

基于地平面的单目视觉辅助激光雷达 SLAM 研究

晏小彬^{**}, 彭道刚^{*}, 戚尔江^{***}

上海电力大学自动化工程学院, 上海 200090

摘要 融合视觉传感器和激光雷达可以实现优于单一传感器的同时定位与建图(SLAM)系统, 现有的视觉和激光雷达融合算法仍然存在计算复杂度高、系统精度及稳定性受错误的深度匹配影响等问题。为了更加高效、鲁棒地融合视觉和激光雷达的信息, 充分利用图像和激光点云中的地平面信息, 提出了高效的视觉辅助激光雷达 SLAM 算法。首先, 从激光点云中分割出地面点云用于提取图像中的地面 ORB 特征点, 并通过单应性变换中的交比不变性校验特征匹配, 从而高效鲁棒地利用单应性矩阵分解实现绝对尺度相机运动估计。然后, 将得到的相机运动估计以李群 SE(3)形式进行插值, 用于校正激光雷达在自身运动过程中产生的点云畸变。最后, 单目相机的运动估计作为初值用于激光里程计的位姿优化。公共数据集 KITTI 和实际环境的测试结果表明, 本文算法可以有效利用相机运动估计对激光点云畸变进行校正, 实时准确地实现里程计和建图。

关键词 遥感; 同时定位与建图; 激光雷达; 单目相机; 单应性变换; 交比不变性

中图分类号 TP242

文献标志码 A

doi: 10.3788/AOS202040.2428002

Research on Ground-Plane-Based Monocular Aided LiDAR SLAM

Yan Xiaobin^{**}, Peng Daogang^{*}, Qi Erjiang^{***}

College of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China

Abstract The fusion of a vision sensor and LiDAR can achieve a simultaneous localization and mapping (SLAM) system superior to a single sensor. However, the existing vision and LiDAR fusion algorithms still have such problems as high computational complexity and the system accuracy and stability susceptible to wrong depth matching. In order to combine vision and LiDAR information more efficiently and robustly, we made full use of ground plane information in the images and LiDAR point clouds, and proposed an efficient SLAM algorithm of vision-assisted LiDAR. Firstly, the ground point cloud was segmented from the laser point cloud to extract the ground ORB feature points in the images, and feature matching was verified by the cross-ratio invariance in the homography transformation. In this way, the absolute scale motion estimation of camera was realized efficiently and robustly via the homography matrix decomposition. Then, the obtained motion estimate of the camera was interpolated in the form of Lie group SE(3) to correct the point cloud distortion generated by the LiDAR during its own motion. Finally, the motion estimate of the monocular camera was taken as the initial value for the position optimization of LiDAR odometry. The test results of KITTI, a public data set, and the actual environment show that the proposed algorithm can effectively employ the motion estimate of the camera to correct the point cloud distortion of LiDAR and achieve odometry and mapping in real time and accurately.

Key words remote sensing; simultaneous localization and mapping; LiDAR; monocular camera; homography transformation; cross-ratio invariance

OCIS codes 280.3640; 150.0155; 110.4153; 110.2970

1 引 言

同时定位与建图(SLAM)是机器人领域的一项关键技术,是机器人能够自主运行的一项基本要求。激光雷达和视觉传感器是 SLAM 中的两种主流传

感器,近年来基于这两种传感器的 SLAM 算法得到广泛的研究和应用^[1-7]。视觉 SLAM 方面,产生了 ORB-SLAM2^[8]、DSO^[9]、VINS^[10] 等为代表的优秀方案。激光 SLAM 中则形成了 LOAM^[11]、IMLS-SLAM^[12]、SegMatch^[13] 等框架。

收稿日期: 2020-07-20; 修回日期: 2020-08-24; 录用日期: 2020-09-15

基金项目: 上海市科学技术委员会工程技术研究中心项目资助(14DZ2251100)

* E-mail: pengdaogang@126.com; ** E-mail: tobelegend@hotmail.com; *** E-mail: xinbdzh@163.com

然而,单一的传感器总是存在某种不足,例如视觉传感器对环境纹理及光照条件敏感,而激光雷达存在由本体运动引起的点云畸变问题以及回环检测能力不足等。因此,视觉传感器与激光雷达的融合成为研究热点。DEMO^[14]是一种通过将激光雷达的深度测量用于恢复图像中特征点深度从而增强视觉里程计(在SLAM中,通过连续地定位估计机器人的运动轨迹的过程称为里程计,而视觉里程计即采用视觉传感器获取环境信息的里程计)的方法。V-LOAM^[15]进一步利用DEMO^[14]的视觉里程计实现高频运动估计来校正激光雷达点云畸变并为激光里程计提供运动估计初值,从而实现更加精确的激光里程计。LIMO^[16]中提出了一种新的将激光雷达测量用于图像特征深度恢复从而实现无尺度漂移的视觉里程计,并使用深度学习算法剔除错误特征点。这些通过激光雷达距离测量恢复图像特征点深度的方法容易受到错误的深度匹配而引入额外的几何残差,降低系统精度和稳定性。而且DEMO^[14]和LIMO^[16]这类方法丢弃了大量有用的激光雷达测量信息。为了最大化利用相机和激光雷达的测量信息,同时避免不佳的深度匹配的影响,TVLO^[17]中提出分别对两种传感器独立地运行里程计和建图流程,最后对它们的残差进行联合优化,以此实现视觉和激光的紧耦合。然而,基于两种传感器同时进行里程计和建图需要很大的系统开销,不适用于嵌入式平台等低算力场合。LIC-Fusion^[18]中提出一种紧耦合的多传感器融合算法,有效地融合了惯性测量单元(IMU)测量值、稀疏的视觉特征和提取的激光点,不过没有对视觉测量数据和激光雷达数据进行关联。

实际上,激光雷达尤其是三维激光雷达在里程计和建图方面可以实现比视觉传感器更高的精度和鲁棒性。如LeGO-LOAM^[19]仅借助IMU对激光点云的自运动畸变进行校正,就实现了很好的定位和建图精度,而且可以在嵌入式平台上实时地运行。因此,分配少量系统资源将视觉传感器用于辅助激光雷达进行同时定位与建图,弥补激光雷达在自运动畸变和回环检测等方面的不足,是一种更加合理的选择。基于上述思想,本文对视觉辅助激光雷达进行同时定位与建图展开了研究,主要贡献如下:

