

融合光度和深度的视觉里程计改进算法

黄宴委, 董文康, 王 俊, 陈少斌

(福州大学电气工程与自动化学院, 福建 福州 350108)

摘要: 针对视觉里程计中对纹理较少的环境难以获得高精度相机位姿的问题, 提出一种融合光度和深度的视觉同步定位与绘图(SLAM)方法. 首先构造基于光度和深度的相机位姿优化函数, 通过 t 分布模型计算每个像素点的光度和深度误差权重, 对每一帧图像引入帧权重系数 λ 来平衡光度和深度, 并采用中值法求解帧权重 λ , 求得相机位姿. 然后对 10 个国际标准数据集进行仿真实验, 结果表明, 对纹理丰富的环境, 本方法能够保持 DVO SLAM 算法的建图精度; 对纹理较少的环境, 本方法的建图精度要高于 DVO SLAM 算法, 绝对路径误差降低 31.6%, 相对位姿误差降低 19.4%.

关键词: 同步定位与绘图(SLAM); 视觉里程计; 光度; 深度; 帧权重

中图分类号: TP242

文献标识码: A

An improved algorithm of visual odometry combining intensity and depth

HUANG Yanwei, DONG Wenkang, WANG Jun, CHEN Shaobin

(College of Electrical Engineering and Automation, Fuzhou University, Fuzhou, Fujian 350108, China)

Abstract: Aiming at the problem that the high-precision camera pose is difficult to obtain in the environment with less texture in the visual odometry, this paper proposes a visual simultaneous location and mapping (SLAM) method that combines intensity and depth. Firstly, the camera pose optimization function based on intensity and depth is constructed. The intensity and depth error weights of each pixel are calculated by the t -distribution model, and the frame weight coefficient λ is introduced for each frame image to balance the intensity and depth. The value method solves the frame weight λ and obtains the camera pose. Simulation experiments on 10 international standard datasets show that the proposed method can maintain the accuracy of DVO SLAM algorithm for texture-rich environments. For environments with less texture, the accuracy of this method is higher. With the DVO SLAM algorithm, the absolute trajectory error (ATE) is reduced by 31.6% and the relative pose error (RPE) is reduced by 19.4%.

Keywords: simultaneous location and mapping (SLAM); visual odometry; intensity; depth; frame weight

0 引言

同步定位与绘图(simultaneous location and mapping, SLAM)算法中的视觉里程计(visual odometry, VO)主要有特征点法和直接法实现^[1]. 相较于特征点法^[2], 直接法对特征较少的环境有较好的鲁棒性, 并且可以构建稠密和半稠密地图. 文[3]利用 RGB-D 相机(深度相机)获取的深度图像, 通过迭代最近点算法估计相机位姿, 构建三维稠密地图. 文[4]提出在假设光度误差服从 t 分布模型, 采用光度误差能够较好地构建纹理丰富环境的地图, 但是对纹理较少而结构较明显的环境建图精度不高. 文[5]在文[4]基础上同时融合了深度和光度, 提出一种 DVO SLAM 方法, 其结果比文[3]中的绝对路径误差(absolute trajectory error, ATE)降低了 83%. 文[6]在保留光度误差的条件下, 通过逆深度误差实现前端视觉里程计的设

收稿日期: 2018-12-25

通讯作者: 黄宴委(1976-), 教授, 主要从事智能控制与信息处理研究, sjtu_huanghao@fzu.edu.cn

基金项目: 福建省科技计划资助项目(2019H0007); 现代精密测量与激光无损检测福建省高校重点实验室基金资助项目(2018XKA005)

计, 比直接采用深度值降低了相对位姿误差 (relative pose error, RPE). 文[7]采用 t 分布模型拟合光度误差, 用高斯噪声模型拟合深度误差, 提高了位姿的准确性. 文[8]提出一种基于概率面元的 RGB-D SLAM 算法提高位姿精度. 本研究提出一种融合光度和深度的改进视觉里程计 (modified visual odometry, MVO) 算法, 提高 SLAM 算法对纹理较少环境的建图精度. 在文[5]基础上, 考虑不同场景下的纹理和结构情况, 采取不同权重构建目标函数, 求解每个像素点的光度和深度误差权重, 提高 VO 估计的位姿的准确性. 在后端算法中, 采用基于 g2o 框架的图优化方法得到全局一致的相机位姿^[9]. 对 10 个标准数据集的性能测试^[10]结果显示, MVO SLAM 具有很高的建图精度, 在纹理较少的环境时优势明显.

