

挑战:

NOT LIO+VIO 紧耦合

统一的地图

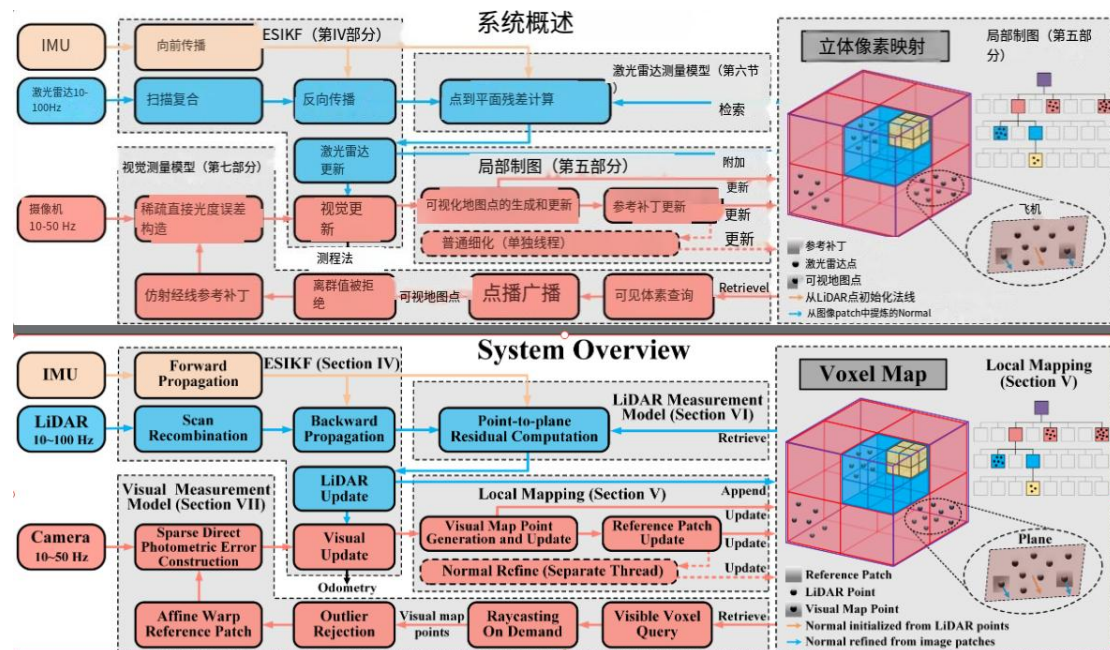
保障彩色点云: 像素级姿态估计, 硬件同步, 外参校准, 实时性, 曝光精度

整体:

IMU 先验, 首先 LIDAR 更新, 后图像更新, 基于单一统一的体素地图

将雷达地图点重新利用为视觉地图点 HOW?

稀疏图像对齐: 使用 LIDAR 平面先验, 无雷达点时按需进行体素光线投射



IMU 坐标系为载体坐标系

起点建立全局坐标系

实现雷达和相机同步：扫描重组 在同一时间更新状态

ESIKF 使用当前所有测量来更新数据，但 LIDAR 和图像感知方式不同，数据维度不匹配，采用**顺序更新策略**，即使用所有测量的标准更新，但要求给定状态向量的雷达和图像统计独立

公式 (5) 表明，总条件分布 可以通过两次顺序贝叶斯更新得到。第一步仅融合LIDAR测量 与IMU传播的先验分布 以获得分布：

$$p(\mathbf{x} | \mathbf{y}_l) \propto p(\mathbf{y}_l | \mathbf{x})p(\mathbf{x}) \quad (6)$$

第二步然后融合摄像头测量 与 以获得最终的状态后验分布：

$$p(\mathbf{x} | \mathbf{y}_l, \mathbf{y}_c) \propto p(\mathbf{y}_c | \mathbf{x})p(\mathbf{x} | \mathbf{y}_l) \quad (7)$$

