

引用格式:李奎霖,魏武,高勇,等. 基于 LDSO 的机器人视觉定位与稠密建图的应用[J]. 微电子学与计算机,2020,37(2):51-56. [LI K L, WEI W, GAO Y, et al. Application of robot visual localization and dense mapping based on LDSO[J]. Microelectronics & Computer, 2020, 37(2): 51-56.]

基于 LDSO 的机器人视觉定位与稠密建图的应用

李奎霖, 魏 武, 高 勇, 李艳杰, 王栋梁

(华南理工大学 自动化科学与工程学院, 广东 广州 510640)

摘 要: 针对基于特征点匹配的 SLAM(Simultaneous Localization and Mapping)系统在缺乏角点的弱纹理区域无法提取足够的特征点而导致位姿估计失败等问题,提出应用直接视觉里程计算法 LDSO(Direct Sparse Odometry with Loop Closure)进行室内机器人视觉定位并结合深度估计或深度相机采集到的关键帧深度图,关键帧相机位姿,原始关键帧图像数据,点云拼接生成三维点云稠密地图,实验结果表明,机器人可在复杂环境中准确快速的定位自身位置,且算法在没有全局 BA(Bundle Adjustment)的情况下通过位姿图优化显著减少了旋转,平移与尺度漂移等累积误差,算法整体性能与基于特征点匹配的 SLAM 系统相媲美,耗时更少,实时性更佳,在缺乏角点区域具有较强的鲁棒性。

关键词: 视觉里程计;LDSO;位姿图优化;点云地图

中图分类号: TP242.6

文献标识码: A

文章编号: 1000—7180(2020)02—0051—06

Application of robot visual localization and dense mapping based on LDSO

LI Kui-lin, WEI Wu, GAO Yong, LI Yan-jie, WANG Dong-liang

(College of Automation Science and Engineering, South China University of Technology,
Guangzhou 510440, China)

Abstract: For the SLAM system based on feature point matching, the problem of pose estimation failure is due to the inability to extract enough feature points in weak texture regions lacking corner points. Applying direct visual odometry method LDSO for indoor robot visual localization and combined with depth estimation or depth camera acquired key frame depth image, key frame camera pose, original key frame image data, point cloud stitching to generate 3D point cloud dense map, experimental results shows that the robot can be in a complex environment Accurate and fast localization, and the algorithm significantly reduces the cumulative error of rotation, translation and proportional drift by pose optimization without global Bundle Adjustment, so that the overall performance is comparable to SLAM systems based on feature point matching, the algorithm takes less time, has better real-time performance and is more robust in areas lacking corner points.

Key words: visual odometry; LDSO; pose graph optimization; point cloud map

收稿日期: 2019—05—27; 修回日期: 2019—07—14

基金项目: 国家自然科学基金(61573148); 广东省科技计划项目(2016A040403012); 广东省省级科技计划项目(2017B090901043)

1 引言

自 20 世纪 80 年代以来,即时定位与建图(SLAM)系统已经成为计算机视觉和机器人技术中的一个活跃研究领域,SLAM 系统已成为许多如移动机器人,自动驾驶,虚拟与增强现实等需要实时定位的应用的基础模块。

随着 V-SLAM 技术的进步,本世纪初 Hartley 等人^[1]认为计算机视觉领域的运动推断结构(Structure From Motion, SFM)问题与 SLAM 问题有许多共同点. SFM 技术中的非线性优化方法-捆集调整^[2](Bundle Adjustment, BA)被引入 SLAM 领域中,并逐步发展到目前视觉 SLAM 领域中的主流方法^[3-7]. Engle 等^[8]提出了一种单目稀疏直接法视觉里程计 DSO(Direct Sparse Odometry),该方法加入了完整的光度标定,同时对透镜衰减、伽马响应和曝光时间进行标定,这些策略使得系统的定位精度和鲁棒性都很高,然而由于 DSO 中并不具备回环检测模块,因此其并非完整的 SLAM 系统. 为了减小 DSO 中的累积误差, X. Gao 等^[9]提出了基于 DSO 方法的 LDSO(Direct Sparse Odometry with Loop Closure)法,该方法中的闭环检测是传统的基于特征的词袋法(Bag-of-Words, BoW),LDSO 改进了 DSO 的点选择策略,并用来选取可用于闭环检测的可重复的特征角点,保留了在特征缺失环境下的鲁棒性,继承了 DSO 的准确性和稳定性,并且通过闭环检测显著减少了累积漂移误差,使得直接法在没有全局 BA 优化的情况下,表现性能与目前最新的基于特征点匹配的 SLAM 方法相当。