1) 创新地提出一种单目视觉和激光雷达紧耦合SLAM方案,利用激光里程计中的地面点云高效提取图像中的地面特征点,基于单应性变换实现高效的绝对尺度相机运动估计,进而将运动估计用于

校正激光里程计中的自运动点云畸变和位姿优化。

2) 针对地面的弱纹理特性导致的错误匹配问题,提出通过共线特征点搜索和交比不变性校验来挑选正确的匹配特征,保证单应性矩阵计算的正确性。

2 算法框架

本节总体描述本文提出的单目视觉与激光雷达的融合方法。假设相机和激光雷达之间的标定已完成,并且传感器数据的时间也已对齐。则通过相机估计出的运动和通过激光雷达估计出的运动在时间上是对齐的,可以通过估计相机的运动对激光点云畸变进行校正并提供一个位姿初值。

图1为算法流程图,本文系统由视觉前端和激光雷达后端构成。

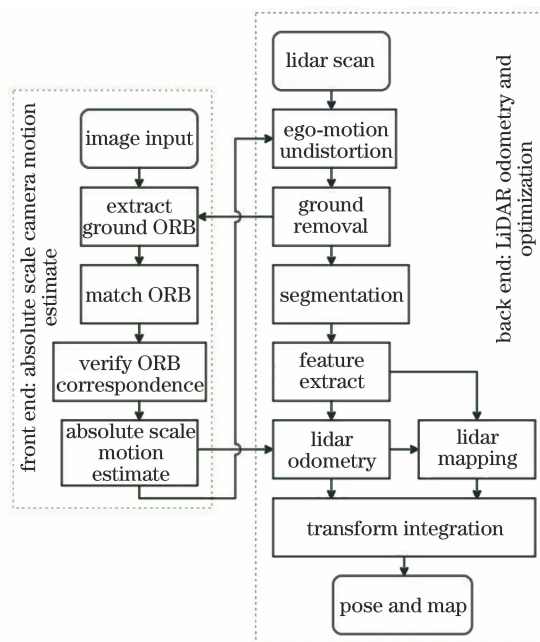


图1 本文算法流程图

Fig. 1 Flow chart of the proposed algorithm

视觉前端对连续的两帧图像进行帧到帧的运动估计。对于每一帧从单目视觉采集的图像,通过选取图像中的兴趣区域或激光点云提供的信息,提取出图像中可能的地面区域。然后在提取的地面区域中提取ORB特征并与上一帧图像进行特征匹配。通过交比不变性验证特征匹配的正确性,保留下来的特征匹配集合用于计算图像之间的单应性变换矩阵。单应性变换矩阵包含了图像间的运动信息和地平面信息。最后对单应性变换矩阵进行分解并利用已知的相机安装高度,即可得到绝对尺度下的帧间运动。

激光雷达后端是基于 LeGO-LOAM^[19]的工作实现的。本文通过视觉前端估计出的运动来校正点云畸变,并为激光雷达里程计部分提供优化初值。具体而言,首先利用来自视觉前端的运动估计,对自身运动造成的点云畸变进行校正。然后通过文献[20]中的方法提取激光点云中的地面点,配准后的地面点将用于加强视觉前端的帧间运动估计。接着采用文献[21]中的方法对移除地面点后的点云进行分割,最后对分割后的点云提取特征并利用视觉前端的初始运动估计加速点云特征匹配的迭代优化,实现更加高效且精确的里程计。

3 前端:绝对尺度相机运动估计

单目相机图像之间的帧间运动估计存在尺度歧义,该尺度由两帧图像对应的相机平移决定。因此,不同图像之间估计的运动将在不同的尺度之下。若直接将不同尺度下估计出的相机运动用于激光点云的校正,将导致两组点云被校正到不同尺度而发生误匹配,造成整个系统失效。为了解决此问题,本文利用图像中的地平线信息和已知的相机安装高度来估计尺度因子,从而将不同时刻的相机运动估计统一在公制尺度下。同时,利用地平线在图像间的单应性变换关系来估计帧间运动。图像中的地平线通过选取感兴趣区域和借助激光点云语义信息来实现高效鲁棒地提取。针对地平面的弱纹理特性,通过交比不变性剔除错误匹配。

3.1 基于地平面的单应性矩阵

通常,相机是固定安装在平台上,其相对于地面的高度是已知的,因此可以利用相机高度这一先验信息来估计尺度。假设已知的相机高度为 h^* ,通过图像估计出的相机高度信息为 h ,则尺度因子 $s=h/h^*$ 。

将前一帧的相机坐标系作为参考坐标系,地平面上任意三维点 $\mathbf{X}=(x,y,z)^T$ 满足下列约束,

$$\mathbf{n}^T \mathbf{X} = h, \quad (1)$$

式中: \mathbf{n} 表示地平面的单位法向量; h 表示相机光心到地面的距离。

假设 \mathbf{X} 在图像 1 的相机坐标系和图像 2 的相机坐标系中的坐标分别为 \mathbf{X}_1 和 \mathbf{X}_2 ,两个相机坐标系之间的变换为旋转 \mathbf{R} 和平移 \mathbf{t} ,则有

$$\mathbf{X}_2 = \mathbf{R}\mathbf{X}_1 + \mathbf{t}. \quad (2)$$

且由(1)式可知

$$\frac{\mathbf{n}^T \mathbf{X}_1}{h} = 1. \quad (3)$$

联立(2)、(3)式可以得到

$$\mathbf{X}_2 = \left(\mathbf{R} + \frac{\mathbf{t}\mathbf{n}^T}{h} \right) \mathbf{X}_1. \quad (4)$$

记相机内参矩阵为 \mathbf{K} ,则 \mathbf{X} 在图像 1 和图像 2 上的像素齐次坐标为

$$\mathbf{x}_i = s_i \mathbf{K}\mathbf{X}_i, i = 1, 2, \quad (5)$$

式中: s 为尺度因子。由(4)、(5)式可得,

$$\lambda \mathbf{x}_2 = \mathbf{K} \left(\mathbf{R} + \frac{\mathbf{t}\mathbf{n}^T}{h} \right) \mathbf{K}^{-1} \mathbf{x}_1, \quad (6)$$

$$\mathbf{H} \triangleq \mathbf{K} \left(\mathbf{R} + \frac{\mathbf{t}\mathbf{n}^T}{h} \right) \mathbf{K}^{-1}, \quad (7)$$

