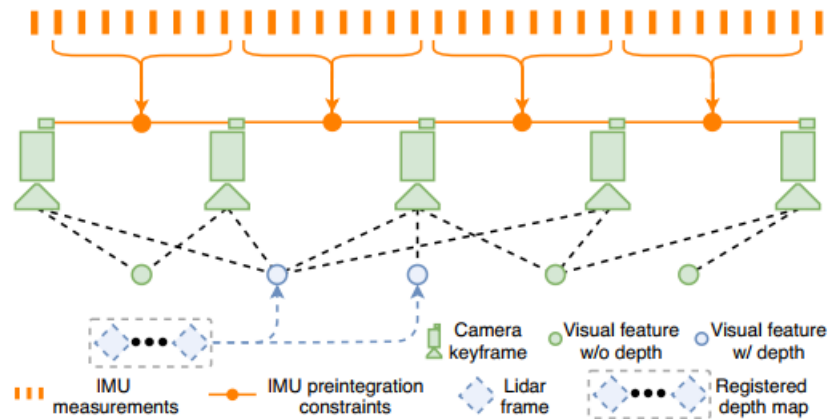

```
1   Fig. 2: The framework of our visual-inertial system. The system optimizes
    residuals from IMU preintegration, visual measurements without depth, and visual
    measurements with depth.

2
3   图 2：我们的视觉惯性系统的框架。该系统优化了来自 IMU 预积分、无深度视觉测量和有深度视觉测量的残差。
```

2.2. *Visual-Inertial System*

We adapt the processing pipeline from [7] for our VIS, which is shown in Figure 2. The visual features are detected using a corner detector [20] and tracked by the Kanade−Lucas−Tomasi algorithm [21]. Upon VIS initializa- tion, we register lidar frames using visual odometry and obtain a sparse depth image for feature depth estimation. The system performs bundle adjustment in a sliding-window setting, where system state **x** ∈ **X** can be written as: **x** = [**R**, **p**, **v**, **b**]. **R** SO(3) is the rotation matrix, **p** R3 is the position vector, **v** is the speed, and **b** = [**b***a,* **b***w*] is the IMU bias **b**a and **b**w are the bias vectors for acceleration and angular velocity respectively. The transformation **T** SE(3) from sensor body frame **B** to world frame **W** is represented as **T** = [**R p**]. In the following sections, we give detailed procedures for improving VIS initialization and feature depth estimation. Due to space limitations, we refer readers to [7] for further details, such as the implementation of residuals.

> 我们为我们的 VIS 调整了 [7] 中的处理管道，如图 2 所示。使用角点检测器 [20] 检测视觉特征，并由 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束调整，其中系统状态 x ∈X 可以写为：x = [R, p, v, b]。 R ∈ SO(3) 是旋转矩阵，p ∈ R3 是位置向量，v 是速度，b = [ba, bw] 是 IMU 偏差。 ba 和 bw 分别是加速度和角速度的偏置向量。从传感器主体框架 B 到世界框架 W 的变换 T ∈ SE(3) 表示为 T = [R|p]。在以下部分中，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

2.2.1. *Initialization:*

Optimization-based VIO usually suffers from divergence due to solving a highly-nonlinear problem at initialization. The quality of initialization heavily depends on two factors: the initial sensor movement and the accuracy of the IMU parameters. In practice, we find that [7] often fails to initialize when the sensor travels at a small or constant velocity. This is due to the fact that metric scale is not observable when acceleration excitation is not large enough. The IMU parameters include a slowly varying bias and white noise, which effect both the raw acceleration and angular velocity measurements. A good guess of these parameters at initialization helps the optimization converge faster.

> 由于在初始化时解决高度非线性问题，基于优化的 VIO 通常会出现发散。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移动时，[7] 经常无法初始化。这是因为当加速度激励不够大时，无法观察到公制尺度。 IMU 参数包括缓慢变化的偏置和白噪声，它们会影响原始加速度和角速度测量值。在初始化时很好地猜测这些参数有助于优化更快地收敛

To improve the robustness of our VIS initialization, we leverage the estimated system state **x** and IMU bias **b** from the LIS. Because depth is directly observable from a lidar, we first initialize the LIS and obtain **x** and **b**. Then we interpolate and associate them to each image keyframe based on the image timestamp. Note that the IMU bias is assumed to be constant between two image keyframes. Finally, the estimated **x** and **b** from the LIS are used as the initial guess for VIS initialization, which significantly improves initialization speed and robustness. A comparison of VIS initialization with and without the help of the LIS can be found in the supplementary video

> 为了提高我们的 VIS 初始化的鲁棒性，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，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对它们进行插值并将它们关联到每个图像关键帧。请注意，假设 IMU 偏差在两个图像关键帧之间是恒定的。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。可以在补充视频中找到使用和不使用 LIS 的 VIS 初始化的比较
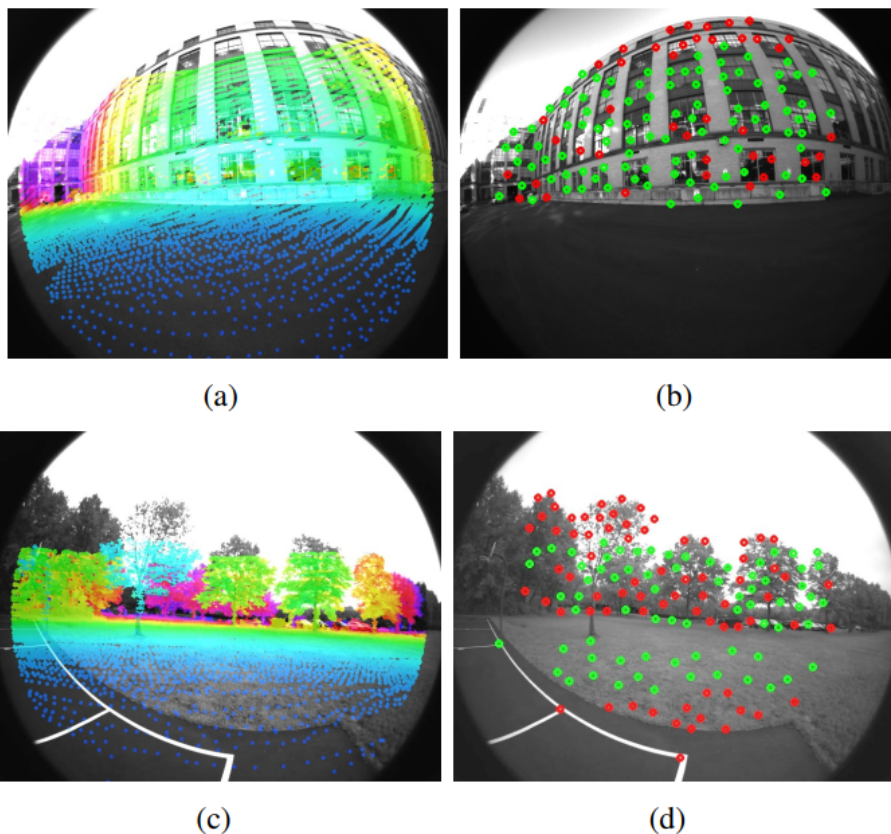
2.2.2. *Feature depth association*

