

文章编号 1004-924X(2018)06-1497-10

采用混合回环检测与闭环优化的 TSDF 地图创建

贾松敏, 李柏杨*, 张国梁

(北京工业大学 信息学部, 北京 100124)

摘要: 由于原始 TSDF (Truncated Signed Distance Function, TSDF) 模型仅考虑相邻时间上的关联, 误差将不可避免的累积到下一时刻, 无法构建全局一致的地图。为了实时精确的建立大场景稠密 3D 地图, 对 TSDF 模型进行了改进。首先, 构筑相机位姿模型和加权融合 3D 点截断信息的 TSDF 模型, 用于准确表示创建物体的表面。其次, 提出一种改进的回环检测方法, 并将其与随机蕨类彩色图像编码化相结合, 进而优化 TSDF 模型, 即混合优化位姿模型。最后, 使用 g2o 图优化库解算约束函数, 建立数据集间的优化边。实验结果表明: 混合优化位姿模型能识别曾到达区域, 特别在较大场景下使用可以得到更加准确的相机轨迹和地图。采用 TUM 数据集中的 fr1/xyz、fr1/room、fr1/desk 对所提算法进行检验, 结果表明该方法能够使相机轨迹的均方根误差分别下降 0.59 cm, 3.14 cm, 0.94 cm。在室内环境和公开数据集上的实验结果证明了所提算法的有效性和准确性。

关键词: 随机蕨类; TSDF; 闭环优化; 图优化

中图分类号: TP242.6 **文献标识码:** A **doi:** 10.3788/OPE.20182606.1497

TSDF map building based on hybrid loopback detection and closed-loop optimization

JIA Song-min, LI Bo-yang*, ZHANG Guo-liang

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

* Corresponding author, E-mail: jsm@bjut.edu.cn

Abstract: Because the original Truncated Signed Distance Function (TSDF) model only considers the correlation of adjacent time, the error will inevitably accumulate to the next time, and it is impossible to build a globally consistent map. In order to establish the dense 3D map of large scenes in real time and accurately, the TSDF model was improved in this paper. First, a camera pose model and weighted fusion 3D point truncation information model were established to represent the surface of the created object. Secondly, an improved loop detection method was proposed, which was combined with the random fern color image coding, and then optimized the TSDF model, that is, the mixed optimization pose model. Finally, the g2o diagram was used to solve the constraint function and established the optimized edges between the datasets. The experimental results showed that the loopback detection method proposed in this paper could obtain more accurate camera trajectory and map in a larger scene. The proposed algorithm was verified by fr1/xyz, fr1/room and fr1/desk in the TUM dataset. The re-

收稿日期: 2017-10-19; 修订日期: 2017-12-10.

基金项目: 国家自然科学基金资助项目 (No. 61175087); 北京工业大学 2017 智能制造领域大科研推进计划资助项目 (No. 040000546317552)

sults show that the root mean square error of the camera trajectory decreased by 0.59 cm, 3.14 cm, 0.94 cm, respectively. The validity and accuracy of the proposed algorithm are verified in the laboratory environment and the open data set.

Key words: random ferns; Truncated Signed Distance Function (TSDF); closed-loop optimization; pose graph optimization

1 引言

SLAM (Simultaneous Localization and Mapping, SLAM) 是机器人实时定位和外部环境地图创建同步完成的系统算法^[1-2], 是机器人实现自主化移动的根本性任务。静态小范围室内环境下, 认为使用激光或声纳创建 2D 地图是已经能够很好解决的 SLAM 问题。但是在动态、复杂的大场景下, 视觉 SLAM 被认为是更好的解决方案^[3]。当视觉作为 SLAM 系统唯一感知外界的传感器时, 该系统被称为 V-SLAM (Vision-only SLAM, V-SLAM)^[4]。