1 基于光度和深度的 VO

深度相机获取的一帧 RGB-D 图像, 包含光度图像 L 和深度图像 D . 已知光度图像 L 上的像素点 x , 像素坐标为 $[x, y]^T$, x, y 分别是像素坐标系下的横、纵坐标. 从深度图像 D 中, 得到像素点 x 的深度值 d . 由针孔相机模型得到像素点 x 在相机坐标系下的相机坐标为 $P = [X \ Y \ Z]^T$, X, Y, Z 表示相机坐标系下的坐标轴.

$$P = \pi^{-1}(x, d) = K^{-1} \cdot d \cdot \tilde{x} \quad (1)$$

式中: K 为相机内参矩阵, 是满秩矩阵, 可通过相机标定得到; K^{-1} 为矩阵 K 的逆矩阵; $\pi^{-1}(\cdot)$ 为像素坐标到相机坐标 P 的映射函数; \tilde{x} 为 x 的齐次坐标 $[x, y, 1]^T$,

$$\tilde{x} = \pi(P) = \frac{KP}{d} \quad (2)$$

式中: $\pi(P)$ 为相机坐标变换到齐次像素坐标的函数.

1.1 相机位姿变换函数

相机的位姿变换如图 1 表示. 对于相邻两个时刻的 RGB-D 图像数据帧 L_1, D_1 和 L_2, D_2 , 其中: L_1, D_1 表示前一时刻的光度图和深度图; L_2, D_2 表示后一时刻的光度图和深度图.

以光度图 L_1, L_2 为例, 用变换矩阵 T 表示 L_1 帧到 L_2 帧的变换,

$$T = \begin{bmatrix} R_{3 \times 3} & t_{3 \times 1} \\ 0 & I \end{bmatrix} \quad (3)$$

式中: $R_{3 \times 3}$ 是旋转矩阵; $t_{3 \times 1}$ 是平移向量; $T \in SE(3)$, $SE(3)$ 是特殊欧氏群. 采用相对于 $SE(3)$ 空间的李代数 $se(3) = [\xi \in \mathbf{R}^6, \xi^\wedge \in \mathbf{R}^{4 \times 4}]$ 表示相机位姿变换, 其中符号 \wedge 表示将向量转换成矩阵. 则 ξ 为,

$$\xi = (v_1, v_2, v_3, w_1, w_2, w_3)^T \quad (4)$$

式中: v_1, v_2, v_3 称为平移线速度; w_1, w_2, w_3 称为旋转角速度. 则式(3) 经指数矩阵变换得:

$$T = \exp(\xi^\wedge) \quad (5)$$

对光度图 L_1 上的任一像素点 x , 根据深度图像得到相应的深度值 $Z = Z(x)$, 其中 $Z(\cdot)$ 表示取像素点的深度值. 首先根据式(1) 将该像素点映射到对应于 L_1 帧的相机坐标 $P = [X \ Y \ Z]$, 然后根据相机变换矩阵 T , 将 L_1 帧下的相机坐标 P 变换到 L_2 帧下的相机坐标 P' .

