

## 基于相机与摇摆激光雷达融合的非结构化环境定位

俞毓锋<sup>1,2</sup> 赵卉菁<sup>1,2</sup>

**摘要** 定位是机器人导航的关键问题,在缺乏结构信息的室外非结构化环境下,精确的三维定位面临更大挑战.本文提出一种基于相机与摇摆激光雷达融合的定位算法,重点解决在光照、地面起伏等因素影响下的机器人定位问题.本文结合激光雷达的深度信息和图像的颜色纹理信息,构建在时序帧间的特征点匹配关系;引入一种置信度评价方法,结合系统误差、数据关联、物体遮挡、特征跟踪等因素对特征点及其匹配关系进行评估,减少低质量特征的影响;最终将定位问题转化为特征点对的加权重投影误差优化问题予以解决.本文利用小型轮式移动机器人在越野和公园等典型非结构化环境下进行数据采集和实验验证.实验结果表明,与前沿的视觉定位算法相比,本文算法可有效提高在非结构化环境中的定位精度.

**关键词** 单目相机, 激光雷达, 多传感器融合, 机器人定位, 非结构化环境

**引用格式** 俞毓锋, 赵卉菁. 基于相机与摇摆激光雷达融合的非结构化环境定位. 自动化学报, 2019, 45(9): 1791–1798

**DOI** 10.16383/j.aas.2018.c170281

## Off-road Localization Using Monocular Camera and Nodding LiDAR

YU Yu-Feng<sup>1,2</sup> ZHAO Hui-Jing<sup>1,2</sup>

**Abstract** As a key issue of unmanned ground robots, localization faces more challenges in an off-road environment, where scenes are lack of structure. This paper proposes a visual localization method developed by combining a nodding LiDAR with a monocular camera, where localization is converted to a problem of re-projection error minimization between LiDAR-camera feature correspondences. In order to establish plenty of scene features while reducing the effects from low-quality ones, two procedures, i.e. LiDAR-to-camera and camera-to-LiDAR projections, are developed to generate feature correspondences between the data of two sensing modalities. Confidence measure is developed to evaluate the reliability of the established correspondences by assessing the factors such as system error, data association, occlusion, feature detection and tracking, so as to improve the robustness in complex situations. An experiment is conducted in two typical off-road environments using a wheeled robot with a nodding LiDAR and a monocular camera. The results show that the proposed method achieves more accurate localization than the state-of-the-art mono or stereo visual localization approach.

**Key words** Monocular camera, LiDAR, multi-sensor fusion, robot localization, off-road environment

**Citation** Yu Yu-Feng, Zhao Hui-Jing. Off-road localization using monocular camera and nodding LiDAR. *Acta Automatica Sinica*, 2019, 45(9): 1791–1798

收稿日期 2017-05-22 录用日期 2018-01-19  
Manuscript received May 22, 2017; accepted January 19, 2018  
国家自然科学基金 (61573027), 国家 863 课题 (2012AA011801) 资助  
Supported by National Natural Science Foundation of China (61573027) and National High Technology Research and Development Program of China (863 Program) (2012AA011801)

1. 北京大学信息科学技术学院 北京 100871 2. 北京大学机器感知与智能教育部重点实验室 北京 100871  
1. School of Electronic Engineering and Computer Science, Peking

面向非结构化环境的轮式移动机器人在民用及国防等领域有着重要的应用需求,其技术发展得到了一些著名赛事的验证和推广(如美国 DARPA 无人车挑战赛<sup>1</sup>, 欧洲 ELROB 室外机器人比赛<sup>2</sup>, 国内跨越险阻地面无人系统挑战赛<sup>3</sup>等),获得了社会各界的广泛关注.

精确定位作为机器人导航的关键,在室外非结构化环境面临巨大挑战.环境复杂多变、光照、遮挡、缺乏结构化信息等,为视觉里程计<sup>[1]</sup>或视觉 SLAM<sup>[2]</sup>算法的特征提取与匹配带来很多不确定性,影响定位精度和鲁棒性.另一方面,激光雷达已被广泛应用于机器人系统.基于三维激光雷达<sup>[3]</sup>的定位系统有较高精度,然而传感器系统昂贵,限制了其广泛的应用.通过电机带动单线激光雷达旋转所构建的三维扫描系统成本较低,但完成一次三维扫描的周期长,如何对周期内的扫描数据进行运动补偿成为其关键因素.文献[3]采取 stop-scan-go 的方式,将数据扫描与机器人运动分离,严重影响了机器人运动的连续性;文献[4]利用航位推算法对周期内扫描数据进行运动补偿,仅适用于平坦地形环境;文献[5]利用空间中直线和平面等强结构信息对激光特征进行约束,无法应用于非结构化环境.为了弥补单一传感器信息的局限,融合相机与单线激光雷达的方案也被研究.文献[6–7]利用激光数据构建三维点云,为图像中的特征点赋予深度信息,从而提高定位精度.然而三维点云的重建精度受到累计定位误差的影响,长时间运行过程中特征点深度信息难以准确估计.另一方面,前述研究中均未考虑到由于系统、光照、遮挡等因素对特征点提取与匹配所带来的不确定性影响.

本文提出一种基于相机与摇摆激光雷达融合的定位算法,重点解决在非结构化环境中的机器人位姿估计问题.本文结合激光雷达的深度信息和图像的颜色纹理信息,构建在时序帧间的特征点匹配关系;引入一种置信度评价方法,结合系统误差、数据关联、物体遮挡、特征跟踪等因素对特征点及其匹配关系进行评估,减少低质量特征的影响;最终将定位问题转化为特征点对的加权重投影误差优化问题予以解决.利用小型轮式移动机器人在越野和公园等典型非结构化环境下进行数据采集和定位实验,验证了本文算法的有效性.实验结果表明,与前沿的双目视觉定位算法相比,本文算法可有效提高在非结构化环境中的定位精度.

本文结构如下:第1节围绕室外机器人的定位问题调研国内外研究现状,并分析关键技术难点;第2节给出本文的问题定义,将定位问题转化为特征点重投影误差优化问题;第3节介绍在时序帧间的特征点匹配关系构建方法,包括从激光到图像,图像到激光等两类匹配方法;第4节针对系统误差、数据关联、物体遮挡、特征跟踪等不确定性因素,建立特征点置信度评价方法;第5节给出实验结果与分析,包括越野和公园等两类典型非结构环境的定位结果,与航位推算法,单目或双目视觉定位算法的精度对比,以及置信度评价的有效性分析等;第6节给出总结和今后的研究展望.

## 1 相关研究

目前的无人车或室外移动机器人大多依赖于高精度卫星定位与惯性导航组合系统(俗称 GPS 惯导)并结合轮速仪等设备实现定位.然而在复杂路面,基于轮速仪的航位推算有

University, Beijing 100871 2. Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871

<sup>1</sup><http://www.darpa.mil/grandchallenge05/>

<sup>2</sup><http://www.elrob.org/>

<sup>3</sup>[http://www.mod.gov.cn/power/2016-10/18/content\\_4748815.htm](http://www.mod.gov.cn/power/2016-10/18/content_4748815.htm)

很大误差, 而卫星信号的遮挡, 干扰等严重影响卫星定位精度. 通过匹配相机, 激光雷达等采集的环境信息进行定位, 可提高无人车或移动机器人在复杂环境中的定位精度.