Kinect 拥有使用三角测量原理的深度摄像机和彩色摄像机, 能够同时捕捉深度信息和彩色纹理信息, 具有精度较高, 价格低廉, 数据采集速度快等特点, 被广泛应用于 V-SLAM 系统中。DVO (Dense Visual Odometry, DVO) 作为一种先进的视觉里程计, 使用 Kinect 作为视觉传感器, 并且采用相邻关键帧进行帧与帧之间的匹配, 提高了机器人地图创建的鲁棒性和精度^[5]。Endres 等^[6-7]提出的 RGB-D SLAM 在视觉里程计的基础上使用 g2o 图优化^[8]而非传统的滤波器^[9], 能够消除累积误差。KinectFusion 系统^[10-11]中 TSDF (Truncated Signed Distance Function, TSDF) 模型使用 GPU (Graphics Processing Unit, GPU) 进行点云融合, 所建模型冗余噪点少, 更符合实际外部环境情况。Keisuke 等^[12]使用 CNN-SLAM 预测外部环境深度值, 进而从一定程度上克服了单目相机的缺陷。Yan Z 等^[13]使用 surfel 模型进行图像融合, 通过 GPU 加速, 可以在一定的时间上得出光滑准确的重建模型。刘宗明等^[14]使用词袋建立闭环检测策略, 对重访目标减少累积误差。

上述方法中, CNN-SLAM 使用单目相机进行地图创建, 实时性和准确性均不如本文所使用的 Kinect 传感器, 且使用 CNN 进行绝对距离估计需要很长的建模时间, 应用场景有局限性。DVO 只采用相邻关键帧进行帧与帧之间的匹配, 不能消除累积误差和识别曾到达区域。RGB-D SLAM 采用 g2o 进行回环检测, 虽然能够消除累积误差但因为没有使用 GPU 进行加速运算, 故而运行速度较低。

由于原始 TSDF 模型仅考虑相邻时间上的关联, 误差将不可避免的累积到下一时刻, 无法构建全局一致的地图。为了实时精确地建立大场景稠密 3D 地图, 对 TSDF 模型进行了改进, 提出一种混合位姿模型优化的 TSDF 系统。首先, 构筑相机位姿模型和加权融合 3D 点截断信息的 TSDF 模型, 用于准确表示创建物体的表面。其次, 提出一种基于 g2o 的模型与模型配准的匹配方法, 使用 GPU 进行并行运算加速, 最后将其与随机蕨类^[15]彩色图像编码化策略相结合, 进而优化 TSDF 模型。实验结果表明, 混合优化位姿模型可以较快建立全局 SLAM 地图, 并有效识别曾到达区域, 消除点云模型的重影与不匹配现象, 进而创建用户体验更好, 鲁棒性更强, 精度更高的 V-SLAM 系统。

2 总体设计

地图创建系统主要由 Kinect 传感器对实验环境进行数据采集, 得到深度图和彩色图。对深度图进行优化并由采集得到的数据信息, 计算旋转平移矩阵进而创建相机位姿模型。使用 g2o 建立并优化局部与全局的回环模型, 利用随机蕨类进行闭环匹配。建立 TSDF 模型表示三维物体表面并进行多帧融合。总体设计图如图 1 所示。