$$P' = TP \quad (6)$$

由式(2)得到 P' 映射到 L_2 帧下的齐次像素坐标 $\tilde{x}' = \pi(P')$, 进一步得到像素坐标 x' . 将从 L_1 帧上任一像素点 x 像素坐标变换到 L_2 帧, 像素坐标的过程用变换函数 τ 表示,

$$x' = \tau(x, T) = \pi(T\pi^{-1}(x, z(x))) \quad (7)$$

1.2 融合光度和深度的位姿目标函数

定义 L_1 帧中的像素点 x 由式(7)中的变换函数 τ 变换到 L_2 帧下的像素点后, 两像素点的光度误差 r_1 为

$$r_1 = I(\tau(x, T)) - I(x) \quad (8)$$

式中: I 为像素点的光度值. 也可得到像素点 x 经变换后的深度误差 r_z 为

$$r_z = Z(\tau(x, T)) - [T\pi^{-1}(x, z(x))]_z \quad (9)$$

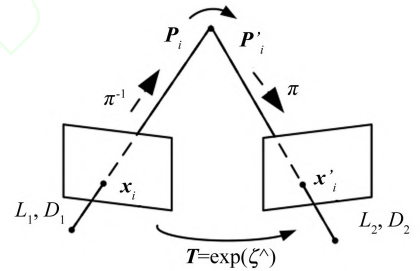


图 1 相邻两帧图像位姿变换

Fig.1 Pose transformation of two adjacent frames

其中: $[\cdot]_Z$ 为某点在相机坐标系 Z 轴坐标值. 式(9) 右边第二项为图像 D_1 的像素点由式(6) 所得的 Z 轴坐标值.

对于相邻两帧 RGB-D 图像, 设第 i 个像素点的光度误差为 $r_{1,i}$, 深度误差为 $r_{Z,i}$, 令 $\mathbf{r}_i = (r_{1,i}, r_{Z,i})^T$, 用向量 $\boldsymbol{\xi} \in \mathbf{R}^6$ 及其矩阵指数函数来表示相机位姿. 文[5] 采用 $v = 5$, 均值 $u = 0$, 协方差为 Σ 的二维 t 分布模型拟合光度和深度误差, 构建目标函数为

$$\boldsymbol{\xi}^* = \arg \min_{\boldsymbol{\xi}} \sum_{i=1}^n \omega_i \mathbf{r}_i^T \Sigma^{-1} \mathbf{r}_i \tag{10}$$

式中: n 表示每帧图像所取的像素点总数. 式(10) 只是将光度和深度以相同的比例进行融合, 并未考虑两者之间的权重. 由于不同场景的结构和纹理不同, 相邻两帧图像的光度误差和深度误差之间的权重应根据环境的结构、纹理的不同而有所区别. 为平衡两帧 RGB-D 图像的光度误差 $\mathbf{r}_1 = (r_{1,1}, \cdots, r_{1,n})$ 和深度误差之间 $\mathbf{r}_Z = (r_{Z,1}, \cdots, r_{Z,n})$ 的权重, 在每帧深度图的深度误差前加入权重系数 λ , 即 $\lambda \mathbf{r}_Z$, 通过最小二乘构建相机位姿的目标函数为

$$\min_{\boldsymbol{\xi}} h(\boldsymbol{\xi}) = \|\mathbf{r}_1\|^2 + \|\lambda \mathbf{r}_Z\|^2 = \sum_{i=1}^n (\mathbf{r}_{1,i}^T \mathbf{r}_{1,i} + \lambda^2 \mathbf{r}_{Z,i}^T \mathbf{r}_{Z,i}) \tag{11}$$

对每个像素的光度误差和深度分别加上权重 $W_{1,i}$ 和 $W_{Z,i}$, 最终构建目标函数为

$$\boldsymbol{\xi}^* = \min_{\boldsymbol{\xi}} \sum_{i=1}^n (\mathbf{r}_{1,i}^T W_{1,i} \mathbf{r}_{1,i} + \lambda^2 \mathbf{r}_{Z,i}^T W_{Z,i} \mathbf{r}_{Z,i}) \tag{12}$$