Upon the initialization of the VIS, we register lidar frames to the camera frame using the estimated visual odometry. Due to the fact that a modern 3D lidar often yields sparse scans, we stack multiple lidar frames to obtain a dense depth map. To associate a feature with a depth value, we first project visual features and lidar depth points on a unit sphere that is centered at the camera. Depth points are then downsampled and stored using their polar coordinates for a constant density on the sphere. We find the nearest three depth points on the sphere for a visual feature by searching a two-dimensional K-D tree using the polar coordinates of the visual feature. At last, the feature depth is the length of the line formed by the visual feature and camera center O_c, which intersects the plane formed by the three depth points in Cartesian space. A visualization of this process can be found in Figure 3(a), where the feature depth is the length of the dashed straight line.

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。由于现代 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。然后使用它们的极坐标对深度点进行下采样和存储，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，找到球体上最近的三个深度点以获取视觉特征。最后，特征深度是视觉特征和相机中心Oc形成的线的长度，该线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，其中特征深度是虚线的长度

(a) Depth association   (b) Depth validation
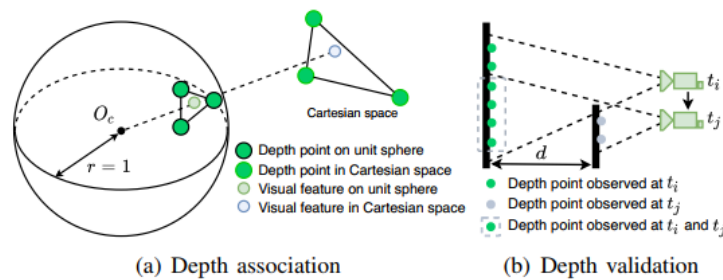
```
1   Fig. 3: Visual feature depth association
2   图 3: 视觉特征深度关联
```

(a)   (b)

(c)   (d)

We further validate the associated feature depth by check- ing the distance among the three nearest depth points. This is because stacking lidar frames from different timestamps may result in depth ambiguity from different objects. An illustration of such a case is shown in Figure 3(b). The depth points observed at time t_i are depicted in green. The camera is moved to a new position at t_j and observes new depth points that are colored gray. However, the depth points at t_i, which are circled by a dashed gray line, may still be observable at t_j due to lidar frame stacking. Associating feature depth using depth points from different objects results in inaccurate estimation. Similar to [16], we reject such estimations by checking the maximum distance among depth points for a feature. If the maximum distance is larger than 2m, the feature has no depth associated.

我们通过检查三个最近的深度点之间的距离来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不同对象的深度模糊。这种情况的说明如图 3(b) 所示。在时间 ti 观察到的深度点用绿色表示。相机移动到 tj 处的新位置并观察新的深度灰色的点。然而，由于激光雷达帧堆叠，在 ti 处的深度点（由灰色虚线圈出）在 tj 处仍然可以观察到。使用来自不同对象的深度点来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这种估计。如果最大距离大于 2m，则要素没有关联深度

A demonstration of the registered depth map and visual features is shown in Figure 4. In Figure 4 (a) and (c), the depth points that are registered using visual odometry are projected onto the camera images. In Figure 4 (b) and (d), the visual features that are successfully associated with depth are colored green. Note that though the depth map covers the majority of the image in Figure 4(a), many features in 4(b) that are located on the corners of windows lack depth association due to validation check failure.

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

2.2.3. *Failure detection---缺陷检测:*

The VIS suffers from failure due to aggressive motion, illumination change, and texture-less environments. When a robot undergoes aggressive motion or enters a texture-less environment, the number of tracked features decreases greatly. Insufficient features may lead to optimization divergence. We also notice that a large IMU bias is estimated when the VIS fails. Thus we report VIS failure when the number of tracked features is below a threshold,or when the estimated IMU bias exceeds a threshold. Active failure detection is necessary for our system so that its failure does not corrupt the function of the LIS. Once a failure is detected, the VIS re-initializes and informs the LIS.

> 由于剧烈运动、光照变化和无纹理环境，VIS 会出现故障。当机器人进行剧烈运动或进入无纹理环境时，跟踪特征的数量会大大减少。特征不足可能会导致优化发散。我们还注意到，当 VIS 出现故障时，估计会产生很大的 IMU 偏差。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 偏差超过阈值时。我们的系统需要主动故障检测，以便其故障不会破坏 LIS 的功能。一旦检测到故障，VIS 重新初始化并通知 LIS。
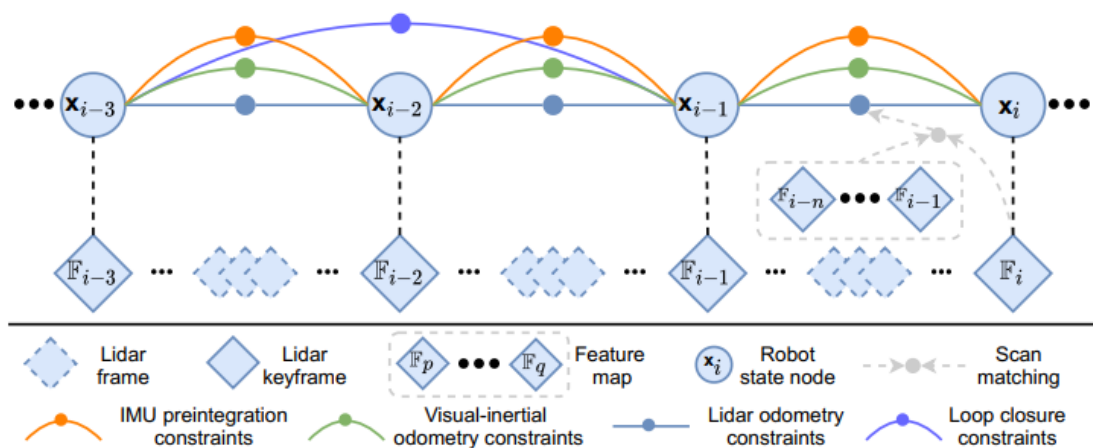

```
1   Fig. 5: The framework of our lidar-inertial system. The system maintains a factor
        graph that has four types of constraints
2   图 5: 我们的激光雷达惯性系统的框架。系统维护一个因子图，它有四种类型的约束
```

