

论文《RGB-D SLAM in Indoor Environments with STING-Based Plane Feature Extraction》补充实验方案

孙沁璇, 2017.9.13

基于位姿图的后端优化

SLAM 系统的前端(frontend)采用基于 PF-RGBD-CPE 方法的视觉里程计(VO), 后端(backend)采用位姿图优化[1][2]的方法。位姿图的节点表示机器人在某一时刻的位姿, 边表示节点之间通过传感器观测获得的约束信息。位姿图优化的目标是最大化观测似然来得到节点的最佳配置。

通过 PF-RGBD-CPE 方法可以得到第 i 帧与第 $i+1$ 帧之间的旋转变换 $\mathbf{R}_{i(i+1)} \in SO(3)$ 和平移变换 $\mathbf{t}_{i(i+1)} \in \mathbf{R}^3$, 令

$$\mathbf{G}_{i(i+1)} = \begin{bmatrix} \mathbf{R}_{i(i+1)} & \mathbf{t}_{i(i+1)} \\ \mathbf{0}_{1 \times 3} & 1 \end{bmatrix} \in SE(3) \quad (1)$$

令 \mathbf{G}_i 为机器人在第 i 帧时的位姿, 并假设机器人 SLAM 过程中的第一帧的传感器坐标系即为全局坐标系, 即 $\mathbf{G}_1 = \mathbf{I}_{4 \times 4}$ 。则有

$$\mathbf{G}_i = \mathbf{G}_{i-1} \mathbf{G}_{(i-1)i} \quad (2)$$

在位姿图优化的过程中, 分别用李代数 $\xi_i = \log_{SE(3)}(\mathbf{G}_i) \in se(3)$ 和 $\xi_{ij} = \log_{SE(3)}(\mathbf{G}_{ij}) \in se(3)$ 表示机器人在第 i 帧的位姿(作为图的节点), 以及由传感器得到的第 i 帧与第 j 帧机器人位姿之间的空间约束(作为图的边), 位姿图的示意图如图 1 所示。

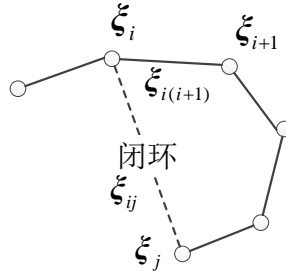


图 1 位姿图示意图

令 $f_{ij}(\xi_i, \xi_j)$ 表示根据节点 ξ_i 与 ξ_j 计算得到的其相对位姿变换, 并由此定义误差 e_{ij} 为

$$e_{ij}(\xi_i, \xi_j) = f_{ij}(\xi_i, \xi_j) - \xi_{ij} \quad (3)$$

假设观测噪声为高斯白噪声, 则第 i 帧与第 j 帧之间的边对应的负对数似然为

$$F_{ij}(\xi_i, \xi_j) = e_{ij}(\xi_i, \xi_j)^T \Omega_{ij} e_{ij}(\xi_i, \xi_j) \quad (4)$$

其中 Ω_{ij} 为描述 ξ_{ij} 不确定性的信息矩阵(本文实验中均设为单位矩阵)。假设每次观测都是独立的, 则有

$$F(\xi) = \sum_{\langle i, j \rangle \in C} F_{ij}(\xi_i, \xi_j) \quad (5)$$

其中 $\xi = [\xi_1^T, \xi_2^T, \dots, \xi_N^T]^T$ 表示图中所有的节点, $C = \{\langle i, j_1 \rangle, \dots, \langle i_M, j_M \rangle\}$ 表示约束 ξ_{ij} 存在的所有下标组合的集合。最大化 $F(\xi)$ 从而得到节点的最佳配置 ξ^*

$$\xi^* = \arg \min_{\xi} F(\xi) \quad (6)$$

式(6)可以通过非线性优化(Gauss-Newton 方法)来迭代求解。对于每一步迭代的过程, 求解 $e_{ij}(\xi_i, \xi_j)$ 在迭代的初值 $\hat{\xi} = [\hat{\xi}_1^T, \hat{\xi}_2^T, \dots, \hat{\xi}_N^T]^T$ 处的一阶泰勒展开

$$e_{ij}(\hat{\xi}_i + \Delta\xi_i, \hat{\xi}_j + \Delta\xi_j) = e_{ij}(\hat{\xi} + \Delta\xi) \approx e_{ij}(\hat{\xi}) + \mathbf{J}_{ij} \Delta\xi \quad (7)$$