式中: $W_{1,i}$ 和 $W_{Z,i}$ 分别表示第 i 个像素点的光度误差和深度误差的权重; n 表示每帧图像所取的像素点总数; λ 是相邻两帧 RGB-D 图像的深度误差相较于光度误差所占的帧权重. 式(12) 是关于位姿 $\boldsymbol{\xi}$ 的非线性函数, 分别将光度误差函数 \mathbf{r}_1 和深度误差函数 \mathbf{r}_Z 进行线性化, 迭代得到最优的 $\boldsymbol{\xi}$. 假定函数 \mathbf{r}_1 和 \mathbf{r}_Z 是局部连续且光滑可导的, 分别将 \mathbf{r}_1 和 \mathbf{r}_Z 关于 $\boldsymbol{\xi}$ 进行一阶泰勒展开, 将所有 n 个像素, 代入式(12) 中, 用雅可比矩阵的形式表示,

$$\mathbf{J}^T \mathbf{W} \mathbf{J} \Delta \boldsymbol{\xi} = -\mathbf{J}^T \mathbf{W} \mathbf{r} \tag{13}$$

式中: \mathbf{J} 是 n (n 是所取的像素点总数) 个像素点的光度误差和深度误差关于 $\boldsymbol{\xi}$ 的雅可比矩阵, 维数为 $2n \times 6$, \mathbf{W} 是 n 个像素点的权重矩阵, 其维数是 $2n \times 2n$, $\Delta \boldsymbol{\xi}$ 是每次迭代的增量, 为 6×1 向量. \mathbf{J} 的具体形式为

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_1 \\ \lambda \mathbf{J}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{J}_1 \mathbf{J}_w \mathbf{J}_T \\ \lambda (\mathbf{J}_Z \mathbf{J}_w \mathbf{J}_T - [\mathbf{J}_T]_Z) \end{bmatrix} \tag{14}$$

式中: \mathbf{J}_1 和 \mathbf{J}_2 分别是 n 个像素点的光度误差和深度误差关于 $\boldsymbol{\xi}$ 的雅可比矩阵; \mathbf{J}_1 和 \mathbf{J}_Z 称为像素点的亮度和深度梯度矩阵, 其维数是 $n \times 2n$; \mathbf{J}_w 称为变换函数 τ 关于相机坐标系下的三维点的导数, 其维数是 $2n \times 3n$; \mathbf{J}_T 称为变换后的三维点对相机位姿 $\boldsymbol{\xi}$ 的导数, 其维数是 $3n \times 6$. 式(13) 通过迭代求解, 在第 k 次迭代 $\boldsymbol{\xi}^{(k)}$, 首先将等式左边经过柯列斯基分解, 然后使用高斯-牛顿迭代算法^[11] 求解位姿增量 $\Delta \boldsymbol{\xi}$, 并根据下式更新相机位姿 $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}(k+1) = \lg(\exp(\boldsymbol{\xi}^{(k)}) \exp(\Delta \boldsymbol{\xi})) \tag{15}$$

当迭代次数大于 100 次, 或者 $\Delta \boldsymbol{\xi}$ 小于 5×10^{-7} 时, 迭代停止, 将此时的 $\boldsymbol{\xi}^{(k)}$ 作为最终估计的相机位姿 $\boldsymbol{\xi}$.

1.3 基于 t 分布的像素点光度和深度误差权重

采用自由度为 v , 方差为 σ^2 的 t 分布误差模型来计算第 i 个像素点相应的 $W_{1,i}$ 和 $W_{Z,i}$. 以光度误差为例, 通过 t 分布模型推导出每个像素点的权重. 设第 i 个像素光度误差为 $r_{1,i}$, 光度误差权重为 $W_{1,i}$, 每个像素点的权重为:

$$W_{1,i} = (v+1) \left(v + \left(\frac{r_{1,i}}{\sigma_1} \right)^2 \right)^{-1} \tag{16}$$

为确定自由度 v 的值, 选择 TUM 提供的 RGB-D 数据集进行实验, 该数据集提供了由 Kinect 相机采集的 RGB 图像序列和深度图像序列, 以及由高精度运动捕捉系统获得的相机真实位姿, 并提供了用于评估所得位姿的绝对轨迹误差 (ATE) 和相对位姿误差 (RPE) 的工具. 分别取 v 为 1~10 十个整数在 fr1, fr2 系列数据集上进行实验, 计算视觉里程计所得位姿漂移量的均方根误差 (root mean square error, RMSE), 单位是 $\text{m} \cdot \text{s}^{-1}$, 该值越小表明估计的位姿越准确, 具体如表 1 所示. 表 1 第一列为 fr1, fr2 系列数据集, 第

2 至 11 列为不同 v 值下位姿漂移量的 RMSE.

