

# 激光雷达与相机自适应紧耦合的同时定位与建图算法

周维超<sup>1,2</sup>, 黄俊<sup>1,2\*</sup>

<sup>1</sup>中国科学院上海高等研究院, 上海 201210;

<sup>2</sup>中国科学院大学, 北京 100049

**摘要** 同时定位与建图(SLAM)是自动驾驶的基本要求之一。多传感器融合,尤其是激光雷达和相机的融合,对于自动驾驶来说是必不可少的。其中,如何针对各种场景调整不同传感器的置信度是关键问题,基于此,提出一种自适应紧耦合的激光雷达相机融合的SLAM(AVLS)算法。首先,所提AVLS算法建立在基于滑动窗口的因子图上,包含提升整体算法精度和鲁棒性的灵活深度关联和弹性初始化等模块。其次,为了充分探索激光雷达和相机在不同环境中的性能,采用一种基于先验知识的动态加权方案。最后,将所提AVLS算法在两个公开的大规模自动驾驶数据集上进行了全面实验,包括与经典算法的对比及消融实验,实验结果表明,AVLS算法的鲁棒性和精确度可以达到目前领先水平。

**关键词** 激光雷达; 同时定位与建图; 传感器融合; 自主导航; 稀疏位姿优化

中图分类号 TP242.6

文献标志码 A

DOI: 10.3788/LOP223209

## Adaptive Tightly Coupled Lidar-Visual Simultaneous Localization and Mapping Framework

Zhou Weichao<sup>1,2</sup>, Huang Jun<sup>1,2\*</sup>

<sup>1</sup>Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China;

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

**Abstract** Simultaneous localization and mapping (SLAM) is one of the basic requirements of autonomous driving. Furthermore, multisensor fusion, more particularly, the fusion of lidar and camera, is essential for autonomous driving, and understanding how to adjust the weights of different sensors for various scenarios is a critical challenge. Therefore, an adaptive tightly coupled lidar-visual SLAM (AVLS) algorithm is proposed. First, AVLS is built on a factor graph based on sliding windows, including modules such as flexible depth correlation and elastic initialization that improve the accuracy and robustness of the overall algorithm. Second, in order to fully explore the performance of lidars and cameras in different environments, a dynamic weighting scheme based on prior knowledge is adopted. Finally, comprehensive experiments are conducted on the proposed AVLS algorithm on two publicly available large-scale autonomous driving datasets, including comparisons with classical algorithms and ablation experiments. The experimental results show that the robustness and accuracy of the AVLS achieves state-of-the-art performance.

**Key words** lidar; simultaneous localization and mapping; sensor fusion; autonomous navigation; sparse pose optimization

## 1 引言

同时定位与建图(SLAM)是自动驾驶汽车需要具备的基本能力之一。在过去的几十年中,激光雷达与相机传感器取得了深入的发展<sup>[1-2]</sup>。基于激光雷达或相机的单传感器SLAM算法<sup>[3-5]</sup>已得到充分研究,可以

在相对稳定的环境中实现准确和鲁棒的性能。但是在很多具有挑战性的场景中,单传感器SLAM具有明显的缺陷。激光雷达SLAM如LeGO-LOAM<sup>[6]</sup>通常需要根据曲率提取特征,导致其在长廊和开阔地面场景这两种典型环境中性能不佳,因为在这些环境中提取足够多的有效特征具有挑战性。此外,文献[7]中提

收稿日期: 2022-11-30; 修回日期: 2023-01-16; 录用日期: 2023-02-06; 网络首发日期: 2023-03-09

基金项目: 国家重点研发计划基金资助项目(2019YFC1521204, 2020YFC1523202)

通信作者: \*huangj@sari.ac.cn

到,视觉SLAM受制于光照和运动条件。光照条件差可能会导致特征提取不正确,快速运动会导致初始化不当和位姿丢失。为了提高自动驾驶的鲁棒性和精度,激光雷达和相机融合的SLAM算法在过去几年中逐渐引起了研究人员的关注。

现有的融合方法可以分为松耦合方法和紧耦合方法。松耦合方法主要基于时间序列,采用滤波器进行融合<sup>[8-10]</sup>。松耦合方法,如DEMO<sup>[11]</sup>和LIMO<sup>[12]</sup>稳定性高,但过于依赖主传感器,无法保证融合后的性能优于单个传感器。紧耦合方法主要采用图优化的技术,并且在大多数情况下,它们的性能优于松耦合方法<sup>[13-14]</sup>。但是现有的大部分紧耦合方法都忽略了传感器性能与环境的关联,不能动态地调整激光雷达和相机的权重以实现更好的性能。因此,能够动态适应环境的激光雷达相机紧耦合SLAM算法值得研究。

基于此,本文提出一种自适应紧耦合的激光雷达与相机融合框架(AVLS)来解决上述问题。首先,AVLS构建于基于滑动窗口的因子图上,融合了激光雷达和相机数据以实现准确和鲁棒的性能。其次,具

有准确深度信息的激光雷达点云被用来提升对应视觉特征的深度精确度。此外,还探讨了融合过程中不同传感器在不同环境中的置信度问题,并提出一种动态加权方案,根据每一帧的激光雷达点云和相机图像的质量动态调整因子图中视觉因子和激光雷达因子的权重,有效提升了融合框架的性能。

## 2 基本原理

所提AVLS算法框架如图1所示。首先,输入包括激光雷达点云和相机图像数据,两者被并行预处理从而实时提取激光雷达与视觉特征。提取的特征存储在本地图中,然后通过深度关联模块为视觉特征分配准确的深度。其次,待优化位姿的初始值由弹性初始化模块生成,并且使用基于滑动窗口的因子图来计算位姿。此外,激光雷达因子和视觉因子的权重根据激光雷达点云和相机图像的质量动态调整。最后,使用scan-context(SC)<sup>[15]</sup>进行回环检测以减少全局漂移。接下来依次介绍算法框架中各个模块的细节。