其中 \mathbf{J}_{ij} 是 $e_{ij}(\hat{\xi} + \Delta\xi)$ 对 $\hat{\xi}$ 的 Jacobian 矩阵。式(7)代入式(4)可得

$$\begin{aligned} F_{ij}(\hat{\xi} + \Delta\xi) &= e_{ij}(\hat{\xi} + \Delta\xi)^T \Omega_{ij} e_{ij}(\hat{\xi} + \Delta\xi) \\ &\approx \left(e_{ij}(\hat{\xi}) + \mathbf{J}_{ij} \Delta\xi \right)^T \Omega_{ij} \left(e_{ij}(\hat{\xi}) + \mathbf{J}_{ij} \Delta\xi \right) \\ &= c_{ij} + 2\mathbf{b}_{ij} \Delta\xi + \Delta\xi^T \mathbf{H}_{ij} \Delta\xi \end{aligned} \quad (8)$$

其中

$$\begin{aligned} c_{ij} &= e_{ij}(\hat{\xi})^T \Omega_{ij} e_{ij}(\hat{\xi}) \\ \mathbf{b}_{ij} &= 2e_{ij}(\hat{\xi})^T \Omega_{ij} \mathbf{J}_{ij} \\ \mathbf{H}_{ij} &= \mathbf{J}_{ij}^T \Omega_{ij} \mathbf{J}_{ij} \end{aligned} \quad (9)$$

再将式(8)代入式(5)可得

$$\begin{aligned} F(\hat{\xi} + \Delta\xi) &= \sum_{\langle i,j \rangle \in C} F_{ij}(\hat{\xi} + \Delta\xi) \\ &\approx \sum_{\langle i,j \rangle \in C} (c_{ij} + 2\mathbf{b}_{ij} \Delta\xi + \Delta\xi^T \mathbf{H}_{ij} \Delta\xi) \\ &= c + 2\mathbf{b}^T \Delta\xi + \Delta\xi^T \mathbf{H} \Delta\xi \end{aligned} \quad (10)$$

其中

$$c = \sum_{\langle i,j \rangle \in C} c_{ij}, \quad \mathbf{b} = \sum_{\langle i,j \rangle \in C} \mathbf{b}_{ij}, \quad \mathbf{H} = \sum_{\langle i,j \rangle \in C} \mathbf{H}_{ij} \quad (11)$$