式中: $\lambda = s_2/s_1$ 为两个尺度因子之比。

矩阵 \mathbf{H} 称为单应性矩阵,表达了空间平面上一点在两个图像上投影点像素坐标之间的映射关系。单应性矩阵有 9 个元素,不过由于尺度不确定性,其自由度为 8。而一对特征匹配点可以构建 2 个关于 \mathbf{H} 矩阵的约束。因此, \mathbf{H} 矩阵的计算最少只需要 4 对匹配点^[22]。从弱纹理地面提取的图像特征含有较多噪声,因此容易产生错误匹配。为了保证鲁棒地通过地面特征点计算 \mathbf{H} 矩阵,本文采用优化的方法。通过最小化由对称转移误差定义的代价函数得到最优的单应性变换,

$$\mathbf{H} = \underset{\mathbf{H}}{\operatorname{argmin}} \sum_{i=1}^N e_i(\mathbf{H}), \quad (8)$$

$$e_i(\mathbf{H}) = \rho [d(\mathbf{x}'_i, \mathbf{H}\mathbf{x}_i)] + \rho [d(\mathbf{x}_i, \mathbf{H}^{-1}\mathbf{x}'_i)], \quad (9)$$

式中: \mathbf{x}_i 和 \mathbf{x}'_i 表示两帧图像上匹配的特征点; $d(\cdot)$ 表示两像素点的欧氏距离; $\rho(\cdot)$ 是为避免较大误差主导优化方向而引入的鲁棒核函数。本文采用 Huber 核,

$$\rho(L_i) = \begin{cases} L_i, & \text{if } L_i < k^2 \\ 2k \cdot \sqrt{L_i} - k^2, & \text{if } L_i \geq k^2 \end{cases}, \quad (10)$$

式中: k 是预先给定的阈值。首先采用归一化的直接线性变换(DLT)算法对满足交比不变性(将在 3.3 节描述)的 4 对特征匹配计算得到的 \mathbf{H} 矩阵进行初始估计,然后基于该初值通过 Levenberg-Marquardt 算法优化(8)式表示的代价函数得到最优的单应性矩阵。

对计算出的 \mathbf{H} 矩阵通过文献[23]提出的方法进行分解,即可得到相机运动的旋转矩阵 \mathbf{R} 和平移向量 \mathbf{t} 、地平面法向量 \mathbf{n} 和相机高度 h 。由于旋转矩阵满足正交矩阵约束,因此只需要校正平移向量 \mathbf{t} 的尺度,

$$\mathbf{t}_{\text{scaled}} = \frac{\mathbf{t}}{\|\mathbf{t}\|} \cdot \frac{h^*}{h}. \quad (11)$$

最后得到绝对尺度的相机运动即为

$$\mathbf{T}_{\text{scaled}}^C = \begin{bmatrix} \mathbf{R} & \mathbf{t}_{\text{scaled}} \\ 0 & 1 \end{bmatrix}. \quad (12)$$

3.2 图像地平面特征点提取

地平面特征点提取的关键在于如何确定图像中的地平面。本文将激光雷达后端流程中提取的地平面激光点云投影到图像上,实现高效且鲁棒地提取图像中的地面特征点。地平面在图像中主要位于下半部分,因此首先在图像的下半部分提取 ORB 特征点。然后,利用相机与激光雷达之间的变换矩阵和相机的投影模型,将激光点云中的地平面点投影到图像上。假设标定好的相机与激光雷达间的变换矩阵为 \mathbf{T}_L^C ,相机的投影矩阵为 \mathbf{P}_0 ,则地面激光点云中给定一点 \mathbf{X}_i 的投影点齐次像素坐标为

$$\mathbf{x}_i = \mathbf{P}_0 \mathbf{T}_L^C \mathbf{X}_i. \quad (13)$$

每一个激光投影点可以确定一个以其为中心,半径为 r 的窗口,所有投影点窗口外的特征点首先被剔除。进一步地,窗口中如果存在多个特征点,则

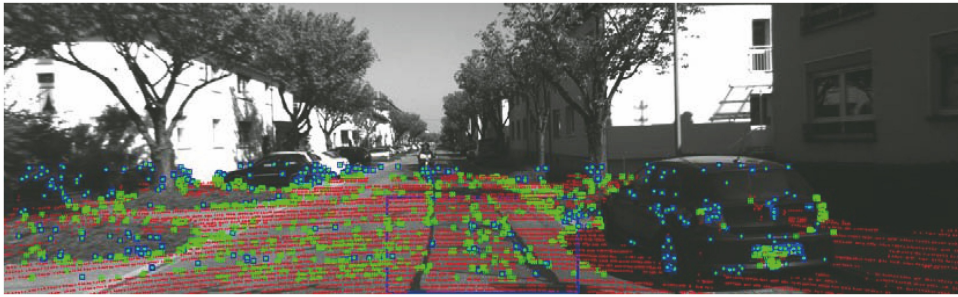


图 2 通过激光点云和兴趣区域提取地平面特征点

Fig. 2 Extract ground features point from LiDAR cloud and ROI

3.3 基于交比不变性的特征匹配验证

ORB 特征^[24]具有方向和尺度不变性,是一种快速鲁棒的图像特征。然而,由于地平面通常缺乏丰富的纹理,各特征点之间的描述子差异并不十分明显,导致误匹配的发生,即所谓的匹配外点。基于随机抽样一致性(RANSAC)的特征匹配外点剔除算法是目前普遍采用的方法。然而,当错误匹配的数量较多时,基于 RANSAC 的方法无法保证正确地剔除外点(从大量错误匹配中找出少量正确匹配)。文献[22]指出,单应性矩阵的计算至少需要 4 对匹配特征,这表明只需要保证找到 4 对以上的正确匹配,即可计算出正确的单应性矩阵。因此本文利用交比^[22]不变性(CRI)进行特征匹配的校验,并且只采用满足交比不变性的特征点对进行 \mathbf{H} 矩阵的计算。

