

基于改进 M-ORB 的视觉 SLAM 直接-闭环检测算法

李伟^{1,2}, 任孟瀚^{1,2}, 黄威豪^{1,3}, 杜晓玉^{1,2}, 周毅^{1,3}

(1. 河南大学计算机与信息工程学院, 河南 开封 475004;

2. 河南省车联网协同技术国际联合实验室, 河南 开封 475004;

3. 鹰驾科技(深圳)有限公司, 广东 深圳 518052)

摘要: 直接法 SLAM 不在前端提取图像特征点, 使得后端无法生成视觉词袋, 这导致大部分直接法 SLAM 无法使用带有词袋模型的闭环检测来消除系统的累积误差。针对此问题, 提出一种基于改进 M-ORB 的视觉 SLAM 直接-闭环检测算法, 生成闭环检测所需的词袋模型, 然后采用词频-逆文档频率算法对视觉词典树各个子节点中的视觉单词进行自适应分配权重, 得到场景信息的准确表述。在 TUM、KITTI 两种公开数据集上进行了对比实验, 实验结果表明, 所提出的算法能够有效检测到闭环, 并在不降低准确性的同时, 提高 SLAM 的实时性与鲁棒性。
关键词: 视觉 SLAM; 闭环检测; 词袋模型; 词频-逆文档频率

中图分类号: TP242

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202148

Improved M-ORB based direct-loop closure detection algorithm for visual SLAM

LI Wei^{1,2}, REN Menghan^{1,2}, HUANG Weihao^{1,3}, DU Xiaoyu^{1,2}, ZHOU Yi^{1,3}

1. School of Computer and Information Engineering, Henan University, Kaifeng 475004, China

2. International Joint Research Laboratory for Cooperative Vehicular Networks of Henan, Kaifeng 475004, China

3. Eagle Drive Technology (Shenzhen) Co., Ltd., Shenzhen 518052, China

Abstract: Most kinds of direct methods do not extract image feature points in the front end of SLAM system, resulting in that they cannot use loop closure detection with bag-of-words models to eliminate the cumulative error of the system. To resolve this problem, an improved mature-oriented fast and rotated BRIEF (M-ORB) based direct-loop closure detection algorithm for visual SLAM was proposed, which designed an improved M-ORB, generated the bag of words model required for loop closure detection, and then used the term frequency-inverse document frequency (TF-IDF) algorithm to adaptively assign weights to the visual words in each sub-node of the dictionary tree. Finally, an accurate representation of the scene information was obtained. In the end, the proposed algorithm and conducted comparative experiments were verified through two public data sets TUM and KITTI. The experimental results show that the algorithm proposed in this paper can effectively detect the loop closure, and has better real-time and robustness performance without reducing the accuracy.

Key words: visual SLAM, loop closure detection, bag of words model, term frequency-inverse document frequency

1 引言

近年来, 随着移动机器人技术的快速发展, 视

觉同时定位与建图 (simultaneous localization and mapping, SLAM) 受到广泛的关注。早期视觉 SLAM 因为计算量大、实用性比较差, 发展受到阻碍, 计

收稿日期: 2021-01-29; 修回日期: 2021-02-26

通信作者: 周毅, zhouyi@henu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61701170); 河南省科技发展计划项目 (No.202102210327, No.202102310198, No.202102210412)

Foundation Items: The National Natural Science Foundation of China (No.61701170), The Program for Science and Technology Development of Henan Province (No.202102210327, No.202102310198, No.202102210412)

算机与非结构化视听觉感知信息的理解能力和海量异构信息的处理效率出现了一定的瓶颈^[1],但是随着理论与科技的发展,计算困难逐渐被克服。为了使移动机器人拥有自主导航的能力,2007年以来,对视觉 SLAM 的研究取得了突破性的进展。与激光 SLAM 相比,视觉 SLAM 感知能力和重定位能力更强,且具有低成本、效率高等优点,逐渐成为研究热点^[2]。视觉 SLAM 通过相机获得图像,通过几何方法估计相邻图像的运动,并在后端对其进行优化。闭环检测是后端优化的关键技术,能够较好地消除前端和后端的累积误差。因此,闭环检测是视觉 SLAM 的重要组成部分,它的存在对于视觉 SLAM 系统在室外大范围场景下的长期运行是十分重要且必要的^[3]。

目前主流的闭环检测算法主要为基于外观的图像方法^[4],此方法不需要调用前端的位姿估计信息,仅通过两帧之间的相似度检测,判断是否出现闭环。此方法脱离了前端的误差累积,使闭环检测成为系统中相对独立的线程。该方法被提出以后,很快成为闭环检测中的主流方法,被广泛应用于各类视觉 SLAM 系统。

基于外观的图像方法主要可分为基于神经网络的图像识别模型^[5-9]和视觉词袋(bag of visual word, BoVW)模型^[10],二者均可解决图像的相似度对比问题。

基于神经网络的图像识别模型能够更好地提取图像中的整体特征,并且部分特征具有较好的环境不变性和语义识别能力。Chen Z T 等人^[5]利用神经网络的卷积特征代替图像的全局特征,Bai D D 等人^[6]提出利用深度学习网络模型的学习技术代替原特征,在上述方法中,图像特征均提取自通用的神经网络模型。Gomez-Ojeda R 等人^[7]设计了一种针对闭环检测所需场景的神经网络模型,在此基础上,Chen Z T 等人^[8]在大量数据集上进行了训练,上述网络架构均依赖于监督式学习,训练集需要大量带标签的图像。Merrill N 等人^[9]设计了一种针对闭环检测的无监督式深度学习网络架构,核心思想是从网络中提取更轻巧、更紧凑的卷积特征。虽然以上方法相较于传统方法有较高的准确性和鲁棒性,但是会有高维度的特征向量,使得计算量大幅度增加,而且神经网络的训练不能被应用在常搭载 SLAM 的移动平台,因为不合适的模型被应用在闭环检测中容易形成过拟合问题。