2 LDSO 简介

LDSO 是少数直接使用纯直接法(Fully Direct)计算视觉里程计的系统之一。

直接法相比于特征点法,有两个非常不同的地方:(1)直接法依赖于最小化光度误差(Photometric Error). 即最小化的目标函数,通常由图像之间的误差来决定。(2)直接法将数据关联(Data Association)与位姿估计(Pose Estimation)放在了一个统一的非线性优化问题中. 在 LDSO 的线程中会一直求解一个比较复杂的优化问题,很难将其分为一步一步的过程,在匹配过程中会尝试把每个点投影到所有关键帧中,计算它在各帧中的残差,而不在意点与点之间的一一对应关系. 从后端来看, LDSO 使用一个由若干系统关键帧组成的滑动窗口(Sliding

Window)作为优化后端,这个窗口通常具有 5-7 个关键帧,并且在前端线程判断某帧符合条件而插入后端时,后端会实时去除某非关键帧,即执行边缘化(Marginalization)过程. 需优化的最小化光度误差可定义如下: 令 $S = \{T_1, T_2 \cdots T_j, p_1, p_2 \cdots p_k\}$, 其中 T_j 表示关键帧 j 的位姿, p_k 表示在关键帧中的像素点坐标, 那么最小化光度误差 $\min_{T_u, T_v, p_w \in S} \sum E_{u,v,w}$ 应为:

$$E_{u,v,w} = \sum_{p \in N_{p_w}} \omega_p \| (I_v[p'] - b_v) - \frac{t_v e^{a_v}}{t_u e^{a_u}} (I_u[p] - b_u) \| \quad (1)$$

式中, t_u, t_v 为图像 I_u, I_v 对应的曝光时间; N_{p_w} 为像素点的集合; a, b 代表光度仿射参数; ω_p 为补偿加权参数; p' 是含有逆深度参数 d_p 的像素坐标 p 在图像 I_v 上的重投影坐标, p' 的计算如下所示:

$$p' = \Pi (R \Pi^{-1}(p, d_p) + t) \quad (2)$$

式中, R 和 t 为旋转矩阵与平移向量, 由变换矩阵 T_u, T_v 计算得来, Π 与 Π^{-1} 分别代表重投影与逆重投影变换。

2.1 系统流程

在单目 SLAM 系统中,所有的地图点在一开始被观测到时,都只有一个二维的像素坐标,其深度值是未知的,这种点在 LDSO 系统中被称为未成熟的地图点(Immature Points),随着相机的运动, LDSO 会在每张地图上追踪每个未成熟的地图点,这个过程实际上是一个沿极线搜索的过程,十分类似于 SVO 中的深度滤波器方法,在追踪的过程系统会确定每个未成熟地图点的逆深度(即深度的倒数)和变化范围,如果此逆深度收敛,即可确定该点的三维坐标,形成正常的地图点. 在整个系统中,所有未成熟点存在于 immature Points 结构体中,每个帧的数据存储于 Frame Hessian 结构体中,与之对应的是,每个确定三维坐标的点均存在于 Point Hessian 结构体中,该结构体存储所有活跃点的信息,即该点存在于相机的拍摄范围内,其残差项仍参与优化计算与迭代,而中当新的图像到来时, LDSO 处理该图像信息的流程如下图 1 所示。