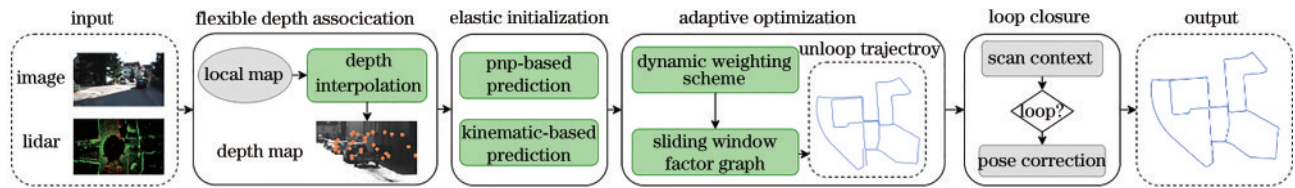


图1 所提算法流程图

Fig. 1 Flow chart of proposed algorithm

### 2.1 输入数据预处理

假设激光雷达和相机的内部参数和外部参数是已知且固定的,并且它们的畸变已经被校正。根据LOAM<sup>[5]</sup>中定义的曲率提取激光雷达特征,曲率较高的激光雷达特征定义为激光雷达边缘特征,曲率较低的激光雷达特征定义为激光雷达平面特征。使用Shi-Tomasi检测视觉特征,并通过KLT稀疏光流算法进行跟踪,最后执行具有基本矩阵模型的随机采样一致算法(RANSAC)来排除异常值<sup>[16]</sup>。所有提取出的激光雷达特征和视觉特征都在局部地图中注册方便后续的处理。

### 2.2 灵活深度关联

深度关联模块是为了更准确地为视觉特征分配深度而设计的,如图2所示,激光雷达点云和视觉特征被投影在同一个单位球面上。

对于每个视觉特征,可以通过KD-tree<sup>[17]</sup>选择3个最近的激光雷达点。根据这些点的深度,进行一次验证来提高后续匹配的准确性。具体的验证过程是计算当前视觉特征最近的3个激光雷达点的欧氏空间距离。如果3点之间的最远距离小于0.5 m,可以通过双线性插值计算深度。否则,应用三角测量来为视觉特征分配深度。图2显示了验证的具体过程。验证过程

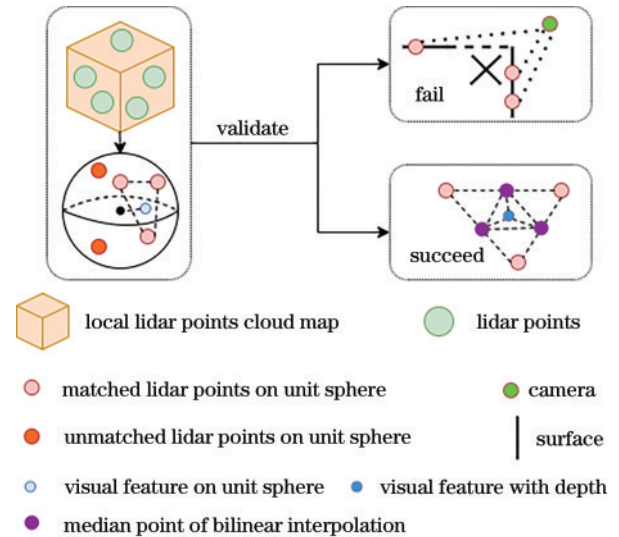


图2 灵活深度关联示意图

Fig. 2 Illustration of flexible depth association

的目的是检查这些点是否在同一平面上,如果深度相差过大,则这些点可能在不同的平面上,因此需要排除,从而减少错误匹配的可能性。

### 2.3 弹性初始化

因子图优化的本质是非线性最小二乘优化,合适

的初始值  $I$  对最终优化结果有很大帮助。视觉 SLAM 的初始值  $I_V$  通常由 perspective-n-point (PnP) 计算, 激光雷达 SLAM 的初始值  $I_L$  通常由运动学模型给出, 例如匀速运动模型。

根据先验知识以及相关实验的结果可知, 如果有足够多的视觉特征, 基于 PnP 的初始值比基于运动学公式的初始值更准确。因此, 弹性初始化可以描述为

$$I = \begin{cases} I_V, & n \geq c \\ I_L, & n < c \end{cases}, \quad (1)$$

式中:  $n$  是当前状态下好的视觉特征的数量, 为了后续在同一实验条件下与 VINS 等算法进行对比, 沿用 VINS 的设置, 将其定义为 Shi-Tomasi 点的数量;  $c$  是阈值, 设置为 200。