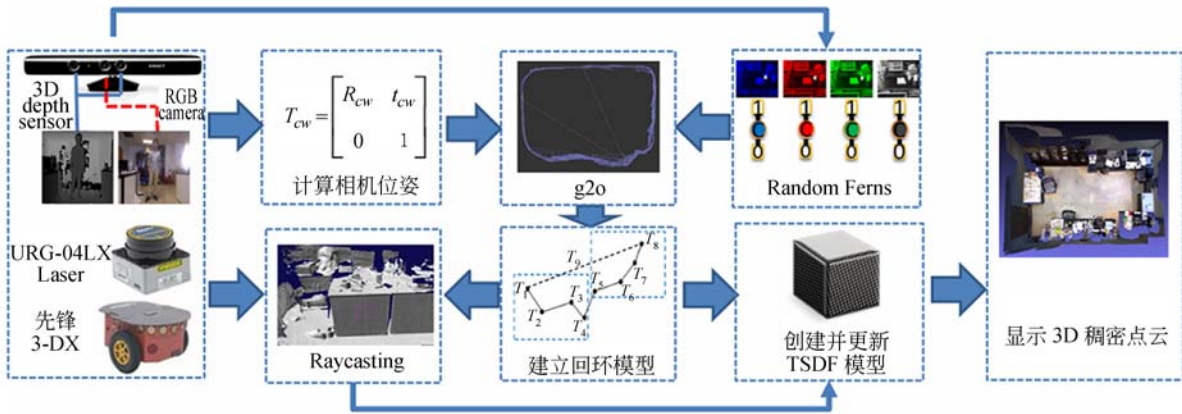


图 1 总体设计图

Fig. 1 Diagram of system design

3 相机位姿模型与 TSDF 模型的建立

在世界坐标系下,相机位姿进行欧氏变换,它的变换矩阵 T 可表示为:

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

其中 \mathbf{R}, \mathbf{t} 为世界坐标系下的旋转平移矩阵。相机坐标系下点 p_c 与世界坐标系下点 p_w 的坐标转换关系为:

$$p_c = \mathbf{K}p_w = \frac{1}{Z_w} \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \left(\frac{f_x X_w + c_x Z_w}{Z_w}, \frac{f_y Y_w + c_y Z_w}{Z_w}, 1 \right)^T, \quad (2)$$

其中: \mathbf{K} 为相机内参数矩阵, f_x, f_y 为相机焦距, c_x, c_y 为相机中心坐标 X_w, Y_w, Z_w 为世界坐标。图片内坐标使用 TSDF 模型融合深度数据,消除深度数据中噪声和孔洞的影响。

三维空间的深度数据被 TSDF 模型划分成三维像素,每个三维像素中存储了相应的数值,其表现形式如图 2 所示。

在 TSDF 模型的每个三维像素 v_i 中存储着数据结构:

$$\Phi_i(v_i) = (f_i(v_i), \omega_i(v_i)), \quad (3)$$

式中: $f_i(v_i)$ 为当前三维像素 v_i 到场景表面的归一化截断距离,权重 $\omega_i(v_i)$ 越大表示整合次数越多, $f_i(v_i)$ 的置信度就越大。如上图表示重建人脸表面(网格模型中值为 0 部分连成的线条表示

-0.9	-0.6	-0.4	0.2	0.8	1	1	1	1	1
-1	-0.9	-0.6	-0.1	0.5	0.9	1	1	1	1
-1	-0.9	-0.6	-0.2	0.2	0.8	1	1	1	1
-1	-0.9	-0.5	-0.1	0.2	0.8	0.9	1	1	1
-1	-1	-0.8	-0.3	0.1	0.6	0.8	1	1	1
-1	-0.9	-0.5	-0.2	0.3	0.7	0.7	1	1	1
-1	-0.8	-0.4	-0.1	0.3	0.8	0.9	1	1	1
-0.9	-0.7	-0.4	0.0	0.4	0.9	0.8	1	1	1
-0.1	-0.1	0.1	0.1	0.4	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1

图 2 TSDF 模型 2D 示意图

Fig. 2 Two-dimensional Illustration of TSDF

重建的表面),重建表面到相机一侧是正值,另一侧是负值,网格点离重建好的表面距离越远绝对值越大,在网格模型中从正到负的穿越点表示重建好的场景的表面。本文使用 GPU 更新网格中的值,每个线程更新一串网格值。其更新策略如算法 1 所示。