2.3. *Lidar-Inertial System----雷达惯性系统*

As is shown in Figure 5, the proposed lidar-inertial system, which is adapted from [13], maintains a factor graph for global pose optimization. Four types of constraints, IMU preintegration constraints, visual odometry constraints, lidar odometry constraints, and loop closure constraints, are added to the graph and optimized jointly. Lidar odometry con- straints are derived from scan-matching, where we match the current lidar keyframe to a global feature map. The candidates for loop closure constraints are first provided by the VIS and then further optimized by scan-matching. We maintain a sliding window of lidar keyframes for the feature map, which guarantees bounded computational complexity. A new lidar keyframe is selected when the change of the robot pose exceeds a threshold. The intermittent lidar frames lying between pairs of keyframes are discarded. Upon the selection of a new lidar keyframe, a new robot state **x** is added to the factor graph as a node. Adding the keyframes in this way not only achieves a balance between memory consumption and map density, but also helps maintain a relatively sparse factor graph for real-time optimization. Due to space limitations, we refer the readers to [13] for implementation details. In the following sections, we focus on new procedures that improve system robustness.

> 如图 5 所示，根据 [13] 改编的拟议激光雷达惯性系统维护了一个用于全局姿态优化的因子图。四种类型的约束，IMU 预积分约束、视觉里程计约束、激光雷达里程计约束和回环约束，被添加到图中并联合优化。激光雷达里程计约束来自扫描匹配，我们将当前的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。我们为特征图维护了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度。当机器人姿态的变化超过阈值时，选择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。在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。以这种方式添加关键帧不仅可以实现内存消耗和地图密度之间的平衡，而且有助于维护一个相对稀疏的因子图以进行实时优化。由于篇幅所限，我们请读者参考[13]了解实现细节。在以下部分中，我们将重点介绍提高系统稳健性的新程序。

2.3.1. *Initial guess:*

We find the initial guess plays a critical role in the success of scan-matching, especially when the sensor undergoes aggressive motions. The source of initial guess is different before and after LIS initialization.

> 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。

Before LIS initialization, we assume the robot starts from a static position with zero velocity. Then we integrate raw IMU measurements assuming the bias and noise are zero-valued. The integrated translational and rotational change between two lidar keyframes produce the initial guess for scan- matching. We find this approach can successfully initialize the system in challenging conditions when the initial linear velocity is smaller than 10 *m/s* and the angular velocity is smaller than 180 ◦**/s*. Once the LIS is initialized, we estimate the IMU bias, robot pose, and velocity in the factor graph. Then we send them to the VIS to aid its initialization.

> 在 LIS 初始化之前，我们假设机器人从零速度的静态位置开始。然后我们整合原始 IMU 测量，假设偏置和噪声为零值。两个激光雷达关键帧之间的综合平移和旋转变化产生了扫描匹配的初始猜测。我们发现这个方法可以成功初始化当初始线速度小于 10 m/s 且角速度小于 180 ◦/s 时，系统处于具有挑战性的条件下。 LIS 初始化后，我们在因子图中估计 IMU 偏差、机器人位姿和速度。然后我们将它们发送到 VIS 以帮助其初始化
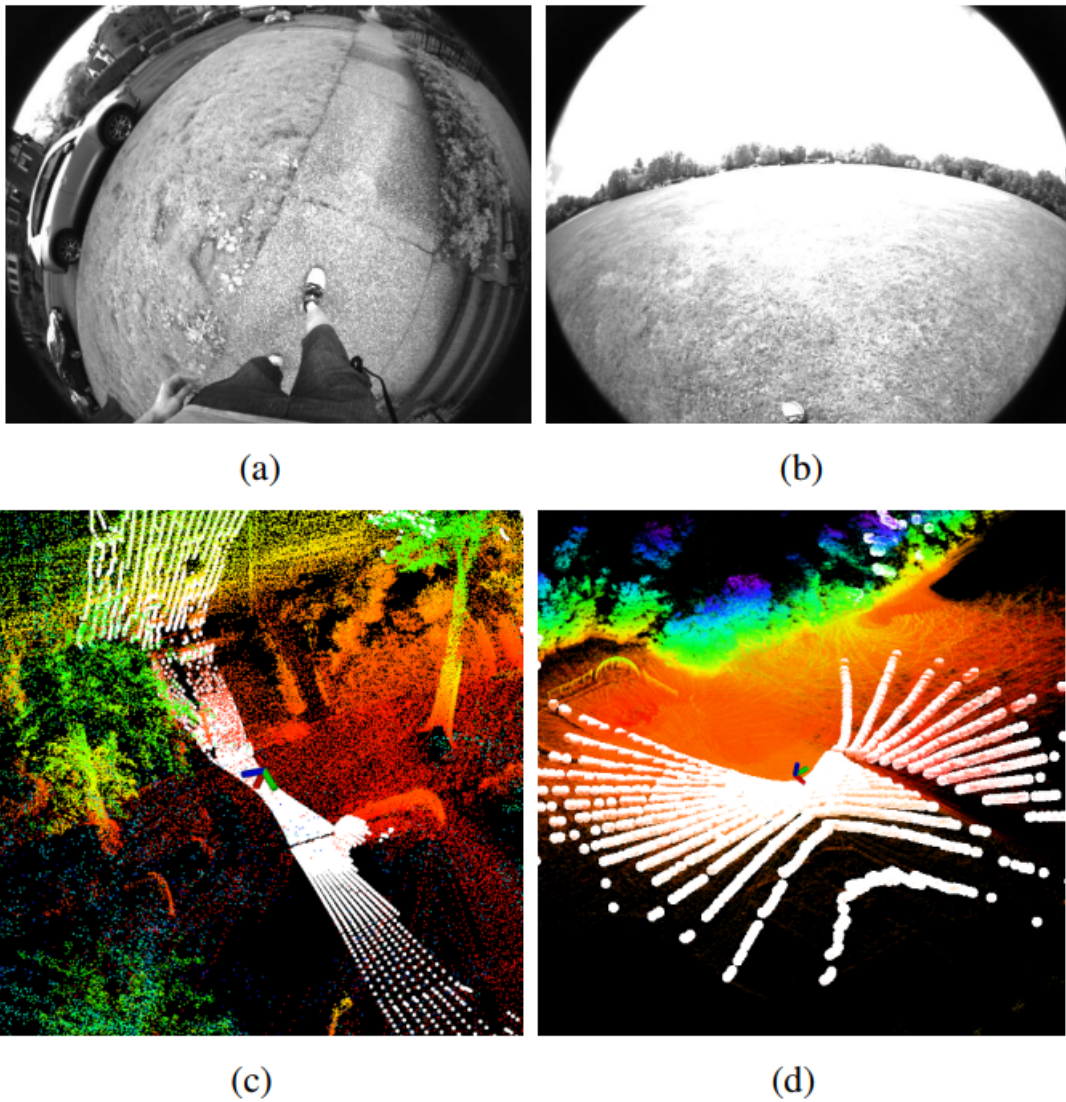
Fig. 6: Degenerate mapping scenarios where scan-matching is ill- constrained. In (a) and (c), the lidar is placed facing the ground. In (b) and (d), the lidar is in a flat, open, and structure-less environment. White points indicate the contents of the lidar scan. Color variation indicates elevation change.