## 2.4 自适应优化

自适应优化旨在计算当前状态  $X$ , 其表达式为

$$X = [R, t]^T, \quad (2)$$

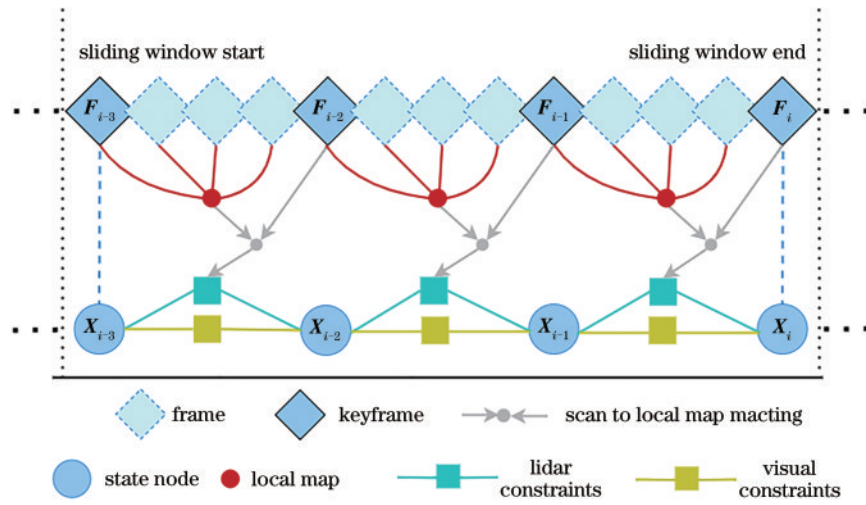


图 3 自适应的滑动窗口因子图

Fig. 3 Adaptive sliding window factor graph

### 2.4.1 视觉因子和激光雷达因子

采用重投影残差来描述视觉因子。如果在第  $m$  个关键帧中观察到第  $i$  个视觉特征, 则可以将重投影残差定义为

$$r_{c_i} = s_i \cdot u_i - KX_m P_i^o, \quad (4)$$

式中:  $u_i$  是像素坐标;  $s_i$  是对应的深度;  $K$  是相机投影矩阵;  $X_m$  是第  $m$  帧的状态;  $P_i^o$  表示视觉特征在世界坐标中的 3D 位置。

可以根据激光雷达边缘特征和激光雷达平面特征将残差分为两种:

激光雷达边缘特征残差为

$$r_{l_j}^{\text{edge}} = \frac{|(P_j^{\text{edge}} - P_u^{\text{edge}}) \times (P_j^{\text{edge}} - P_v^{\text{edge}})|}{|P_u^{\text{edge}} - P_v^{\text{edge}}|}, \quad (5)$$

式中:  $P_j^{\text{edge}}$  是当前激光雷达边缘特征;  $P_u^{\text{edge}}$ 、 $P_v^{\text{edge}}$  是离  $P_j^{\text{edge}}$  的两个最近的边缘特征。

式中:  $R$  表示旋转矩阵;  $t$  是平移向量

当前帧的状态估计问题可以建模为最大后验 (MAP) 问题。在高斯噪声的假设下, MAP 问题求解等价于非线性最小二乘问题计算:

$$\operatorname{argmin}_x \left\{ \sum_{i \in c} \omega_{c_i} \rho(\|r_{c_i}\|^2) + \sum_{j \in l} \omega_{l_j} \rho(\|r_{l_j}\|^2) \right\}, \quad (3)$$

式中:  $r_{c_i}$  和  $r_{l_j}$  分别是视觉和激光雷达因子的残差;  $\rho(\cdot)$  是 Huber 损失函数;  $\omega_{c_i}$  和  $\omega_{l_j}$  分别是视觉和激光雷达的自适应权重。

用来求解位姿的基于滑动窗口的因子图如图 3 所示。窗口大小是固定的, 关键帧的选择策略与 VINS 相同, 将非关键帧的激光雷达点云合并到局部地图中从而与关键帧的点云进行扫描配准。因子图中有两种因子, 分别是视觉因子和激光雷达因子, 两者都通过动态加权方案实时地调整置信度。接下来分别介绍视觉因子、激光雷达因子和动态加权方案。

激光雷达平面特征残差为

$$r_{l_j}^{\text{surface}} = \frac{|(P_j^{\text{surf}} - P_u^{\text{surf}})|}{|(P_u^{\text{surf}} - P_v^{\text{surf}}) \times (P_u^{\text{surf}} - P_w^{\text{surf}})|}, \quad (6)$$

式中:  $P_j^{\text{surf}}$  是当前激光雷达平面特征;  $P_u^{\text{surf}}$ 、 $P_v^{\text{surf}}$ 、 $P_w^{\text{surf}}$  是离  $P_j^{\text{surf}}$  最近的 3 个平面特征。

在实验中, 如果当前状态有过多的平面残差时, 里程计会显示严重的漂移。因此, 平面残差的数量受  $a < 2b$  的限制。其中,  $b$  是边缘残差的数量,  $a$  是平面残差的数量。

### 2.4.2 动态加权方案

相机或激光雷达数据在具有挑战性的环境 (例如低纹理和开阔地面场景) 中会产生异常的数据, 这可能导致不准确的位姿估计。因此, 实时调整传感器权重是必不可少的。



受 GR-SLAM<sup>[18]</sup> 的启发,设计了一种基于当前状态下激光雷达点云和相机图像质量的动态加权方案,即式(7)。相比于相机,激光雷达不易受外界干扰,因此激光雷达的短期位姿估计结果更准确,所以采用激光雷达位姿估计的增量值作为参考来计算先验因子  $f_{\text{prior}}$ ,  $f_{\text{prior}}$  的分子由激光雷达和视觉的位姿增量之差计算得到。较大的  $f_{\text{prior}}$  表示视觉的位姿预测增量发生了较大的偏移,即意味着视觉中可能存在不可靠的运动,需要减小视觉因子的权重;反之同理。权重比因子  $\frac{\omega_{c_i}}{\omega_{l_i}}$  采用平方的形式,可以更敏锐地感知环境。在开阔平坦的场景中,由于没有足够的激光雷达边缘和平面特征,激光雷达因子的性能并不理想。因此,引入地面特征数量  $N_g$  作为补充因子来调整传感器的权重。

AVLS\_weight:

$$\left\{ \begin{array}{l} f_{\text{prior}} = \left( \frac{\| \mathbf{P}_{l_i l_{i+1}}^C \| - \| \mathbf{P}_{l_i l_{i+1}}^L \|}{\mathbf{P}_{l_i l_{i+1}}^L} \right)^2 \\ \frac{\omega_{c_i}}{\omega_{l_i}} = \left\| \frac{(a \times N_g + b \times N_v)}{(c \times N_c + d \times N_s)} \right\|^2 \times |1 - f_{\text{prior}}| \end{array} \right. , \quad (7)$$