表 1 TSDF 更新策略

Tab. 1 Updating strategy of TSDF

算法 1	
if v_i 在相机视野下 then	
$d_i \leftarrow \ v_i - t_i\ - D_i(u_i, v_i)$	
if ($d_i > 0$) then	
$f_i \leftarrow \min(1, d_i / \text{MaxTruncation})$	
else	
$f_i \leftarrow \max(-1, d_i / \text{MinTruncation})$	
$\omega_i \leftarrow \min(\text{MaxWeight}, \omega_{i-1} + 1)$	
$f_i^{\text{ave}} \leftarrow \frac{f_{i-1} \omega_{i-1} + f_i \omega_i}{\omega_{i-1} + \omega_i}$	
储存 ω_i 和 d_i^{ave} 于 TSDF 模型中	

算法 1 中 v_i 为更新的全局坐标下的三维像素点, t_i 表示相机光心全局的坐标, $D_i(u_i, v_i)$ 表示对应像素位置的深度值, f_i 为第 i 次得到当前三维像素 v_i 到场景表面的归一化截断距离, w_i 为该三维像素的整合次数。 d_i^{avg} 表示未归一化前的截断距离。 MaxTruncation, MinTruncation 分别为最大最小截断距离阈值, MaxWeight 最大整合次数阈值。

4 创建回环模型与闭环检测

4.1 使用 g2o 构造代价函数

对于经典 SLAM 模型, 它由一个状态方程和运动方程组成, 形式如下:

$$\begin{cases} \mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_k) + \mathbf{w}_k \\ \mathbf{z}_{k,j} = h(\mathbf{y}_j, \mathbf{x}_k) + \mathbf{v}_{k,j} \end{cases}, \quad (4)$$

其中: $\mathbf{x}_k, \mathbf{y}_j$ 为机器人或相机位姿, \mathbf{u}_k 为机器人运动量, $\mathbf{z}_{k,j}$ 为相机观测量, $\mathbf{w}_k, \mathbf{v}_{k,j}$ 为噪音项。传统滤波器, 即 EKF(Extended Kalman Filter, EKF) 虽然对上述两个非线性方程进行一阶泰勒展开^[16], 但 EKF 有很大的局限性, 在一定程度上假设了马尔科夫性, 只考虑了上一状态和当前状态的关系, 每一个状态之间的估计必定有或大或小的偏差, 当机器人运行轨迹很长时则会出现较大偏差。本文使用 g2o 对机器人状态进行图优化, 不仅使用机器人的前一状态数据, 更会把历史数据加入状态估计, 在同等计算量的情况下, 在同等计算量的情况下, 非线性优化会取得更好的效果。若只考虑 Kinect 相机得到的观测数据, 使用 Lie 代数对相机位姿和路标点进行转换, 以便于进行无约束优化。观测误差 e 为:

$$e = \mathbf{z} - h(\boldsymbol{\xi}, \mathbf{p}), \quad (5)$$

其中 \mathbf{z} 为相机观测数据, $\boldsymbol{\xi}, \mathbf{p}$ 为李代数表示的相机参数与位姿, 函数 h 为观测方程。其中 m, n 为对应矩阵的行坐标与列坐标, i, j 分别为由 1 到 m 和 n 的数值, $e_{ij}, \mathbf{z}_{ij}, \mathbf{p}_{ij}$ 为观测误差, 相机观测数据与位姿矩阵中对应坐标的变量。把关键状态的观测量加入, 得到代价函数为:

$$\sum_{i=1}^m \sum_{j=1}^n \|e_{ij}\|^2 = \sum_{i=1}^m \sum_{j=1}^n \|\mathbf{z}_{ij} - h(\boldsymbol{\xi}_i, \mathbf{p}_j)\|^2. \quad (6)$$

设代价函数为 $s_e(x_e)$, \mathbf{x}_e 为 $[\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_m, \mathbf{p}_1, \dots, \mathbf{p}_n]^T$ 并进行泰勒展开:

$$\begin{aligned} \|s_e(\mathbf{x}_e + \Delta \mathbf{x}_e)\|^2 &\approx \\ \|s_e(\mathbf{x}_e)\|^2 + J(\mathbf{x}_e) \Delta \mathbf{x}_e + \frac{1}{2} \Delta \mathbf{x}_e^T H \Delta \mathbf{x}_e, \end{aligned} \quad (7)$$