表 1 不同 v 值下视觉里程计算法所得相机位姿漂移量的 RMSE
Tab.1 RMSE of camera pose drift obtained by VO under different v values

Dataset	1	2	3	4	5	6	7	8	9	10
fr1/360	0.891	0.672	0.531	0.473	0.429	0.482	0.543	0.628	0.743	0.947
fr1/desk	0.774	0.529	0.446	0.374	0.327	0.381	0.442	0.525	0.642	0.866
fr1/floor	0.958	0.726	0.642	0.576	0.521	0.579	0.654	0.743	0.842	1.053
fr1/room	0.719	0.604	0.516	0.461	0.413	0.467	0.523	0.616	0.731	0.957
fr2/desk	0.593	0.456	0.364	0.291	0.245	0.304	0.382	0.469	0.591	0.889
fr2/rpy	0.147	0.106	0.079	0.065	0.058	0.067	0.082	0.101	0.135	0.242
fr2/xyz	0.129	0.097	0.075	0.058	0.049	0.061	0.083	0.108	0.149	0.194

注: 黑体部分表示性能指标的优势.

从表 1 中可以看出, 当自由度 $v=5$ 时, 视觉里程计得到的相机位姿漂移量的 RMSE 在各数据集下达到最小, 因此, 选择自由度 $v=5$. 计算方差 σ_1^2 的方法为:

$$\sigma_1^2 = \frac{1}{n} \sum_{i=1}^n r_{1,j}^2 \frac{v+1}{\left(\frac{r_{1,i}}{\sigma_1}\right)^2}$$

(17)

将初始 σ_1 设为 σ_1^0 , 取值为 5, 通过式(16) 求解 σ_1 . 当取 k 个像素点后, 若 $|1/\sigma_1^k - 1/\sigma_1^0| < \varepsilon$ (本实验选择 ε 为 10^{-3}), 则迭代停止, 此时的 σ_1^k 作为最终的 σ_1 . 同样可得第 i 个像素的深度误差权重 $W_{z,i}$ 为:

$$W_{z,i} = \frac{v+1}{v + \left(\frac{r_{z,i}}{\sigma_z}\right)^2}$$

(18)

1.4 基于中值法的帧权重 λ

由式(13)得到矩阵 H :

$$H = J^T W J = J_1^T W_1 J_1 + \lambda^2 J_2^T W_z J_2$$

(19)

从式(19)中可以看出, λ 值起着平衡光度误差和深度误差的作用, 该值对相机位姿估计的准确性有重要的影响. 当环境的纹理较多时, 应该增加光度误差所占的权重; 当环境的纹理较少时, 应增加深度误差所占的权重. 通过中值动态地获取 λ , 即将每一帧光度图像的光度中值和深度图像的深度中值的比值作为 λ 的值, 即

$$\lambda = \frac{\text{median}(\boldsymbol{I})}{\text{median}(\boldsymbol{Z})}$$

(20)

式中: median 表示取中值. 从每帧光度图像的光度直方图得到光度中值, 范围是 $[0 \sim 255]$, 从深度图像得到深度数据值并等比到 $[0 \sim 255]$ 的值域, 并构建深度直方图得到深度中值. 实验表明该方法得到的 λ 值与人工选择的最佳 λ 值最为接近.

2 实验结果与分析

为验证本算法的有效性, 采用 10 个 TUM 标准数据集进行实验, 计算机 CPU 为 Intel core i7 八核 3.40 GHz, 内存 8 GB, 运行 Ubuntu 16.04 操作系统.

2.1 标准数据集实验

采用 fr1, fr2, 和 fr3 系列 RGB-D 数据集进行建图实验, 如图 2 所示. 各个数据集的结构、纹理状况不相同, 图 2(a) 是纹理丰富的环境, 图 2(b) 是纹理较少的环境, 比如墙壁、地面等地方.

