

基于混合位姿估计模型的移动机器人三维地图创建方法

王可¹, 贾松敏¹, 徐涛^{1,2}, 李秀智¹

(1. 北京工业大学 电子信息与控制工程学院, 北京 100124; 2. 河南科技学院 机电学院, 河南 新乡 453003)

摘要: 针对室内复杂环境下的稠密三维建模问题, 提出一种基于RGB-D相机的移动机器人同时定位与三维地图创建方法. 该方法利用架设在移动机器人上的RGB-D相机获取环境信息, 根据点云和纹理加权模型建立结合局部纹理约束的混合位姿估计方法, 确保定位精度的同时减小失败率. 在关键帧选取机制下, 结合视觉闭环检测方法, 运用树结构网络优化(TORO)算法最小化闭环误差, 实现三维地图的全局一致性优化. 在室内环境下的实验结果验证了所提出算法的有效性和可行性.

关键词: 移动机器人; 关键帧; 混合位姿估计; 三维地图创建; 闭环

中图分类号: TP341.1

文献标志码: A

Mobile robot 3D map building based on hybrid pose estimation model

WANG Ke¹, JIA Song-min¹, XU Tao^{1,2}, LI Xiu-zhi¹

(1. College of Electronic Information and Control Engineering, Beijing University of Technology, Beijing 100124, China; 2. School of Mechanical and Electrical Engineering, He'nan Institute of Science and Technology, Xinxiang 453003, China. Correspondent: WANG Ke, E-mail: kwangs@sina.com)

Abstract: A real-time dense method to address the problem of mobile robot simultaneous localization and 3D mapping(3D SLAM) in complex indoor environment is proposed. In this approach, the environmental data is captured by using a RGB-D camera which is fixed on the robot. Combining with the local texture association, a hybrid algorithm model is established to ensure the pose estimation accuracy and concurrently decrease the failure rate during mapping by using the point cloud and image texture. By taking advantage of the keyframe selection mechanism, a visual-based loop detection algorithm and tree-based network optimizer(TORO) are used to achieve a global consistency map. Experimental results show the feasibility and effectiveness of the proposed algorithm in the indoor environment.

Keywords: mobile robot; keyframe; hybrid pose estimation; 3D map building; loop closure

0 引言

随着同时定位与地图创建(SLAM)研究的不断深入, 环境三维立体化建模已趋于成熟, 广泛应用于移动机器人场景重建^[1-2]和自主导航^[3-4]等领域, 并引起了众多学者的关注. Henry等^[5-6]利用Kinect体感相机获取RGB-D信息, 采用图像特征匹配作为数据关联方式, 运用RGBD-ICP算法实现对机器人的位姿更新, 在此基础上结合深度信息创建三维环境地图. 尽管该方法具有较高的精度, 但实时性差, 易受光照影响, 无法满足实时三维地图创建的需要. Newcombe等^[7]在Kinect获取深度图的基础上, 利用分层迭代最近点(ICP)算法实现了Kinect精确定位, 结合距离标记函数(SDFs)法实现对环境的实时三维建模. 但

算法对于GPU硬件配置要求较高, 且基于投影方式的ICP算法精度较低. Steinbrücker等^[8]提出了一种基于RGB-D相机的实时稠密视觉里程计算法, 该方法利用相机投影建立能量函数, 结合一阶泰勒公式求解上述算法模型, 实现相机的精确定位, 并通过稠密图像误差建立当前三维地图. 但该方法采用帧到帧(CL)的位姿估计方法, 随着时间的推移, 位姿估计误差过分累计, 导致所生成的地图产生较大漂移. 综上所述, 目前基于图像特征和三维点云的三维地图创建方法在一定程度上存在算法复杂度较高以及易受环境因素影响等问题, 与之相比, 基于稠密图像匹配方法计算量小, 易于实现并行处理, 且能有效克服光照影响^[9-10], 是一种鲁棒性更强的算法. 因此, 该方法在

收稿日期: 2014-04-14; 修回日期: 2014-10-15.