其中 $\Delta \mathbf{x}_e$ 的一阶、二阶导数分别为雅可比矩阵和海塞矩阵。构建迭代函数使用 g2o 库优化求得误差最小的观测状态。

4.2 进行关键帧匹配与构建模型

4.2.1 基于 ICP 匹配的投影算法

使用 g2o 进行回环检测并优化时, 会对邻近的关键帧和历史数据集中的关键帧进行匹配并优化校正全局地图。其中关键帧之间的匹配方式为 ICP^[17] 匹配, 匹配过程如图 3 所示。

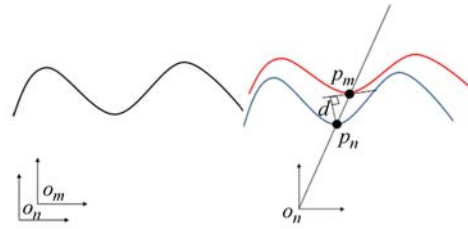


图 3 ICP 匹配

Fig. 3 Matching of ICP

其中 o_n 和 o_m 为当前帧历史数据集中第 n 和 m 帧的相机坐标, 黑色曲线为环境中的物体。 P_1 和 P_2 点为两个相机位姿下投影在 2D 平面上的对应点, 通过对对应点的欧氏距离和法向量夹角优化相机位姿的变换矩阵。 ICP 匹配中每个点的投影和匹配操作均在 GPU 中完成, GPU 分出模块中的每一个线程分别计算一对匹配点。本文把 GPU 分成 grid. x 乘以 grid. y 个 block, 每个 block 中有 block. x 乘以 block. y 个线程, 每个线程的计算独立。对于本文所用的 GPU, 每个 block 有 512 个线程。本文使用 Lie 代数表示变换矩阵, 将该问题转变为无约束优化问题, 公式如下:

$$E(\boldsymbol{\xi}_{g,k}) = \sum_{\Omega(p) \neq \emptyset} \|(\boldsymbol{\xi}_{g,k} V_k(p - V_{k-1}^g(p)))^T \mathbf{N}_{k-1}^g(p)\|_2, \quad (8)$$

其中 $\boldsymbol{\xi}_{g,k}$ 为李代数表示的变换矩阵, $\Omega(p) \neq \emptyset$ 表示两个关键帧之间存在对应点。 $V_k(p)$ 表示第 k 帧的顶点坐标, $V_{k-1}^g(p)$ 表示顶点坐标的预测值, $\mathbf{N}_{k-1}^g(p)$ 为对应点的法向量。使用 ICP 进行投影算法匹配得到图 4。

4.2.2 构造并优化回环模型

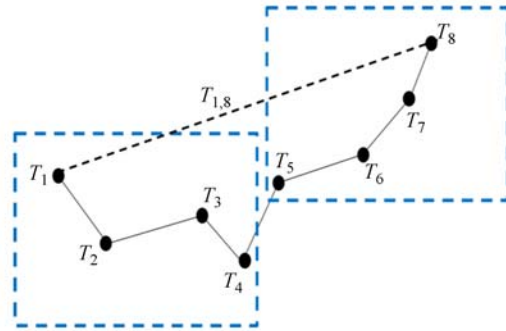
g2o 构造回环对关键帧的选取策略是随机



图 4 投影算法效果图

Fig. 4 Effect of projection algorithm

的、无序的。如文献[7]使用的 g2o 构造回环优化方法建立帧与帧之间的数据约束时会发生如图 5(a)所示帧与帧之间约束过密,造成运算量大,降低系统实时性;或者如图 5(b)所示只有邻近帧之间建立约束,造成旋转平移矩阵校正效果不佳,出现累积误差。故而建立如图 5(c)所示模型,在多次实验测试后,选定 4~7 个帧为一个数据模型,建立模型与模型之间的约束关系,这样既减少了运算量又增加了校正精度。并在模型与模型匹配时加入随机的局部回环和大回环配准,保证全局 SLAM 的建立。