$\Delta\xi$ 可以通过解线性方程来求解

$$\mathbf{H} \Delta\xi^* = -\mathbf{b} \quad (12)$$

则下一步迭代的初值 $\hat{\xi}^*$ 可由式(13)计算得到。

$$\hat{\xi}^* = \hat{\xi} + \Delta\xi^* \quad (13)$$

实验结果

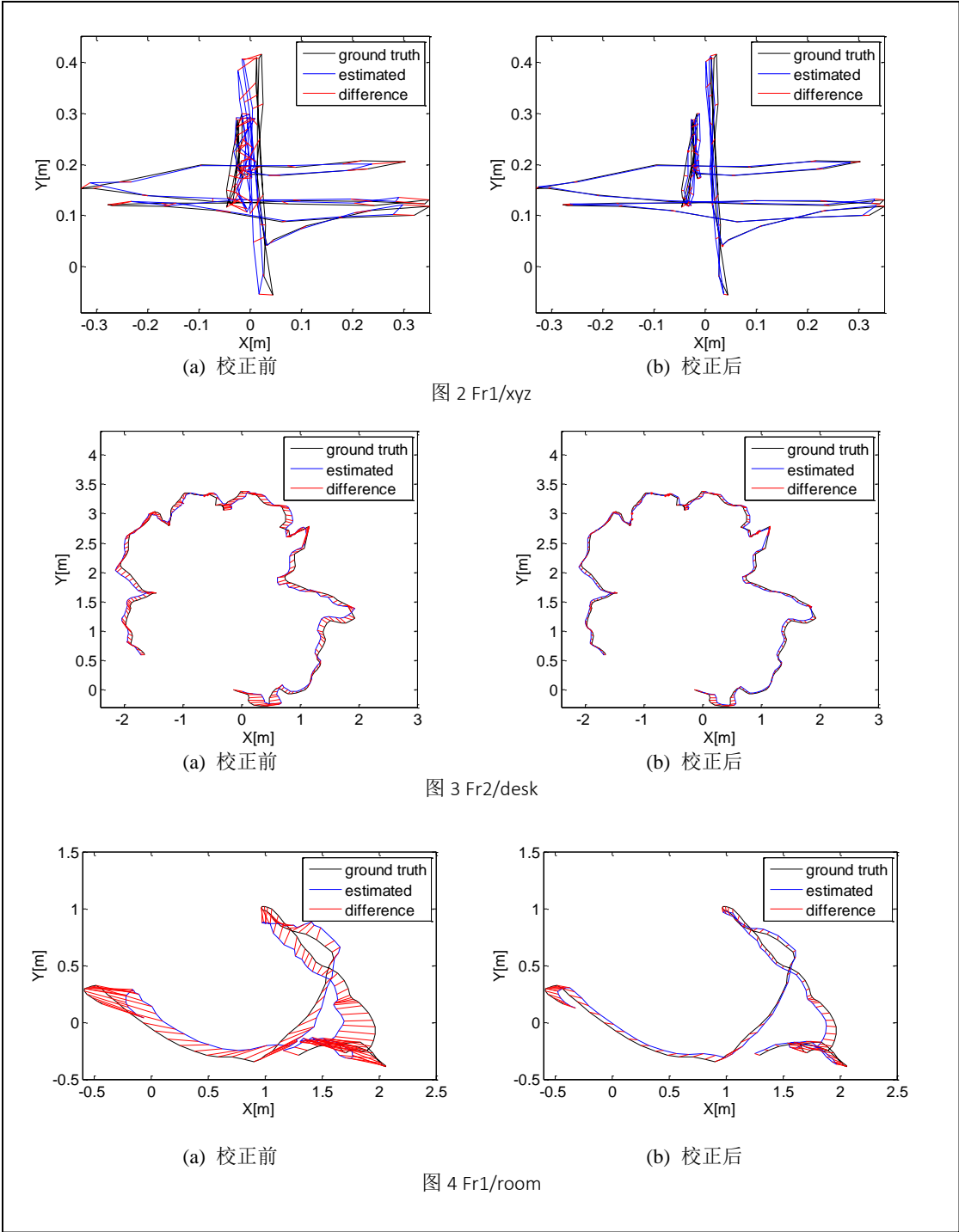
在实验中, SLAM 系统采用基于 PF-RGBD-CPE 方法的视觉里程计(VO)进行前端位姿估计, 通过 g2o[3]来实现地图校正作为整个 SLAM 系统的后端优化, 并将实现结果与另外两个 SLAM 系统——ElasticFusion[4]以及基于 RGBD-ICP 和 SBA 的 SLAM 系统[5], 在 Freiburg 数据集上运行的结果进行对比, 其 ATE 的对比结果如表 1 所示。

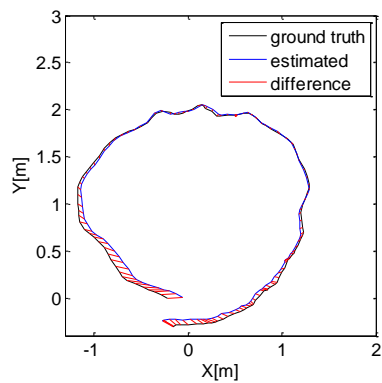
表 1 三个 SLAM 系统的 ATE 对比结果

	PF-RGBD-CPE + g2o	ElasticFusion	RGBD-ICP + SBA
Fr1/xyz	0.011m	0.011m	0.014m
Fr2/desk	0.053m	0.071m	0.113m
Fr1/room	0.083m	0.068m	1.322m
Fr3/cabinet	0.032m	Failed	1.152m
Fr2/pioneer360	0.209m	Failed	Failed