对于 ROI 中提取的特征点和激光投影点邻域提取的特征点,本文均通过交比不变性来验证其特

比较它们的描述子差异,选择描述子差异最大的特征点作为该窗口的地面特征点。

然而,在初始状态下还没有任何的相机运动估计,无法提供需要的激光点云中的地面点。针对这一问题,初始状态下本文通过提取图像兴趣区域(ROI)来保守地提取地面特征点。对于固定安装的相机,地平面通常成像于图像下半部分的中间位置。因此,本文假设图像下 1/3 和中间 1/5 的区域(如图 2 中间矩形框内)是对应到前方地平面的,初始状态仅从该区域中提取特征点。地平面特征提取效果如图 2 所示,其中地平面激光点云投影点用灰色圆点表示,特征点用包含圆点的方框表示。浅色方框浅色圆点表示通过激光投影点提取的地面特征点,深色方框浅色圆点表示通过激光投影点排除的非地面特征点。从图 2 可以看出,算法能排除大量非地面特征点同时保留大量正确地面特征点。

征匹配的正确性。首先通过文献[25]提出的方法找出所有共线点数量在 4 点以上的特征点。然后对每 4 个共线特征点及它们的匹配特征点计算交比,一旦两个交比相等就接受所有的 4 个匹配点。对于不满足交比不变性的特征点并不会立即剔除,因为它们之中可能存在正确的匹配点。本文在所有大于 4 点的共线特征验证结束之后剔除未通过交比不变性校验的特征点对。如图 3 所示,经过交比不变性校验之后,得到的特征匹配虽然数量不多,但是绝大多数为正确的匹配,足以用于后续计算。

4 后端:激光雷达里程计及优化

本文的激光里程计部分是基于 LeGO-LOAM^[19]实现的,并结合前端相机帧间运动估计做出了改进。

LeGO-LOAM 是 Shan 等^[19]基于 LOAM^[11]框架提出的轻量级激光 SLAM 算法,可以在嵌入式系

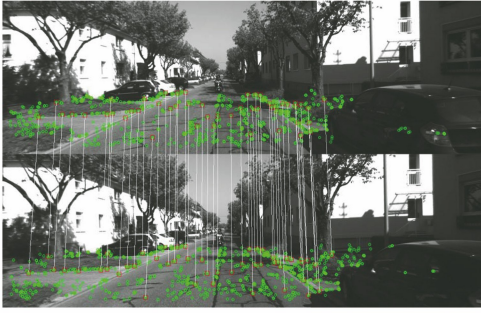


图 3 交比不变性校验后的匹配特征

Fig. 3 Feature correspondences after CRI verify

统上实时运行。该算法首先将单次扫描的点云投影为深度图像,分别采用文献[20]和文献[21]中的方法提取出点云中的地面点以及分割点云,并滤除比较小的点云簇。然后采用 LOAM^[11]中的方法从分割后的点云和地面点中提取直线和平面特征,通过最小化点到线和平面特征的距离对两帧点云进行匹配从而实现位姿估计,并利用 Levenberg-Marquardt 方法进行位姿优化。最后,激光点云中提取的直线和平面特征与环境点云地图进行匹配可以得到更加精确的位姿。通过最终位姿将直线和平面特征配准到点云地图上。

4.1 激光点云的校正

扫描式激光雷达通过内部机构的机械旋转实现 360°的环境测量。这一工作方式使得其完成一次扫描得到的点云不是在同一时间采集的,当激光雷达本身运动时,将产生所谓的自运动点云畸变。本文利用相机帧间估计得到的位姿来校正点云。激光雷达工作频率一般在 10 Hz 甚至更高,因此其在一次扫描中的运动可以近似建模为线性运动。假设激光雷达第 k 次扫描从 τ_k 时刻到 τ_{k+1} 时刻,且这两个时刻得到的两帧图像间已经估计得到位姿为 $T_{k+1,k}^C$,并假设相机到激光雷达之间已标定的固定

变换矩阵为 T_L^C 。则可通过相机位姿粗略估计激光雷达第 k 次扫描起止时刻之间运动为

$$T_{k+1,k}^L = T_L^C T_{k+1,k}^C \quad (14)$$

给定该次扫描任意激光测量点 $p_i(x_i, y_i, z_i)$,其时间戳为 τ_i 。则 τ_i 时刻激光雷达相对于扫描起始点 p_k 的运动通过李群 SE(3)上的位姿插值得到:

$$\alpha = \frac{\tau_i - \tau_k}{\tau_{k+1} - \tau_k}, \quad (15)$$

$$T_{i,k} = (T_{k+1,k}^L)^\alpha \quad (16)$$

通过该位姿校正后的激光点 \bar{P}_i 齐次坐标表示为

$$\bar{P}_i = T_{i,k}^{-1} P_i \quad (17)$$

4.2 激光里程计

在对激光点云进行自运动畸变校正后,通过激光点云跟踪可以实现更加精确的激光里程计。前端估计出的相机位姿除了校正激光点云的自运动畸变,还可以提供位姿初值加速激光里程计中特征匹配的优化过程。

激光里程计估计两帧连续激光点云之间的运动,通过执行点到直线和点到平面的扫描匹配,并最小化所有匹配之间的几何距离求取两帧点云之间的变换。LeGO-LOAM^[19]中将上一帧激光点云中的特征点组织为三维 K 维树的结构,通过最近邻搜索当前点云中直线和平面特征的匹配。搜索特征匹配的过程如图 4 所示。对于直线特征中的点 i ,通过最近邻搜索上一帧中的最近点 j ,并在该最近点的相邻线激光扫描中搜索次近邻点 k 。平面特征中的点 i' 匹配的最近邻点 j' 和次近邻点 k'_1, k'_2 用同样的方法搜索,唯一区别是需要搜索两个次近邻点。最后,通过最小化所有直线特征点 i 与 j, k 所在直线的距离和 i' 与 j', k_1 和 k'_2 所形成平面的距离,得到两帧点云之间的最优变换。