图 6: 扫描匹配受到不良约束的退化映射场景。在 (a) 和 (c) 中,激光雷达面向地面放置。在 (b) 和 (d) 中,激光雷达处于平坦、开放且无结构的环境中。白点表示激光雷达扫描的内容。颜色变化表示海拔变化。

After LIS initialization, we can obtain initial guesses from two sources: integrated IMU measurements with corrected bias, and the VIS. We use visual-inertial odometry as an initial guess when it is available. If the VIS reports failure, we then switch to IMU measurements for the initial guess. These procedures increase the initial guess accuracy and robustness in both texture-rich and texture-less environments.

LIS 初始化后,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 测量值和 VIS。当可用时,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 测量以进行初始猜测。这些程序在纹理丰富和纹理少的环境中提高了初始猜测的准确性和鲁棒性。

2.3.2. *Failure detection*

Though lidar can capture the fine details of an environment at long ranges, it still encounters degraded scenarios where scan-matching is ill-constrained. Such scenarios are depicted in Figure 6. We adapt the method from [24] for LIS failure detection. The non-linear optimization problem in scan-matching can be formulated as solving a linear problem iteratively: $\min \|\mathbf{A}\mathbf{T} - \mathbf{b}\|_2\|$. where $\mathbf{A}$ and $\mathbf{b}$ are obtained from linearization at $\mathbf{T}$. The LIS reports failure when the smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$ is smaller than a threshold at the first iteration of optimization. Lidar odometry constraints are not added to the factor graph when failure happens. We refer the reader to [24] for the detailed analysis upon which these assumptions are based.

> 缺陷检测：尽管激光雷达可以在远距离捕获环境的精细细节，但它仍然会遇到扫描匹配受到不良约束的退化场景。图 6 描述了此类场景。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：minT||AT −b ||2其中 A 和 b 是从 T 处的线性化获得的。当 ATA 的最小特征值在第一次优化迭代时小于阈值时，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] 以了解这些假设所依据的详细分析。

3. EXPERIMENTS---实验

We now describe a series of experiments to validate the proposed framework on three self-gathered datasets, which are referred to as *Urban*, *Jackal*, and *Handheld*. The details of these datasets are provided in the following sections. Our sensor suite for data-gathering includes a Velodyne VLP-16 lidar, a FLIR BFS-U3-04S2M-CS camera, a MicroStrain 3DM-GX5-25 IMU, and a Reach RS+ GPS (for ground truth). We compare the proposed framework with open-sourced solutions, which include VINS-Mono, LOAM, LIO-mapping, LINS, and LIO-SAM. All the methods are implemented in C++ and executed on a laptop with an Intel i7-10710U in Ubuntu Linux. Our implementation of LVI- SAM and datasets are available at the link below2.