MVO SLAM 在 fr2/xyz 数据集上构建的三维地

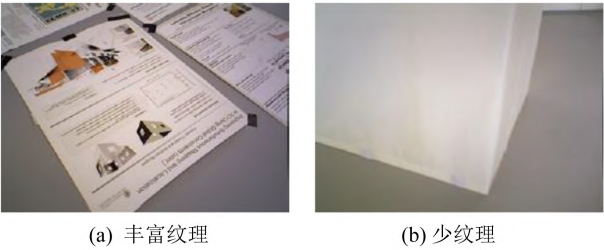


图 2 不同纹理环境示例

Fig.2 Example images of the datasets with varying texture

图如图 3 所示, 该数据集是对一个桌面的重构. 由图 3 可见, MVO SLAM 较好地构建了桌面的三维稠密点云地图, 并且能够还原较小的物件如鼠标, 可乐瓶等. MVO SLAM 在 fr2/xyz 数据集下形成的轨迹和真实轨迹的比较如图 4 所示. 从图 4 可见, MVO SLAM 较准确地估计了相机的运动轨迹, 尤其是在相机旋转较小的地方轨迹拟合的更好, 但是当相机旋转较大时, 仍会存在相对较大的误差.



图 3 fr2/xyz 稠密点云图
Fig.3 Dense point cloud of fr2/xyz

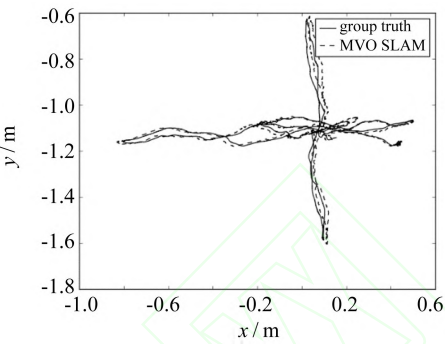


图 4 fr2/xyz 估计轨迹与真实轨迹对比
Fig.4 Trajectory of MVO plotted against ground truth for the fr2/xyz

MVO SLAM 与 DVO SLAM 两种方法建图性能指标对比如表 2 所示. 性能指标有 2 个: 1) 相对位姿误差 (RPE, [m/s]) 的均方根, 用来评估在一定的时间间隔内估计位姿的局部精度, 该值越小表示漂移越少; 2) 绝对路径误差 (ATE, [m]) 的均方根, 该值是估计的位姿与真实位姿之间的误差, 用于评估整个 SLAM 系统的精度, 该值越小表示估计的位姿与真实位姿越接近.

表 2 DVO SLAM 与 MVO SLAM 的性能指标 ATE 和 RPE
Tab.2 ATE and RPE between DVO SLAM and MVO SLAM

Dataset	ATE			RPE		
	DVO SLAM	MVO SLAM	Improvement	DVO SLAM	MVO SLAM	Improvement
fr1/360	0.238	0.231	2.94%	0.356	0.402	-12.92%
fr1/desk	0.169	0.187	-10.65%	0.240	0.251	-4.58%
fr1/desk2	0.148	0.153	-3.38%	0.210	0.225	-7.14%
fr1/floor	0.292	0.103	64.73%	0.409	0.307	24.94%
fr1/room	0.219	0.237	-8.22%	0.349	0.375	-7.45%
fr2/flowerbouquet	0.334	0.246	26.35%	0.581	0.473	18.59%
fr2/brownbackground	0.682	0.422	38.12%	0.924	0.631	31.71%
fr2/desk	0.134	0.125	6.72%	0.217	0.172	20.74%
fr2/rpy	0.028	0.019	32.14%	0.044	0.024	45.45%
fr3/office	0.101	0.112	-10.89%	0.159	0.177	-11.32%

注: 黑体部分表示性能指标的优势.