2.2 残差的构成

在后端优化过程中,算法试图将每个先前关键帧中的地图点投影到新关键帧中,形成残差项,整个优化后端内部构成了一个非线性最小二乘问题,表示成文献[10]中的因子图(或图优化)形式如下图 2 所示。

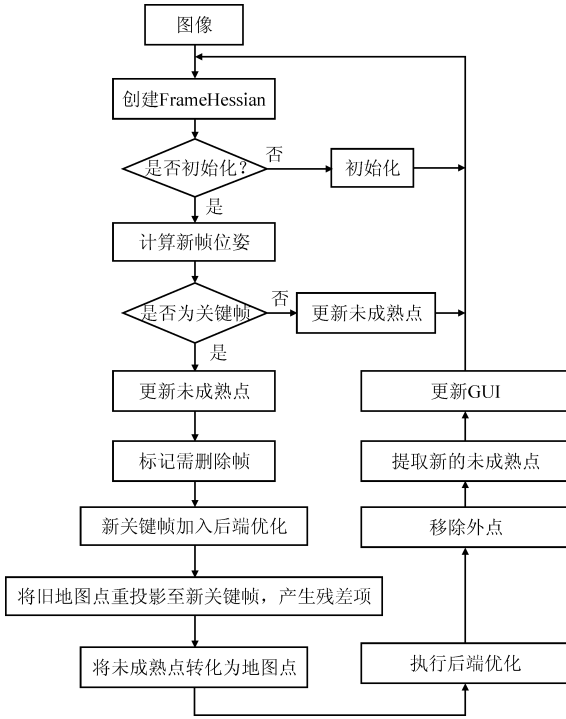


图 1 算法流程

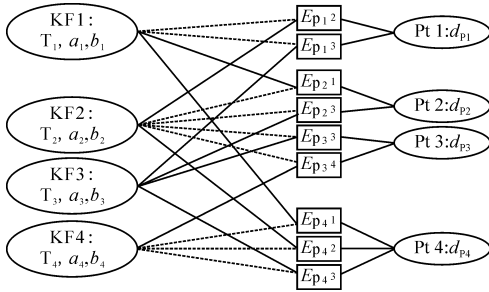


图 2 后端优化因子图模型

每个残差项(或能量项) E_p 将关联两个关键帧 KF 和一个以深度参数 d_p , 下面将说明残差项的计算, 设有两个图像帧 I_H, I_T , 各自到世界坐标的变换矩阵记为 T_{HW}, T_{TW} , 变换矩阵又可拆分为旋转矩阵 R 和平移向量 t , 相机模型由针孔模型描述, 其内参矩阵为:

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix},$$

$$K^{-1} = \begin{bmatrix} f_x - 1 & 0 & -f_x - 1c_x \\ 0 & f_y - 1 & -f_y - 1c_y \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

令 $x_H = [u_T, v_T, 1]^T$ 为 I_H 帧中的一个像素点, u_T, v_T 为该点的像素坐标, 该点的深度值估计为 d_H , 则其逆深度为

$$\rho_H = d_H - 1 = \frac{1}{d_H} \quad (4)$$

则像素点 x_H 在 I_T 帧中的投影为:

$$\rho_T - 1x_T = K T_{TW} T_{HW}^{-1} K^{-1} \frac{1}{\rho_H} x_H = K(R_{TH} p_W + t_{TH}) \quad (5)$$

定义中间变量点的世界坐标 p_W 与 I_T 到 I_H 的相对变换矩阵 T_{TH} , 该矩阵又可拆分为 R_{TH} 和 t_{TH} . 由此定义该点的残差:

$$e = I_H(x_H) - I_T(x_T) \quad (6)$$

在整个后端优化线程中, 为了优化若干个关键帧和与之相关联的地图点组成的残差项, 采用 Gauss-Newton 或 Levenberg-

Marquardt 方法进行迭代, 在迭代过程中, 所有的残差项可以拼成一个大型的线性方程:

$$J^T W J \delta x = -J^T W r \quad (7)$$

式中, J 为拼接后的雅可比矩阵, 主要包括图像梯度、相对几何量、相对光度参数等信息; W 和 r 代表权重与残差矩阵, δx 为整体的优化更新变量, 而方程左侧可记为 $J^T W J = H$, 即记录为 Hessian 矩阵, 该矩阵将会在优化线程中一直存在, 不断更新, 即先前的优化为下一步优化提供了先验知识.