> 我们现在描述了一系列实验，以在三个自收集数据集上验证所提出的框架它们被称为 Urban、Jackal 和 Handheld。以下部分提供了这些数据集的详细信息。我们用于数据收集的传感器套件包括 Velodyne VLP-16 激光雷达、FLIR BFS-U3-04S2M-CS 相机、MicroStrain 3DM-GX5-25 IMU 和 Reach RS+ GPS （用于地面实况）。我们将提议的框架与开源解决方案进行比较，其中包括 VINS-Mono、LOAM、LIO-mapping、LINS 和 LIO-SAM。所有方法都是用 C++ 实现的，并在带有 Intel i7-10710U 的笔记本电脑上在 Ubuntu Linux 上执行。我们对 LVI-SAM 和数据集的实施可在以下链接中获得
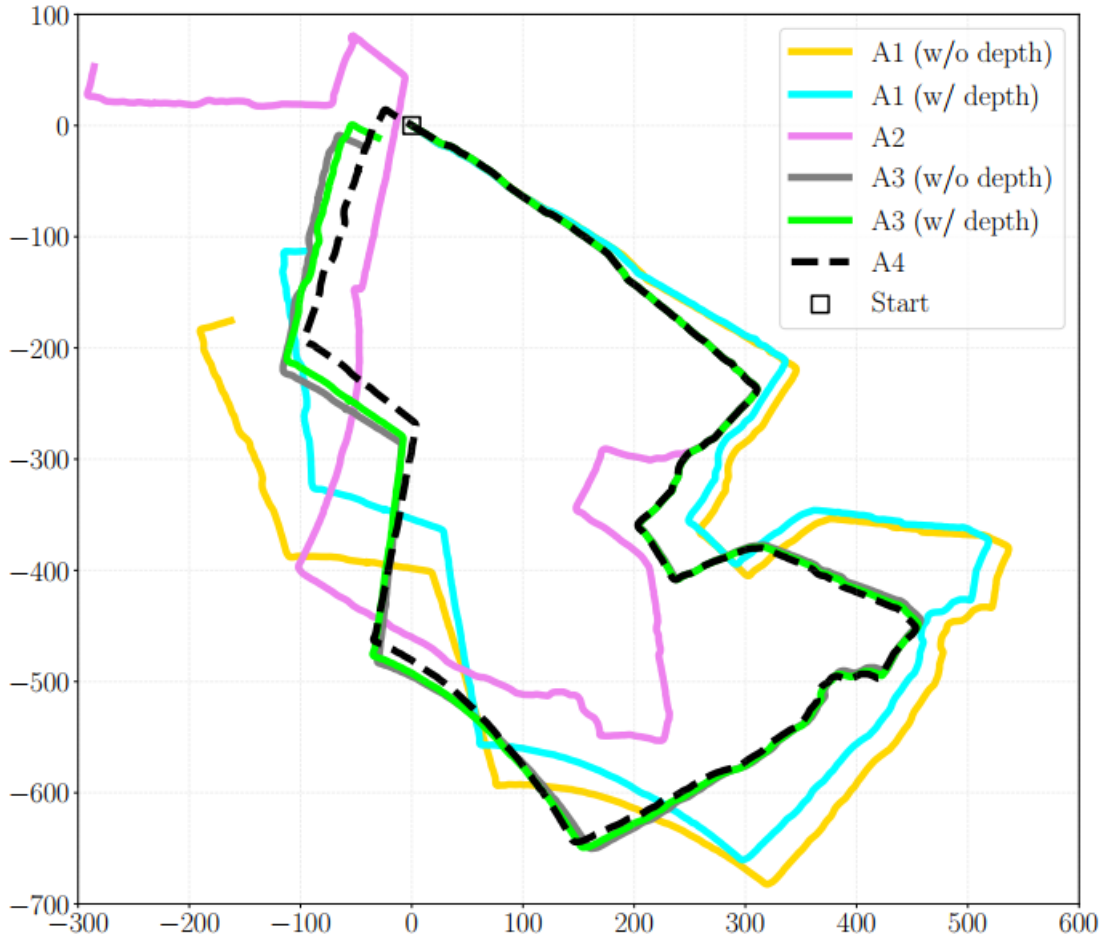
Error type	A1 (w/o depth)	A1 (w/ depth)	A2	A3 (w/o depth)	A3 (w/o depth)	A4
Translation (m)	239.19	142.12	290.43	45.42	32.18	0.28
Rotation (degree)	60.72	39.52	116.50	6.41	7.66	5.77

3.1. *A. Ablation Study*---消融研究

We show how the design of each module in the system affects the performance of the proposed framework using the *Urban* dataset. This dataset, which features buildings, parked and moving cars, pedestrians, cyclists, and vegeta- tion, is gathered by the operator walking and carrying the sensor suite. We also deliberately place the sensor suite in challenging positions (Figure 6(a)) to validate the robustness of the system in degraded scenarios. Due to dense overhead vegetation, the region is GPS-denied. We start and end the data-gathering process at the same position to validate end- to-end translation and rotation errors, provided in Table I.

我们展示了系统中每个模块的设计如何使用城市数据集影响所提出框架的性能。该数据集以建筑物、停放和移动的汽车、行人、骑自行车者和植被为特征，由步行和携带传感器套件的操作员收集。我们还特意将传感器套件放置在具有挑战性的位置（图 6（a）），以验证系统在降级场景中的稳健性。由于头顶植被茂密，该地区无法使用 GPS。我们在同一位置开始和结束数据收集过程，以验证端到端的平移和旋转错误，如表 I 所示。

3.1.1. *A1 - Effect of including feature depth information from lidar for visual-inertial odometry:*---包括来自激光雷达的特征深度信息用于视觉惯性测距的效果

We disable the scan- matching in the LIS and perform pose estimation solely relying on the VIS. The resulting trajectories with and without enabling depth registration are labeled as A1 in Figure 7. The trajectory direction is clock-wise. The end- to-end pose errors, shown in Table I, are greatly reduced when associating depth with visual features.

我们禁用LIS 中的扫描匹配并仅依靠 VIS 执行姿态估计。启用和不启用深度配准的结果轨迹在 2 https://git.io/lvi-sam (a) Jackal 数据集环境 (b) 手持数据集环境中标记为 A1 图 8：卫星图像，其中 Jackal 和手持数据集被聚集。白点表示数据集中的 GPS 可用性。图 7. 轨迹方向为顺时针方向。当将深度与视觉特征相关联时，表 I 中显示的端到端姿势错误大大减少。

3.1.2. *A2 - Effect of including visual-inertial odometry:*---A2 - 包括视觉惯性里程计的影响：

We disable the VIS and use only the LIS for pose estimation. The trajectory, which is labeled A2 in Figure 7, diverges a couple of times when a degraded scenario is encountered.

我们禁用 VIS 并仅使用 LIS 进行姿态估计。在图 7 中标记为 A2 的轨迹在遇到降级场景时会发散几次。

3.1.3. *A3 - Effect of including feature depth information from lidar for lidar-visual-inertial odometry:*---将来自激光雷达的特征深度信息用于激光雷达-视觉-惯性里程计的效果：

We now use the VIS and LIS together, toggling the depth registration module in the VIS to compare the resulting LVIO trajectories. With the help of depth for visual features, the translation error is further reduced by 29%, from 45.42 m to 32.18 m. Note that the loop closure detection is disabled in this test to validate the system in a pure odometry mode.

我们现在将 VIS 和 LIS 一起使用，切换 VIS 中的深度注册模块来比较生成的 LVIO 轨迹。在视觉特征深度的帮助下，平移误差进一步降低了 29%，从 45.42 m 减少到 32.18 m。请注意，此测试中禁用回环检测以在纯里程计模式下验证系统。

3.1.4. *A4 - Effect of including visual loop closure detection:*---包含视觉闭环检测的效果：

We eliminate the drift of the system by enabling the loop closure detection function in the VIS. The final trajectory when every module is enabled in the framework is labeled as A4 in Figure 7.