Cummins M 等人^[11]较早把视觉词袋模型引入闭环检测中,通过计算视觉单词之间的联合概率分布,得到闭环的后验概率,从而判断闭环是否可能发生。Williams B P 等人^[12]考虑了不同视觉单词之间的联系,提出基于 Chow-Liu 树的闭环检测概率模型;Mathieu L 等人^[13]利用离散贝叶斯估计等方法,估算出不同场景中闭环出现的概率,从而只需要对其中发生闭环概率较大的场景进行检测,提高了实时性;李博等人^[14]提出了一种基于视觉词典树的金字塔词频-逆文档频率(term frequency-inverse document frequency, TF-IDF)权重自适应分配方法,有效区分了视觉词典树的不同层次节点的影响。Gao X 等人^[15]提出了一种基于闭环检测的直接稀疏里程计(direct sparse odometry, DSO)方法,将直接稀疏里程计扩展到有闭环检测和位姿图优化的单目视觉 SLAM 系统。

上述方法在进行闭环检测时,均需要提取图像中的特征点。直接法不在前端提取图像特征点信息,而对整张图像的像素点进行提取和处理,通过优化像素点的光度误差来完成相机的位姿估计和地图构建。因为不提取图像特征点,所以直接法在鲁棒性和实时性上比特征点法表现优异,但由于没有生成视觉词典,需要对每帧图像进行额外的特征提取,增加了闭环检测的计算量,降低了系统的实时性。

针对上述问题,本文提出一种基于改进 M-ORB (mature-oriented fast and rotated BRIEF) 的视觉 SLAM 直接-闭环检测(direct-loop closure detection, D-LCD)算法。首先通过改进 M-ORB 特征点提取方法,让后端线程能够充分调用前端线程处理后的像素点信息,在后端生成闭环检测所需的视觉词典;然后采用 TF-IDF 算法对视觉词典树各个子节点中的视觉单词进行自适应分配得分权重,得到场景信息的准确表述;最后通过闭环检测算法级实验和系统级实验进行验证,证明本文提出的算法的有效性。

2 系统框架

视觉 SLAM 系统的经典框架包含 5 个线程,如图 1 所示。根据前端对图像信息的处理方式,可将现行 SLAM 系统分为两大类:基于特征点法的 SLAM 系统、基于直接法(不提取特征点)的 SLAM 系统。

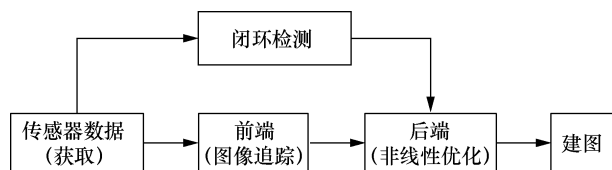


图1 视觉SLAM系统的经典框架

由图1可知，传统的闭环检测算法从传感器数据模块获取图像信息，此种做法的优点是使得闭环检测的图像数据来源独立于前端线程，数据来源不受前端准确率的干扰，缺点是对于整个系统来说，数据处理存在重叠，计算量较大，不利于整个系统的实时性。一般而言，基于直接法的SLAM系统基于灰度不变假设，前端通过对像素点进行追踪与处理，具有较高的实时性与鲁棒性，但无法直接生成词袋模型，若闭环检测还采用传统的数据获取方式，将极大地增加计算量，对整个系统的实时性造成影响。

本文提出的闭环检测算法采用了一种新型的数据获取方式，如图2所示。通过让闭环检测与前端线程紧密耦合，直接获取前端处理后的图像信息，较大地节约了系统的计算资源，提高了实时性；也加强了前端、后端和闭环检测3个主要线程之间的耦合度，在一定程度上提升了系统的鲁棒性。

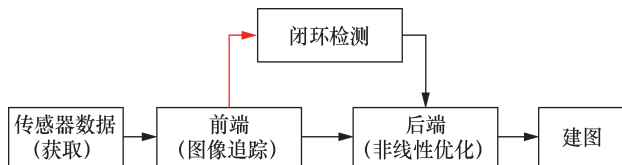


图2 系统框架

3 基于改进 M-ORB 的直接-闭环检测算法

本文提出的基于改进 M-ORB 的视觉 SLAM 直接-闭环检测算法由 M-ORB 特征点提取、BoVW 模型生成、TF-IDF 自适应加权以及闭环验证环节（时间连续性验证、空间一致性验证）组成，算法流程如图3所示，主要分为以下3个步骤。

- **M-ORB 特征点提取：**提取目标关键帧上的 M-ORB 特征，并得到特征点。
- **BoVW 模型生成：**通过对特征点进行聚类，创建目标帧的视觉词典。
- **TF-IDF 自适应加权：**利用 TF-IDF 算法对视觉词典树各个子节点中的视觉单词进行自适应分配权重。

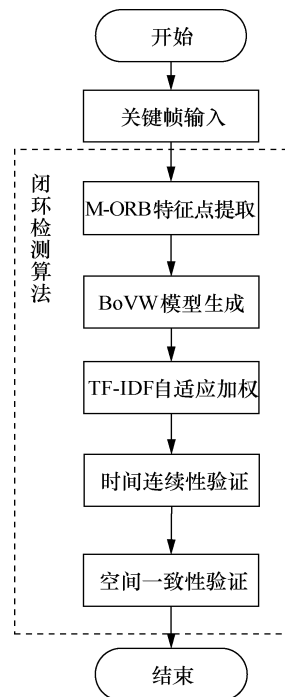


图3 算法流程

3.1 M-ORB 特征点提取

基于直接法的 SLAM 系统首先对图像中的像素点进行提取，然后在相邻帧之间通过最小化光度误差对同一像素点进行追踪，并在此过程中解算出相机的运动姿态，最后将追踪到的成熟特征点 (mature point) 在局部地图中激活。因此整个系统无论是前端还是后端，均不涉及图像特征点提取。若直接在后端对所有关键帧进行特征提取，必将造成计算量激增，对整个系统的实时性造成挑战。

为了解决此问题，本文在 ORB 特征点^[16]的基础上，提出了 M-ORB 特征点提取算法。相比 ORB 特征点，M-ORB 特征点通过与前端追踪线程紧密耦合，调用前端线程中的已经成熟的地图点信息，并在此基础上提取 ORB 特征点。此举在很大程度上减轻了闭环检测线程的计算负荷，并且为后续的 BoVW 模型的构建提供了基础。前端线程通过对像素点提取、追踪、匹配来完成相机位姿与地图点信息的估算，因此前端可分为未成熟点 (immature point) 提取、未成熟点追踪、未成熟点激活3个步骤。