视觉里程计<sup>[1]</sup> 作为一个通用视觉定位算法, 利用前后帧特征点的检测跟踪结果, 解算机器人实际行走里程. 由于缺乏深度信息, 单目视觉里程计无法有效地解算尺度, 成为定位误差的重要原因之一. 双目相机系统<sup>[8]</sup> 在两个相机标定准确的前提下, 利用左右图像生成三维点. 图像中的二维特征点有唯一的三维点进行对应, 减小了求解三维深度的误差, 但相机之间的对应点匹配受到环境光照、阴影等因素影响, 在光照条件受限的情况下, 容易产生定位偏差.

不同于相机, 激光雷达传感器主动发出激光束进行测距, 受到光照的干扰小. 在地面平坦的区域, 直接使用水平放置的二维激光雷达即可进行定位<sup>[9]</sup>. 在地面起伏明显区域, 三维激光雷达可通过点云匹配实现较高精度的定位<sup>[3]</sup>, 但传感器价格昂贵限制了这类方法的应用. 为了构建低成本三维扫描系统, 一些研究者将二维激光雷达安装在旋转电机上构建低成本三维扫描系统. 然而这类系统完成一次三维扫描的周期长, 需对周期内的扫描数据进行运动补偿. Stop-scan-go 的方式将数据扫描与机器人运动分离, 虽避免了扫描周期中的运动影响, 但严重影响了机器人的运行速度<sup>[4]</sup>; 航位推算算法对运动的补偿可在平坦路面得到较高精度, 但无法适用于非平坦路面<sup>[5]</sup>; 空间中直线和平面等强结构信息可对激光特征点进行约束来估计扫描周期内的运动, 但环境的结构假设过强, 无法应用于非结构化环境<sup>[10]</sup>.

在单一传感器无法有效定位的情况下, 多传感器融合的定位方法可提高传感器信息量, 成为解决定位问题的一个可行途径. 一些研究者使用多个不同朝向的单目相机进行融合定位<sup>[11]</sup>, 将一个相机向下倾斜观察地面, 一定程度上减少了光照的影响. 另一些则在单目或双目视觉定位系统中加入惯性导航单元, 有效减少角度误差<sup>[12-13]</sup>, 从而提高定位精度, 但仍然无法避免光照、遮挡等因素干扰下, 低质量特征点对定位结果的影响. 相机与激光雷达两类传感器的融合可有效提高特征的信息维度. 传统相机与激光雷达融合定位算法主要依靠激光传感器生成的三维点云对图像中特征点的深度进行估计, 再使用视觉特征进行定位<sup>[6-7]</sup>, 提高定位精度. 然而, 受到非结构化场景中地面颠簸的影响, 三维点云重建结果随着时间的推移存在明显累计误差, 长期运行的偏差很大程度影响定位结果, 需考虑直接在局部帧间进行相机和激光雷达特征层上的匹配. 另一方面, 相机和激光雷达在融合时引入了更多系统误差, 容易造成错误匹配, 需要分析特征的置信度给出权重, 剔除误差较大的特征点.

综上, 面向轮式移动机器人室外非结构化环境的精确定位, 本文提出一种基于相机和摇摆单线激光融合的视觉定位算法, 构建在时序帧间的特征点匹配关系, 并引入一种结合系统误差、数据关联、物体遮挡、特征跟踪等因素的置信度评价方法, 最终将定位估计转化为特征点对的加权重投影误差优化问题予以解决.

## 2 问题定义

机器人与相机为刚性连接, 两者的坐标系可以通过外部标定参数进行转换. 因此, 机器人的定位求解等价于相机的定位求解. 定义相机坐标系为基准坐标系, 图像帧率作为基准帧率, 第  $k$  帧图像所对应的相机坐标系记为  $C_k$ , 对应的图像坐标系为  $I_k$ . 由于激光与相机的帧率不一致, 将激光帧与最接近的图像帧进行对应. 假设激光帧  $l$  上的一个激光点所

对应的图像帧为  $k$ , 令三维激光点在基准坐标系 (相机坐标系) 下定义为  $\mathbf{P}_k$ , 对应的二维图像上的点为  $\mathbf{p}_k = K\mathbf{P}_k$ , 其中  $K$  为相机的内参矩阵.

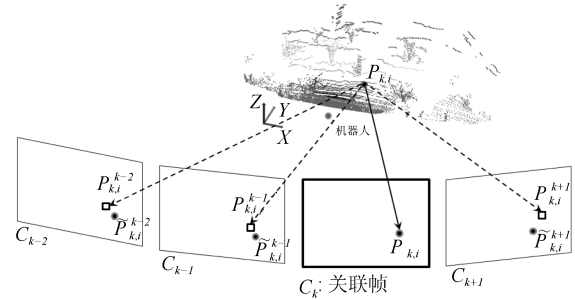


图1 问题定义

Fig.1 Problem statement

如图1所示,  $(\mathbf{P}_{k,i}, \mathbf{p}_{k,i})$  表示在第  $k$  帧下第  $i$  个特征匹配点对. 一方面, 通过跟踪图像点  $\mathbf{p}_{k,i}$ , 可获得其在相邻帧上的图像跟踪点  $\tilde{\mathbf{p}}_{k,i}^j, j = \dots, k-2, k-1, k, k+1, \dots$ . 另一方面, 三维激光点  $\mathbf{P}_{k,i}$  可通过坐标变换投影至对应帧  $j$ . 令  $M_k$  表示第  $k$  帧的相机坐标系  $C_k$  到世界坐标系的坐标变换矩阵

$$M_k = \begin{bmatrix} R_k & \mathbf{t}_k \\ \mathbf{0} & 1 \end{bmatrix} \quad (1)$$

其中,  $R_k$  为  $3 \times 3$  的旋转矩阵,  $\mathbf{t}_k$  为平移向量.  $\mathbf{P}_{k,i}$  可先转换至世界坐标系  $M_k\mathbf{P}_{k,i}$ , 再转换至第  $j$  帧的相机坐标系  $C_j$  下  $M_j^{-1}M_k\mathbf{P}_{k,i}$ , 最后投影到第  $j$  帧的图像坐标系  $I_j$ , 得  $\mathbf{p}_{k,i}^j = KM_j^{-1}M_k\mathbf{P}_{k,i}$ . 若定位无误差, 且特征对应与跟踪准确, 则第  $j$  帧图像上, 基于图像特征点跟踪的结果  $\tilde{\mathbf{p}}_{k,i}^j$  和基于坐标系转换的投影结果  $\mathbf{p}_{k,i}^j$  为同一点. 由于定位误差的存在, 或特征对应与跟踪时的误差影响, 跟踪点与投影点之间存在偏差. 定义第  $k$  帧中的第  $i$  个特征在第  $j$  帧中的跟踪点与投影点之间的偏差为特征的重投影误差, 即

$$d\left(KM_j^{-1}M_k\mathbf{P}_{k,i}, \tilde{\mathbf{p}}_{k,i}^j\right)^2 \quad (2)$$