包含视觉闭环检测的效果：我们通过在 VIS 中启用闭环检测功能来消除系统的漂移。在框架中启用每个模块时的最终轨迹在图 7 中标记为 A4

3.2. *Jackal Dataset*

The *Jackal* dataset is gathered by mounting the sensor suite on a Clearpath Jackal unmanned ground vehicle (UGV). We manually drive the robot in a feature-rich environment, beginning and ending at the same position. The environment, shown in Figure 8(a), features structures, vegetation, and various road surfaces. The regions where GPS reception is available are marked with white dots.

> Jackal 数据集是通过将传感器套件安装在 Clearpath Jackal 无人地面车辆 (UGV) 上来收集的。我们在功能丰富的环境中手动驱动机器人，在同一位置开始和结束。图 8(a) 所示的环境具有结构、植被和各种路面特征。可以接收 GPS 的地区用白点标记。

We compare various methods and show their trajectories in Figure 9(a). We further validate the accuracy of methods that have loop closure functionality by manually disabling and enabling it. The benchmarking results are shown in Table II. LVI-SAM achieves the lowest average root mean square error (RMSE) with respect to the GPS measurements, which are treated as ground truth. The lowest end-to-end translation error is achieved by LINS, which is adapted from LeGO- LOAM [10] and specifically designed for UGV operations. The lowest end-to-end rotational error is again achieved by LVI-SAM.

> 我们比较了各种方法并在图 9(a) 中显示了它们的轨迹。我们通过手动禁用和启用它来进一步验证具有闭环功能的方法的准确性。基准测试结果如表二所示。 LVI-SAM 实现了相对于 GPS 测量的最低平均均方根误差 (RMSE)，这些测量被视为地面实况。最低的端到端翻译错误是由 LINS 实现的，它改编自 LeGO-LOAM [10]，专为 UGV 操作而设计。 LVI-SAM 再次实现了最低的端到端旋转误差。
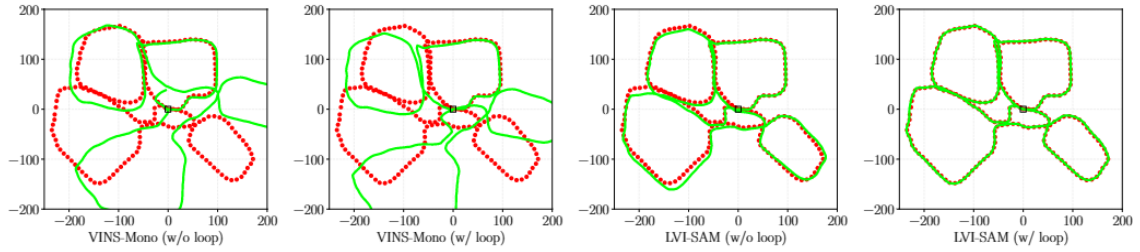
3.3. *Handheld Dataset*

The *Handheld* dataset is gathered by an operator carrying the sensor suite walking around several open fields, which are shown in Figure 8(b). The dataset also starts and ends at the same position. We increase the challenge of this dataset by passing an open baseball field, which is located at top center of the image. When passing this field, the main observations gathered by the camera and lidar feature grass and a ground plane, respectively (Fig. 6 (b) and (d)). Due to the aforementioned degeneracy problem, all the lidar-based methods fail to generate meaningful results. The proposed framework, LVI-SAM, successfully finishes the test with or without loop closures enabled, achieving the lowest errors over all three benchmarking criteria featured in Table II.

> 手持式数据集由操作员携带传感器套件在几个开阔的田野中走动收集，如图 8(b) 所示。数据集也在同一位置开始和结束。我们通过经过位于图像顶部中心的开放棒球场来增加该数据集的挑战。当经过这个领域时，相机和激光雷达收集的主要观察结果分别是草和地平面（图 6（b）和（d））。由于上述退化问题，所有基于激光雷达的方法都无法产生有意义的结果。所提议的框架 LVI-SAM 在启用或不启用闭环的情况下成功完成了测试，在表 II 中的所有三个基准测试标准中实现了最低的错误
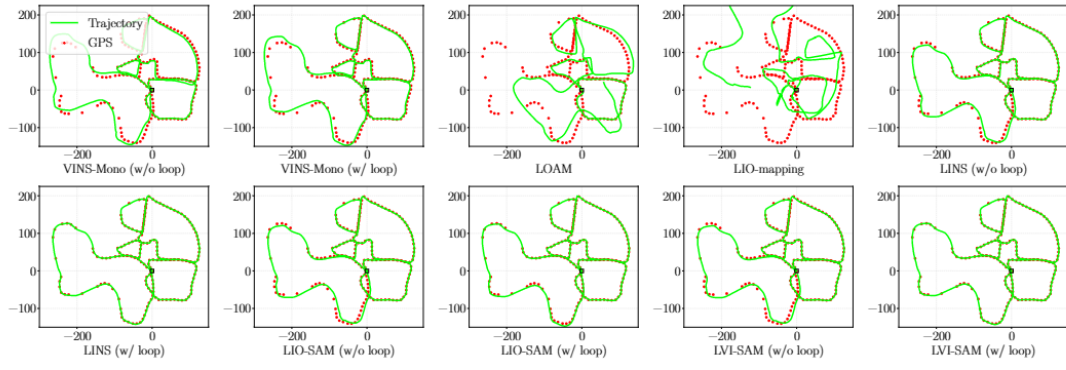
(a) Trajectories of Jackal dataset

a）Jackal 数据集的轨迹

(b) Trajectories of Handheld dataset

（b）手持数据集的轨迹

Fig. 9: Trajectories of various methods using the Jackal and Handheld datasets. The green line depicts the resulting trajectory from each compared method. The GPS positioning measurements, which serve as the ground truth, are shown with red dots. Trajectories of LOAM, LIO-mapping, LINS, and LIO-SAM are not shown due to their failure to generate meaningful results.