3.1.1 未成熟点提取

由于每帧图像上像素点数量巨大，不可能对所有像素点进行提取与追踪。像素点的提取需要代表图像特征信息，同时也要考虑到后端优化的便捷性，该点应具有一定的像素梯度。对此，像

素点的提取应遵循以下两个标准：

- 分布均匀良好；
- 具有足够的像素梯度。

为了满足上述要求，本节使用一种自适应梯度阈值提取法来完成对像素点的提取操作，并将提取到的像素点命名为未成熟点，提取流程如下。

第一步：自适应阈值计算。将图像划分成 $M \times M$ 个区域（本文中 M 取值为 32），对每个区域计算阈值 g ：

$$g = \bar{g} + g_{th} \quad (1)$$

其中， $\bar{g}$ 为该区域所有像素点的中值绝对梯度， g_{th} 为全局梯度，一般为常量，本文中设为 7。

第二步：梯度计算。在每个 $M \times M$ 区域内选取像素梯度值最大的前 N 个点（本文中 N 取值为 5），本文中图像梯度值定义如式（2）所示：

$$G(x, y) = dx(i, j) + dy(i, j) \quad (2)$$

其中， $\begin{cases} dx(i, j) = I(i+1, j) - I(i, j) \\ dy(i, j) = I(i, j+1) - I(i, j) \end{cases}$ ， $dx(i, j)$ 为 x 轴相邻像素值的差值， $dy(i, j)$ 为 y 轴相邻像素值的差值， $I(i, j)$ 为坐标为 (i, j) 的像素点的像素值。

若这些像素点的梯度超过阈值 g ，则将其选为未成熟点，否则不在该区域内提取像素点。该方法可以保证在一些像素点梯度不够明显的场景中（如缺少纹理的白墙上）依然能够提取出一定的像素点，避免出现像素点缺失的情况。

3.1.2 未成熟点追踪

完成未成熟点的提取工作后，需要在相邻帧之间进行追踪与处理。由于未成熟点的深度未知，系统会在相邻帧之间对未成熟点沿极线方向进行离散化搜索匹配，如图 4 所示。

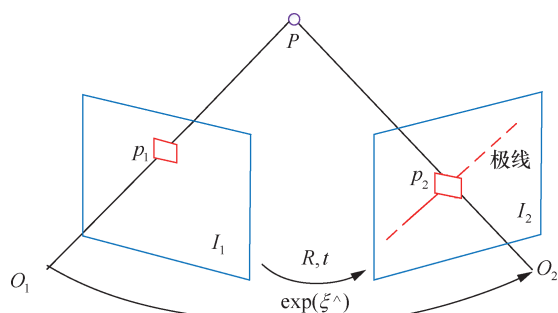


图 4 未成熟点追踪

令相邻帧 I_1 、 I_2 中的两个像素点在非齐次像素坐标系中坐标为 p_1 、 p_2 ； P 为三维空间中的真实

点，在 I_1 和 I_2 中的深度为 Z_1 和 Z_2 。 I_1 和 I_2 之间的旋转和平移为 R 和 t ，对应的李代数为 ξ ，相机内参数为 K ， u 和 v 为像素点在 x 轴、 y 轴上的运动速度，可得 p_1 和 p_2 如下：

$$p_1 = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_1} KP \quad (3)$$

$$p_2 = \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_2} K(RP + t) = \frac{1}{Z_2} K(\exp(\xi^{\wedge})P) \quad (4)$$

由于直接法没有特征匹配，无法准确确定 p_2 的像素位置，直接法通过相机位姿优化来寻找最优的 p_2 与 p_1 相匹配，此过程也是未成熟点 p_1 的追踪收敛过程。直接法基于灰度不变假设，认为同一像素点在不同时刻的灰度值相同。 p_2 与 p_1 之间的光度误差可表示为：

$$e = I_1(p_1) - I_2(p_2) \quad (5)$$

其中， e 为向量，优化目标为该光度误差的二范数。对于 N 个空间点 p_i ，整个相机位姿优化问题可表示为：

$$\begin{cases} \min_{\xi} J(\xi) = \sum_{i=1}^N e_i^T e_i \\ e_i = I_1(p_1, i) - I_2(p_2, i) \end{cases} \quad (6)$$

光度误差相对于李代数形式的雅可比矩阵为：

$$J = -\frac{\partial I_2}{\partial u} \frac{\partial u}{\partial \delta \xi} \quad (7)$$

其中， $\frac{\partial I_2}{\partial u}$ 为 u 处的像素梯度， $\frac{\partial u}{\partial \delta \xi}$ 为相机坐标系下的三维点的导数。

对于每个关键帧上的 N 个像素点，通过该方法计算优化目标的雅可比矩阵，然后使用高斯-牛顿法迭代求解相机位姿和成熟地图点。

3.1.3 未成熟点激活

在未成熟点提取线程中，像素梯度值没有超过阈值的点将被永久丢弃，其余的点被选为未成熟点。未成熟点经过追踪线程后，其中深度信息能够收敛的未成熟点会被插入局部建图中激活，转化为成熟地图点，深度信息未能收敛的点会随着后端优化窗口的滑动，被边缘化、移除，如图 5 所示。成熟地图点具有确切的三维坐标信息，能够与相机位姿相匹配，准确反映周围的环境信息。随着相机的

运动,越来越多的成熟地图点被插入局部建图中,构成了局部三维点云地图,局部三维点云地图经过后端优化后,最终得到全局三维点云地图。

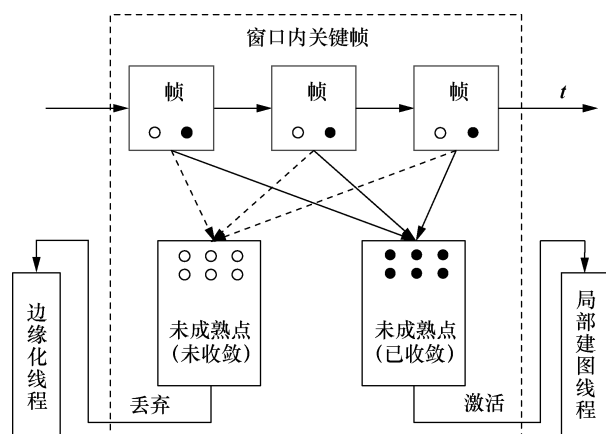
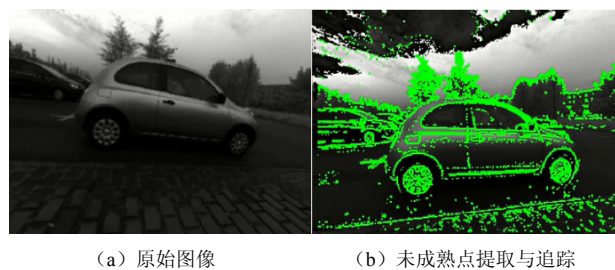