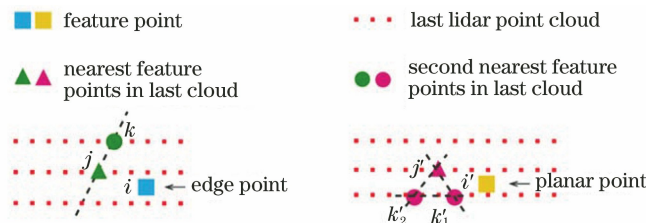


图 4 激光里程计搜索特征点匹配过程

Fig. 4 Process of feature point correspondence searching in Lidar odometry

当两帧点云之间的运动较大时,上述方法需要迭代多次才能找到正确匹配。利用相机的帧间运动估计先验,可以将新一帧校正后的激光点云变换到前一

帧点云的坐标系下,再通过 LeGO-LOAM^[19]中的方法进行特征匹配和位姿优化,并将相机帧间运动先验作为优化初值。由于经过变换的激光点云与上一帧

点云重合度更高,通过最近点搜索更容易找到正确的匹配特征,因此算法可以更加快速地收敛。实际效果如图 5 所示,其中图 5(a)为直接搜索特征匹配,图 5(b)为方框的局部放大,图 5(c)为利用相机帧间运动

先验变换后的特征分布,图 5(d)为方框的局部放大。图中尺寸较大的浅色圆点和深色圆点为待匹配特征,其余为上一帧中的直线和平面特征。可以看出,经过变换后的特征与上一帧点云特征已基本重合。

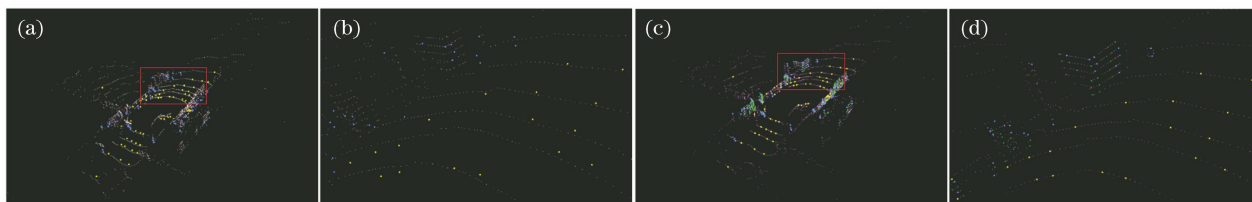


图 5 特征点匹配过程对比。(a)直接搜索匹配;(b)直接搜索过程的局部细节;
(c)经相机运动先验变换后搜索匹配;(d)变换后搜索过程的局部细节

Fig. 5 Comparison of feature point correspondence process. (a) Search the correspondence directly; (b) local details of direct search process; (c) search correspondence after transformed by camera motion *a priori*; (d) local details of search process after transform

5 实验分析

算法的运行平台为研华公司的工控机 ARK-3500,CPU 为 Intel Core i7-3610QE 2.3 GHz,4 GB 运行内存,64 位 Linux 操作系统。首先使用 KITTI 数据集^[26]进行实验,与 LeGO-LOAM^[19]算法进行对比。然后使用实验室机器人平台进行实际场景实验。

5.1 KITTI 数据集实验

本文通过 KITTI 数据集^[26]进行仿真实验,验证算法的有效性和性能。该数据集包含 1 个 360° 三维激光雷达、2 个灰度相机、2 个彩色相机和 1 个高精度全球定位系统(GPS)/IMU 的数据,并提供各

传感器之间的标定参数以及传感器数据的时间戳。实验采用数据集中原始数据子集序列 2011_10_03_drive_0027 ([http://www. cvlibs. net/datasets/kitti/raw_data. php](http://www.cvlibs.net/datasets/kitti/raw_data.php)),该数据子集采集环境为居住区室外场景,轨迹全长约 3.7 km,包含房屋、道路、汽车、植被等物体[图 6(b)]。该原始数据子集中的图像数据和激光雷达数据都是未校正、未同步的。本文通过同步读取其中的激光点云序列(velodyne_points 数据包)和单目灰度相机的图像序列(camera_0 数据包),利用时间戳数据进行同步,利用标定参数进行图像畸变校正和激光点云投影。



图 6 轨迹对比及真实场景图。(a)本文算法、LeGO-LOAM-IMU、LeGO-LOAM-noIMU 估计轨迹与真实轨迹对比;
(b)轨迹所对应位置的卫星地图

Fig. 6 Comparison of trajectory and the real scene. (a) Comparison of real trajectory and trajectories estimated by LeGO-LOAM-IMU, LeGO-LOAM-noIMU, and proposed, respectively; (b) satellite map of the trajectory

与 LeGO-LOAM^[19]算法的两种版本进行对比,一种是通过 IMU 数据校正激光点云畸变并提供位姿初值的版本(以下简称 LeGO-LOAM-

IMU),另一种是无 IMU 传感器数据支持的 LeGO-LOAM 算法(以下简称 LeGO-LOAM-noIMU)。

定性对比结果如图 6、7 所示,图中轨迹采用

evo(<https://github.com/MichaelGrupp/evo>) 工具进行数据对齐。图 6(a) 为 LeGO-LOAM-IMU 算法估计轨迹、LeGO-LOAM-noIMU 算法估计轨迹、本文算法估计轨迹与真实轨迹的对比,图 6(b) 为轨迹所在区域的卫星地图。图 7 为它们在 x 轴(向右)、 y 轴(向上)、 z 轴(向前)的数据对比。容易看出,LeGO-LOAM-IMU 算法估计的轨迹始终与真实轨迹较为符合,在 x 、 y 、 z 三轴方向的估计数据都与真实值相差不大。LeGO-LOAM-noIMU 算法估

计的轨迹逐渐偏离,与真实轨迹相差越来越大,尤其在 z 轴和 x 轴方向偏差很大。这是由于没有 IMU 数据校正点云畸变,激光里程计中的误差增大并且随着运行距离积累。本文算法估计的轨迹也存在误差逐渐增大的问题,不过与 LeGO-LOAM-noIMU 算法相比有了极大的改善,特别是在 x 轴和 z 轴方向上的估计精度有了很大的提高。上述分析表明,本文算法能够有效利用相机的帧间运动估计校正激光雷达的自运动点云畸变,从而提高激光里程计的精度。