当特征对应与跟踪准确时, 通过最小化重投影误差, 可进行定位, 即

$$\arg \max_M \sum_k \sum_i \sum_j d\left(KM_j^{-1}M_k\mathbf{P}_{k,i}, \tilde{\mathbf{p}}_{k,i}^j\right)^2 \quad (3)$$

针对特征对应与跟踪的误差影响, 假定误差噪声为白噪声, 通过置信度估计方法调整各个特征的权重, 减少甚至消除特征误差的影响. 因此, 在最小化重投影误差函数中加入权重参数, 定义为

$$\arg \max_M \sum_k \sum_i \sum_j v_{k,i}^j \cdot d\left(KM_j^{-1}M_k\mathbf{P}_{k,i}, \tilde{\mathbf{p}}_{k,i}^j\right)^2 \quad (4)$$

其中,  $v_{k,i}^j$  为第  $k$  帧中第  $i$  个特征在第  $j$  帧中的特征对应与跟踪误差.

上式中所有的坐标系变换矩阵  $M$  都需进行优化. 随着帧数的逐渐增加, 优化也会越来越慢. 实际上, 在线定位系统无需对过期的信息进行优化, 只需优化当前一小段时间内的定位结果. 因此, 最终所需优化的代价函数如下

$$\arg \max_{M_{k_1}, \dots, M_{k_2}} \sum_{k=k_1}^{k_2} \sum_i \sum_{j=k_1}^{k_2} v_{k,i}^j \cdot d\left(KM_j^{-1}M_k\mathbf{P}_{k,i}, \tilde{\mathbf{p}}_{k,i}^j\right)^2 \quad (5)$$

优化的过程为非线性优化, 可使用 LM 算法 (Levenberg-Marquardt<sup>[14]</sup>) 进行迭代优化. 综上, 本文定位算法整体框架如图 2 所示.

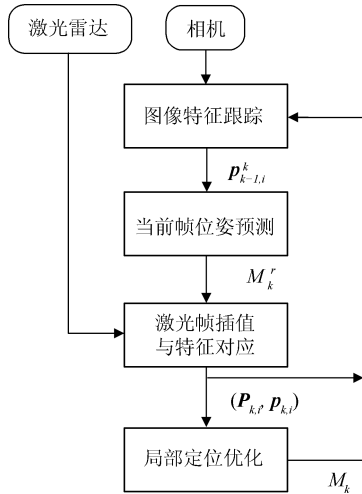


图 2 系统流程

Fig. 2 System overview

- 1) 对于第  $k$  帧图像, 已知  $k-1$  帧图像的特征点集  $\{p_{k-1,i}\}$  与对应三维点集  $\{P_{k-1,i}\}$ , 通过跟踪图像特征点得到当前帧图像特征点集  $\{p_{k-1,i}^k\}$ ;
- 2) 通过最小化当前帧的三维投影点  $\{P_{k-1,i}^k\}$  和二维跟踪点  $\{p_{k-1,i}^k\}$  的重投影误差函数 (2), 获得当前帧  $k$  的相机位姿预测  $M_k^r$ ;
- 3) 通过插值生成  $k$  帧图像相邻激光帧的位姿, 并将激光点投影至对应帧, 构造最新的激光点与图像点特征对应;
- 4) 使用局部若干帧图像, 通过优化代价函数 (5), 优化当前时刻附近的相机位姿.

在上述四个模块中, 模块一图像特征跟踪部分可使用任意的特征点跟踪方法; 模块二当前帧的位姿预测使用当前帧的最小化重投影误差方法, 本质上与模块四一致, 而优化方法可使用开源的非线性优化方法. 因此, 本文的重点内容在于求解模块三的特征对应与构造模块四的优化函数 (5). 这两部分内容分别在第 3 节特征对应求解和第 4 节置信度估计中详述.

### 3 特征对应求解

三维激光点与二维图像点的特征对应关系包含以下难点. 1) 三维激光点在对应该帧的投影很可能是一个无特征或特征较少的区域, 投影点可能无法跟踪或跟踪误差很大; 2) 三维激光点云非常稀疏, 且一帧图像所对应的激光数据也只集中在图像中的一小块区域内, 如果仅使用三维点到二维点的投影, 则特征过少且过于集中. 因此, 在特征对应中不仅需要考虑到三维激光点到二维图像点的正向对应, 还需要考虑到二维图像点到三维激光点的逆向对应. 本文特征对应同时包含激光雷达到图像 (三维到二维) 和图像到激光雷达 (二维到三维) 两部分内容.

#### 3.1 激光雷达到图像的对应

给定一个在第  $k$  帧的三维激光点  $P_{k,i}$ , 通过标定参数与局部运动估计, 可投影至相邻某帧  $j$  帧中, 得到投影点  $p_{k,i}^j$ . 假定标定与同步的参数准确, 其对应帧  $k$  的图像投影点  $p_{k,i}^k$

即为真实  $k$  帧图像的特征对应点  $p_{k,i}$ . 如果对应帧投影点为图像特征点  $p_{k,i}$ , 则可直接进行图像特征跟踪, 获取在  $j$  帧的跟踪点  $\tilde{p}_{k,i}^j$ . 跟踪点  $\tilde{p}_{k,i}^j$  与投影点  $p_{k,i}^j$  即为投影点  $P_{k,i}$  在第  $j$  帧的特征对应, 代入式 (5) 中进行定位.

如果对应帧投影点  $p_{k,i}$  本身所包含的特征信息较弱, 直接对其进行图像特征跟踪会产生较大误差. 为了解决这一问题, 可采用如图 3 所示的处理方法. 假定图像特征点  $p_{k,1}, \dots, p_{k,n}$  为激光投影点  $p_{k,i}$  的邻近点, 通过邻近点的跟踪, 可获得投影点的跟踪. 假定  $\tilde{p}_{k,1}^j, \dots, \tilde{p}_{k,n}^j$  为这些邻近点在第  $j$  帧的跟踪点, 投影点  $p_{k,i}$  在第  $j$  帧的跟踪点可以通过线性插值求解, 得  $\tilde{p}_{k,i}^j$ , 同时建立在第  $j$  帧的投影点与跟踪点的对应关系  $(p_{k,i}^j, \tilde{p}_{k,i}^j)$ .