图 9: 使用 Jackal 和 Handheld 数据集的各种方法的轨迹。绿线描绘了每个比较方法的结果轨迹。作为地面实况的 GPS 定位测量结果用红点显示。 LOAM、LIO-mapping、LINS 和 LIO-SAM 的轨迹由于未能产生有意义的结果而未显示。

TABLE II: Quantitative comparison of Jackal and Handheld datasets using various methods.

表 II: 使用各种方法对 Jackal 和 Handheld 数据集进行定量比较。

Dataset	Error type	VINS (w/o loop)	VINS (w/ loop)	LOAM	LIO-mapping	LINS (w/o loop)	LINS (w/ loop)	LIO-SAM (w/o loop)	LIO-SAM (w/ loop)	LVI-SAM (w/o loop)	LVI-SAM (w/ loop)
Jackal	RMSE w.r.t GPS (m)	8.58	4.49	44.92	127.05	3.95	0.77	3.54	1.52	4.05	**0.67**
	Translation (m)	10.82	11.86	61.73	123.22	7.37	**0.09**	5.48	0.12	4.69	0.11
	Rotation (degree)	9.98	12.79	61.41	139.23	4.80	2.07	2.18	2.64	2.28	**1.52**
Handheld	RMSE w.r.t GPS (m)	87.53	73.07	Fail	Fail	Fail	Fail	53.62	Fail	7.87	**0.83**
	Translation (m)	40.65	1.87	Fail	Fail	Fail	Fail	58.91	Fail	7.57	**0.27**
	Rotation (degree)	53.48	52.09	Fail	Fail	Fail	Fail	20.17	Fail	24.82	**3.82**

4. CONCLUSIONS

We have proposed LVI-SAM, a framework for tightly- coupled lidar-visual-inertial odometry via smoothing and mapping, for performing real-time state estimation and map- ping in complex environments. The proposed framework is composed of two sub-systems: a visual-inertial system and a lidar-inertial system. The two sub-systems are designed to interact in a tightly-coupled manner to improve system robustness and accuracy. Through evaluations on datasets over various scales, platforms, and environments, our sys- tem shows comparable or better accuracy than the existing publicly available methods. We hope that our system will serve as a solid baseline which others can easily build on to advance the state of the art in lidar-visual-inertial odometry.

> 我们提出了 LVI-SAM，这是一种通过平滑和映射进行紧耦合激光雷达-视觉-惯性测距的框架，用于在复杂环境中执行实时状态估计和映射。所提出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密耦合的方式进行交互，以提高系统的稳健性和准确性。通过对不同规模、平台和环境的数据集的评估，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可靠的基线，其他人可以轻松地建立它以推进激光雷达视觉惯性里程计的最新技术。

5. ACKNOWLEDGEMENT---致谢

6. REFERENCES

[1] J. Graeter, A. Wilczynski, and M. Lauer, "LIMO: Lidar-Monocular Visual Odometry," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 7872–7879, 2018.

[2] Y.-S. Shin, Y. S. Park, and A. Kim, "DVL-SLAM: Sparse Depth Enhanced Direct Visual-LiDAR SLAM," *Autonomous Robots*, vol. 44, no. 2, pp. 115–130, 2020.

[3] A. I. Mourikis and S. I. Roumeliotis, "A Multi-state Constraint Kalman Filter for Vision-aided Inertial Navigation," *IEEE International Con- ference on Robotics and Automation (ICRA)*, pp. 3565–3572, 2007.

[4] M. Bloesch, S. Omari, M. Hutter, and R. Siegwart, "Robust Visual Inertial Odometry using A Direct EKF- based Approach," *IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pp. 298–304, 2015.

[5] P. Geneva, K. Eckenhoff, W. Lee, Y. Yang, and G. Huang, "OpenVINS: A Research Platform for Visual-Inertial Estimation," *IROS Workshop on Visual-Inertial Navigation: Challenges and Applications*, 2019.

[6] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based Visual-Inertial Odometry using Nonlinear Optimiza- tion," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[7] T. Qin, P. Li, and S. Shen, "VINS-Mono: A Robust and Versatile Monocular Visual-Inertial State Estimator," *IEEE Transactions on Robotics*, vol. 34, no. 4, pp. 1004–1020, 2018.

[8] J. Delmerico and D. Scaramuzza, "A Benchmark Comparison of Monocular Visual-Inertial Odometry Algorithms for Flying Robots," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2502–2509, 2018.

[9] J. Zhang and S. Singh, "Low-drift and Real-time Lidar Odometry and Mapping," *Autonomous Robots*, vol. 41, no. 2, pp. 401–416, 2017.

[10] T. Shan and B. Englot, "LeGO-LOAM: Lightweight and Ground- Optimized Lidar Odometry and Mapping on Variable Terrain," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4758–4765, 2018.

[11] C. Chen, H. Zhu, M. Li, and S. You, "A Review of Visual-Inertial Simultaneous Localization and Mapping from Filtering-based and Optimization-based Perspectives," *Robotics*, vol. 7, no. 3, p. 45, 2018.

[12] H. Ye, Y. Chen, and M. Liu, "Tightly Coupled 3D Lidar Inertial Odometry and Mapping," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3144–3150, 2019.

[13] T. Shan, B. Englot, D. Meyers, W. Wang, C. Ratti, and D. Rus, "LIO- SAM: Tightly-coupled Lidar Inertial Odometry via Smoothing and Mapping," *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4758–4765, 2020.

[14] C. Qin, H. Ye, C. E. Pranata, J. Han, S. Zhang, and M. Liu, "LINS: A Lidar-Inertial State Estimator for Robust and Efficient Navigation," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 8899–8906, 2020.

[15] C. Debeunne and D. Vivet, "A Review of Visual-LiDAR Fusion based Simultaneous Localization and Mapping," *Sensors*, vol. 20, no. 7, p. 2068, 2020.

[16] J. Zhang and S. Singh, "Laser-Visual-Inertial Odometry and Mapping with High Robustness and Low Drift," *Journal of Field Robotics*, vol. 35, no. 8, pp. 1242–1264, 2018.

[17] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision Meets Robotics: The KITTI Dataset," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1231–1237, 2013.

[18] X. Zuo, P. Geneva, W. Lee, Y. Liu, and G. Huang, "LIC-Fusion: LiDAR-Inertial-Camera Odometry," *arXiv preprint arXiv:1909.04102*, 2019.