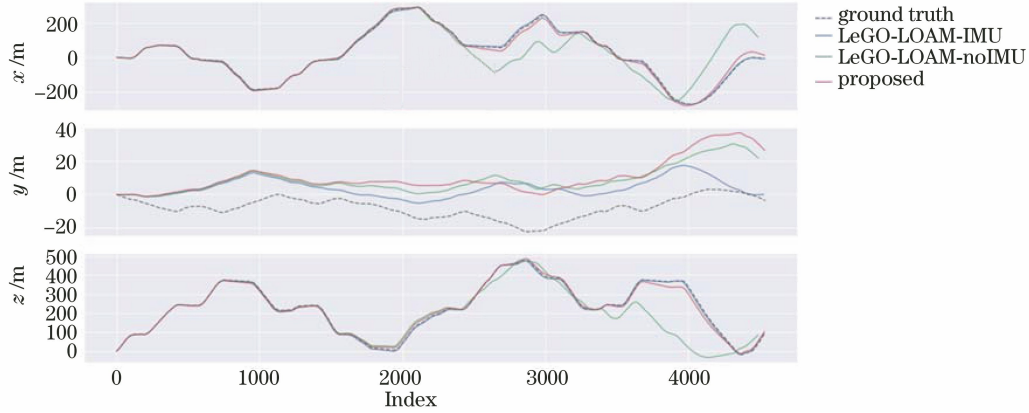


图 7 x, y, z 轴方向上本文算法、LeGO-LOAM-IMU、LeGO-LOAM-noIMU 估计值与真实值对比

Fig. 7 Comparison of the estimated value from proposed algorithm, LeGO-LOAM-IMU, LeGO-LOAM-noIMU and the real value on $x, y,$ and z axis

为了进一步评估本文算法的性能,在仿真实验中采用绝对轨迹误差 (ATE) 作为精度性能衡量指标,对比算法的定位精度。记 $\mathbf{T}_{gt,i}$ 为真实轨迹的第 i 个位姿, $\mathbf{T}_{esti,i}$ 为算法估计轨迹的第 i 个位姿,则 ATE 可计算为

$$E_{ATE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\log_e(\mathbf{T}_{gt,i}^{-1} \mathbf{T}_{esti,i})^v\|_2^2} \quad (18)$$

该式实际上是每个位姿李代数^[27]的均方根误

差(RMSE)。为了检验算法的稳定性,以及排除实验中偶然因素的影响以保证算法对比的公平,本文对每种算法执行 10 次重复实验,统计各算法每次的 ATE 及去掉最大值和最小值后的平均值,统计结果如表 1 所示。

从表 1 可以看出,本文算法与没有 IMU 提供点云畸变校正的 LeGO-LOAM 算法相比,平均绝对轨迹误差减小了 86.92%,证明了本文算法的有效性。本文算法定位误差略高于 LeGO-LOAM-

表 1 不同算法绝对轨迹误差比较

Table 1 Comparison of ATE of different algorithms

unit: m

Sequence	Proposed algorithm	LeGO-LOAM-noIMU	LeGO-LOAM-IMU
1	9.7763	72.3448	5.4013
2	12.9146	89.8510	37.6222
3	9.7763	86.6547	5.3066
4	9.7762	102.4516	6.2184
5	10.9626	86.0318	5.2739
6	9.7762	75.7008	6.0007
7	8.7765	67.9349	11.4841
8	12.9145	84.7132	6.7188
9	9.8208	92.2212	5.2530
10	12.9146	60.0927	7.7874
Total average	10.7147	81.9316	6.7739

IMU,但平均绝对轨迹误差只多 3.9708 m,定位准确度与其相当。此外,单目相机可以提供环境的视觉信息,为提取语义信息、回环检测等任务提供支持,而且实验中使用的单目相机(FLIR FL2-14S3M-C)成本较IMU(OXTS RT3003)低40%,这使得本文算法在工程上的综合性能更优。

统计各算法处理一帧点云过程中不同模块的平均运行时间如表2所示,其中本文算法对Ego-motion undistortion模块耗时的统计包含整个前端

表2 不同算法运行时间统计

Table 2 Runtime statistics of different algorithms

unit: ms

Module	Proposed algorithm	LeGO-LOAM-noIMU	LeGO-LOAM-IMU
Ego-motion undistortion	32.6	0	2.2
Ground removal and segmentation	21.7	22.1	21.6
Feature extract	4.8	6.2	4.8
Odometry	3.3	15.1	7.6
Total	62.4	43.4	36.2

5.2 实际场景实验

实际环境测试实验采用课题组自主研发的移动机器人[图8(a)]。移动机器人上搭载了研华ARK-3500工控机、velodyne激光雷达VLP-16、Fotric红外可见光双视云台相机、WI-FI模块、电池等。

相机运动估计过程的耗时。由于本文算法需要估计相机帧间运动,在Ego-motion undistortion模块耗时相比其他两种算法有所增加(LeGO-LOAM-noIMU没有畸变校正过程,因此耗时为0),但在Odometry模块的耗时明显减少,这正是相机运动先验加速里程计算法收敛的结果。本文算法平均每帧总耗时为62.4ms,平均每帧处理速度达到约16 frame/s,而激光雷达工作频率为10 frame/s,算法满足实时性要求。

VLP-16为本文算法提供周围环境的激光点云数据,云台中的可见光相机提供环境的图像数据。在定位和建图过程中,云台是固定不动的。由于实验平台没有IMU设备,实际场景实验只与LeGO-LOAM-noIMU算法进行对比。