式中:  $f_{\text{prior}}$  是先验运动因子;  $\omega_{c_i}$  是视觉因子的权重;  $\omega_{l_i}$  是激光雷达边缘和平面因子的权重;  $\| \mathbf{P}_{l_i l_{i+1}}^C \|$ 、 $\| \mathbf{P}_{l_i l_{i+1}}^L \|$  分别是基于 PnP 和基于运动学的预测计算的位姿增量值; 由于不同特征点的数量存在一定的差异,且在类似的环境中,数量比相对稳定,因此定义常数  $a$ 、 $b$ 、 $c$ 、 $d$  来调整不同特征点的数量比;  $N_g$ 、 $N_v$ 、 $N_c$  和  $N_s$  分别是地面点数量、视觉特征数量、激光雷达边缘特征数量和激光雷达平面特征数量。实验中设  $a = 0.5$ 、 $b = 0.5$ 、 $c = 0.7$ 、 $d = 0.3$ 、 $\omega_{l_i} = 1$ 。在实验中,由于激光雷达因子权重参数  $\omega_{l_i}$  设置为常数 1,因此权重调整因子  $\frac{\omega_{c_i}}{\omega_{l_i}}$  所计算得出的值等于视觉因子的权重。

## 2.5 回环检测及优化

回环检测的目的是减少累积漂移。随着时间的推移,不可避免会出现轨迹漂移,回环检测可以有效地缓解这个问题。回环检测是耗时的,因此关键帧对于实时性能是必要的。在实验中,当运动距离或运动时间超过阈值时,将当前帧视为关键帧,所有关键帧都存储在一个位姿图中以供进一步优化。

对于每个关键帧,使用 SC 作为全局描述符。与其他激光雷达或视觉描述符相比,SC 对旋转变化的鲁棒性,此外应用 ISC-LOAM<sup>[19]</sup> 的几何一致性验证策略来检查候选帧的相似性。实验中基于 GTSAM<sup>[20]</sup> 构建位姿图进行回环约束优化,将相似度最高的候选帧定义为回环候选对,在候选对之间通过 iterative closest point (ICP) 约束构建边,并将其添加到位姿图中以矫正漂移。

## 3 实验结果分析与讨论

通过基于 KITTI 数据集<sup>[21]</sup> 和 nuScenes 数据集<sup>[22]</sup> 的一系列实验来验证所提 AVLS 算法的性能。KITTI 数据集是目前自动驾驶领域规模最大的公开数据集之一,其中有 11 个序列包含地面真值(序列 00~10),总长共计 22 km,环境丰富多样,包含乡村、城市、高速公路以及其他混合场景。为了公平地评估鲁棒性和准确性,所提框架在所有具有地面真值的 KITTI 序列进行了验证。这些序列的数据输入仅包含激光雷达和图像数据,并且激光雷达和摄像头的频率已预同步到 10 Hz。为了丰富实验数据集的场景,nuScenes 是补充的另一个数据集。nuScenes 数据集的规模与 KITTI 数据集的规模接近,区别在于 nuScenes 数据集包含更多的高动态场景,这对于自动驾驶来说是非常具有挑战性的。与处理 KITTI 数据集的方式类似,对于 nuScenes 数据集,实验中仅使用激光雷达和图像的数据作为输入,并在数据输入之前执行时间同步。采用 KITTI 数据集提供的轨迹评估工具评估 KITTI 数据集的结果。由于 nuScenes 的轨迹很短,因此使用均方根误差(RMSE)作为另一个评估指标,由 EVO<sup>[23]</sup> 计算得出。

### 3.1 KITTI 数据集的实验结果

输入为 KITTI 数据集的双目相机图像和激光雷达点云。KITTI 数据集的评测指标主要为平均平移误差(A TE),辅助指标为绝对旋转误差(A RE),单位为  $10^{-2}(\circ)/\text{m}$ 。VINS 是 KITTI 数据集里表现出色的一个视觉 SLAM,F-LOAM<sup>[24]</sup> 是 KITTI 榜单中性能最好的激光雷达里程计之一。因此实验选择 VINS 和 F-LOAM 作为基线来验证所提 AVLS 算法的有效性。此外,还补充了与其他配备回环的最先进 SLAM 的比较结果,以进一步说明 AVLS 的精确度和鲁棒性。

#### 3.1.1 里程计定位精度评估

里程计是 SLAM 的重要组成部分,其精度主要反映在无回环轨迹的精度。这里将 AVLS 的里程计性能与 VINS 和 F-LOAM 进行比较,三者都具有准确和鲁棒的里程计。定量比较结果如表 1 和表 2 所示。其中,AVLS-VLO 指的是 AVLS 框架的里程计,VLO 指的是 AVLS 框架去除动态加权方案后的里程计。相比于基线算法,AVLS-VLO 获得了最佳性能,其 ATE 为 1.12%,ARE 为  $0.48 \times 10^{-2}(\circ)/\text{m}$ 。

为了进一步分析所提框架的优势,选择序列 00、02、07 进行定性分析,对比结果如图 4 所示。在这些场景中,AVLS 的性能优于 F-LOAM 和 VINS。与基线相比,AVLS-VLO 的精度提高在海拔轴(y 轴)上很明显,这主要得益于自适应优化策略。灰色矩形框标记了 AVLS-VLO 具有明显改进效果的区域,F-LOAM