2.3 闭环检测

在理想情况下, 基于 DSO 的闭环检测应使用光度误差进行全局捆集调整(BA), 这很好的匹配了 DSO 中的后端优化方法, 但因光度误差需要在图像中计算, 这需要在线程中保存全部的图像, 因此 LDSO 中的闭环检测使用的是位姿图优化, LDSO 在滑动窗口中加了一个全局位姿来维持关键帧之间的连接, 取滑动窗口中关键帧的相对 3D 位姿变换作为位姿配对的测量数据. 基于词袋法(Bag of Word)提出一种将 ORB 特征与原始点选择策略结合的匹配方法, 当选择了环路匹配子后, 则计算其关于当前关键帧的 Sim(3) 约束并将其添加到全局姿态图中, 随着优化获得准确的相机位姿估计, 闭环检测框架如图 3 所示:

在闭环线程中, 算法为每个关键帧计算 ORB 描述符, 使用 DBoW3^[11] 构建词袋数据库(BoW), 通过查询数据库来为当前关键帧选出闭环匹配子, 匹配子从被边缘化的关键帧中选取, 对于每个匹配子, 算法尝试将其 ORB 特征与当前帧相匹配, 然后使用 RANSAC PnP 计算出一个初始的 SE(3) 变换, 之后使用 Gauss-Newton 法最小化 3D 与 2D 的几何约束, 以此来优化 SE(3) 变换. 定义闭环匹配子的

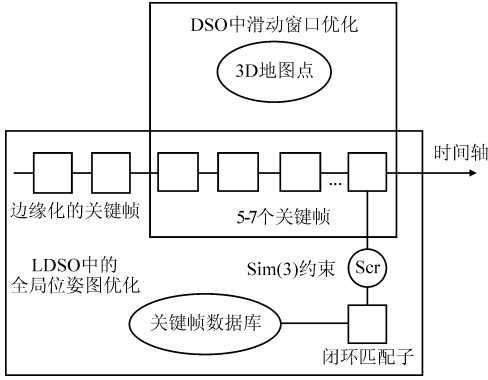


图3 闭环检测框架

重建特征为 p_i , 对应的逆深度为 d_{p_i} , 构成的集合为 P , 在当前关键帧中与之匹配的特征为 q_i , 构成集合 O , 令 $O_1 \subseteq O$ 为不具有深度的 q_i 集合, 而 $O_2 \subseteq O$ 为具有深度的 q_i 集合, 那么闭环匹配子与当前关键帧的 $Sim(3)$ 变换 S_{σ} 就可以通过最小化下面的代价函数来估计:

$$E_l = \sum_{q_i \in O_1} \omega_1 \| S_{\sigma} \Pi^{-1}(p_i, d_{p_i}) - \Pi^{-1}(q_i, d_{q_i}) \|_2 + \sum_{q_j \in O_2} \omega_2 \| \Pi(S_{\sigma} \Pi^{-1}(p_j, d_{p_j})) - q_j \|_2 \quad (8)$$

式中, Π 与 Π^{-1} 为之前定义的重投影与逆重投影变换; ω_1 和 ω_2 是为了平衡不同测量单元的权重。

3 深度估计与稠密建图

稠密点云地图的建立一般需要三种数据, 当前帧原始 RGB 图像, 当前帧估计得出的相机位姿, 与当前帧原始 RGB 图像像素对应的深度值, 即深度图, 基于此三类数据, 即可根据坐标系变换拼接出点云地图, 设图像 I 上某点像素坐标为 (u, v) , 深度为 d , 该点在相机坐标系下为 $p_c = [x, y, z]$, 则有:

$$d \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K p_c \quad (9)$$

$$\begin{aligned} x &= \frac{u - c_x}{f_x} z \\ y &= \frac{v - c_y}{f_y} z \\ z &= d \end{aligned}$$