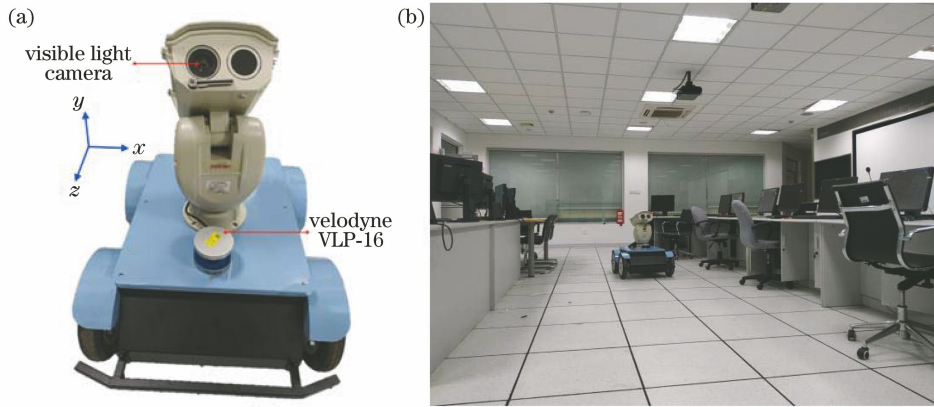


图8 机器人及室内实验环境。(a)机器人;(b)实验室环境

Fig. 8 Robot and laboratory environment. (a) Robot; (b) laboratory environment

本实验分别测试了室内场景和室外场景。图8(b)所示为实验室环境,包含弱纹理的地板、桌椅、白墙、玻璃墙等物体。实验时控制机器人绕办公桌行驶一圈,通过尺规工具测量,实际运行轨迹近似为长11.4m、宽4.2m的长方形轨迹,本文算法和LeGO-LOAM-noIMU算法估计的轨迹如图9所示。从图9可以看出,两种算法估计的轨迹与尺规工具测量的轨迹基本符合,LeGO-LOAM-noIMU估计的轨迹有更明显的误差漂移,本文算法估计的轨迹明显更接近真实轨迹(长方形轨迹)。

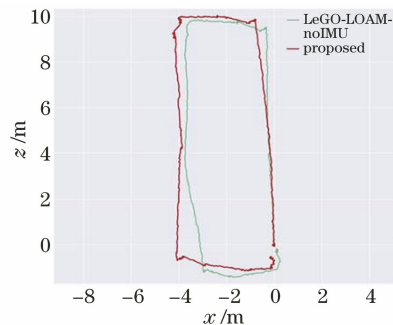


图9 实验室环境实验的估计轨迹

Fig. 9 Estimated trajectory of laboratory environment

室外场景为校园内的道路环境。由于室外场景的真实轨迹不容易获得,也不容易通过尺规测量,本实验将本文算法和 LeGO-LOAM-noIMU 估计轨迹与卫星地图进行对比。实验结果如图 10 所示,图 10(a)为本文算法和 LeGO-LOAM-noIMU 估计轨迹的对比,图 10(b)为点云地图,图 10(c)为室外实验环境的卫星地图,其中的手绘路线大致为实验过程中

机器人的行驶路线。从图 10(a)可以看出,LeGO-LOAM-noIMU 估计的轨迹(LeGO-noIMU)逐渐偏离,估计轨迹的终点未与起点(坐标(0,0)处)重合,整条轨迹与卫星图中标示轨迹相差较大。而本文算法估计的轨迹(proposed)始终与卫星图中给出的实际运行轨迹基本吻合,表明本文算法优于 LeGO-LOAM-noIMU 算法,具有足够的精度和场景鲁棒性。

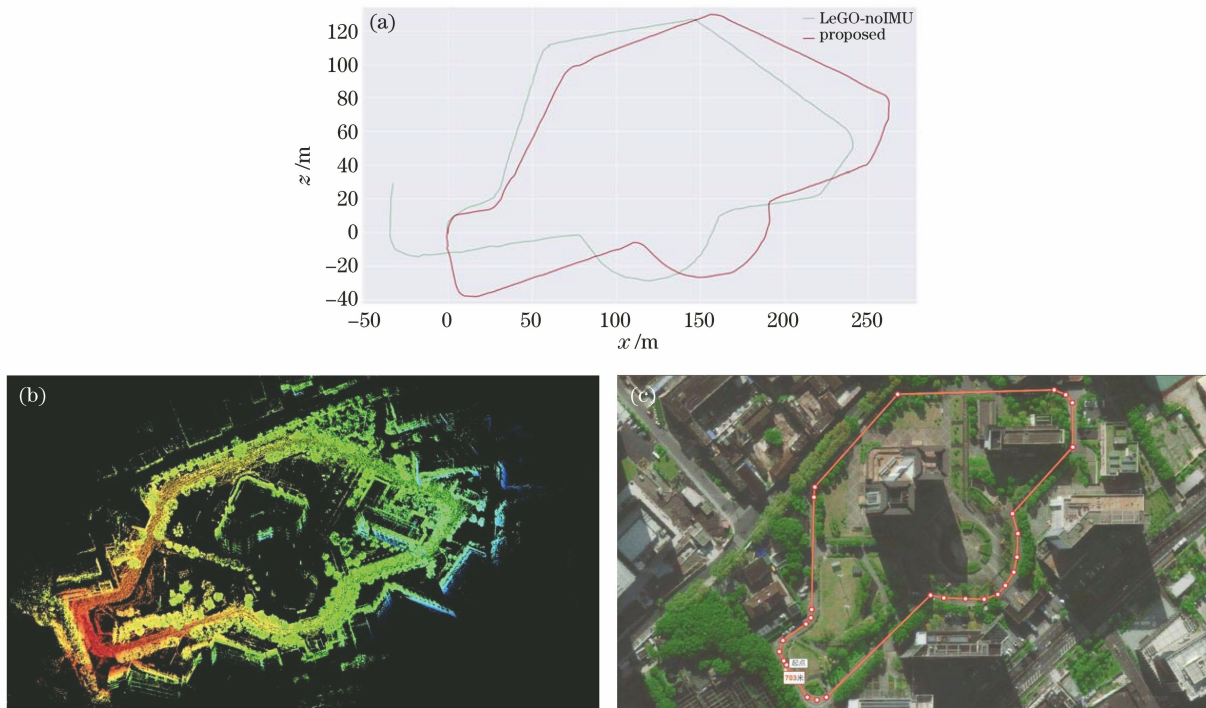


图 10 室外环境实验结果。(a)算法估计轨迹对比;(b)点云地图;(c)卫星地图

Fig. 10 Experimental results of outdoor environment. (a) Comparison of trajectory estimated by algorithms; (b) point cloud map; (c) satellite map