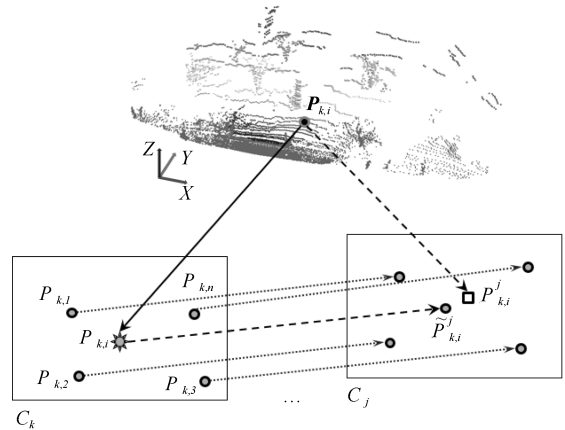


图 3 激光特征点到图像的对应求解

Fig. 3 LiDAR-to-camera correspondence generation

#### 3.2 图像到激光雷达的对应

由于三维激光点的稀疏性, 三维激光点到二维点的对应数量较少, 仅使用这些对应点很难进行定位求解. 而图像的特征点更为稠密, 利用图像到激光雷达的特征对应, 可有效地提高特征点数目.

如图 4 所示, 给定第  $k$  帧图像上的一个特征点  $p_{k,i}$ , 需找到其在三维上的对应点. 通常情况下, 由于激光点的稀疏性, 图像特征点无法直接与激光点对应, 因此需通过插值求解. 对于第  $k$  帧图像, 假设相邻一段时间内  $[k_1, k_2]$  的激光点投影到图像  $I_k$  上, 根据其扫描规律, 不同帧的激光点处于不同的区域. 图 4 所示的一帧激光的投影点可形成一个区域. 在此区域内, 可忽略激光点之间的同步误差. 对于图像  $I_k$  上的点  $p_{k,i}$ , 其邻近点均为第  $j$  帧所对应的激光点投影  $p_{j,1}^k, \dots, p_{j,n}^k$ , 对应的三维点为  $P_{j,1}, \dots, P_{j,n}$ . 通过拟合三维点形成平面或曲面, 再根据平面与图像点所对应射线的角点, 即可求解出  $p_{k,i}$  到第  $j$  帧的三维对应点  $P_{j,i}$ . 将其投影到其他帧  $k'$ , 可以得到对应的投影点  $p_{j,i}^{k'}$ . 另一方面, 图像特征点  $p_{k,i}$  也可跟踪至帧  $k'$ , 得到对应的跟踪点  $\tilde{p}_{j,i}^{k'}$ . 最终建立第  $k$  帧图像特征点在第  $k'$  帧的跟踪点与投影点之间的对应, 代入式 (5) 中进行定位.

### 4 置信度估计

特征对应存在不同程度的误差, 给定一个特征对应, 其误差来源可分为以下四个方面:

- 1) 硬件系统误差 ( $v_h$ ), 包括激光雷达与相机的标定与时钟同步误差;

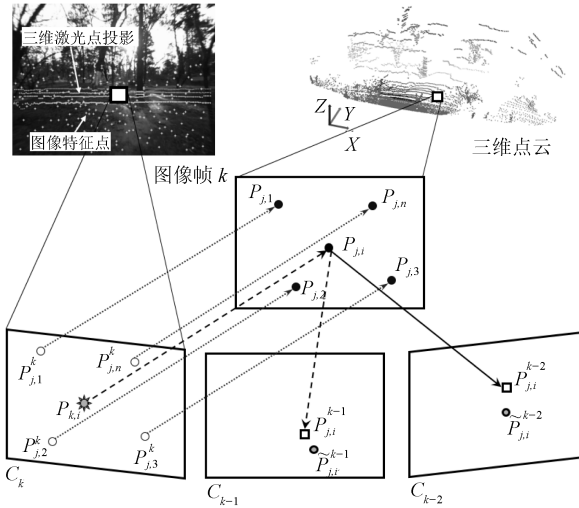


图4 图像特征点到激光的对应求解

Fig. 4 Camera-to-LiDAR correspondence generation

2) 数据关联误差 ( $v_f$ ), 包括数据插值误差与不同源数据间投影的误差;

3) 物体遮挡影响 ( $v_o$ ), 为两个传感器视角不同产生遮挡导致的误差;

4) 特征跟踪误差 ( $v_t$ ), 包括单个图像特征点跟踪误差与多个特征点联合跟踪误差.

通过以上四个方面的置信度分析, 定义特征对应综合置信度如下:

$$v = v_h \cdot v_f \cdot v_o \cdot v_t \quad (6)$$

#### 4.1 系统标定与同步

对于一个使用多传感器的系统, 标定误差和时钟同步误差无法避免. 因此, 系统置信度  $v_h$  包含相机与激光之间的标定产生的置信度  $v_{h(c)}$  与时钟同步产生的置信度  $v_{h(s)}$ .

$$v_h = v_{h(c)} \cdot v_{h(s)} \quad (7)$$

相机与激光之间的标定参数包含旋转矩阵  $R$  和平移向量  $t$ , 其相对误差分别为  $e_R$  和  $e_t$ . 对于一个激光扫描结果, 投影到电机所在坐标系下为  $P = [X, Y, D]^T$ , 转换到相机坐标系下的过程中, 旋转和平移误差产生不同的效果. 旋转标定误差所带来的实际误差与距离  $D$  成正比, 而平移标定误差所带来的实际误差为定值. 因此, 标定置信度函数建立如下:

$$v_{h(c)} = \frac{1}{1 + \lambda_c |e_P|} \quad (8)$$

$$e_P = e_R \cdot D + e_t \quad (9)$$

激光雷达与相机有不同的扫描频率, 为了将两个数据进行融合, 需要进行插值. 在插值的过程中, 两者时间相差越大, 所带来的误差也越大. 因此, 同步置信度定义如下:

$$v_{h(s)} = \frac{1}{1 + \lambda_s |\delta_t|} \quad (10)$$

#### 4.2 数据关联

本文算法所述的特征对应求解包含激光雷达到图像 (三维到二维) 和图像到激光雷达 (二维到三维) 两部分. 一方面为了将两者放入同一个目标函数下, 另一方面由于激光点的

稀疏性所带来的插值误差以及图像投影点的低纹理性带来的跟踪误差, 需进行数据关联上的置信度估计  $v_f$ .

在三维激光点到二维的对应求解上, 给定一个第  $k$  帧的激光点  $P_{k,i}$ , 如果对应投影点  $p_{k,i}$  为图像特征点, 那么两者进行直接关联, 置信度为 1. 如果对应投影点  $p_{k,i}$  不是图像特征点, 需找到邻近区域的特征点  $p_{k,1}, p_{k,2}, \dots, p_{k,n}$ , 用这些特征点的跟踪实现投影点的跟踪. 在数据关联置信度上, 通过计算投影点周围一小块图像块的纹理特征显著度来决定置信度的大小. 我们以  $3 \times 3$  区域内的梯度与亮度值作为描述子, 比较投影点  $p_{k,i}$  与投影点周围像素之间的相似度, 加权平均得到平均相似度  $\bar{s}$ . 置信度定义如下:

$$v_f = \max(\epsilon, 1 - \bar{s}) \quad (11)$$

其中相似度  $\bar{s}$  越高, 投影点与周围点的区分度越低, 置信度  $v_f$  越低.  $\epsilon$  为大于 0 的一个小量, 保证置信度为正.