有趣的是，两个融合步骤 (6) 和 (7) 遵循相同的形式：

$$q(\mathbf{x} | \mathbf{y}) \propto q(\mathbf{y} | \mathbf{x})q(\mathbf{x}) \quad (8)$$

为了进行公式 (8) 中的融合，无论是LiDAR还是图像测量，详细介绍先验分布 和测量模型 如下。对于先验分布，将其表示为 且 。在LiDAR更新的情况下（即第一步）， 和 是从传播步骤中获得的 状态和协方差。在视觉更新的情况下（即第二步）， 和 是从LiDAR更新中收敛得到的状态和协方差。

IMU->LIDAR->CAREMA

局部建图

叶体素代表局部平面，存储平面特征（平面中心、法线向量、不确定性）与平面上的雷达原始点，其中一部分带有三层图像补丁，称为视觉地图点，收敛的视觉地图点仅附有参考补丁，而未收敛的点附有参考补丁和其

他可见补丁。

超出局部地图范围的区域所占内存被重置以防止地图尺寸无限增长

局部到全局

对经过 ESIKF 的雷达更新之后，将所有点注册到全局坐标系，查找根体素，若无则新建哈希索引，若有则添加为叶体素。分配完所有扫描点后进行几何构建和更新

创建新体素后，确定雷达扫描点**是否在同一平面，通过奇异值分解实现**。如果是，则计算提取平面特征（中心点、法线向量、平面协方差矩阵——>代表由于姿态估计不确定性和噪声引起的平面不确定性）。如果不是，则划分更小体素，**直到位于同一平面或达到最大层数（可给定）**，若达到最大层数，则丢弃扫描点。故体素中只有平面体素

对于新添加的扫描点，确定与原先的点是否在同一平面，若不是则重新细分，若是则更新平面特征参数。**一旦平面收敛，则不再添加新点，并且固定平面特征参数。**

添加到平面上的雷达扫描点**将用于生成视觉地图点**，成熟平面中的**最近的 50 个雷达点**作为**生成视觉的候选点**，未成熟平面将**所有雷达点**作为候选。在候选点中选择一部分作为视觉地图点，并将**图像补丁**附加其上用于**图像对齐**。

视觉地图点生成和更新

在图像方面处理

什么特征？雷达的特征如何用于图像？点云如何作用于图像？如何

投影候选点？补丁包含什么，如何实现？

选取地图**当前帧视野中可见**的候选点，以及在图像中表现出**显著灰度梯度**的点。在视觉更新后将候选点投影到当前图像中，为每个体素的局部平面**保留深度最小的候选**。将图像划分为**网格单元**，**若无候选投影**，则将使用**灰度梯度最高的候选点**生成新的视觉地图点，**？**

并将当前图像补丁、估计的当前状态（帧姿态和曝光时间）与平面法线向量投影附加到该点。

图像补丁具有三个相同尺寸的图层，每层是上层的二分采样（会降低分辨率），形成补丁**金字塔**。**若有候选投影**，则在以下情况是为候选点添加图像补丁：**1.距上次添加补丁超过 20 帧 2.当前帧像素位置与上次添加补丁的像素位置偏离超过 40 像素。**经上述处理后，地图点可具备**观测视角均匀分布的有效图像补丁**。为地图点附加补丁和估计姿态，不附加平面法线。