式中, K 为(3)中定义的相机内参矩阵。

在单目直接视觉里程计中, 由于依靠弹幕相机无法直接获得像素点的深度信息, 因此需要极线搜索和块匹配技术^[12]对深度进行估计, 设相机在 a 处得到帧 O_1 , 在 b 处得到帧 O_2 , O_1 上的某像素点 p_1 , 该点对应的深度未知, 分布范围为 d , 投影至 O_2 得到对应的极线范围, 示意图如图4。

接下来就是对极线上存在的像素点与 p_1 匹配, 一旦确定, p_1 的深度值也就确定。由于噪声的存在,

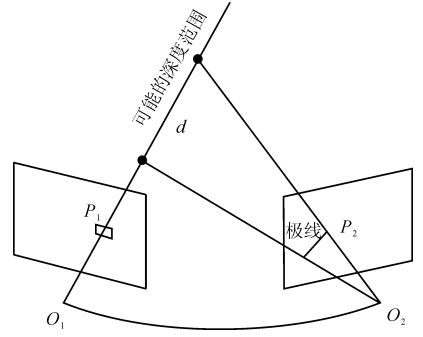


图4 极线搜索示意图

每次观测必与真实值之间有一定的误差, 针对此不可避免的误差项, 选择使用滤波器对其进行优化估计, 假设某像素点 p 的深度 d 为高斯分布, 假设对每次观测有:

$$P(d_o) = N(\mu_o, \sigma_o^2) \quad (10)$$

当有新的数据经过观测后其深度 $P(d_m)$ 也服从如(10)所示的高斯分布, 则两项数据融合后的分布为:

$$P(d_f) = N(\mu_f, \sigma_f^2)$$

$$\mu_f = \frac{\sigma_{o_2}^2 \mu_{o_1} + \sigma_{o_1}^2 \mu_{o_2}}{\sigma_{o_1}^2 + \sigma_{o_2}^2}, \sigma_f^2 = \frac{\sigma_{o_1}^2 \sigma_{o_2}^2}{\sigma_{o_1}^2 + \sigma_{o_2}^2} \quad (11)$$

深度估计的步骤如下, (1) 假设像素点深度符合高斯分布, 对像素点通过极线搜索和匹配确定投影点位置; (2) 计算深度与不确定性 σ ; (3) 将新数据的观测融合至上述深度估计中, 若 σ 收敛则停止计算, 否则返回极线搜索继续迭代。

4 实验结果与分析

为了验证算法的有效性, 本文进行了两组实验:

(1) 通过分析算法在 TUM 和 KITTI 数据集中算法估计的位姿轨迹效果, 验证算法的快速性与准确性。

(2) 利用 Turtlebot2 机器人在实地环境进行视觉定位与稠密点云地图的构建, 验证算法的有效性。

4.1 数据集验证

基于本文算法在开源的 TUM-Mono 数据集上估计出的轨迹与 DSO 方法估算出的轨迹作对比, 比较算法准确性与鲁棒性。TUM-Mono 数据集是一个单目数据集, 它提供了光度标定图, 但没有完整的实际地面轨迹, 每个数据集序列的摄像机始终返回起点, 因此十分适合评估算法的累计漂移, 其中两个序列的轨迹对比结果如图5所示:

在 TUM-Mono-Seq10 数据集中可以明显看出 DSO 方法由于累积误差导致的轨迹漂移, 即放大部

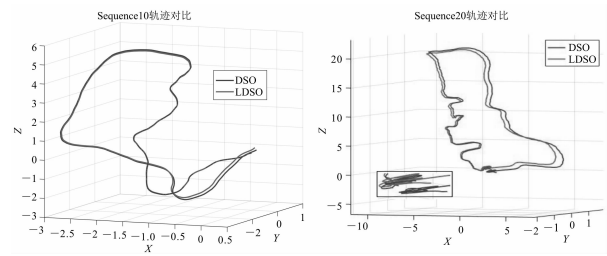


图 5 TUM 数据集轨迹对比

分起点与终点部分估计轨迹无法拟合,而本文方法则通过闭环检测消除了累积误差,轨迹在重点与起点很好的拟合在了一起,具有较高的准确性。

进一步使用 KITTI 数据集中的 Visual Odometry 部分来评估本分文算法的准确率, Visual Odometry 部分共有 21 个图像序列,其中 0-10 序列提供经运动抓捕系统捕获的真实位姿,可是为正确轨迹,将该轨迹与本文算法评估得出的轨迹以及 DSO 算法产生的轨迹同时对比,基于此来评估本文算法的准确性与鲁棒性。

结果表明,缺乏回环检测的 DSO 方法在这种大型复杂环境下进行位姿估计的乏力,该方法评估出的轨迹与标准轨迹相差甚远,而本文方法由于同时具备直接法的鲁棒性与回环检测的准确性,即使在特征点缺失的高速路面也能精准的实现位姿轨迹,整体表现较为优异。

最后在 KITTI 数据集上,比较本文算法与基于特征点的 ORB_SLAM2 算法的绝对轨迹误差,结果如表 1 所示:

表 1 绝对轨迹误差对比

KITTI 数据集	尺度 (m×m)	序列 帧数	算法每 帧耗时 (ms)	轨迹误差(m)	
				LDSO	ORB-SLAM2
Sequence00	564×496	4 540	295.39	9.16	8.45
Sequence01	1 157×1 827	1 100	103.84	11.37	—
Sequence02	599×946	4 660	303.63	30.67	25.43
Sequence05	479×426	2 760	227.12	5.34	8.79
Sequence07	191×209	1 100	187.67	2.63	3.26
Sequence09	465×568	1 590	241.87	22.83	75.96
Sequence10	671×177	1 200	203.69	18.13	7.12

从上表可知,无需计算关键点与描述子的本文算法与当前应用广泛的基于特征点的 ORB_SLAM 在 KITTI 数据集的评估准确性上表现相当,且在特征点缺失的序列中表现出更强的鲁棒性,如在 Se01 中由于存在许多缺乏角点与重复纹理的图像帧, ORB_SLAM2 由于无法顺利提取足够的特征点而导致评估位姿失败,本文算法表现良好,同样的情况

也在 Se09 中出现。

4.2 实地环境试验

为了进一步验证本文算法的有效性与稠密建图效果,将本文算法移植到移动机器人平台 Turtlebot 2 在室内楼道环境中进行了实地实验验证算法有效性 与准确性。

由于本文实验平台并无运动抓捕系统,为了分析 机器人定位精度,将采取在环境中设置路标点的方法, 实地测量路标点二维坐标并经过坐标系转换后与估 计出的数据相对比,以此来计算本文算法的均方根 误差 $RMSE = [(\sum_{i=1}^n (t_i - \bar{t}_i)^2)/n]^{1/2}$, 其中 $t_i, \bar{t}_i$ 分别 为真实值与测量值,结果见下表:

表 2 均方根误差对比

测量点	真实值(m)	估计值(m)	均方根误差 (m)
起点	(0,0)	(0,0)	0
路标点 1	(2.3, 0.8)	(2.043,0.673)	0.202 7
路标点 2	(3.3,1.5)	(3.186,1.324)	0.148 3
路标点 3	(2.5,1.8)	(2.962,1.422)	0.421 1
路标点 4	(1.7,1.5)	(1.526,1.112)	0.300 7
终点	(-0.1,0.2)	(-0.396,0.083)	0.225 1