[19] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, and

F. Dellaert, "iSAM2: Incremental Smoothing and Mapping using the Bayes Tree," *The International Journal of Robotics Research*, vol. 31, no. 2, pp. 216–235, 2012.

[20] J. Shi *et al.*, "Good Features to Track," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593−600, 1994.

[21] B. D. Lucas, T. Kanade *et al.*, "An Iterative Image Registration Technique with an Application to Stereo Vision," 1981.

[22] D. Ga´lvez-Lo´pez and J. D. Tardos, "Bags of Binary Words for Fast Place Recognition in Image Sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188−1197, 2012.

[23] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "Brief: Binary robust independent elementary features," *European conference on computer vision*, pp. 778−792, 2010.

[24] J. Zhang, M. Kaess, and S. Singh, "On Degeneracy of Optimization- based State Estimation Problems," *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 809−816, 2016.

[1] J. Graeter、A. Wilczynski 和 M. Lauer，"LIMO：激光雷达-单目视觉里程计"，IEEE/RSJ 智能机器人和系统国际会议 (IROS)，第 7872-7879 页，2018 年。

[2] Y.-S。 Shin、Y. S. Park 和 A. Kim，"DVL-SLAM：稀疏深度增强型直接视觉激光雷达 SLAM"，自主机器人，第一卷。 44，没有。 2，第 115−130 页，2020 年。

[3] AI Mourikis 和 SI Roumeliotis，"用于视觉辅助惯性导航的多状态约束卡尔曼滤波器"，IEEE 机器人与自动化国际会议 (ICRA)，第 1 页。 3565−3572, 2007.

[4] M. Bloesch、S. Omari、M. Hutter 和 R. Siegwart，"Robust Visual Inertial Odometry using A Direct EKF-based Approach"，IEEE/RSJ 智能机器人和系统国际会议(IROS)，第 298-304 页，2015 年。

[5] P.Geneva、K. Eckenhoff、W. Lee、Y. Yang 和 G. Huang，"OpenVINS：视觉惯性估计研究平台"，IROS视觉惯性导航研讨会：挑战与应用，2019 年。

[6] S. Leutenegger、S. Lynen、M. Bosse、R. Siegwart 和 P. Furgale，"使用非线性优化的基于关键帧的视觉惯性里程计，"国际机器人研究杂志，第一卷。 34，没有。 3, pp. 314−334, 2015.

[7] T. Qin、P. Li 和 S. Shen，"VINS-Mono：一种鲁棒且通用的单目视觉惯性状态估计器"，IEEE 机器人学报，卷。 34，没有。 4, pp. 1004−1020, 2018.

[8] J. Delmerico 和 D. Scaramuzza，"飞行机器人单目视觉惯性测距算法的基准比较"，IEEE 机器人与自动化国际会议 (ICRA)，第2502−2509, 2018.

[9] J. Zhang 和 S. Singh，"低漂移和实时激光雷达测距和测绘，"自主机器人，卷。 41，没有。 2, pp. 401−416, 2017.

[10] T. Shan 和 B. Englot，"LeGO-LOAM：轻量级和地面优化激光雷达测距和可变地形测绘"，IEEE/RSJ 智能机器人和系统国际会议(IROS), pp. 4758−4765, 2018.

[11] C. Chen、H. Zhu、M. Li 和 S. You，"基于过滤和优化的视觉惯性同时定位和映射回顾 -基于的观点，"机器人学，卷。 7，没有。 3，第。 45, 2018.

 [12] H. Ye, Y. Chen, and M. Liu, "Tightly Coupled 3D Lidar Inertial Odometry and Mapping"，IEEE 机器人与自动化国际会议 (ICRA)，第 3144-3150 页，2019 年。

[13] T. Shan、B. Englot、D. Meyers、W. Wang、C. Ratti 和 D. Rus，"LIO-SAM：通过平滑和映射的紧密耦合激光雷达惯性里程计"，IEEE/RSJ 国际会议关于智能机器人和系统 (IROS)，第 4758-4765 页，2020 年。

[14] C. Qin、H. Ye、CE Pranata、J. Han、S. Zhang 和 M. Liu，"LINS：A Lidar- Inertial State Estimator for Robust and Efficient Navigation，IEEE 机器人与自动化国际会议 (ICRA)，第 8899-8906 页，2020 年。

[15] C. Debeunne 和 D. Vivet，"基于视觉-LiDAR 融合的同时回顾定位和测绘，"传感器，卷。 20，没有。 7，第。 2068, 2020.

[16] J. Zhang 和 S. Singh，"具有高鲁棒性和低漂移的激光-视觉-惯性测距和测绘"，现场机器人学杂志，第一卷。 35，没有。 8, pp. 1242–1264, 2018.

[17] A. Geiger、P. Lenz、C. Stiller 和 R. Urtasun，"Vision Meets Robotics: The KITTI Dataset"，《国际机器人研究杂志》，第一卷。 32，没有。 11, pp. 1231–1237, 2013.

[18] X. Zuo, P. Geneva, W. Lee, Y. Liu, and G. Huang, "LIC-Fusion: LiDAR-Inertial-Camera Odometry," arXiv preprint arXiv :1909.04102, 2019.

[19] M. Kaess、H. Johannsson、R. Roberts、V. Ila、JJ Leonard 和 F. Dellaert，"iSAM2：使用贝叶斯树的增量平滑和映射"，国际机器人学杂志研究，卷。 31，没有。 2, pp. 216–235, 2012.

[20] J. Shi 等人，"Good Features to Track"，IEEE 计算机视觉和模式识别会议，第 593–600 页，1994 年。

[21] BD Lucas， T. Kanade 等人，"一种应用于立体视觉的迭代图像配准技术"，1981。

[22] D. G´alvez-L´opez 和 JD Tardos，"用于图像序列中快速位置识别的二进制词袋，"" IEEE 机器人学汇刊，卷。 28，没有。 5, pp. 1188–1197, 2012.

[23] M. Calonder、V. Lepetit、C. Strecha 和 P. Fua，"Brief：Binary稳健独立的基本特征"，欧洲计算机视觉会议，第 778 页792, 2010.

[24] J. Zhang、M. Kaess 和 S. Singh，"基于优化的状态估计问题的退化"，IEEE 机器人与自动化国际会议 (ICRA)，第 809-816 页，2016 .