表 2 中有 10 个标准数据集, 其中 fr1/360, fr1/desk, fr1/desk2, fr1/room, fr3/office 等五个数据集含有丰富的纹理; fr1/floor, fr2/flowerbouquet, fr2/brownbackground, fr2/desk, fr2/rpy 等五个数据集纹理较少, 结构较丰富. 从表 2 可见, MVO SLAM 在纹理较丰富的环境下如 (fr1/room 等五个数据集) 所估计的位姿比 DVO SLAM 稍差, 但劣化程度不多, 基本上能够达到 DVO SLAM 算法的精度, 可以满足导航和路径规划的要求; 而对纹理较少的环境 (如 fr1/floor 等五个数据集) 有更高的精度, 五个数据集的绝对路径误差 (ATE) 相较 DVO SLAM 平均降低了 31.6%, 相对位姿误差 (RPE) 平均降低了 19.4%.

MVO SLAM 算法与一些先进算法在不同数据集下 ATE 的比较如表 3 所示. MRSSMap^[12] 是一种基于面元的 SLAM 算法, 该算法根据全局地图结构进行优化, RGB-D SLAMV2^[13] 是一种基于特征点的 SLAM

算法, Kintinuous^[14] 和 ElasticFusion^[15] 是基于 GPU 的稠密 SLAM 算法. 从表 3 中可见, 本算法在多数数据集上的 ATE 相比其他算法要低, 即精度更高, 说明算法具有更高的性能.

表 3 MVO SLAM 与先进算法在 ATE 上的比较

Tab.3 Comparison of MVO with respect to the existing state-of-the-art SLAM approaches in ATE					
Algorithm	fr1/360	fr1/floor	fr2/desk	fr2/rpy	fr3/office
DVO SLAM	0.238	0.292	0.134	0.028	0.101
RGB-SLAMV2	0.227	0.214	0.247	0.032	0.099
MRSMap	0.241	0.302	0.410	0.046	0.118
Kintinuous	0.234	0.158	0.239	0.026	0.124
ElasticFusion	0.219	0.173	0.208	0.024	0.115
MVO SLAM	0.231	0.103	0.125	0.019	0.112

注: 黑体部分表示性能指标的优势.

2.2 实际建图实验

为进一步验证本算法的有效性, 在真实室内环境下进行建图实验. 实验采用搭载 Xtion2 深度摄像头的小车, 小车实验平台如图 5 所示, 控制小车在实验室中运动.



图 5 小车实验平台
Fig.5 Experiment platform

采用本方法构建的实验室地点云地图如图 6 所示. 从图 6 可见, 本方法对实际环境建图仍然有较好的效果, 构建的三维稠密点云地图与实际环境相比并无太大偏差, 并且环境的边缘拼接较好, 未出现变形、重叠的情况, 对于电脑主机等物件也能较好地构建出来. 在精度方面, 由于没有高精度的运动捕捉装置, 因此无法得到确切的精度, 但是从构建的点云图可见, 本算法完全可以满足后续导航或路径规划等要求.