基金项目: 国家自然科学基金项目(61175087, 61105033).

作者简介: 王可(1988—), 男, 博士生, 从事机器视觉的研究; 贾松敏(1964—), 女, 教授, 博士生导师, 从事移动机器人、机器视觉等研究.

环境重建方面更具研究和应用价值。

针对上述研究现状, 本文提出一种基于RGB-D相机的移动机器人实时三维重建方法, 以实现在室内复杂环境下的三维建模. 该方法是在关键帧选取机制下, 通过权衡基于ICP和纹理的位姿估计方法建立混合位姿估计模型, 实现移动机器人的位姿估计. 在关键帧到帧(CG)位姿估计策略的基础上, 引入局部纹理关联约束, 保证位姿估计精度的同时减少跟踪失败率. 同时, 为了提高所生成地图的质量, 运用视觉闭环检测和树结构网络优化(TORO)^[11-12]算法最小化闭环误差, 以实现所生成三维地图的一致性优化.

1 相机的位姿及其投影模型

在世界坐标系下, 相机的位姿可表示为

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}, \quad (1)$$

其中 R 和 t 分别为相机在世界坐标系下的旋转和平移关系. 该相机位姿矩阵 T 建立了世界坐标系下点 p 到相机坐标系下点 p_c 的对应关系, 即

$$p_c = Tp. \quad (2)$$

由Lie代数知识可知, T 为特殊李群SE(3)中的元素, 且 T 可由六元组 $\xi = (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6)$ 表示, 即有

$$T = e^{\hat{\xi}}. \quad (3)$$

其中: (μ_1, μ_2, μ_3) 为相机在世界坐标系下的坐标值, μ_4, μ_5 和 μ_6 分别为世界坐标系下相机在 x, y 和 z 轴方向上的旋转量, $\hat{\xi}$ 可表示为

$$\hat{\xi} = \begin{bmatrix} 0 & \mu_6 & -\mu_5 & \mu_1 \\ -\mu_6 & 0 & \mu_4 & \mu_2 \\ \mu_5 & -\mu_4 & 0 & \mu_3 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (4)$$

在相机坐标系下, 根据相机模型, 相机平面点与三维空间点的投影关系可定义为

$$\pi(u, d) = \begin{bmatrix} \frac{u_x - c_x}{f_x} d & \frac{u_y - c_y}{f_y} d & d & 1 \end{bmatrix}^T. \quad (5)$$

其中: $u = (u_x, u_y)$ 为图像纹理坐标, f_x 和 f_y 为相机焦距, c_x 和 c_y 为相机中心坐标, d 为深度图 u 处的深度值. 式(5)的逆变换可表示为

$$\pi^{-1}(p) = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{x}{f_x} + c_x \\ \frac{y}{f_y} + c_y \\ z \end{bmatrix}^T, \quad (6)$$

其中 $p = (x, y, z)$ 为三维点云坐标.

2 混合位姿估计算法

2.1 混合位姿估计模型

基于CG的位姿估计方法是通过建立关键帧与当前帧图像间的映射关系实现对三维信息的配准的.

由于数据间重合程度直接关系到当前三维配准算法的有效性, 在位姿估计过程中, 该位姿估计方法能有效减少因配准误差的过分累计而产生的影响, 但其可靠性将随着数据重合率的减少而降低, 且收敛速度也将有所下降. 而基于CL的位姿估计策略是建立在相邻两帧数据的基础上, 由于其重合率高, 该方法可靠性高, 且收敛速度快. 但在地图创建的过程中, 位姿估计误差的过分累计使得该方法产生较大的漂移. 在室内环境中, 基于稠密纹理位姿估计方法比ICP算法具有更高精度^[8]. 当环境中存在较多重复纹理(如地面)时, 基于稠密纹理的位姿估计方法将产生较大的配准误差. 而基于ICP算法的位姿估计方法与环境中的纹理信息无关, 能有效克服基于图像纹理的位姿估计方法的不足.