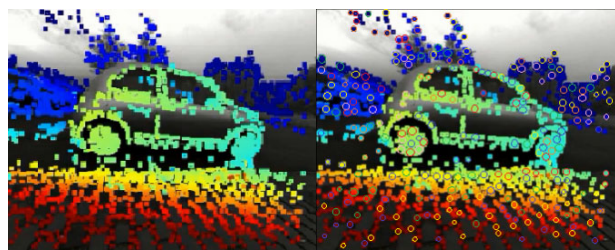
图5 地图点激活

M-ORB 特征点通过与前端线程紧密耦合,调用其中的成熟地图点信息,并在此基础上进行 ORB 特征点提取,整个过程如图 6 所示。其中原始图像图 6(a)经过未成熟点的提取与追踪后得到图 6(b),未成熟点被激活后得到成熟点地图,即图 6(c),在此基础上提取 M-ORB 特征点信息,最后得到图 6(d)。



(a) 原始图像

(b) 未成熟点提取与追踪



(c) 成熟点地图

(d) M-ORB 特征点

图6 M-ORB 特征点提取过程

为了检验此方法在复杂场景下能否顺利调取前端信息,提取 M-ORB 特征点,本文在 TUM^[17]数据集 Sequence_01~Sequence_30 系列中进行了大量实验。实验结果表明,在室内(外)各类场景下,

此方法均能够顺利提取 M-ORB 特征点,如图 7 所示。

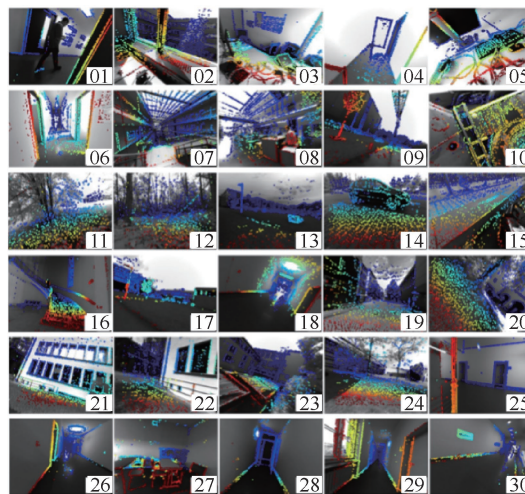


图7 Sequence_01~Sequence_30 场景提取实验

3.2 视觉词典的创建

视觉词典模型由于具有简单、高效等优点,被广泛应用于图像检索和计算机视觉领域。视觉词典的生成过程如图 8 所示,一般分为以下 3 个步骤。

步骤 1: 特征提取,即从图像中提取具有代表性的局部特征信息。

步骤 2: 视觉词典创建,即对特征信息进行聚类,生成视觉词典。视觉词典由大量视觉单词组成,每个视觉单词可看作图像中某一类特征的集合,视觉词典模型的生成类似于一个聚类过程。使用 K-Medoids 聚类算法对其图像特征进行聚类,通过迭代计算,将样本中的 N 个特征点划分为 K 个聚类,所有聚类均满足同一聚类中相似度较高、不同聚类相似度较低的要求。最后得到一个包含 K 个视觉单词的视觉词典。

步骤 3: 相似度计算,即通过视觉词典量化图像间相似程度。在步骤 2 中通过聚类得到了视觉词典,视觉词典中含有 4 个视觉单词,这些视觉单词是输入图像 M-ORB 特征点的近似表示。系统通过视觉词典检索出每个视觉单词(特征)在图像中的出现频率,并将此绘制成一个簇类 $K=4$ 的视觉单词直方图。

至此,系统通过视觉词典模型将 3 幅图 A、B、C 的图像信息转化成 BoVW 分数(向量),如图 8 所示,分别为 A: [6,3,2,3]; B: [2,7,2,4]; C: [2,1,7,9]。图像间相似度对比也转换成了 BoVW 分数对比。