参考补丁更新

由于视觉地图点可能具有多个补丁，需**选择一个参考补丁**用于视觉更新中的**图像对齐**。根据光度相似性和视角进行补丁评分。

$$\text{NCC}(\mathbf{f}, \mathbf{g}) = \frac{\sum_{x,y} [\mathbf{f}(x,y) - \bar{\mathbf{f}}][\mathbf{g}(x,y) - \bar{\mathbf{g}}]}{\sqrt{\sum_{x,y} [\mathbf{f}(x,y) - \bar{\mathbf{f}}]^2 \sum_{x,y} [\mathbf{g}(x,y) - \bar{\mathbf{g}}]^2}}$$

$$c = \frac{\mathbf{n} \cdot \mathbf{p}}{\|\mathbf{p}\|}, \quad \omega_1 = \frac{1}{1 + e^{\text{tr}(\Sigma_{\mathbf{n}})}}$$

$$S = (1 - \omega_1) \cdot \frac{1}{n} \sum_{i=1}^n \text{NCC}(\mathbf{f}, \mathbf{g}_i) + \omega_1 \cdot c$$

最佳补丁评分策略：

NCC 代表归一化系数，衡量补丁 \mathbf{f} 和 \mathbf{g} 在补丁金字塔最高层的相似性， $\bar{\mathbf{f}}$ 和 $\bar{\mathbf{g}}$ 代表补丁均值， c 代表法线向量 \mathbf{n} 和视角方向 $\frac{\mathbf{p}}{\|\mathbf{p}\|}$ 的余弦相似度。当补丁正对地图点所在平面时, c 为 1。总分由目标补丁 \mathbf{f} 和其他补丁 \mathbf{g} 的平均相似性和法线向量的余弦相似度加权求和得到， $\text{tr}(\Sigma_{\mathbf{n}})$ 代表法线向量协方差矩阵的迹。

通过此策略，可选择静态物体补丁 视角正交于平面的补丁，保持高分辨率

法线优化

普通算法假定补丁中的所有像素具有相同深度，误差较大

使用雷达计算法线，初始化平面，后进一步细化，提高对齐准确性

LIDAR 测量模型

模型构建

计算扫描点到最近平面的距离，如果点在平面上，可近似为点到平面法线的

垂直距离

在 ESIKF 中，通过最小化所有平面残差的平方和来估计位姿增量

平面残差？

配准

1. 使用上次的位姿估计将雷达点**变换到全局坐标**
2. 计算当前残差，构建测量残差向量，最小化残差以求解最优位姿增量
3. 通过迭代最小二乘法优化位姿增量，直至残差平方和收敛到设定阈值以下，每次迭代使用更新后的位姿重新计算残差并继续优化

测量不确定性

将平面残差的协方差纳入 ESIKF，假设测量噪声高斯分布。

协方差大小取决于**雷达测量的准确性**和**地图平面的几何形状**

可提高鲁棒性，增大精度

视觉测量模型

用历史地图补全当前视野

体素地图中提取当前帧的可见视觉地图点，确定体素是否在当前可见，提取可见点用于图像对齐

将选取点投影到当前帧，计算像素光度误差

最小化光度误差以优化位姿增量

金字塔由低到高对齐，底层粗对齐

为**减轻光照变化对对齐准确性的影响**，引入曝光时间补偿机制，根据实时更新的曝光时间调整图像光度值，在不同光照条件下**获得更加一致的光度测量值**，提高对齐准确性

在图像对齐中引入**平面先验**。假设视觉地图点位于某个地图平面上，引入 LIDAR 测得的平面特征来初始化。为姿态提供额外的几何约束，减少纹理或光照不足引起的误差。