针对上述问题, 本文提出一种基于结合局部纹理约束的混合位姿估计方法. 该方法通过结合当前帧与关键帧之间的纹理与点云信息间建立位姿估计模型, 并利用上一帧数据与当前帧间的纹理信息建立局部约束条件, 从而在保证位姿估计精度的同时, 提高算法可靠性. 该算法模型可定义为

$$\begin{aligned} E &= \min \sum (\alpha r_{cl}^2, \beta r_{cg}^2, \lambda r_p^2)^T, \\ r_{cl} &= I_{k-1}(u) - I_k(\pi^{-1}(T(\xi)\pi(u, d_{k-1}))), \\ r_{cg} &= I_{ref}(u) - I_k(\pi^{-1}(T(\xi)T_{k-1}\pi(u, d_{ref}))), \\ r_p &= (p_k - T(\xi)T_{k-1}p_{ref})^T n_k. \end{aligned} \quad (7)$$

其中: α, β 和 λ 为权重系数, r_{cl} 为 k 与 $k-1$ 之间的纹理误差值, r_{cg} 和 r_p 分别为 k 时刻与参考帧间的纹理和点云误差, $T(\xi)$ 为 $k-1$ 时刻到 k 时刻的位姿增量, T_{k-1} 为 $k-1$ 时刻相机位姿. 由欧拉方程可得能量函数的极值条件为

$$\frac{\partial E}{\partial \xi} = \sum J_r^T J_r r + \sum J_r \xi = 0. \quad (8)$$

其中

$$\begin{aligned} r &= [r_{cl} \ r_{cg} \ r_p]^T, \\ J_r &= [J_{cl} \ J_{cg} \ J_p] = \left[\frac{\partial r_{cl}}{\partial \xi} \ \frac{\partial r_{cg}}{\partial \xi} \ \frac{\partial r_p}{\partial \xi} \right]. \end{aligned}$$

此时, 能量函数的极值条件可变形为

$$\xi = \left(\sum J_r^T J_r \right)^{-1} \sum J_r^T r. \quad (9)$$

为了有效地实现上述能量函数的优化, 通过迭代的方式不断地对式(9)进行求解, 当能量函数值满足收敛条件或者达到最大迭代次数后, 位姿估计算法过程结束.

通过实验可知, 当 $\alpha = 0.12, \beta = 0.08, \lambda = 1$ 时, 算法精度最优^[13]. 3组标准库数据如图1所示.

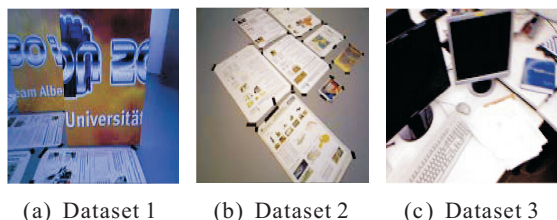


图 1 3组典型标准数据

将该数据作为位姿估计的输入, 所获取的估计位姿与真实值之间平移量的均方根误差 (RMSE) 如表 1 所示.

Dataset	Method				m
	ICP	CL	CG	CGL	
1	0.5834	0.0414	0.0251	0.0244	0.0108
2	1.4706	0.0301	0.0309	0.0278	0.0243
3	0.5715	0.1134	0.4676	0.1108	0.0849

通过对比标准 ICP 算法, 混合 frame-to-frame 与 keyframe-to-frame (CGL) 算法, CL 算法, CG 算法, 以及混合位姿估计 (GLP) 算法可知, 本文所提方法在精度上均具有较高优势.