(c)模型与模型之间建立约束

(c) Establishes a constraint between two models

图 5 优化回环模型

Fig. 5 Optimized loopback model

模型与模型之间的匹配迭代算法如表 2 所示,其中 λ 为拉格朗日乘子,将算法转化为无约束优化问题, μ 为 Δx 增量半径, \mathbf{D} 为非负数对角阵,可将增量区域变更为更加合理的椭圆形区域,MaxTH ρ 为变量 ρ 近似可行的阈值,本方法中设定为 0.9。在迭代过程中,如果 $\rho > 3/4$,那么 $\mu \leftarrow 2\mu$; 如果 $\rho < 1/4$,那么 $\mu \leftarrow 0.5\mu$; 如果 $\rho > \text{MaxTH}\rho$,那么 $x_{k+1} \leftarrow x_k + \Delta x_k$; 不断迭代直到 $f(x_k)$ 收敛时, x_k 的最优值。 $f(x)$ 为模型与模型间的匹配函数, $J(x)$ 为 $f(x)$ 的雅可比矩阵。

表 2 模型间的迭代算法

Tab. 2 Iterative algorithm between models

算法 2

$x_0 \leftarrow [\xi_1, \dots, \xi_m, p_1, \dots, p_n]^T, \mu \leftarrow \mu_0$

while 算法不收敛 **do**

$$\min_{\Delta x_k} \frac{1}{2} \| f(x_k) + J(x_k) \Delta x_k \|^2 + \frac{\lambda}{2} \| \mathbf{D} x_k \|^2$$

$$\rho \leftarrow \frac{f(x_k + \Delta x_k) - f(x_k)}{J(x_k) \Delta x_k}$$

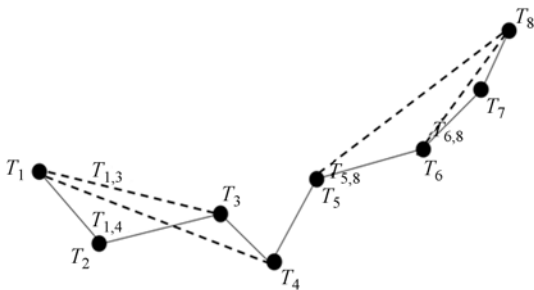
if $\rho > 3/4$ **then** $\mu \leftarrow 2\mu$

if $\rho < 1/4$ **then** $\mu \leftarrow 0.5\mu$

if $\rho > \text{MaxTH}\rho$ **then**

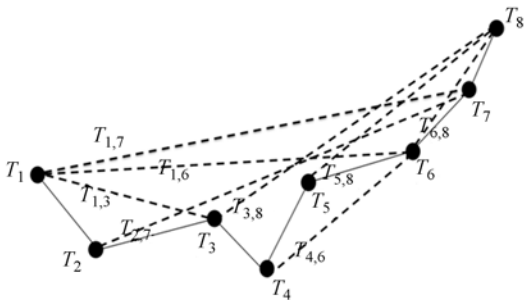
$$x_{k+1} \leftarrow x_k + \Delta x_k$$

得到 $f(x)$ 收敛时, x 的最优值



(a)只有邻近帧之间建立约束

(a) Matching only for close frames



(b)帧与帧之间约束过密

(b) Constraint between frames is too dense

4.2.3 使用随机簇类实现闭环匹配

只使用 g2o 构造如上模型对模型的回环优化时,并不能有效地识别曾到达区域,完成最终的全局闭环。在反复建立某一局部区域或者回到曾经到达区域时,会出现如图 6 所示的现象,出现重投

影模型不匹配导致的重叠和错位的现象。



(a) 3D 模型不匹配 (b) 3D 彩色渲染图
(a) Not matching between modles (b) Three-dimensional color rendering