表 1 里程计定位精度平移误差指标评测

Table 1 Translational error index of odometry localization accuracy unit: %

Seq. No.	00	01	02	03	04	05	06	07	08	09	10	Average
VINS	1.30	2.30	1.43	1.48	1.33	1.38	1.33	1.07	1.75	1.78	1.81	1.52
F-LOAM	0.98	2.20	1.29	1.52	0.81	0.74	<b>0.54</b>	1.22	1.37	<b>1.00</b>	1.81	1.16
AVLS-VLO	<b>0.97</b>	<b>1.21</b>	<b>1.27</b>	<b>1.45</b>	<b>0.60</b>	<b>0.65</b>	1.04	<b>0.52</b>	<b>1.28</b>	1.28	<b>1.74</b>	<b>1.12</b>

表 2 里程计定位精度旋转误差指标评测

Table 2 Rotational error index of odometry localization accuracy unit:  $10^{-2} (^{\circ})/m$

Seq. No.	00	01	02	03	04	05	06	07	08	09	10	Average
VINS	0.63	0.35	0.53	<b>0.49</b>	0.70	0.71	0.69	0.83	0.67	0.53	0.85	0.61
F-LOAM	0.60	0.56	0.65	0.91	<b>0.45</b>	0.55	<b>0.33</b>	1.03	0.56	<b>0.47</b>	0.85	0.60
AVLS-VLO	<b>0.51</b>	<b>0.30</b>	<b>0.47</b>	0.85	0.47	<b>0.40</b>	0.36	<b>0.46</b>	<b>0.48</b>	0.50	<b>0.73</b>	<b>0.48</b>

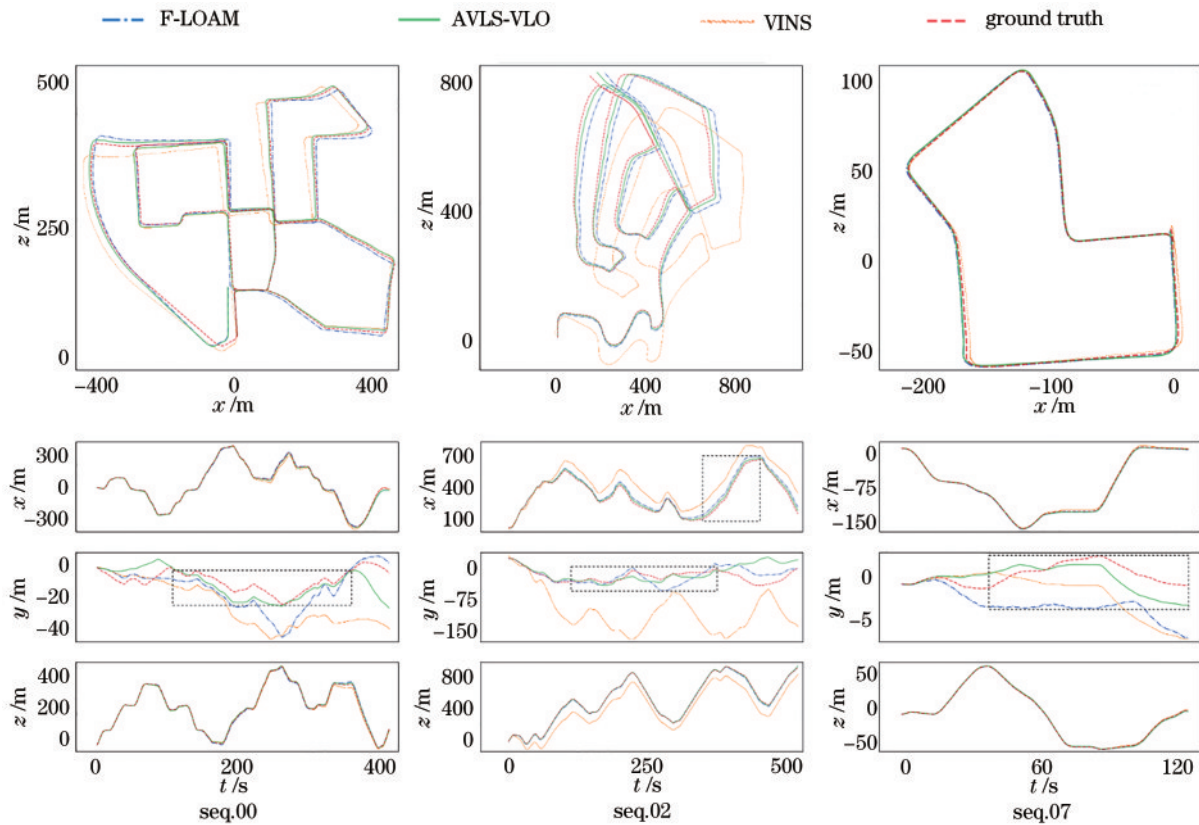


图 4 里程计轨迹和地面真值的对比图

Fig. 4 Comparison of odometer trajectory and ground truth

和 VINS 均发生了明显的漂移,但 AVLS-VLO 持续保持和地面真值轨迹的拟合,证明所提算法具备极高的鲁棒性。

### 3.1.2 自适应权重研究及消融实验

为了全面验证所提动态加权方案的有效性,进行了自适应加权研究,并设计了一个指标来更准确地表达权重:

$$\omega_{\text{Avg}} = \text{Avg}(\omega_{c_i}/\omega_{l_j}) = \frac{\sum_{i \in F} (\omega_{c_i}/\omega_{l_j})}{N_F}, \quad (8)$$