2.2 基于混合位姿估计模型的求解

本文采用由粗到精逐层细化的金字塔策略求解混合位姿估计模型. 通过低分辨率图像层时, 获取相机的初始位姿初始估计, 并逐层优化到原始图像层, 从而实现高精度位姿估计. 构建 $k-1$ 时刻和 k 时刻与关键帧处数据金字塔模型, 针对图像金字塔中每一层数据, 利用相机位姿变换关系和投影模型, 分别以 $k-1$ 时刻与关键帧处数据为参考, 获取当前图像的合成图像信息, 运用式 (9) 计算当前极值 ξ 并更新相机位姿. 进行迭代运算, 直到能量函数收敛, 从而完成相机位姿估计, 具体步骤如下.

Step 1: 构建图像 I_{k-1} 、 I_k 、 I_{ref} 、 P_k 、 P_{ref} 和金字塔模型 F_{k-1} 、 F_k 、 F_{ref} 、 F_{pk} 、 F_{pref} , 迭代初始层 $\text{level} = \text{maxlevel}$;

Step 2: 如果 $\text{level} \geq 0$, 则进入 Step 3, 否则结束位姿估计过程;

Step 3: 分别以 $F_{k-1}(\text{level})$ 和 $F_{\text{ref}}(\text{level})$ 图像为基准, 利用空间投影关系获取 $F_k(\text{level})$ 在当前位姿下的合成图像, 建立点云集 $F_{pk}(\text{level})$ 与 $F_{pref}(\text{level})$ 之间点的对应关系;

Step 4: 分别对 J_p 、 J_{cl} 和 J_{cg} 进行求解, 将结果代入式 (9) 求解 ξ , 并根据位姿增量公式 $T = e^{\xi}T$ 更新当前位姿;

Step 5: 计算当前误差值, 如果误差变化小于设定阈值或者达到最大迭代次数后, $\text{level} = \text{level} - 1$, 则返回 Step 2.

3 地图更新与闭环优化

3.1 地图更新与关键帧的选取

在准确的位姿估计的基础上, 通过建立关键帧选取机制, 结合式 (5) 建立当前环境的三维地图, 并运用局部坐标与全局坐标间的位姿关系实现环境地图更新. 在位姿估计过程中, 配准算法精度一定程度上依赖于输入数据间的重合程度. 鉴于此, 本文考虑采用关键帧间的重合程度作为关键帧的选取依据之一, 如下式所示:

$$s = \frac{\sum G(r_{cg}(\xi))}{N}. \quad (10)$$

其中: $G(x)$ 为当前深度图中有效值的判别函数, 若图像 x 处深度为有效值, 则 $G(x) = 1$, 否则 $G(x) = 0$; N 为图像 I_{ref} 中有效深度值的数量; s 为第 k 帧与关键帧间有效数据的重合率, 当 s 小于阈值时, 添加该帧为关键帧. 同时, 为了保证相机在位姿估计过程中的连续性, 当相机移动距离超过阈值条件后, 将确定当前帧为关键帧, 并执行三维地图的更新. 关键帧选取的数学描述如下:

$$\{I_k \in I_{kf} | s \leq s_0 \text{ or } \text{distance}(I_k, I_{\text{ref}}) \geq d_0\}. \quad (11)$$

其中: I_{kf} 为关键帧集合, s_0 为图像最小重合率, d_0 为相机平移最大距离.

3.2 闭环检测与优化

针对因配准误差的不断累计而产生的漂移现象, 本文采用视觉信息闭环检测方法, 并结合 TORO 算法实现对所生成的三维模型的优化. 该闭环检测方法在获取关键帧后, 利用位姿关系确定当前可能闭环节点, 并求解该关键帧与可能闭环节点处关键帧间的相对位姿, 从而完成对闭环信息的获取. TORO 算法利用相机全局位姿和当前观测相对位姿构成相机位姿关系树, 通过最小化闭环处相机位姿推算偏移量与当前观测偏移量的差异, 从而实现对全局位姿的优化, 其误差函数为

$$T = \min_{\mathbf{T}} F(\mathbf{T}),$$

$$F(\mathbf{T}) = \sum_n e(T_i, T_j, T_{ij})^T \Psi_{ij} e(T_i, T_j, T_{ij}). \quad (12)$$