图 6 重叠及错位现象

Fig. 6 Appearance of overlap and dislocation

为了解决上述问题,在回环优化模型中使用随机蕨类组成混合回环优化模型。它使用如公式(9)、(10)所示的方式,编码化关键帧。每幅图像 P 的编码由 m 个 block 组成,每个 block 由 n 个 Ferns 组成,其中 n 个 ferns 组成一个 block, m 个 block 表达一个关键帧。每个 ferns 由每个通道 x 处的像素值 θ 确定。

$$P = \{block_k\}_{k=1}^m \rightarrow block = \{Ferns_i\}_{i=1}^n, \quad (9)$$

$$f(I, \theta, \tau) = \begin{cases} 1 & \text{if } I(\theta) \geq \tau \\ 0 & \text{if } I(\theta) < \tau \end{cases} \quad \theta = \{c, x\}. \quad (10)$$

应用公式(11)如图 7 所示,对关键帧进行编码化处理。对于一个新获取的关键帧 P,分别比较归一化的 $I(\theta)$ 与阈值 τ 的大小,来计算每一个 block 里的 Fern 列表中的值。其中列表的表现形式如表 3 所示:

表 3 图片编码结果

Tab. 3 Result of encoding of the picture

图像编码	Ferns 编号
...	...
{1001}	(145)
{0010}	(146)
{1110}	(147)
{0011}	(148)
...	...

当新获取一个关键帧 P 时,对 P 进行同样的编码化处理,并比对临近关键帧和历史关键帧。通过对每一个 Fern 的比对,计算两帧之间的相似度。即对比图 7 的 x 处的值,若相同则相似度加

1。通过相似度判别,决定当前关键帧是否加入回环(与邻近关键帧相似度大于一定阈值),或者存在回环(和历史关键帧相似度大于一定阈值),并通过和相似度大的图像配准,进行重定位。

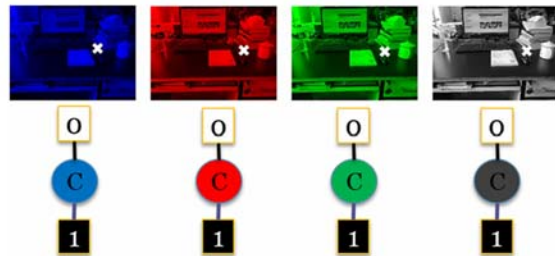


图 7 图像编码示意图

Fig. 7 Illustration of image coding

5 实验结果

本文使用的硬件平台主要包含三个部分。该系统如图所示。机器人是 Pioneer3-DX,由于其多功能性和灵活性,被用户广泛认可。RGB-D 相机是 Kinect,是由 Microsoft 于 2010 年 6 月发布的体感传感器。在实验中,笔记本电脑配置运行 64 位 Ubuntu 16.04 系统,配备 Intel Core i7 2.9 GHz CPU,8GB DDR4 2400 和 NVIDIA GeForce GTX 960 4 GB GPU。

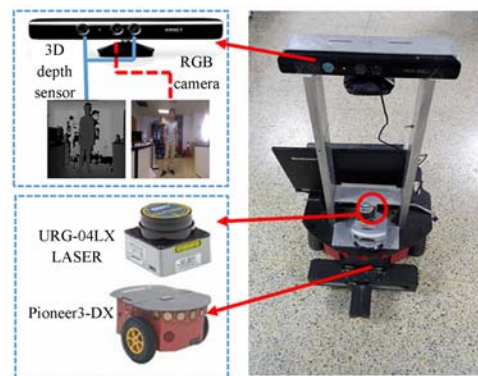


图 8 硬件平台

Fig. 8 Hardware platform

5.1 验证混合回环模型的有效性

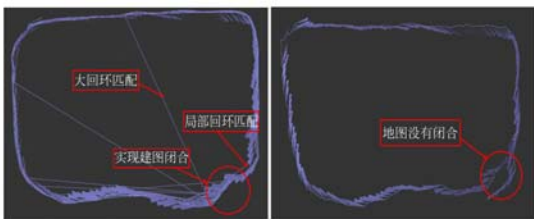
在如图 8 所示的实验室环境下,使用移动机器人携 Kinect 传感器并将机器人速度设置为 0.15 m/s,角速度为 0.15 rad/s,图像频率为 30 Hz,机器人在实验室中运行并建立了三维的地图。