光度误差、图像金字塔、曝光时间补偿、平面先验

实现高精度的视觉——IMU 组合

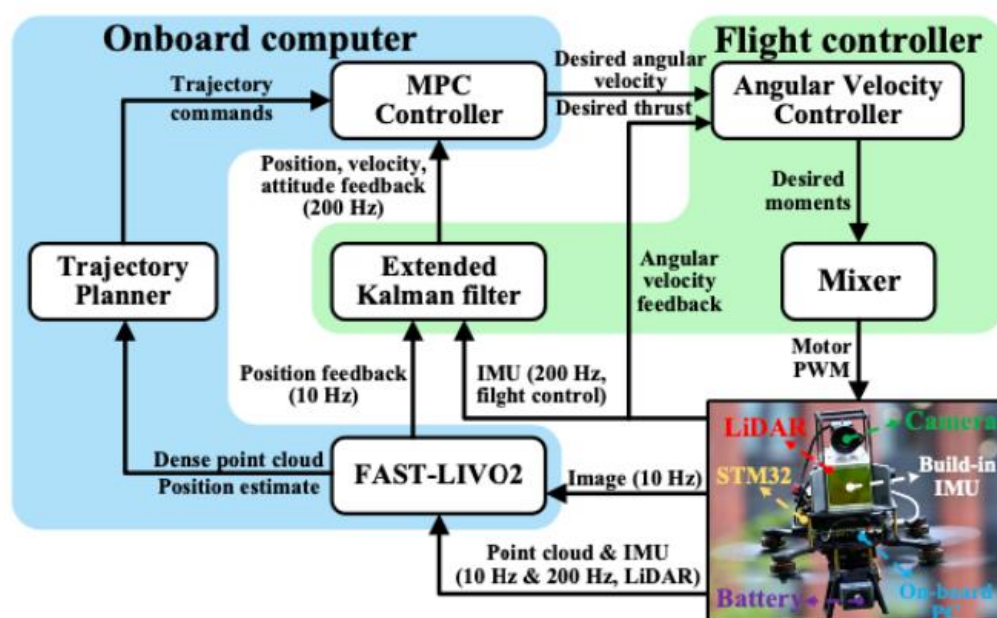


Fig. 12: Fully onboard UAV navigation algorithm flowchart.

摘要:

顺序更新以解决不匹配问题，直接融合 LIDAR 和相机，雷达不提取边缘或平面，直接注册原始点，视觉不提取 ORB 或 FAST 角点特征，最小化直接光度误差。二者融合基于统一的体素地图。雷达构建几何结构，注册新的扫描点，视觉为扫描点添加图像补丁，生成视觉地图点，实现图像对齐。通过平面先验以提高对齐准确性，在对齐后动态更新参考补丁。为增强对齐稳定性，采用按需光线投射，实时计算并补偿曝光时间。基于网格和 NERF 的 3D 模型渲染。

INTRODUCTION

LIVIO 处理大量点云和高分辨率图像，计算效率大

现有系统分开使用 LIO 和 VIO，基本都需要提取特征点，在某些环境下（结构、纹理缺失）时提取困难，工程适配难度大

如何建立统一的地图，同时包含相机和雷达的异构测量，降低计算需求，实现更紧密的集成，同时管理点云和高分辨率图像

为保证重建的彩色点云的精度，姿态估计需像素级精度。

先行工作：硬件同步，外参校准，曝光时间计算与回复，实时高精度融合策略

基于上述问题：

顺序更新的误差状态迭代 KF，IMU->LIDAR->CAREMA 使用直接法

相较于 LIVIO1:

顺序更新，添加 ESIKF，雷达点平面先验，选取最优图像补丁，计算补偿曝光时间，按需体素光线投射——激光雷达盲区问题

理论

直接法

相机：最小光度误差，雷达：点到平面残差

速度更快，需要相当准确的先验以避免陷入局部最小，视觉采用稀疏直接法，雷达的直接法通过在扫描中实时构建姿态优化的约束实现。

LIDAR 改自 VoxelMAP, 相机选取的关键点是雷达扫描点。

传感器融合

融合分类：状态估计层面和原始测量层面

一个传感器的估计是否是另一个的优化目标 原始数据是否结合

（一些系统（例如，DEMO[23]、LIMO[24]、CamVox[25]、[26]）使用 3D 激光雷达点为视觉模块[1, 4, 27]提供深度测量 分辨率不匹配，雷达点和特征点不能对应，通过深度关联中插值解决，但误差较大