式中:  $F$  是序列  $k (0 \leq k \leq 11)$  中所有帧的集合;  $N_F$  是集合  $F$  的大小。

此外,补充几种权重调整方案进行消融实验,分别记为 VLO\_1、VLO\_2、VLO\_3,方案整体形式和所提 AVLS 的方案一样,区别在于权重比因子  $\frac{\omega_{c_i}}{\omega_{l_j}}$  不同:

$$\frac{\omega_{c_i}}{\omega_{l_i}} = \begin{cases} \left\| \frac{a \times N_g + b \times N_v}{c \times N_c} \right\|^2 \times |1 - f_{\text{prior}}|, & \text{VLO\_1} \\ \left\| \frac{a \times N_g + b \times N_v}{d \times N_s} \right\|^2 \times |1 - f_{\text{prior}}|, & \text{VLO\_2} \\ \left\| \frac{a \times N_g}{c \times N_c + d \times N_s} \right\|^2 \times |1 - f_{\text{prior}}|, & \text{VLO\_3} \end{cases} \quad (9)$$

动态加权消融实验的定量比较结果如表 3 所示。AVLS-VLO 代表所提动态加权方案, VLO 代表没有动态加权的方案, VLO\_1 代表不考虑激光平面特征影响的动态加权方案, VLO\_2 代表不考虑激光边缘特征影响的动态加权方案, VLO\_3 代表不考虑视觉特征影响的动态加权方案。AVLS\_VLO 获得了最佳的性能, 充分说明了 AVLS 中的动态权重调整方案的有效性。

表 3 动态加权方案消融实验对比结果(ATE/ARE)

Table 3 Ablation study of the dynamic weighting scheme (ATE/ARE)

unit: %/[10<sup>-2</sup>(°)/m]

Seq. No.	VLO_1	VLO_2	VLO_3	VLO	AVLS_VLO
00	0.98/0.51	0.96/0.51	0.97/0.51	1.08/0.55	0.97/0.51
01	1.23/0.29	1.32/0.37	1.50/0.36	1.02/0.23	1.21/0.30
02	1.24/0.45	1.28/0.47	1.27/0.47	1.38/0.50	1.27/0.47
03	1.42/0.85	1.42/0.85	1.70/0.92	1.28/0.59	1.45/0.85
04	1.11/0.90	0.66/0.51	0.61/0.48	0.68/0.56	0.60/0.47
05	0.63/0.39	0.63/0.39	0.66/0.40	0.77/0.44	0.65/0.40
06	1.42/0.68	0.99/0.39	0.70/0.31	0.69/0.39	1.04/0.36
07	1.42/0.68	0.51/0.43	0.54/0.47	0.64/0.59	0.52/0.46
08	1.29/0.48	1.28/0.48	1.28/0.48	1.42/0.55	1.28/0.48
09	3.35/1.09	1.44/0.59	1.25/0.50	1.50/0.54	1.28/0.50
10	1.50/0.60	1.67/0.69	4.12/0.73	1.57/0.60	1.74/0.73
Average	1.24/0.50	1.13/0.49	1.19/0.51	1.20/0.50	<b>1.12/0.48</b>

表 4 显示了 KITTI 数据集中不同序列的差异条件。Env 记录了各个序列的真实环境类别, U 代表城市环境, H 代表高速公路环境, C 代表乡村环境, U+C 代表城市乡村混合环境。V<sub>Avg</sub> 代表对应于各个序列中车辆的平均速度。ω<sub>Avg</sub> 是自定义描述权重的因子,

由式(8)给出。Seq. 01 的 ω<sub>Avg</sub> 是最大的, 因为 Seq. 01 是高速公路环境, 是一个开阔平坦的场景。Seq. 03 的 ω<sub>Avg</sub> 最小, 因为在该场景中可以提取足够多的激光雷达边缘和表面特征。

表 4 KITTI 数据集环境差异对比

Table 4 Comparison of environmental differences in KITTI dataset

Seq. No.	00	01	02	03	04	05	06	07	08	09	10
Env	U	H	U+C	C	C	U	U	U	U+C	U+C	U+C
ω <sub>Avg</sub>	1.3	2.5	2.0	0.8	2.1	1.1	1.9	0.9	1.1	1.4	1.5
V <sub>Avg</sub> /(m/s)	7.9	21.5	10.5	6.8	14.0	7.7	10.8	6.1	7.6	10.3	7.4

当 ω<sub>Avg</sub> 在 0.9 和 2.2 之间, 并且移动速度相对温和的情况下, 例如序列 00、02、04、05、07、08、09, 使用动态加权方案的结果更好。与没有动态权重方案的定位精度相比, AVLS 的平移误差从 1.20% 缩小到 1.12%, 旋转误差从 0.50 × 10<sup>-2</sup>(°)/m 缩小到 0.48 × 10<sup>-2</sup>(°)/m, 这表明了所提出的动态加权方案的有效性。

### 3.1.3 与现有完整 SLAM 的性能对比

回环检测对于 SLAM 系统来说是必不可少的。与其他具有回环检测及后端优化功能的 SLAM 系统相比, AVLS 取得了最优的性能。比较结果如表 5 所示, VINS\* 指的是 VINS 带有回环检测的结果, 用来和表 1 和表 2 中不带有回环检测的 VINS 结果作区分,

DEMO 的结果直接引用原论文中的数据, LeGO-LOAM, VINS\*, ISC-LOAM, LITAMIN<sup>[25]</sup> 和 AVLS 的结果均为本机实测给出。从表 5 可以看出, 由于 KITTI 的复杂环境, 多个算法发生明显的失效现象, LeGO-LOAM 在序列 01 中的 ATE 达到了 28.03, 意味着该算法在序列 01 的情况下每百米行驶距离发生 28 m 的漂移, 即该算法在序列 01 的场景下完全失效, 因此实验中不计算其平均精度值, LITAMIN 在序列 01 和序列 04 的情况下, 也发生了类似的失效现象, 在上述导致基线算法失效的全部场景中, AVLS 均实现了最优的性能, 兼具了强鲁棒性与高精度度。在整个 KITTI 数据集中, AVLS 取得了最佳结果, 其 ATE 为 1.01%, ARE 为 0.42 × 10<sup>-2</sup>(°)/m, 优于其他先进的 SLAM 系统。