其中: $\mathbf{T} = \{T_1, T_2, \dots, T_n\}$ 为相机位姿的描述集合; T_{ij} 为通过闭环检测获取的第 i 和第 j 时刻相机的相对位姿; Ψ_{ij} 为在 i 与 j 处相机位姿间观测相关性矩阵; $e(T_i, T_j, T_{ij})$ 为在 i 和 j 时刻相机推算相对位姿与观测值间的误差, 即闭环误差. 若 $e(T_i, T_j, T_{ij}) = 0$, 则从第 i 到第 j 时刻相机的位姿满足闭环条件, 认为此时相机位姿漂移量为零.

4 实验结果

本文采用的深度相机为 Kinect-XBOX 360, RGB

图像分辨率为 640×480 , 最高帧频为 30 fps, 水平视角为 52° . 在实验过程中, Kinect 架设在 Pioneer3-DX 移动机器人平台上, 如图 2 所示.



(a) 移动机器人平台 (b) 实验环境

图 2 移动机器人平台和实验环境

本文所使用的 PC 机配备有 Intel i5 CPU, 其主频为 2.8 GHz. 在该硬件平台上, 位姿估计算法平均处理时间为 35 ms, 满足实时性要求. 为了验证算法的有效性, 在室内真实环境下进行地图创建实验.

为了验证混合位姿估计算法的精度, 离线采集 100 帧 Kinect 数据帧. 以离线数据中的第一帧作为关键帧, 分别采用 CL、CG 和本文所提方法进行地图创建实验, 上述方法所生成的三维地图如图 3(a)~图 3(c) 所示.

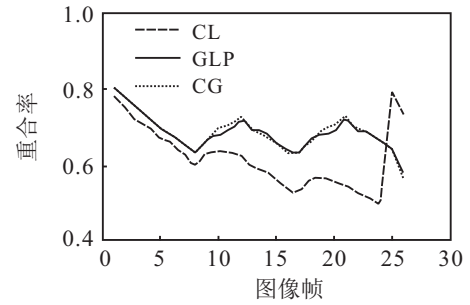


(a) GLP (b) CG (c) CL

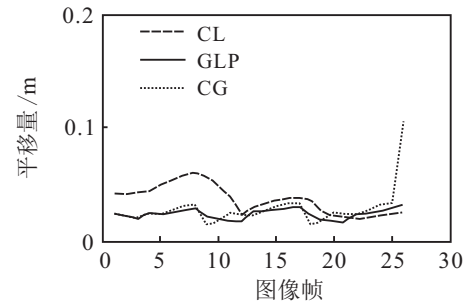
图 3 GLP、CG 和 CL 方法

由图 3 可以看出, 图像间数据重合程度的减小使得基于 CG 的位姿估计方法在第 3 个关键帧进入时跟踪失败, 如图 3(b) 中圆所包含的区域; 基于 CL 的位姿估计算法在跟踪性能上与前者相比更具优势, 但位姿估计误差的不断累计, 导致测试结束后产生一定的偏差, 如图 3(c) 中圆所包含的区域; 而本文算法在跟踪性能上与 CL 方法保持一致, 且所生成的地图最优. 同时, 本文将 CL 中第 2 个关键帧与第 3 个关键帧间的图像序列作为测试数据, 分别记录使用 3 种方法生成的平移量、旋转量和帧间重合率 s , 如图 4 所示.

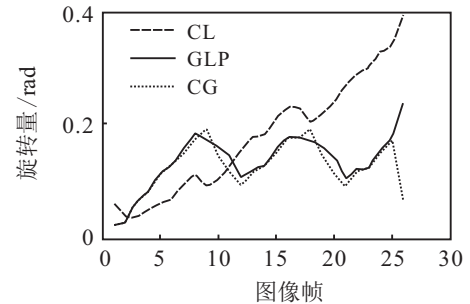
由图 4 可以看出, 本文算法与基于 CG 方法在跟踪精度上具有较高的一致性, 当第 26 帧图像进入后, 基于 CG 的位姿估计方法旋转值与平移值出现突变, 并与其他两种方法趋势相反, 此时位姿估计失败. 上述实验结果表明, 本文算法有效结合了 CG 和 CL 位姿估计策略的优势, 在避免相机移动过程中位姿估计失败的同时, 保证了位姿估计精度.