3.3 TF-IDF 自适应加权

通过 BoVW 模型将图像的特征信息转化为 BoVW 分数,即所有视觉单词所代表的图像信息。为了对视觉单词的重要性进行评估,并赋予不同

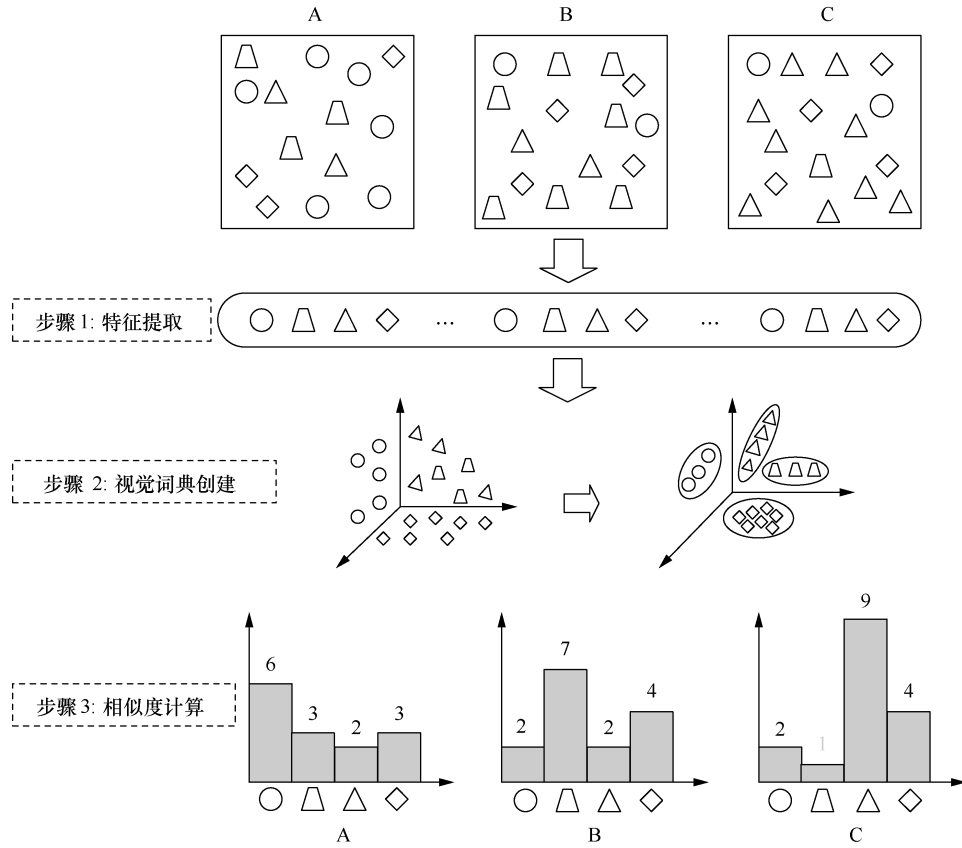


图8 视觉词典的生成

的权值,这里采用 TF-IDF 算法来分配每个视觉单词的权重。TF-IDF 由两部分组成,TF 表示频率部分,代表某个视觉单词在单个图像中出现的频率。对于图像 I,出现的单词总数为 n ,视觉单词 w_i 出现了 n_i 次,TF 为:

$$TF_i = \frac{n_i}{n} \quad (8)$$

IDF 表示数量部分,代表视觉词典中某视觉单词 w_i 中的特征数量与所有特征数量的比例。若某一视觉单词在图像中出现频率越高,则区分度就越高。对于图像 I,特征数量为 n ,视觉单词 w_i 的特征数量为 n_i ,IDF 为:

$$IDF_i = \log \frac{n}{n_i} \quad (9)$$

则可得视觉单词 w_i 的综合权重:

$$\eta_i = TF_i \times IDF_i \quad (10)$$

图像 I 中所有特征点经过聚类后对应 N 个单词,则视觉词典模型可表述为:

$$A = \{(w_1, \eta_1), (w_2, \eta_2), \dots, (w_N, \eta_N)\} \triangleq \mathbf{v}_A \quad (11)$$

通过式 (11),使用向量 $\mathbf{v}_A$ 表示图像 I。对于给定的 $\mathbf{v}_A$ 和 $\mathbf{v}_B$,本文通过定义 L1 范数的形式,量化二者间的相似程度:

$$S(\mathbf{v}_A - \mathbf{v}_B) = 2 \sum_{i=1}^N (|\mathbf{v}_{A_i}| + |\mathbf{v}_{B_i}| - |\mathbf{v}_{A_i} - \mathbf{v}_{B_i}|) \quad (12)$$

式 (12) 只考虑了两幅图像的绝对相似值评分,但是在不同场景下,绝对相似值评分并不能很准确地描述场景信息。因此,本文使用一种相对相似值评分,在式 (12) 的基础上引入先验相似度 $S(\mathbf{v}_t, \mathbf{v}_{t-\Delta t})$,表示 t 时刻关键帧与 $t-\Delta t$ 时刻关键帧的相似程度,并将所有的评分参照此值进行归一化:

$$S'(\mathbf{v}_t, \mathbf{v}_{t_j}) = \frac{S(\mathbf{v}_t, \mathbf{v}_{t_j})}{S(\mathbf{v}_t, \mathbf{v}_{t-\Delta t})} \quad (13)$$

其中, $\mathbf{v}_{t_j}$ 表示其他时刻。相对相似值评分的使用提升了系统在复杂环境下的鲁棒性。

4 实验与分析

为了验证本文提出的 D-LCD 算法的性能,本

节的实验分为闭环检测算法级实验与系统级实验。闭环检测算法级实验主要对基于 M-ORB 特征点的 D-LCD 与其他主流的闭环检测算法进行实验对比, 给出算法级别的实验验证与定量分析; 闭环检测系统级实验则将 D-LCD 应用于具体的基于直接法的 SLAM 系统, 给出系统级别的实验验证与定量分析。

本节实验选用视觉 SLAM 领域的权威数据集: 慕尼黑工业大学的 TUM 数据集^[17]与 KITTI 数据集^[18]。TUM 数据集主要用于 D-LCD 算法级的实验验证与室内外建图验证, KITTI 数据集主要用于 D-LCD 系统级的实验验证, 验证其能否较好地适配基于直接法的 SLAM 系统, 在不降低实时性的同时, 为系统提供正确的闭环信息, 有效减少误差累计, 遏制轨迹与地图点的漂移现象。

4.1 闭环检测算法级实验

本节对本文提出的基于改进 M-ORB 的视觉 SLAM 直接-闭环检测算法进行算法层次的实验对比。对比对象选择了特征点法中的主流闭环检测算法 ORB-SLAM2^[19]与直接法中的主流闭环检测算法 LSD-SLAM^[20]。本节实验选用 TUM 数据集, 分别在室内场景与室外场景下进行实验对比。

4.1.1 闭环检测率对比

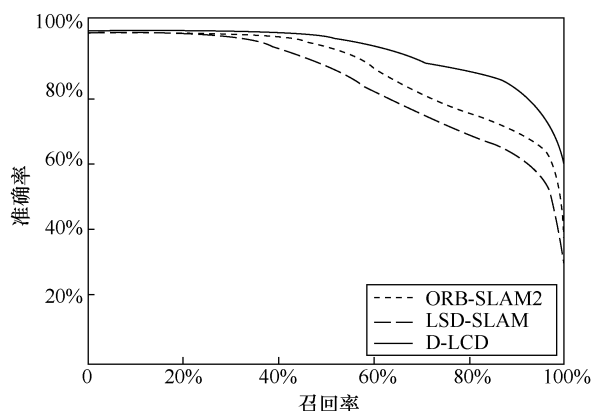
在 TUM 数据集中挑选出室内与室外场景, 分别运行 D-LCD、ORB-SLAM2 以及 LSD-SLAM, 设纵轴为准确率(precision), 横轴为召回率(recall), 得到 3 种算法的准确率/召回率 (precision-recall, PR) 曲线, 如图 9 所示。