表 5 带有回环的完整 SLAM 系统的性能评测 (ATE/ARE)

Table 5 Performance evaluation of a complete SLAM system with loopback (ATE/ARE) unit: %/[ $10^{-2}$ (°)/m]

Seq. No.	Length /m	LeGO-LOAM	VINS*	DEMO	ISC-LOAM	LITAMIN	AVLS
00	3724	1.38/0.57	1.18/0.63	1.05/	1.02/0.42	0.95/0.41	<b>0.88</b> /0.38
01	2453	28.03/2.69	2.16/0.34	1.87/	2.92/0.63	11.3/0.45	<b>1.17</b> /0.28
02	5067	2.14/0.84	1.72/0.72	<b>0.93</b> /	1.67/0.54	1.25/0.45	1.08/0.44
03	560	1.21/0.79	1.46/0.49	<b>0.99</b> /	1.15/0.72	1.17/0.56	1.17/0.52
04	393	1.27/0.84	1.33/0.69	1.23/	1.50/0.56	18.7/0.47	<b>0.62</b> /0.47
05	2205	0.91/0.48	1.40/0.67	1.04/	0.81/0.37	0.70/0.35	<b>0.49</b> /0.32
06	1232	0.80/0.43	0.86/0.56	0.96/	0.76/0.41	0.63/0.32	<b>0.61</b> /0.31
07	694	0.74/0.54	0.67/0.54	1.16/	0.56/0.43	<b>0.45</b> /0.33	0.53/0.44
08	3222	1.40/0.54	1.73/0.67	1.24/	1.20/0.50	<b>1.03</b> /0.37	1.32/0.52
09	1705	1.25/0.60	1.85/0.56	1.17/	1.40/0.59	<b>1.06</b> /0.43	1.29/0.45
10	919	1.70/0.69	1.81/0.85	<b>1.14</b> /	1.87/0.62	3.78/0.90	1.63/0.65
Average			1.56/0.66	1.16/	1.35/0.52	1.59/0.43	<b>1.01</b> /0.42

### 3.2 nuScenes 数据集的实验结果

为了丰富实验并进一步验证所提 AVLS 算法框架,采用 nuScenes 数据集作为补充实验场景。nuScenes 数据集的规模与 KITTI 数据集类似,区别在于场景中包含更多的动态对象,这对于自动驾驶的定位鲁棒性提出了巨大的挑战。考虑到 nuScenes 数据集中的序列以短程城市场景为主,单个序列中均不包含重复访问区域,因此本实验中不启用回环检测模块,即

仅对里程计精度进行评测。由于 nuScenes 仅包含单目相机,并且其相机设置无法构成双目相机,为了比较的公平性,VINS 被排除在本次评测之外。测试集是 nuScenes 的一个子集,包含 150 个小场景序列。测试集一半的场景在序列号 500 之前,另一半在序列号 500 之后。为了公平性,以序列 500 为分界线,从中各随机抽取 20 个序列。实验结果如表 6 所示。与 F-LOAM 相比,AVLS 在这些序列中取得了更好的性能。

表 6 定位精度评测 (RMSE)

Table 6 Evaluation of localization accuracy (RMSE)

unit: m

Seq. No.	77	80	84	89	111	115	119	144	265	279
F-LOAM	<b>0.68</b>	1.01	0.81	2.58	2.45	<b>0.82</b>	1.54	0.91	1.42	1.32
AVLS	1.25	<b>0.58</b>	<b>0.62</b>	<b>0.54</b>	<b>0.98</b>	1.02	<b>1.16</b>	<b>0.86</b>	<b>0.78</b>	<b>0.99</b>
Seq. No.	281	309	311	314	335	343	481	486	489	494
F-LOAM	1.24	<b>4.53</b>	1.03	<b>0.63</b>	<b>0.74</b>	0.84	<b>0.73</b>	1.06	1.37	0.88
AVLS	<b>0.84</b>	4.88	<b>0.51</b>	1.01	1.13	<b>0.81</b>	0.77	<b>0.68</b>	<b>1.00</b>	<b>0.39</b>
Seq. No.	547	549	604	610	615	621	827	830	834	841
F-LOAM	<b>1.26</b>	1.11	2.45	1.06	1.10	0.95	4.54	2.16	1.77	<b>1.52</b>
AVLS	1.50	<b>0.42</b>	<b>1.59</b>	<b>0.84</b>	<b>0.34</b>	<b>0.57</b>	<b>4.52</b>	<b>1.61</b>	<b>1.52</b>	1.61
Seq. No.	933	939	941	1027	1028	1030	1032	1037	1039	1043
F-LOAM	3.60	1.38	0.93	1.59	<b>1.11</b>	1.67	0.81	<b>1.76</b>	1.52	<b>1.72</b>
AVLS	<b>3.46</b>	<b>1.29</b>	<b>0.68</b>	<b>1.10</b>	1.41	<b>1.08</b>	<b>0.32</b>	1.97	<b>0.98</b>	1.80

## 4 结 论

提出一种激光雷达与相机自适应紧耦合的 SLAM 算法框架,实现了鲁棒和准确的性能。通过加入灵活深度关联、弹性初始化和自适应优化等模块,缓解了目前激光雷达与相机紧耦合的 SLAM 领域中存在的无法根据环境动态调整传感器置信度的问题。为了提升视觉特征点深度的精度,文中设计的灵活深度关联方案,使得视觉特征点可以更准确地获取深度值。此外为了提升因子图非线性优化的性能,设计了一种弹性初始化方案。为了使激光雷达和相