(a) 3 种方法的重合率



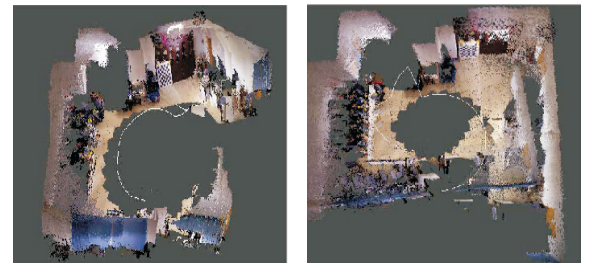
(b) 3 种方法的平移量



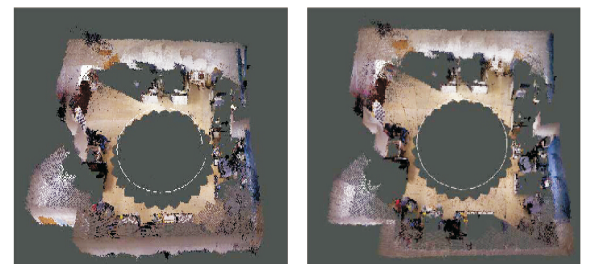
(c) 3 种方法的旋转量

图 4 GLP、CG 和 CL 方法比较

为了进一步验证所提出算法的有效性, 本文利用移动机器人平台在真实环境下进行实时地图创建实验. 室内三维地图创建实验结果如图 5 所示.



(a) 三维地图 ($\lambda=0, \alpha=\beta=1$) (b) 三维地图 ($\lambda=1, \alpha=\beta=0$)



(c) 未经闭环优化的三维地图 ($\lambda=1, \alpha=0.12, \beta=0.08$) (d) 经闭环优化的三维地图 ($\lambda=1, \alpha=0.12, \beta=0.08$)

图 5 室内三维地图创建实验

在实验过程中, PC 机采集到图像帧后, 利用位姿估计算法实现当前移动机器人的定位, 若当前帧数据与关键帧数据间重合程度 s_0 小于 0.5 或者最大距离值 d_0 大于 0.3, 则进入新的关键帧, 并根据式 (5) 更新当前三维地图。

图 5(a) 为 $\lambda = 0, \alpha = \beta = 1$ 时所生成的三维地图。由于此时该方法缺乏空间点云约束, 且环境中存在大量相似纹理 (如地面), 会产生大幅度的下沉现象。

图 5(b) 为 $\lambda = 1, \alpha = \beta = 0$ 所生成的三维地图。此时, 该模型为标准 ICP 的位姿估计方法, 该 ICP 位姿估计方法会因在空间结构过于简单或者复杂时陷入局部最优而产生匹配错误的现象。

图 5(c) 和图 5(d) 为 $\lambda = 1, \alpha = 0.12, \beta = 0.8$ 时所生成的三维地图。位姿估计过程中的误差不断累计, 使得在所生成的地图产生了漂移, 如图 5(c) 所示。为了有效实现地图优化, 本算法在新增关键帧后进行闭环检测, 该闭环检测方法通过建立当前关键帧与已获取的关键帧间位姿关系, 实现闭环检测, 若位姿估计成功则表明当前相机观察到闭环。在观察到闭环后, 运用 TORO 算法实现对位姿的优化, 并重新建立全局地图, 最终经优化后的地图如图 5(d) 所示。