在二维图像点到三维点的对应求解上, 给定一个第  $k$  帧的图像点  $p_{k,i}$ , 如果需要通过插值第  $j$  帧的激光点  $P_{j,1}, \dots, P_{j,n}$  得到对应三维点, 那么就包含同步以及插值误差. 同步误差包含在系统置信度中, 因此只需考虑插值所产生的误差. 在对应点求解的过程中, 我们采用平面拟合的方式确定三维点所在平面, 并通过平面与图像点对应射线的交点来确定三维点. 因此, 平面拟合的准确度就决定了三维点的置信度. 我们假设在拟合平面后, 这  $n$  个三维激光点与平面的距离分别为  $d_1, \dots, d_n$ , 则数据关联的置信度与距离的均方相关, 定义为:

$$v_f = \frac{1}{1 + \lambda_{f1} \bar{d}^2} \quad (12)$$

$$\bar{d}^2 = \frac{1}{n} \sum_{i=1}^n d_i^2 \quad (13)$$

#### 4.3 物体遮挡

由于激光扫描仪与相机的视角不同, 以及环境内移动物体的干扰, 激光与相机的数据关联上可能产生错误匹配. 图 5 为数据示例. 左上图为相机图像, 左下图为激光投影到图像中的结果. 在正确标定的基础上, 观察立柱上的激光点, 有部分点偏离立柱, 原因是激光与相机在水平位置上有偏差 (激光偏右), 导致两个传感器所观察到的内容有所不同. 图 5 右图对应示意. 其中圆形点为激光和相机正确匹配的点, 方形点为错误匹配的点, 包含激光扫描到背景, 而相机看到前景的点; 以及激光扫描到背景, 相机看到前景的点. 从图中可以看到, 物体边界处所带来的遮挡影响较大, 因此以局部深度差作为置信度估计的相关量.

给定一个图像点  $p$ , 令其对应激光点为  $P$ , 并且深度值为  $D$ . 假设  $\hat{D}$  是图像点周围一小块区域内激光点的平均深度值, 物体遮挡的置信度  $v_o$  与深度差相关, 定义为:

$$v_o = \frac{1}{1 + \lambda_o |\hat{D} - D|} \quad (14)$$

#### 4.4 特征跟踪

对于一个三维激光点  $P_{k,i}$ , 其对应帧的二维投影点为  $p_{k,i}$ . 如果  $p_{k,i}$  是一个特征点, 那么在第  $j$  帧上的跟踪点  $\tilde{p}_{k,i}^j$  是直接通过特征跟踪得到. 跟踪所得的匹配度  $sc$  即为特征跟踪置信度  $v_t$ .

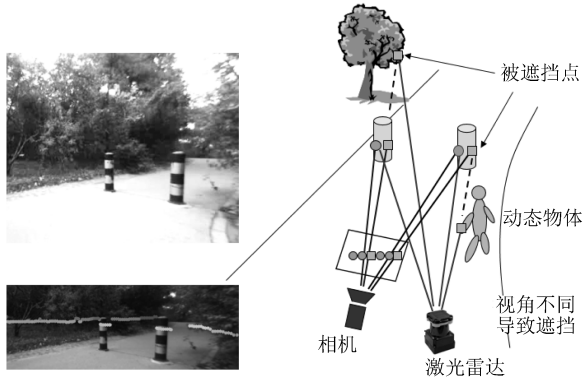


图5 数据融合时, 传感器视点不同造成的物体遮挡示意

Fig. 5 Example of occlusion in camera-LiDAR fusion caused by different point of view

如果投影点  $p_{k,i}$  所在区域为非特征点区域, 需由投影点附近的特征点进行跟踪. 令  $p_{k,1}, \dots, p_{k,n}$  为周围的图像特征点, 在第  $j$  帧中, 跟踪这些图像特征点, 并插值生成的投影点  $p_{k,i}$  在第  $j$  帧上的跟踪点  $\hat{p}_{k,i}^j$ , 如图 3 所示. 影响跟踪置信度的因素包含邻近点的跟踪置信度以及邻近点相对位置的稳定性两方面. 定义  $sc_x$  为第  $x$  个邻近特征点的跟踪匹配度,  $d_{x,k}$  为第  $x$  个邻近点在  $k$  帧中与投影点的距离. 由于在机器人运动过程中, 图像特征点尺度变化明显, 邻近点的绝对位置变化较大, 但相对位置稳定, 因此定义相对距离比例

$$\hat{d}_{x,k} = \frac{d_{x,k}}{\sum_{x=1}^n d_{x,k}} \quad (15)$$

跟踪置信度  $v_t$  定义如下:

$$v_t = \bar{sc} \cdot \frac{1}{1 + \lambda_t \bar{d}^2} \quad (16)$$

$$\bar{d}^2 = \frac{1}{n} \sum_{x=1}^n (\hat{d}_{x,k} - \hat{d}_{x,j})^2 \quad (17)$$

其中,  $\bar{sc}$  是投影点的邻近点的平均跟踪误差,  $\bar{d}^2$  为邻近点与投影点在第  $k$  帧和第  $j$  帧的相对距离的方差.

## 5 实验结果

### 5.1 概述

实验使用小型轮式机器人作为实验平台, 在越野和公园两个典型非结构化环境进行实验, 如图 6 所示. 平台所含设备包括一个 Bumblebee 双目相机, 一个 Hokuyo 单线激光 (UTM-30LX), 一个旋转电机, 一个单点 GPS, 一个低精度 IMU 和轮速编码器等. 其中, 单点 GPS 在所述实验环境无法收到卫星信号, 因此未在实验中使用. Bumblebee 双目相机的分辨率为  $1024 \times 768$ , 最高帧率为 18 Hz. 单线激光水平扫描视场角为 180 度, 角分辨率为 0.5 度, 帧率约为 37 Hz. 电机带动单线激光进行上下周期性摇摆运动, 周期约为 2 秒. 实验过程中, 机器人通过手动遥控的方式进行控制. 考虑到电机摇摆周期的限制, 实验过程中机器人速度不超过 1.2 m/s, 在转弯时, 机器人速度不超过 0.5 m/s.

实验采用本文提出的相机和摇摆激光的融合定位算法, 并考虑了式 (6) 所述的四项置信度. 其中系统和数据关联置信度联合定义为  $S$ , 物体遮挡置信度定义为  $O$ , 特征跟踪置

信度定义为  $T$ . 实验比较了本文算法与双目 libviso<sup>[15]</sup> 视觉里程计算法, 单目或双目 orb-SLAM<sup>[2]</sup> 算法和航位推算法 (odom) 等定位算法的精度误差, 说明定位结果的有效性. 其中双目 libviso 通过左右相机的特征点匹配, 生成单帧的三维点云, 并通过前后帧的特征点匹配, 计算帧间位姿估计. 算法只包含前后帧的匹配定位, 不包含多帧捆集调整. orb-SLAM 通过 orb 特征, 先匹配前后帧特征点生成帧间位姿估计, 再通过局部的捆集调整优化位姿估计, 最后间隔一定帧数选取关键帧, 生成词袋模型用于闭环检测. 本文提出的融合定位算法通过相机与激光雷达的特征对应, 最小化局部重投影误差进行帧间位姿估计. 算法利用激光雷达的距离信息生成局部三维点云, 并将特征点投影至相邻图像帧上用于特征对应与位姿估计, 不显式生成三维点云地图. 因此, 算法只进行局部帧间位姿估计, 不通过地图进行闭环检测.