机在动态环境中发挥最佳性能,设计了一种动态权重调整方案。为了验证所提出的算法框架的鲁棒性和准确性,在 KITTI 数据集和 nuScenes 数据集上进行了大量的实验。结果表明,与现有的先进方法相比,所提出的算法框架具有竞争力,并取得了领先的性能。在未来的工作中,将在现有的算法框架的基础上融合更多的特征,例如线特征、圆柱体特征和语义特征。

### 参 考 文 献

- [1] Leng L M, Zeng Z B, Wu G H, et al. The phase

- calibration for integrated optical phased arrays using an artificial neural networks with resolved phase ambiguity [J]. *Photonics Research*, 2022, 10(2): 347-356.
- [2] Hu C Y, Huang H, Chen M H, et al. FourierCam: a camera for video spectrum acquisition in a single shot[J]. *Photonics Research*, 2021, 9(5): 701-713.
- [3] Teng J, Hu C Y, Huang H, et al. Single-shot 3D tracking based on a polarization multiplexed Fourier-phase camera [J]. *Photonics Research*, 2021, 9(10): 1924-1930.
- [4] Mur-Artal R, Tardós J D. ORB-SLAM2: an open-source SLAM system for monocular, stereo, and RGB-D cameras[J]. *IEEE Transactions on Robotics*, 2017, 33(5): 1255-1262.
- [5] Zhang J, Singh S. LOAM: lidar odometry and mapping in real-time[J]. *Robotics: Science and Systems*, 2014, 2(9): 1-9.
- [6] Shan T X, Englot B. LeGO-LOAM: lightweight and ground-optimized lidar odometry and mapping on variable terrain[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2019: 4758-4765.
- [7] Qin T, Li P L, Shen S J. VINS-mono: a robust and versatile monocular visual-inertial state estimator[J]. *IEEE Transactions on Robotics*, 2018, 34(4): 1004-1020.
- [8] Zhang J, Singh S. Laser-visual-inertial odometry and mapping with high robustness and low drift[J]. *Journal of Field Robotics*, 2018, 35(8): 1242-1264.
- [9] Zuo X, Geneva P, Lee W, et al. LIC-fusion: LiDAR-inertial-camera odometry[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), November 3-8, 2019, Macau, China. New York: IEEE Press, 2020: 5848-5854.
- [10] Zuo X, Yang Y L, Geneva P, et al. LIC-fusion 2.0: LiDAR-inertial-camera odometry with sliding-window plane-feature tracking[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 24-January 24, 2021, Las Vegas, NV, USA. New York: IEEE Press, 2021: 5112-5119.
- [11] Zhang J, Kaess M, Singh S. Real-time depth-enhanced monocular odometry[C]//2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, September 14-18, 2014, Chicago, IL, USA. New York: IEEE Press, 2014: 4973-4980.
- [12] Graeter J, Wilczynski A, Lauer M. LIMO: lidar-monocular visual odometry[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2019: 7872-7879.
- [13] Shan T X, Englot B, Ratti C, et al. LVI-SAM: tightly-coupled lidar-visual-inertial odometry via smoothing and mapping[C]//2021 IEEE International Conference on Robotics and Automation (ICRA), May 30-June 5, 2021, Xi'an, China. New York: IEEE Press, 2021: 5692-5698.
- [14] Jia Y P, Luo H Y, Zhao F, et al. Lvio-fusion: a self-adaptive multi-sensor fusion SLAM framework using actor-critic method[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 27-October 1, 2021, Prague, Czech Republic. New York: IEEE Press, 2021: 286-293.
- [15] Kim G, Kim A. Scan context: egocentric spatial descriptor for place recognition within 3D point cloud map [C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2019: 4802-4809.
- [16] Hartley R, Zisserman A. Multiple view geometry in computer vision[M]. 2nd ed., Cambridge: Cambridge University Press, 2003
- [17] Rusu R B, Cousins S. 3D is here: point cloud library (PCL) [C]//2011 IEEE International Conference on Robotics and Automation, May 9-13, 2011, Shanghai, China. New York: IEEE Press, 2011.
- [18] Su Y, Wang T, Yao C, et al. GR-SLAM: vision-based sensor fusion SLAM for ground robots on complex terrain [C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 24-January 24, 2021, Las Vegas, NV, USA. New York: IEEE Press, 2021: 5096-5103.
- [19] Wang H, Wang C, Xie L H. Intensity scan context: coding intensity and geometry relations for loop closure detection[C]//2020 IEEE International Conference on Robotics and Automation (ICRA), May 31-August 31, 2020, Paris, France. New York: IEEE Press, 2020: 2095-2101.
- [20] Dellaert F. Factor graphs and GTSAM: a hands-on introduction[R]. Atlanta: Georgia Institute of Technology, 2012.
- [21] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset[J]. *International Journal of Robotics Research*, 2013, 32(11): 1231-1237.
- [22] Caesar H, Bankiti V, Lang A H, et al. nuScenes: a multimodal dataset for autonomous driving[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11618-11628.
- [23] Grupp M. EVO: a python package for the evaluation of odometry and slam[EB/OL]. [2022-11-12]: <https://github.com/MichaelGrupp/evo>.
- [24] Wang H, Wang C, Chen C L, et al. F-LOAM: fast LiDAR odometry and mapping[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), September 27-October 1, 2021, Prague, Czech Republic. New York: IEEE Press, 2021: 4390-4396.
- [25] Yokozuka M, Koide K, Oishi S, et al. LiTAMIN: LiDAR-based tracking and mapping by stabilized ICP for geometry approximation with normal distributions[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 24-January 24, 2021, Las Vegas, NV, USA. New York: IEEE Press, 2021: 5143-5150.