对于 PR 曲线来说, 本文更关心在准确率为 80% 和 100% 情况下的召回率分布情况。由图 9 (a) 可知, 在低召回率时, D-LCD 与 ORB-SLAM2 以及 LSD-SLAM 曲线大致重合, 说明此时 3 种算法的准确率相差不大; 但随着召回率升高, D-LCD 依然能够保持较高的准确率, 而 ORB-SLAM2 和 LSD-SLAM 两种算法的准确率下降得较快。并且相较于室内场景, D-LCD 在室外场景下的优势更加明显, 在保证召回率为 95% 的条件下, D-LCD 的准确率为 75%, 而 ORB-SLAM2 和 LSD-SLAM 仅为 55% 和 37%, 这得益于 M-ORB 特征点提取算法和分层词典树, 二者的使用使得闭环检测在具有较高准确率的同时提升了算法鲁棒性。

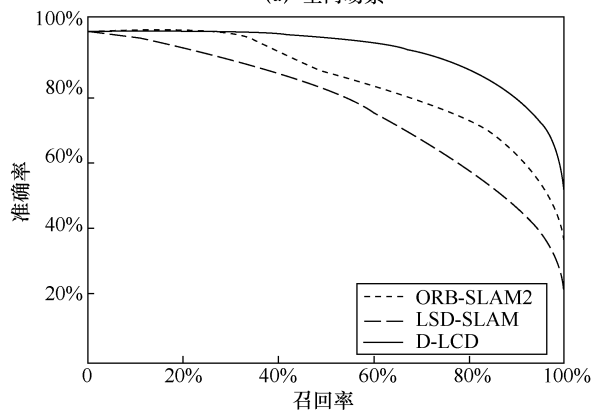
4.1.2 实时性对比

在 TUM 数据集中挑选出 20 个具有闭环的室内

和室外场景, 命名为 Sequence_01~Sequence_20, 其中 01~10 为室内场景, 11~20 为室外场景。为了使实验对比更为全面, 本节分析了单场景下的定量对比与多场景下的整体对比。



(a) 室内场景



(b) 室外场景

图 9 PR 曲线对比

(1) 单场景下的定量对比与分析

分别在 Sequence_05 (室内, 共 603 帧) 和 Sequence_15 (室外, 共 1 198 帧) 两个场景上运行 D-LCD、ORB-SLAM2 和 LSD-SLAM, 三者的闭环检测线程对场景中每帧图像的处理时间对比如图 10 所示。

由图 10 可知, D-LCD 在 Sequence_05 和 Sequence_15 上对每帧图像的平均耗时均最短。在 Sequence_05 中, D-LCD、ORB-SLAM2 和 LSD-SLAM 平均耗时分别为 0.015 s、0.052 s、0.038 s; 在 Sequence_15 中, 三者的平均耗时分别为 0.032 s、0.072 s、0.108 s。另外可以看出在 Sequence_05 中, LSD-SLAM 的性能要优于 ORB-SLAM2, 而在 Sequence_15 中, ORB-SLAM2 的性能优于 LSD-SLAM, 这表明 LSD-SLAM 更适合于室内运行, ORB-SLAM2 则更适合于室外运行。

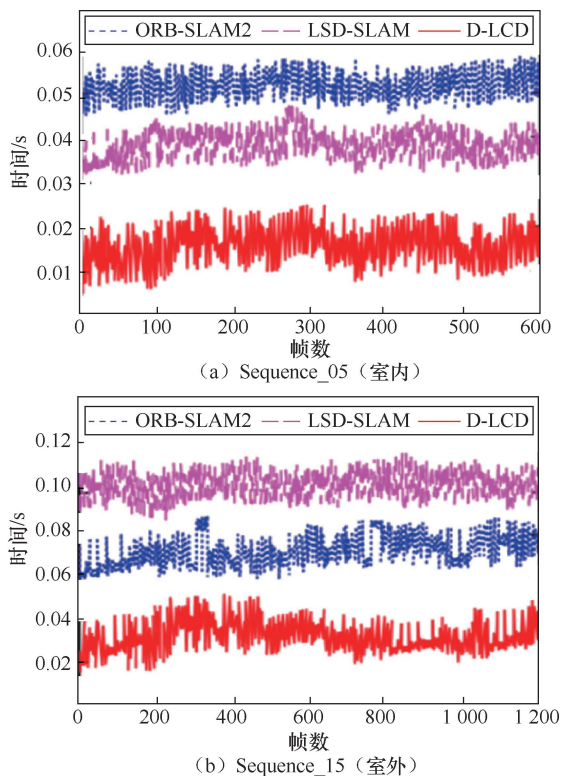


图10 室内外实时性对比

(2) 多场景下的整体对比与分析

为了更全面地测试 D-LCD 在室内外运行时的实时性,使实验结果更具有普遍性,将 D-LCD、ORB-SLAM2 以及 LSD-SLAM 在 Sequence_01~Sequence_20 共计 20 个场景下的运行所耗时间进行了对比,如图 11 所示。

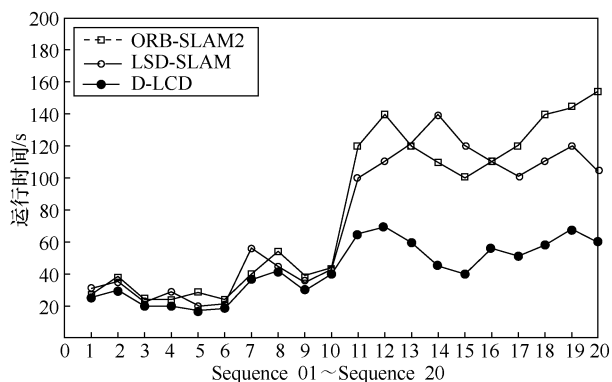


图11 室内外实时性对比

由图 11 可知,在室内场景下,D-LCD 与 LSD-SLAM 以及 ORB-SLAM2 在耗时上不相上下。但以 Sequence_10 为分界点,在室外大范围复杂场景下,D-LCD 展现出巨大优势,在面临海量数据处理需求时,M-ORB 特征点提取算法和分层词典树的使用使得 D-LCD 依然能够保持较低的运行时间,