上述实验结果表明, 所生成的三维模型与真实场景一致, 进一步验证了本文算法的有效性和可行性。

5 结 论

本文提出了一种基于 RGB-D 相机的移动机器人实时稠密三维重建方法, 实现了在室内复杂环境下的三维建模。该方法利用 Kinect 获取环境中纹理和深度信息, 在关键帧选取机制下, 通过引入局部和全局约束建立混合位姿估计方法, 在确保位姿估计精度的同时减少定位失败率。结合视觉闭环检测算法和 TORO 算法实现对所生成的三维地图的全局一致性优化。在室内真实环境下, 利用移动机器人平台进行地图创建, 实验结果表明, 本算法能有效建立与真实场景一致的三维地图, 验证了所提出算法的有效性和可行性。在未来工作中, 将进一步采用机器学习等方式建立鲁棒的权值选取规则, 以提高三维地图创建过程中位姿估计的鲁棒性。

参考文献(References)

- [1] Nüchter A, Lingemann K, Hertzberg J, et al. 6D SLAM-3D mapping out door environments[J]. J of Field Robotics, 2007, 24(8/9): 699-722.
- [2] 杨鸿, 钱堃, 戴先中, 等. 基于 Kinect 传感器的移动机器人室内环境三维地图创建[J]. 东南大学学报: 自然科学版, 2013, 43(1): 183-187.
(Yang H, Qian K, Dai X Z, et al. Kinect-based 3D indoor environment map building for mobile robot[J]. J
- of Southeast University: Natural Science Edition, 2013, 43(1): 183-187.)
- [3] 于振中, 郑为湊, 刘鑫, 等. 基于 Kinect 的移动机器人实时局部路径规划[J]. 计算机工程, 2013, 39(4): 243-247.
(Yu Z Z, Zheng W C, Liu X, et al. Real-time local path planning for mobile robot based on Kinect[J]. Computer Engineering, 2013, 39(4): 243-247.)
- [4] Abraham B, Samuel P, He R, et al. Estimation, planning, and mapping for autonomous flight using an RGB-D camera in GPS-denied environments[J]. The Int J of Robotics Research, 2012: 31(11): 1320-1343.
- [5] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using kinect-style depth cameras for dense 3D modeling of indoor environments[J]. Int J of Robotics Research, 2012, 12(2): 647-663.
- [6] Henry P, Krainin M, Herbst E, et al. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments[C]. The 12th Int Symposium on Experimental Robotics. Berlin: Springer Berlin Heidelberg, 2014: 477-491.
- [7] Newcombe Richard A, Davison Andrew J, Izadi S, et al. Kinectfusion: Real-time dense surface mapping and tracking[C]. The 10th IEEE Int Symposium on Mixed and Augmented Reality. New York: IEEE Press, 2011: 127-136.
- [8] Steinbrucker F, Sturm J, Cremers D. Real-time visual odometry from dense RGB-D images[C]. IEEE Int Conf on Computer Vision Workshops. New York: IEEE Press, 2013: 719-722.
- [9] Tommi T, Cedric A, Andrew I C, et al. Direct iterative closest point for real-time visual odometry[C]. 2011 IEEE Int Conf on Computer Vision Workshops. New York: IEEE Press, 2011: 2050-2056.
- [10] Mester R. Recursive live dense reconstruction: Some comments on established and imaginable new approaches[C]. 2011 IEEE Int Conf on Computer Vision Workshops. New York: IEEE Press, 2011: 712-714.
- [11] Giorgio G, Cyrill S, Wolfram B. Non-linear constraint network optimization for efficient map learning[J]. IEEE Trans on Intelligent Transportation Systems, 2009, 10(3): 428-439.
- [12] Grisetti G, Slawomir G, Cyrill S, et al. Efficient estimation of accurate maximum likelihood maps in 3D[C]. IEEE/RSJ Int Conf on Intelligent Robots and Systems. New York: IEEE Press, 2007: 3472-3478.
- [13] Jürgen Sturm. RGB-D SLAM dataset and benchmark[DB/OL]. (2012-10-7). <https://vision.cs.tum.edu/data/datasets/rgbd-dataset>.