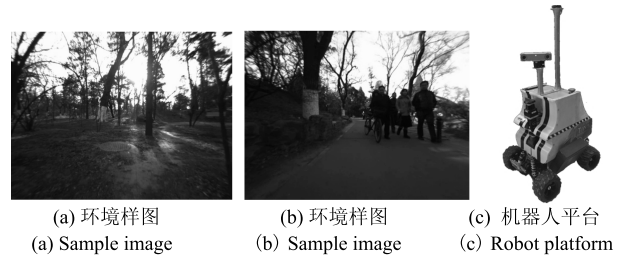


图6 非结构化环境和机器人平台

Fig. 6 Off-road environments and robot platform

为保证各个定位算法所使用数据的一致性, 实验采用数据仿在线回放的方式进行算法测试. 通过数据回放模块, 利用同步采集的相机、激光雷达、电机、惯导、编码器等数据, 模拟数据采集过程中的传感器数据流, 对定位算法进行测试. 该测试方式一方面保证了各个定位算法所使用数据一致性, 另一方面通过数据回放模块和算法模块的分离, 达到仿在线测试目的.

此外, 为了说明本文提出的置信度定义对特征点的筛选和定位的有效作用, 实验中还对比了添加各类置信度的后特征点数目与定位结果比较. 在此, 我们定义不加置信度定义的方法为 our, 添加本文所有置信度内容的方法为 our + SOT. 在置信度的参数  $\lambda$  上, 所有包含  $\lambda$  参数的置信度定义公式, 如式 (8) 等, 形式上均为:

$$v = \frac{1}{1 + \lambda \cdot \delta} \quad (18)$$

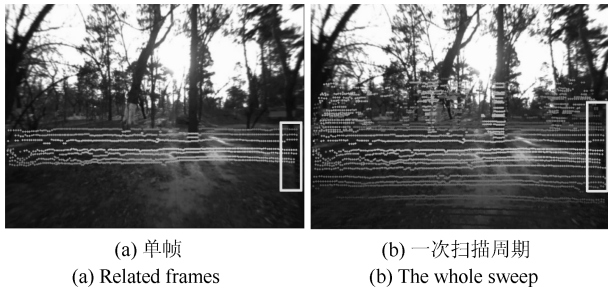
当  $\delta = 0$  时,  $v = 1$  置信度最大. 随着  $\delta$  增加,  $v$  不断减小. 当  $\lambda \cdot \delta = 1$  时,  $v = 0.5$ . 在本实验中, 采用手工调整  $\lambda$  使得  $v = 0.5$  时,  $\delta$  取得了合适的经验值. 具体定义如下:

- 1) 式 (8) 中  $\lambda_c$  的经验值为距离  $D = 30$  m;
- 2) 式 (10) 中  $\lambda_c$  的经验值为时间差  $\delta_t = 100$  ms;
- 3) 式 (12) 中  $\lambda_{fI}$  的经验值为距离方差  $\bar{d}^2 = 0.2$  m<sup>2</sup>;
- 4) 式 (14) 中  $\lambda_o$  的经验值为深度差  $|D - \hat{D}| = 0.5$  m;
- 5) 式 (16) 中  $\lambda_t$  的经验值为相对距离方差  $\bar{d}^2 = 0.2$ .

### 5.2 特征跟踪与置信度

典型非结构化环境如图 6 (a) 和图 6 (b) 所示, 环境无明显结构特性, 地面崎岖不平, 且光照条件一般. 由于车载电脑配置受限, 相机的帧率降至 5 Hz, 相邻的激光帧数据将直接关联到对应图像帧. 单数据投影如图 7 (a) 所示. 由于激光和相机视角不同, 激光的部分观测在图像中被前景遮挡, 但在

单帧激光数据中无法有效提取 (如图 7 (a) 中矩形框区域内, 部分激光观测被图像中的树干遮挡). 当一次电机上下摇摆周期内的激光点全部投影至同一帧图像中时, 可有效地去除被遮挡区域 (如图 7 (b) 矩形框区域所示), 生成物体遮挡置信度图  $O$ .



(a) 单帧 (a) Related frames  
(b) 一次扫描周期 (b) The whole sweep

图 7 激光点投影到视频结果  
Fig. 7 Projection results of LiDAR points

特征置信度估计是本方法的另一重要因素. 图 8 所示为特征跟踪与置信度估计结果. 所示特征点的大小表示跟踪时间的长 (越大表示跟踪越久), 而亮度则表示置信度短 (越亮表示置信度越高). 从图中可以看出, 特征点主要集中在图像中下方区域, 也是机器人导航的最重要区域.



图 8 特征跟踪与置信度结果  
Fig. 8 Feature tracking and confidence measure results

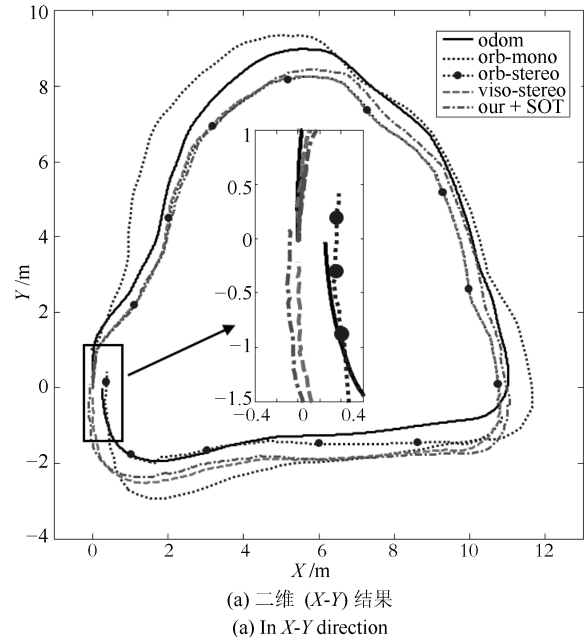
### 5.3 定位结果

#### 5.3.1 越野环境

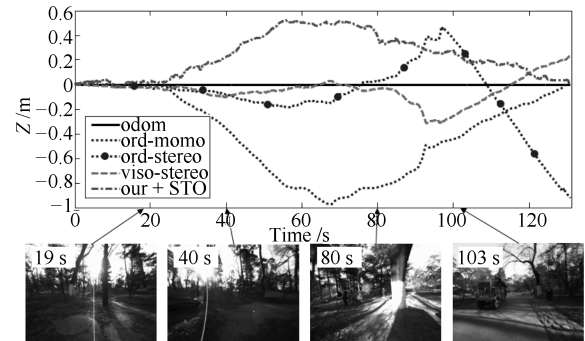
越野实验环境如图 6 (a) 所示, 越野环境地面起伏不平, 光照条件较差. 由于非结构化环境下轨迹真值标注的困难性, 实验顺时针行驶一段约 35 米的闭合轨迹, 以轨迹首尾的闭合误差作为度量方式. 比较航位推算法 (odom), orb-SLAM (含单目 orb-mono 和双目 orb-stereo), 双目 libviso (viso-stereo) 和本文含置信度估计的融合定位算法 (our + SOT) 等方法结果, 如图 9 所示.