其余两种经典算法耗时则急剧上升。

4.2 闭环检测系统级实验

本节实验将 D-LCD 算法应用于具体的基于直接法的 SLAM 系统,给出系统级的实验验证与定量分析。对比算法中用到的是视觉 SLAM 领域内较为先进的 DSO^[21]。DSO 是 2016 年慕尼黑大学计算机视觉实验室发布的一个直接稀疏法的视觉里程计,它的优点是可以生成相当稠密的点云,而且正常运行时速度很快。这里将融合了 D-LCD 的 DSO 命名为 D-SLAM (Direct-SLAM),实验选用 KITTI 数据集,在 KITTI 数据集上对 D-SLAM 进行建图轨迹的定量对比与分析,验证能否在不降低实时性的同时,为系统提供正确的闭环信息,消除轨迹与地图点的漂移现象。实验共分为误差定量分析和误差可视化分析两个部分。

本实验采用 SLAM 领域内常用的衡量指标:相对旋转误差 (relative rotation error, RRE)、相对平移误差 (relation translation error, RTE) 和绝对轨迹误差 (absolute trajectory error, ATE) 进行定量对比与分析。

4.2.1 误差定量分析

实验选用 KITTI 数据集中具有闭环的 4 个场景:KITTI_00、KITTI_05、KITTI_07、KITTI_09,真实轨迹如图 12 所示。

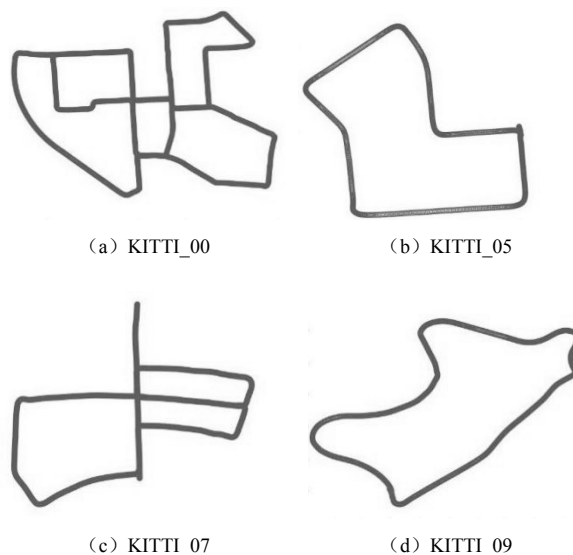


图12 KITTI 数据集 (真实轨迹)

在本文实验平台上载入上述 4 个场景,分别运行 DSO、D-SLAM 以及 ORB-SLAM2,得到各自对应的 RRE、RTE 和 ATE,见表 1、表 2、表 3。

表 1 RRE

KITTI 序列	DSO	ORB-SLAM2	D-SLAM
KITTI_00	0.298 4%	0.246 6%	0.215 3%
KITTI_05	0.187 7%	0.151 3%	0.114 9%
KITTI_07	0.287 4%	0.289 8%	0.159 2%
KITTI_09	0.216 4%	0.183 9%	0.164 9%

表 2 RTE/ (deg/100m)

KITTI 序列	DSO	ORB-SLAM2	D-SLAM
KITTI_00	48.278 1	1.297 5	0.621 7
KITTI_05	41.519 7	0.951 7	0.607 4
KITTI_07	40.476 2	0.769 1	0.445 7
KITTI_09	52.175 9	1.459 2	0.683 7

表 3 ATE/m

KITTI 序列	DSO	ORB-SLAM2	D-SLAM
KITTI_00	2.981 6	1.459 2	0.627 4
KITTI_05	2.274 3	1.297 5	0.601 7
KITTI_07	1.671 4	0.951 7	0.445 7
KITTI_09	2.999 4	1.679 1	0.694 2

由表 1 和表 2 可知, D-SLAM 在 RRE 方面改进较小, 在 RTE 方面具有较大改进。这是由于相机均被固定在数据采集车上, 车辆本身的姿态旋转并不明显, 反而由于误差累积造成的平移现象比较明显。由表 3 可知, 针对上述 4 个场景下的系统的 ATE, D-SLAM 相比于 DSO 和 ORB-SLAM2 均有明显减少。上述实验结果验证了本文提出的 D-SLAM 在应对室外大范围复杂场景时具备主动消除累积误差的能力。

4.2.2 误差可视化分析

选择上述场景 KITTI_07, 将真实轨迹转化为经纬度坐标, 导入谷歌地图, 得到真实运动轨迹, 如图 13 所示。该场景为某城市某一街道。KITTI 数据采集车从 A 点出发, 行驶路径为 A-A₁-A₂-A₃-A₄-A₂-A₃-B-A-C, 车辆在行驶过程中轨迹出现多次重合, 具备丰富的闭环信息, 其中 A、A₁、A₂、A₃、A₄ 为闭环信息路标点, 这些路标点之间的线段为闭环信息路段。在此场景下运行 DSO、D-SLAM 以及 ORB-SLAM2, 并对数据进行可视化, 如图 14 所示。

由图 14 可知, 数据采集车从 A 点出发, 经过闭环信息路标点, 途经 B 点, 最终到达终点 C。在此运动过程中可明显看到: DSO 由于不具有闭环检测线程, 其轨迹随时间出现了较大漂移; ORB-SLAM2 虽然具备闭环检测线程, 但系统各线