原始数据紧耦合，状态估计松耦合）

松耦合：只提供数据，不联合优化（原始测量层面）

紧耦合：LVI R3LIVE（像素级）

顺序更新的 ESIKF

事先同步时间 IMU 的高斯噪声和随机游走

从前一时刻（接收到相机和雷达数据，雷达分割成相机采样时刻，以相机周期为准）获得状态和协方差，作为先验分布。后向传播以补偿失真，确保雷达扫描点在前面出现过。

标准 ESIKF 使用所有测量值，但相机和雷达**维度不匹配**，采用**顺序更新策略**。

假设雷达和相机测量在给定状态向量下统计独立。

地图结构和几何结构

体素地图添加 雷达扫描点注册到全局坐标系下 查找有无（确定其在哈希图中的根体素），只保留有限大小区域以避免无限内存溢出

雷达体素添加平面特征（中心点，法线，不确定性（姿态估计不确定性和测量噪声））

对于新创建的体素，确定新体素的点是否在同一平面，不在则继续八叉划分，最多划分三层并丢弃仍不共面的点

对有新点追加的体素，评估与原先点是否共面。如果不是则下划，如果是则直接更新平面参数。计算平面参数是否收敛，**若收敛则固定参数，丢弃新点**。成熟平面中最近（**空间**

OR 时间？）50 个点为视觉地图候选点；未成熟平面所有雷达点都是候选。

视觉地图点 顺序更新最后结果

选拔候选点要求：1.当前帧可见 2.在当前图像中灰度梯度显著

视觉更新后投影到当前图像（**如何实现？**），保留平面深度最小的候选点。

将图像划分为均匀的网格单元，**判断单元中是否有投影的地图点**。若无，则将灰度梯度最高的候选点生成新地图点，投影到图中，包括：（**雷达候选点+当前图像补丁+当前帧姿态和曝光时间+雷达点的平面法线向量**）
图像补丁为图像金字塔格式。

为视觉地图点添加图像补丁要求：1）20 帧之内没有添加过图像补丁 2）

地图点像素位置距离上次添加补丁的位置偏离超过 40 像素

可使得地图点具有均匀分布视角的图像补丁。为视觉点添加估计的当前状态（姿态和曝光时间）。

图像补丁更新

图像补丁=图像块+.....+.....

一个视觉点可能具有多个补丁，**选取一个用于视觉更新中的图像对齐。**

规则上述

选取 1) 与其他补丁最接近的补丁，**避免选择动态对象上的图像补丁**，2)

视图方向与平面垂直，可在高分辨率下保持纹理细节。

LIVO1 中直接选择与当前帧视图方向差距最小的图像补丁，非常接近当前帧，对当前姿态更新约束较弱

法线细化

现有方法假设图像补丁所有像素具有相同深度，误差较大

通过最小化与附加到同一视觉点的其他图像补丁的**光度误差**来细化图像补丁的平面法线

根据针孔或鱼眼相机实现不同的投影矩阵

一旦平面法线收敛（深度确定），该视觉地图点的参考图像补丁和法线向量将被固定，不再进一步细化，删除其他图像补丁

激光雷达模型

获得未畸变的扫描点后，将雷达更新的第 k 次迭代的先验估计投影到全局坐标系，确定该投影是否有体素。若没有或体素不包含平面，则丢弃该点；**否则使用体素平面建立雷达点的测量方程。**

然后我们确定 κ 次迭代中 $G\hat{p}_j^\kappa$ 所在的根体素或子体素。如果在哈希图中没有找到体素或体素不包含平面，则丢弃该点。否则，我们使用体素中的平面建立激光雷达点的测量方程。具体来说，我们假设真实的激光雷达点 Lp_j^{gt} ，在给定精确的激光雷达姿态 GT_I 的情况下，应该位于体素中具有法线 n_j^{gt} 和中心点 q_j^{gt} 的平面上。即，