6 结 论

本文提出的单目视觉辅助激光雷达同步定位与建图算法,能够有效利用单目相机估计相对运动,对自运动畸变的激光点云进行校正,从而提高激光里程计精度。进一步地,利用视觉前端的相对运动可以加速激光里程计的优化求解过程。公开数据集的仿真实验中,算法的平均绝对轨迹误差为 10.7147 m,与无点云畸变校正的算法相比精度提高了 86.92%,验证了算法的有效性。实际场景实验表明了本文算法的可行性和场景鲁棒性。

视觉传感器与 IMU 相比具有成本低、利于回环检测、提供环境视觉信息等优势。然而,本文算法目前只实现了与激光雷达同频率(10 Hz)的相机运动估计,这种低频估计在位姿插值上会引入更多误

差,导致精度低于 IMU 的相对运动估计。下一步工作将研究在相机高频(≥ 30 Hz)工作下与激光雷达的融合,并引入图像信息进行回环检测,通过全局优化提高定位精度。

参 考 文 献

- [1] Wang R D, Li H, Zhao K, et al. Robust localization based on kernel density estimation in dynamic diverse city scenes using lidar[J]. *Acta Optica Sinica*, 2019, 39(5): 0528003.
王任栋, 李华, 赵凯, 等. 基于核密度估计的城市动态密集场景激光雷达定位[J]. *光学学报*, 2019, 39(5): 0528003.
- [2] Lin H C, Lü Q, Wei H, et al. Quadrotor autonomous flight and three-dimensional dense reconstruction based on VI-SLAM[J]. *Acta Optica Sinica*, 2018, 38(7): 0715004.
林辉灿, 吕强, 卫恒, 等. 基于 VI-SLAM 的四旋翼

- 自主飞行与三维稠密重构[J]. 光学学报, 2018, 38(7): 0715004.
- [3] Zhang Y, Xiang Z Y, Chen S Y, et al. Optimization on visual odometry under weak texture environment [J]. *Acta Optica Sinica*, 2018, 38(6): 0615001. 张易, 项志宇, 陈舒雅, 等. 弱纹理环境下视觉里程计优化算法研究[J]. 光学学报, 2018, 38(6): 0615001.
- [4] Liu H M, Zhang G F, Bao H J. A survey of monocular simultaneous localization and mapping[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2016, 28(6): 855-868. 刘浩敏, 章国锋, 鲍虎军. 基于单目视觉的同时定位与地图构建方法综述[J]. 计算机辅助设计与图形学学报, 2016, 28(6): 855-868.
- [5] Di K C, Wan W H, Zhao H Y, et al. Progress and applications of visual SLAM[J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(6): 770-779. 邸凯昌, 万文辉, 赵红颖, 等. 视觉SLAM技术的进展与应用[J]. 测绘学报, 2018, 47(6): 770-779.
- [6] Zong W P, Li G Y, Li M L, et al. A survey of laser scan matching methods[J]. *Chinese Optics*, 2018, 11(6): 914-930. 宗文鹏, 李广云, 李明磊, 等. 激光扫描匹配方法研究综述[J]. 中国光学, 2018, 11(6): 914-930.
- [7] Singandhupe A, La H M. A review of SLAM techniques and security in autonomous driving[C]//2019 Third IEEE International Conference on Robotic Computing (IRC), February 25-27, 2019, Naples, Italy. New York: IEEE Press, 2019: 602-607.
- [8] Mur-Artal R, Tardos 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.
- [9] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(3): 611-625.
- [10] 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.
- [11] Zhang J, Singh S. LOAM: lidar odometry and mapping in real-time [C]// *Robotics: Science and Systems X*, Robotics: Science and Systems Foundation, 2014.
- [12] Deschaud J E. IMLS-SLAM: scan-to-model matching based on 3D data [C]//2018 IEEE International Conference on Robotics and Automation (ICRA), May 21-25, 2018, Brisbane, QLD, Australia. New York: IEEE Press, 2018: 2480-2485.
- [13] Dubé R, Dugas D, Stumm E, et al. SegMatch: segment based place recognition in 3D point clouds [C]//2017 IEEE International Conference on Robotics and Automation (ICRA), May 29 - June 3, 2017, Singapore, Singapore. New York: IEEE Press, 2017: 5266-5272.
- [14] 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.
- [15] Zhang J, Singh S. Visual-lidar odometry and mapping: low-drift, robust, and fast[C]//2015 IEEE International Conference on Robotics and Automation (ICRA), May 26-30, 2015, Seattle, WA, USA. New York: IEEE Press, 2015: 2174-2181.
- [16] 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, 2018: 7872-7879.
- [17] Seo Y, Chou C C. A tight coupling of vision-lidar measurements for an effective odometry [C]//2019 IEEE Intelligent Vehicles Symposium (IV), June 9-12, 2019, Paris, France. New York: IEEE Press, 2019: 1118-1123.
- [18] Zuo X 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, Macao, China. New York: IEEE Press, 2019: 5848-5854.
- [19] 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, 2018: 4758-4765.
- [20] Himmelsbach M, Hundelshausen F V, Wuensche H J. Fast segmentation of 3D point clouds for ground vehicles [C] // 2010 IEEE Intelligent Vehicles Symposium, June 21-24, 2010, San Diego, CA, USA. New York: IEEE Press, 2010: 560-565.
- [21] Bogoslavskyi I, Stachniss C. Fast range image-based segmentation of sparse 3D laser scans for online operation [C] // IEEE/RSJ International Conference on Intelligent Robots & Systems. IEEE, 2016:163-169.
- [22] Hartley R, Zisserman A. Multiple view geometry in computer vision [M]. Cambridge: Cambridge University Press, 2004.
- [23] Zhang Z, Hanson A R. 3D reconstruction based on

- homography mapping[J]. Proceedings of ARPA96, 1996: 1007-1012.
- [24] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C] // 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE Press, 2011: 2564-2571.
- [25] Edelsbrunner H, Guibas L J. Topologically sweeping an arrangement[J]. Journal of Computer and System Sciences, 1989, 38(1): 165-194.
- [26] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset [J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [27] Varadarajan V S. Lie groups, lie algebras, and their representations[M]. New York: Springer, 1984.