程间的耦合不够优秀, 在终点 C 处仍然存在一定的累计误差; 本文提出的 D-SLAM 在运动过程中几乎做到了与真实轨迹实时吻合。

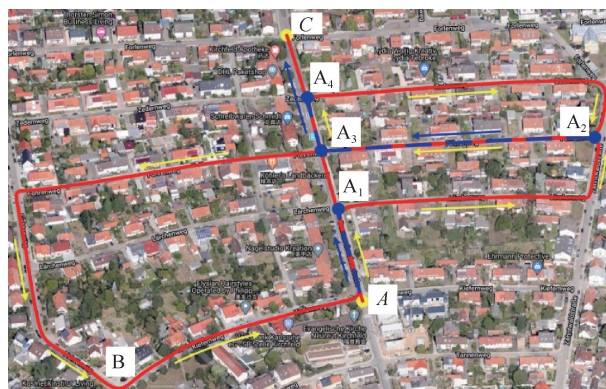


图 13 KITTI_07 数据采集车真实运动轨迹

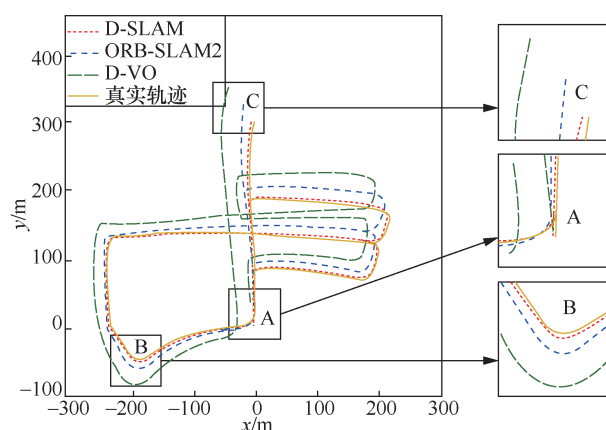


图 14 D-SLAM 与各主流算法及真实轨迹对比

5 结束语

本文提出的基于改进 M-ORB 的视觉 SLAM 直接-闭环检测算法无论是在算法级上的实验, 还是在具体的 SLAM 系统上都达到了预期效果。D-LCD 能够较好地适配于基于直接法的 SLAM 系统, 在不降低实时性的同时, 为系统提供正确的闭环信息, 有效减少误差累积, 遏制轨迹与地图点的漂移现象, 准确地构建全局一致性地图。但近两年, 基于神经网络的闭环检测方法研究逐渐被研究人员所重视。与传统方法不同, 神经网络模型具有学习的能力, 能够在大量数据中训练出较优的图像特征, 甚至可直接将其转换为 (机器) 学习问题。因此本文的后续工作将考虑如何将神经网络与闭环检测进行深度融合, 并通过降维减少深度学习对 SLAM 移动平台的计算需求, 设计出基于神经网络模型并具有自适应特征提取的闭环检测算法。

参考文献:

- [1] 郑南宁. 人工智能新时代[J]. 智能科学与技术学报, 2019, 1(1): 1-3.
- [2] 刘强, 段富海, 桑勇, 等. 复杂环境下视觉 SLAM 闭环检测方法综述[J]. 机器人, 2019, 41(1): 112-123, 136.
- [3] HESS W, KOHLER D, RAPP H H. Systems and methods of detecting loop closure in simultaneous localization and mapping (SLAM) applications: U. S. Patent Application, 10318659B1[P]. 2019-6-11.
- [4] KLOKOV R, LEMPITSKY V. Escape from cells: deep Kd-networks for the recognition of 3D point cloud models[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 863-872.
- [5] CHEN Z T, LAM O, JACOBSON A, et al. Convolutional Neural Network-based Place Recognition[C]//Proceedings of the 16th Australasian Conference on Robotics and Automation. [S.l.:s.n.], 2018.
- [6] BAI D D, WANG C Q, ZHANG B, et al. Sequence searching with CNN features for Robust and fast visual place recognition[J]. Computers and Graphics, 2018, 70: 270-280.
- [7] GOMEZ-OJEDA R, LOPEZ-ANTEQUERA M, PETKOV N, et al. Training a convolutional neural network for appearance-invariant place recognition[J]. Pattern Recognition Letters, 2015.
- [8] CHEN Z T, JACOBSON A, SUNDERHAUF N, et al. Deep learning features at scale for visual place recognition[C]//Proceedings of 2017 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2017: 3223-3230.
- [9] MERRILL N, HUANG G Q. Lightweight unsupervised deep loop closure[C]//Proceedings of Robotics: Science and Systems. [S.l.:s.n.], 2018.
- [10] SIVIC J, ZISSERMAN A. Video google: a text retrieval approach to object matching in videos[C]//Proceedings of the 9th IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2003.
- [11] CUMMINS M, NEWMAN P. Probabilistic appearance based navigation and loop closing[C]//Proceedings of 2007 IEEE International Conference on Robotics and Automation. Piscataway: IEEE Press, 2007: 2042-2048.
- [12] WILLIAMS B P, CUMMINS M, JOSÉ NEIRA, et al. An image-to-map loop closing method for monocular slam[C]//Proceedings of 2008 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2006: 2053-2059.
- [13] MATHIEU L, MICHAUD F. Memory management for real-time appearance-based loop closure detection[C]//Proceedings of 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2011: 1271-1276.
- [14] 李博, 杨丹, 邓林. 移动机器人闭环检测的视觉字典树金字塔 TF-IDF 得分匹配方法[J]. 自动化学报, 2011, 37(6): 665-673.
- [15] LI B, YANG D, DENG L. Visual vocabulary tree with pyramid TF-IDF scoring match scheme for loop closure detection[J]. Acta Automatica Sinica, 2011, 37(6): 665-673.
- [16] GAO X, WANG R, DEMMEL N, et al. LDSO: direct sparse odometry with loop closure[C]//Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway: IEEE Press, 2018: 2198-2204.
- [17] RUBLEE E, RABAU D V, KONOLIGE K, et al. ORB: an efficient alternative to SIFT or SURF[C]//Proceedings of the 13th IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2011: 2564-2571.
- [18] ENGEL J, USENKO V, CREMERS D. A photometrically calibrated benchmark for monocular visual odometry[J]. arXiv preprint, 2016, arXiv:1607.02555.
- [19] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 3354-3361.
- [20] 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.
- [21] ENGEL J, SCH T, CREMERS D. LSD-SLAM: large-scale direct monocular SLAM[C]//Proceedings of European Conference on Computer Vision. [S.l.:s.n.], 2014: 834-849.
- [22] ENGEL J, KOLTUN V, CREMERS D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611-625.

[作者简介]



李伟(1979-), 女, 河南大学计算机与信息工程学院副教授, 主要研究方向为机器人导航、运动规划和控制。



任孟瀚(1997-), 男, 河南大学计算机与信息工程学院硕士生, 主要研究方向为机器人导航控制、路径规划。



黄威豪(1996-), 男, 河南大学计算机与信息工程学院硕士生, 主要研究方向为机器人导航控制、图像处理。



杜晓玉(1979-), 女, 博士, 河南大学计算机与信息工程学院副教授, 主要研究方向为无线传感器网络定位及覆盖技术。



周毅(1981-), 男, 博士, 河南大学计算机与信息工程学院教授、博士生导师, 河南省车联网协同技术国际联合实验室主任, 主要研究方向为机器人导航控制、车联网技术。