实验平台总体运动轨迹距离为 10 m,而上述 6 点的总均方根误差为 0.252 4,存在 2.5%的误差,总 体位姿估计准确,这也在拼接点云地图的结果中得到 验证,本文算法在实地实验中表现优异.良好实现机 器人视觉定位与稠密建图功能。

5 结束语

为了解决在缺乏角点的弱纹理区域无法有效提 取特征点而导致的基于特征点皮配的 SLAM 系统表 现不佳的问题,本文提出应用直接法 LDSO 来进行机 器人视觉定位,以及进行稠密建图,首先在公开的数据 集中验证本文算法的有效性与准确性,结果表明本 文算法在缺乏角点的弱纹理区域仍能有效的进行光 度误差匹配且表现良好,之后利用实验平台进行了实 地环境试验,通过计算本文算法估计出的位姿与真实 路标点坐标的均方根误差 RMSE 为 0.252 4,来验证 本文算法在实地环境中的有效性与准确性,同时结合 本文算法估计出的位姿与 KINECT 相机得到的深度 图与彩色图 3D 稠密点云地图的建立,最终结果表明, 本文算法可以实现机器人视觉定位与建图功能。

参考文献:

[1] HARTLY R, ZISSERMAN A. Multiple view geometry in computer vision[J]. Kybernetes, 2008, 30(9/

- 10);1865-1872.
- [2] TRIGGS B, HARTLEY R I, et al. Bundle Adjustment-A Modern Synthesis[C]// Proceedings of the International Workshop on Vision Algorithms : Theory and Practice, 1999. Springer-Verlag, 1999: 298-372.
- [3] SIBLEY G, MEI C, REID I, et al. Vast-scale Outdoor Navigation Using Adaptive Relative Bundle Adjustment[J]. International Journal of Robotics Research, 2010, 29(8):958-980.
- [4] KLEIN G, MURRAY D. Parallel Tracking and Mapping for Small AR Workspaces[C]// Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on. ACM, 2007.
- [5] MUR-ARTAL R, MONTIEL J M M, TARDOS J D. ORB-SLAM: a Versatile and Accurate Monocular SLAM System[J]. IEEE Transactions on Robotics, 2015, 31(5):1147-1163.
- [6] ENGEL J, JÜRGEN STURM, CREMERS D. Semi-dense Visual Odometry for a Monocular Camera[C]// Proceedings of the 2013 IEEE International Conference on Computer Vision. IEEE, 2013.
- [7] ENGEL J, THOMAS SCHÖPS, CREMERS D. LSD-SLAM: Large-Scale Direct Monocular SLAM[M]. Berlin, Germany: Springer, 2014: 834-849.
- [8] ENGEL J, KOLTUN V, CREMERS D. Direct Sparse Odometry[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2018, 40(3): 611- 625.
- [9] GAO X, WANG R, DEMMEL N, et al. LDSO: Direct Sparse Odometry with Loop Closure[C]// International Conference on Intelligent Robots and Systems . Madrid, Spain 2018.
- [10] FORSTER C, PIZZOLI M, DAVIDE SCARAMUZZA *. SVO: Fast Semi- Direct Monocular Visual Odometry[C]// IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, 2014. IEEE, 2014.
- [11] GALVEZ-LOPEZ D, TARDOS J D. Bags of Binary Words for Fast Place Recognition in Image Sequences [J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.
- [12] PIZZOLI M, FORSTER C, SCARAMUZZA D. Re-mode: Probabilistic, monocular dense reconstruction in real time[C]. // Proc of IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE Press, 2014: 2609-2616.

作者简介:

李奎霖 男,(1996-),硕士研究生.研究方向为计算机视觉,SLAM算法.

魏 武 男,(1970-),博士,教授.研究方向为机器人控制、智能控制、模式识别、人工智能.

高 勇(通信作者) 男,(1993-),博士研究生.研究方向为仿生机器人及智能控制. E-mail:andygao_scut@163.com

李艳杰 女,(1991-),硕士.研究方向为机器人运动规划与控制.

王栋梁 男,(1993-),博士研究生.研究方向为移动机器人.