在图 9 (a) 所示二维闭合误差上, 航位推算法在地面起伏不平的路面下精度较低, 闭合误差大; 单目视觉定位算法由于不包含尺度信息, 闭合误差较大; 本文相机与激光融合定位整体轨迹与双目视觉定位算法类似, 在轨迹首尾闭合精度上略高于双目视觉定位算法. 考虑到越野环境的三维运动, 仅使用二维闭合误差无法准确衡量定位的精度. 因此, 实验

还对比了各个方法之间的高度 ( $Z$ ) 值, 如图 9 (b) 所示. 从图中结果来看, 本文算法在最终高度闭合估计上优于其他算法 (如图中点虚线所示).



(a) 二维 ( $X$ - $Y$ ) 结果  
(a) In  $X$ - $Y$  direction



(b) 高度 ( $Z$ ) 结果  
(b) In  $Z$  direction

图 9 越野环境定位结果

Fig. 9 Localization results in off-road environment

#### 5.3.2 公园环境

公园环境如图 6 (b) 所示, 相比于越野环境, 公园环境路面平整度增加, 加强了运动约束, 但行人较多, 给特征点的对应带来更多干扰. 实验车逆时针行驶一段约 250 米的闭合轨迹, 结果如图 10 所示.

从图 10 (a) 中的二维闭合误差分析, 航位推算法由于地面起伏, 闭合精度无法有效提高; 单目视觉定位算法无法有效保证全局尺度的一致性, 带来较大定位偏差; viso-stereo 算法在第一次拐弯后误差一直增加, 最终的闭合误差超过 10 米; orb-stereo 算法虽然在闭合上精度上较高, 但第一次拐弯处的形状与其他方法有很大偏差; 本文方法结果在整体形状上与其他的方法类似, 最终闭合精度略高于 orb-stereo 算法.

从图 10 (b) 中高度信息进行分析, 整体轨迹起伏明显, 达到 10 米以上. 本文算法与 orb-stereo 虽然在闭合处高度接近, 但机器人在  $p_1$  至  $p_2$  间行驶的过程中大致在平路上, 而 orb-stereo 高度结果产生明显起伏.

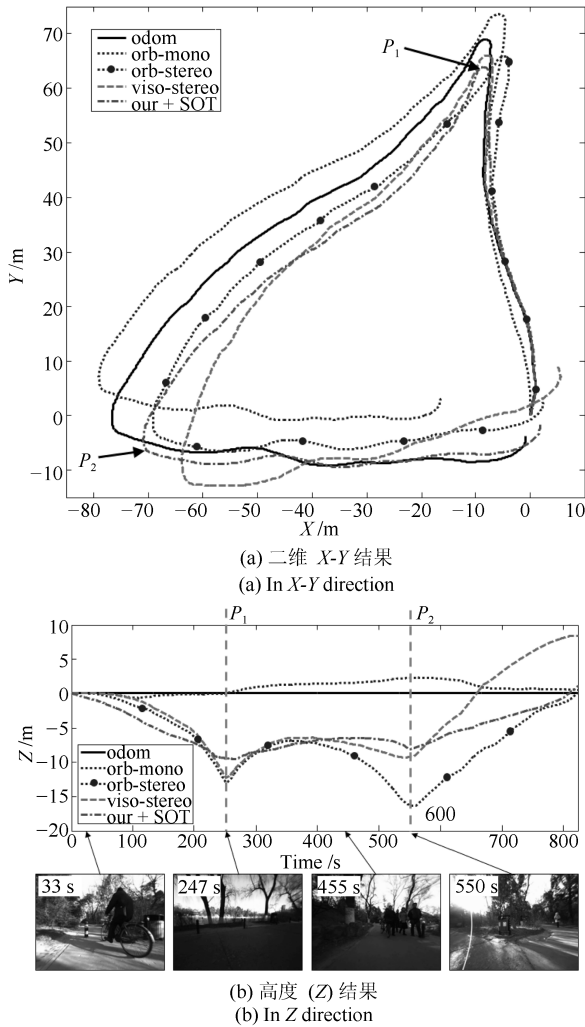


图 10 公园环境定位结果  
Fig. 10 Localization results in park area

5.3.3 数值结果分析

通过比较闭环起始点和终止点之间的偏差, 定量分析定位精度, 两组实验的数值结果如表 1 所示. 航位推算算法无法估计三维精度, 在二维精度上与双目定位算法相似. 单目视觉定位算法由于尺度估计不准确, 偏差较大, 尤其是公园环境. 在越野环境中, viso-stereo 精度较高; 在公园环境中, orb-stereo 精度较高, 原因可能是 libviso 的特征描述子受到光照干扰较小, 而 orb-SLAM 具有三维重建与定位优化, 行人的影响较小. 本文算法在两处实验的精度均高于双目视觉定位算法, 说明了算法的有效性.

表 1 不同定位算法闭环误差 (m)

Table 1 Results of different localization algorithms (m)

	越野环境闭环误差		公园环境闭环误差	
	二维	三维	二维	三维
odom	0.2528	—	3.9004	—
orb-mono	0.9937	0.9937	16.9097	16.9231
orb-stereo	0.5565	1.1419	3.8488	3.8708
viso-stereo	0.2501	0.3644	10.1943	13.1358
our+SOT	<b>0.1031</b>	<b>0.1094</b>	<b>2.3492</b>	<b>2.5889</b>

5.4 重投影误差与定位结果相关性

本文使用最小化重投影误差的方式估计机器人位姿, 并使用特征的置信度估计对特征进行约束. 为了证明重投影误差与定位结果的相关性, 在越野环境实验中, 我们对不同算法的重投影误差和定位结果进行整理, 结果如表 2 所示. 此处, 硬件系统和数据关联置信度统一为系统置信度, 记做  $S$ , 物体遮挡置信度记做  $O$ , 特征跟踪置信度记做  $T$ .