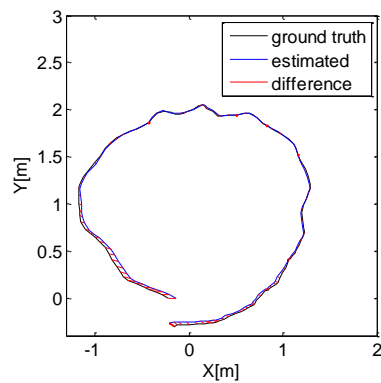
由表 1 可以看出, 本文提出的 SLAM 系统在除了 Fr1/room 之外的其他四个图像序列上都得到了最小 ATE 结果。在图像序列 Fr1/room 上 ElasticFusion 得到最小 ATE, 然而, ElasticFusion

在 Fr3/cabinet 和 Fr2/pioneer360 上都运行失败，因为 Fr3/cabinet 场景简单，纹理信息较少，导致 ElasticFusion 中基于光度误差(photometric error)的数据关联失败。Fr2/pioneer360 的采集场景面积较大，有多处超出了 Kinect 传感器的有效深度测量范围，而 ElasticFusion 是基于稠密地图创建的 SLAM 方法，在这种情况下运行失败。对于本文提出的 SLAM 方法，由于平面特别的利用，对于上述极端场景具备一定的鲁棒性，在 Fr3/cabinet 和 Fr2/pioneer360 上均成功运行并且取得了较好的结果。图 2 到图 6 显示了五个图像序列上后端地图校正前与校正后轨迹与 groundtruth 的对比情况。



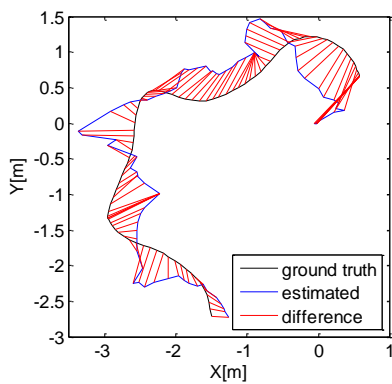


(a) 校正前

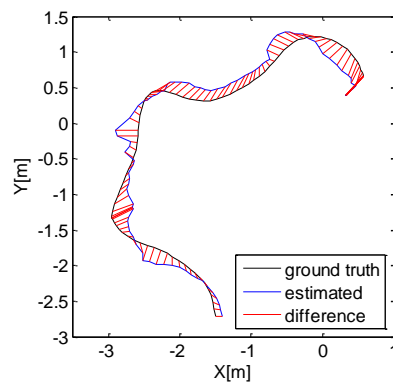


(b) 校正后

图 5 Fr3/cabinet



(a) 校正前



(b) 校正后

图 6 Fr2/pioneer360

参考文献

- [1] E. Olson, J. Leonard, and S. Teller, “Fast iterative optimization of pose graphs with poor initial estimates,” in Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA), 2006, pp. 2262 – 2269.
- [2] G. Grisetti, C. Stachniss, and W. Burgard, “Non-linear constraint network optimization for efficient map learning,” IEEE Trans. on Intelligent Transportation Systems, 2009.
- [3] R. Kummerle, G. Grisetti, H. Strasdat, K. Konolige and W. Burgard, “G2o: A general framework for graph optimization,” in Proc. IEEE Int. Conf. Robot. Autom., Shanghai, China, 2011, pp. 3607-3613.
- [4] T. Whelan, R.F. Salas-Moreno, B. Glocker, A.J. Davison and S. Leutenegger, “ElasticFusion: Real-time dense SLAM and light source estimation,” Int. J. Robot. Res., vol. 35, no. 14, pp. 1697-1716, 2016.
- [5] P. Henry, M. Krainin, E. Herbst, X. Ren and D. Fox, “RGB-D mapping: Using kinect-style depth cameras for dense 3D modeling of indoor environments,” Int. J. Robot. Res., vol. 31, no. 5, pp. 647-663, 2012.