$$0 = (n_j^{gt})^T (GT_I^I T_L Lp_j^{gt} - q_j^{gt}) \quad (18)$$

由于真实的地面点 Lp_j^{gt} 被测量为 Lp_j ，并且具有测距和方位噪声 $\delta^L p_j$ ，我们有 $Lp_j^{gt} = Lp_j - \delta^L p_j$ 。同样，平面参数 (n_j^{gt}, q_j^{gt}) 被估计为 (n_j, q_j) ，其协方差为 $\Sigma_{n,q}$ （第V-B节），因此我们有： $n_j^{gt} = n_j \oplus \delta n_j$ ， $q_j^{gt} = q_j - \delta q_j$ 。因此，

$$\underbrace{0}_{y_l} = \underbrace{(n_j \oplus \delta n_j)^T (GT_I^I T_L (Lp_j - \delta^L p_j) - (q_j - \delta q_j))}_{h_l(x, v_l)} \quad (19)$$

其中测量噪声 $v_l = (\delta^L p_j, \delta n_j, \delta q_j)$ 包括与激光雷达点、法线向量和平面中心相关的噪声。

考虑光束发散

；三种不确定性：TOF 测距不确定性 编码器方向方位不确定性 激光束发散角不确定性

视觉测量模型

视觉地图点用来计算图像对齐，即十四讲中只提取关键点，不计算描述子的直接法用以计算图像变化的关键点。LIVO2 使用稀疏直接法，只有标准的选择一部分像素点作为关键点。

选择合适的视觉点以执行稀疏图像对齐

提取当前视场内可见的地图点集（视觉子图），执行**查询体素和按需光线投射**，从点集中选择视觉地图点并排除异常值，产生精炼视觉点以用于构建视觉光度误差

查询获得可见体素：调查 **1）**当前扫描中雷达点涉及的体素（空间），

注意确保相机和雷达视场大部分重叠。则视觉地图点很有可能位于获得的体素中。**2）**调

查前一帧中通过相同体素查询和光线投射确定为可见地图点的体素

（前向传播类似，时间），假设两个连续图像帧有大的视场重叠。可通过**检查二者视场**获得当前视觉子图。

按需光线投射：当雷达过于靠近物体时存在近场盲区，可能没有返回点。若相机视场没有被雷达完全覆盖，要想办法得到更多的视觉地图点。将图像划分为均匀的网格单元，将从体素查询中得到的视觉地图点投影到网格地图中。若当前网格没有可用地图点，则沿着中心像素向后投射光线，其中**样本点**沿着光线在深度方向在 $[d_{min}, d_{max}]$ 之间均匀分布。预先计算每束光线上样本点在相机载体坐标系中的位置，对每个样本点评估相应体素（**样本点所在的体素，如果有的话**）的状态：若体素包含了执行后向投影之后位于该网格的视觉点（即包含视觉地图点），则将地图点纳入视觉子图并停止对此光线的处理；否则继续处理下一个样本点。可获得分布在整个图像中的视觉地图点。

异常值排除：

前两步得到的视觉地图点可能存在当前帧中被遮挡、深度不连续、图像补丁在大角度下拍摄或在当前帧中有大角度（形变大）等问题，可导致对齐精度差。

遮挡问题：使用激光雷达更新后的姿态，将所有视觉地图点投射到当前帧中，保留网格图像单元中的最低深度点。

深度不连续：使用雷达扫描中的雷达点投影到当前帧，产生深度图。比较视觉地图点和深度图中 9×9 领域的深度，确定遮挡和深度变化。

大角度问题：移除与参考补丁或当前补丁夹角（从点到补丁光学中心的法线向量和方向之间的夹角）过大（ 80° ）的点。

将视觉子图中的地图点使用雷达更新后的姿态投影到当前帧，保留网格地图中的最低深度点。

稀疏直接视觉测量模型

基本原理：将视觉地图点转到真实姿态的当前图像时，参考图像补丁和当前图像

补丁直接的光度误差应该为 **0**。

估算逆曝光时间