表 2 重投影误差与定位结果相关性  
Table 2 Correlation of re-projection errors and localization error

	特征点数	投影误差 (pixel)	定位误差 (m)
orb-stereo	627.00	6.33	1.1419
viso-stereo	621.72	7.69	0.3644
our	612.09	4.32	0.5126
our + $S$	568.07	4.23	0.4299
our + $S$ + $O$	318.08	3.77	0.3047
our + $S$ + $T$	295.92	3.74	0.3100
our + $S$ + $O$ + $T$	<b>244.52</b>	<b>3.40</b>	<b>0.1094</b>

首先, 对于双目定位结果 (orb-stereo 和 viso-stereo), 由于激光与相机本身存在标定与配准误差, 而定位结果 (双目) 和用于重投影误差解算的传感器 (单目 + 激光) 不一致, 导致重投影误差较大. 对于本文基准算法, 即不加任何置信度估计的算法, 特征点数目与双目结果接近, 平均重投影误差比双目小, 但定位误差却更大, 原因是激光雷达与视觉融合的系统误差比双目系统高. 在加入系统误差置信度分析后 (our +  $S$ ), 特征点数目少量减少, 重投影误差也相应减少, 定位误差也随之略微减少. 通过逐渐加入遮挡置信度 ( $O$ ) 和跟踪置信度 ( $T$ ), 整体特征点数目减少至 244.52, 重投影误差减少至 3.40, 定位误差减少至 0.109 米, 与双目系统相比有了显著的提高.

综上, 尽管激光雷达与相机融合定位系统的硬件误差比双目相机大, 在加入置信度分析后, 基于相机与激光的定位算法减少了无效的特征点数据, 有效地减少了重投影误差, 提高了定位精度.

6 总结与展望

针对非结构化环境的定位难点, 本文提出了一种基于相机和摇摆激光雷达融合的定位算法. 算法结合激光雷达的深度信息以及图像中的颜色纹理信息, 通过激光雷达到图像, 图像到激光雷达的两种映射, 构建在时序帧间的特征点匹配关系; 引入一种结合系统误差、数据关联、物体遮挡、特征跟踪等因素的置信度评价方式, 量化评价特征点匹配中的不确定性, 最终将定位问题转化为特征点对的加权重投影误差优化问题予以解决.

实验利用小型轮式移动机器人在越野和公园等典型非结构化环境下进行的数据采集和结果分析, 对提出的基于相机和摇摆激光雷达融合的定位算法进行验证. 结果表明, 与前沿的单目或双目视觉定位算法相比, 本文算法可提高在非结构化环境中的定位精度. 与此同时, 实验还进行置信度与定位精度的相关性分析, 说明本文置信度评价方法可有效地抑制数据融合, 特征提取与匹配等不确定性因素对定位精度的影响.

考虑到在不同环境 (包括路面、卫星定位、环境特征等条件) 不同定位方案各有其优劣, 后续的工作将研究基于环境

评估的环境自适应融合定位解决方案。另一方面, 由于激光摇摆周期过长, 一定程度上限制了机器人的移动速度, 尤其是转弯速度, 今后的研究也将尝试摇摆激光的控制中加入反馈机制, 确保激光雷达的扫描范围处于最重要的区域内。

### References

- 1 Nistér D, Naroditsky O, Bergen J R. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 2006, **23**(1): 3–20
- 2 Mur-Artal R, Montiel J M M, Tardós J D. ORB-SLAM: a versatile and accurate monocular slam system. *IEEE Transactions on Robotics*, 2015, **31**(5): 1147–1163
- 3 Kuramachi R, Ohsato A, Sasaki Y, Mizoguchi H. G-ICP SLAM: an odometry-free 3D mapping system with robust 6dof pose estimation. In: Proceedings of the 2015 IEEE International Conference on Robotics and Biomimetics (RO-BIO). Zhuhai, China: IEEE, 2015. 176–181
- 4 Weingarten J, Siegwart R. EKF-based 3D SLAM for structured environment reconstruction. In: Proceedings of the 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Edmonton, AB, Canada: IEEE, 2005. 3834–3839
- 5 Cole D M, Newman P M. Using laser range data for 3D SLAM in outdoor environments. In: Proceedings of the 2006 IEEE International Conference on Robotics and Automation (ICRA). Orlando, FL, United States: IEEE, 2006. 1556–1563
- 6 Zhang J, Kaess M, Singh S. Real-time depth enhanced monocular odometry. In: Proceedings of the 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. Chicago, IL, United States: IEEE, 2014. 4973–4980
- 7 Zhang J, Singh S. Visual-lidar odometry and mapping: low-drift, robust, and fast. In: Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA). Seattle, WA, United States: IEEE, 2015. 2174–2181
- 8 Grimes M, LeCun Y. Efficient off-road localization using visually corrected odometry. In: Proceedings of the 2009 IEEE International Conference on Robotics and Automation (ICRA). Kobe, Japan: IEEE, 2009. 2649–2654
- 9 Harrison A, Newman P M. High quality 3D laser ranging under general vehicle motion. In: Proceedings of the 2008 IEEE International Conference on Robotics and Automation (ICRA). Pasadena, CA, United States: IEEE, 2008. 7–12
- 10 Zhang J, Singh S. LOAM: lidar odometry and mapping in real-time. In: Proceedings of the 2014 Robotics: Science and Systems Conference. Berkeley, CA, USA, 2014
- 11 González R, Rodríguez F, Guzmán J L. *Autonomous Tracked Robots in Planar Off-Road Conditions: Modelling, Localization, and Motion Control*. Cham: Springer, 2014.
- 12 Forster C, Carlone L, Dellaert F, Scaramuzza D. On-manifold preintegration for real-time visual-inertial odometry. *IEEE Transactions on Robotics*, 2017, **33**(1): 1–21
- 13 Mur-Artal R, Tardós J D. Visual-inertial monocular SLAM with map reuse. *IEEE Robotics and Automation Letters*, 2017, **2**(2): 796–803
- 14 Moré J J. The levenberg-marquardt algorithm: implementation and theory. In: Proceedings of the 1978 Biennial Conference Held at Dundee. Berlin, Heidelberg: Springer, 1978. 105–116
- 15 Geiger A, Ziegler J, Stiller C. StereoScan: Dense 3d reconstruction in real-time. In: Proceedings of the 2011 IEEE Intelligent Vehicles Symposium (IV). Baden-Baden, Germany: IEEE, 2011. 963–968

俞毓锋 北京大学信息科学技术学院博士研究生. 2011 年获得北京大学信息科学技术学院学士学位. 主要研究方向为智能车与计算机视觉.

E-mail: yuyufeng@pku.edu.cn

(YU Yu-Feng Ph.D. candidate at the School of Electronics Engineering and Computer Science, Peking University. He received his bachelor degree from School of Electronics Engineering and Computer Science, Peking University in 2011. His research interest covers intelligent vehicle and computer vision.)

赵卉菁 北京大学机器感知与智能教育部重点实验室研究员. 1999 年获得日本东京大学土木工程专业博士学位. 主要研究方向为智能车, 机器感知与移动机器人. 本文通信作者.

E-mail: zhaohj@cis.pku.edu.cn

(ZHAO Hui-Jing Professor at the Key Laboratory on Machine Perception (Ministry of Education), Peking University. She received her Ph.D. degree from the University of Tokyo, Japan, in 1999. Her research interest covers intelligent vehicle, machine perception and mobile robot. Corresponding author of this paper.)