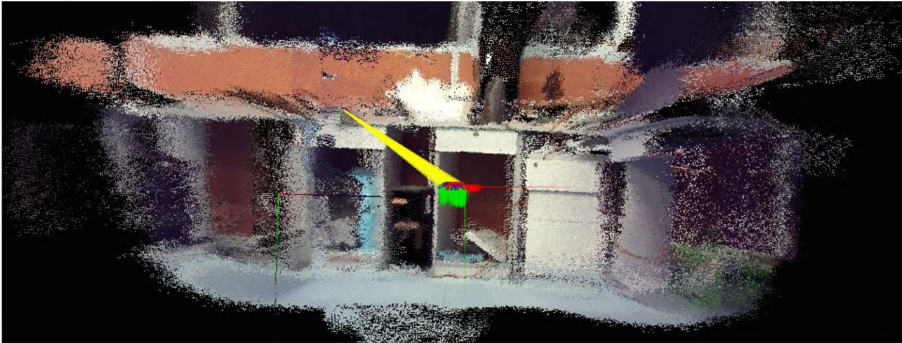


图 6 构建的实验室稠密点云图
Fig.6 Constructed dense point cloud map of laboratory

3 结语

提出一种融合光度和深度的视觉 RGB-D SLAM 方法. 为平衡两误差权重, 加入帧权重系数 λ 构建目

标函数,在视觉里程计部分同时最小化光度误差和深度误差.通过 t 分布模型得到每个像素点的光度和深度误差权重,并采用中值法得到帧权重 λ ,最终得到相机位姿.通过 TUM 标准数据集实验和实际实验验证 MVO SLAM 算法能得到较准确的相机位姿,尤其对纹理较少的环境比经典的 DVO SLAM 有更好的精度.

参考文献:

- [1] DAVIDE S, FRIEDRICH F. Visual odometry[J]. IEEE Robotics & Automation Magazine, 2011, 18(4): 80-92.
- [2] CADENA C, CARLONE L, CARRILLO H, *et al.* Past, present, and future of simultaneous localization and mapping: toward the robust-perception age[J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [3] NEWCOMBE R A, IZADI S, HILLIGES O, *et al.* KinectFusion: real-time dense surface mapping and tracking[C]//IEEE International Symposium on Mixed and Augmented Reality. Basel: IEEE, 2011: 127-136.
- [4] STEINBRÜCKER F, STURM J, CREMERS D. Real-time visual odometry from dense RGB-D images[C]//IEEE International Conference on Computer Vision Workshops. Barcelona: IEEE, 2011: 719-722.
- [5] KERL C, STURM J, CREMERS D. Dense visual SLAM for RGB-D cameras[C]//IEEE International Conference on Intelligent Robots and Systems. Tokyo: IEEE, 2014: 2100-2106.
- [6] GUTIERREZ-GOMEZ D, MAYOL-CUEVAS W, GUERRERO J J. Inverse depth for accurate photometric and geometric error minimisation in RGB-D dense visual odometry[C]//IEEE International Conference on Robotics and Automation. Seattle: IEEE, 2015: 83-89.
- [7] BABU B W, KIM S, YAN Z, *et al.* σ -DVO: sensor noise model meets dense visual odometry[C]//IEEE International Symposium on Mixed and Augmented Reality. Merida: IEEE, 2016: 18-26.
- [8] YAN Z, YE M, REN L. Dense visual SLAM with probabilistic surfel map[J]. IEEE Transactions on Visualization & Computer Graphics, 2017, 23(11): 2389-2398.
- [9] KÜMMERLE R, GRISETTI G, STRASDAT H, *et al.* g2o: a general framework for graph optimization[C]//IEEE International Conference on Robotics and Automation. Shanghai: IEEE, 2011: 3607-3613.
- [10] STURM J, ENGELHARD N, ENDRES F, *et al.* A benchmark for the evaluation of RGB-D SLAM systems[C]//IEEE International Conference on Intelligent Robots and Systems. Vilamoura: IEEE, 2012: 573-580.
- [11] PRESS W H, TEUKOLSKY S A, VETTERLING W T, *et al.* Numerical recipes in C: the art of scientific computing[J]. IEEE Concurrency, 1992, 6(4): 79-79.
- [12] JÖRG S, BEHNKE S. Multi-resolution surfel maps for efficient dense 3D modeling and tracking[J]. Journal of Visual Communication and Image Representation, 2014, 25(1): 137-147.
- [13] ENDRES F, HESS J, STURM J, *et al.* 3-D mapping with an RGB-D camera[J]. IEEE Transactions on Robotics, 2014, 30(1): 177-187.
- [14] WHELAN T, KAESSE M, JOHANSSON H, *et al.* Real-time large-scale dense RGB-D SLAM with volumetric fusion[J]. International Journal of Robotics Research, 2014, 34(4/5): 598-626.
- [15] WHELAN T, LEUTENEGGER S, RENATO F, *et al.* Elasticfusion: dense slam without a pose graph[C]//Proceedings of Robotics: Science and Systems. Rome: MIT Press, 2015. DOI: 10.15607/RSS.2015.XI.001.

(责任编辑:沈芸)