图 9 实验环境

Fig. 9 Experimental environment

在实验中,比较了使用基于 g2o 的混合位姿模型与不使用它的系统效果差异。图 10(a)和图 10(b)是由 g2o_viewer 建立的机器人姿态图。从图 10 可以看出,使用混合位姿模型可以增加当前机器人位姿与历史机器人位姿之间的联系,增加局部回环和大回环。通过这种方式可以有效消除系统的累积误差,充分利用相机得到的数据信息,校正机器人姿态,提高系统的鲁棒性。

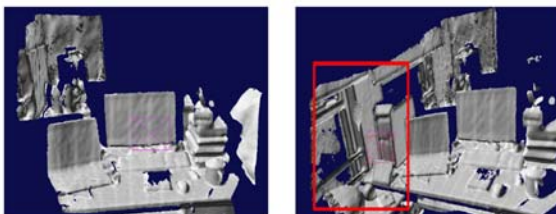


(a)使用回环模型配准 (b)没有回环模型配准
 (a) Using loopback model (b) Without loopback model

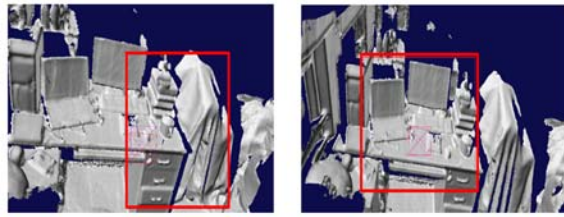
图 10 回环模型配准试验

Fig. 10 Experiment of loopback model registration

实验使用手持 Kinect 摄像头运行系统,加入随机蕨类混合回环检测模型后,验证局部回环的可行性,并且有效识别了重复到达区域。如图 11(a)所示对实验桌进行建图,然后向左旋转摄像头,增加建图区域如图 11(b),然后向右旋转移摄像头如图 11(c)增加建图区域。最后如图四向后拉伸摄像头,可见反复创建的图 11(d)红色区域并没有出现明显的重影和误差,表明系统可以



(a)对实验桌建图 (b)向左增加建图区域
 (a) To build a map of table (b) Add mapping area to left



(c)向右增加建图区域 (d)回到曾到达区域

(c) Add mapping area to right (d) Back to arrival area

图 11 随机蕨类识别曾到达区域

Fig. 11 Random ferns identify the area where is arrived

识别到达过该区域,并且和右方区域建立了较好的拼接(彩图见期刊电子版)。

如图 12 扩大实验区域,手持 Kinect 旋转一周对 1.8 m×1.8 m 的实验区域,以 0.15 rad/s 的角速度进行地图创建。该模型可以准确识别曾到达区域并实现闭环建图。

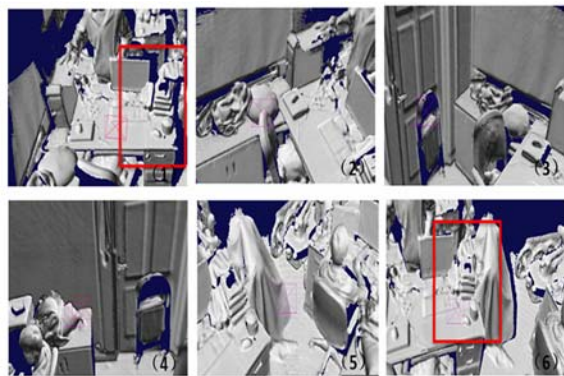


图 12 旋转 Kinect 一周验证随机蕨类效果

Fig. 12 Rotate Kinect a circle to verify effects of random ferns

最终的 3D 效果图如图 13 所示。

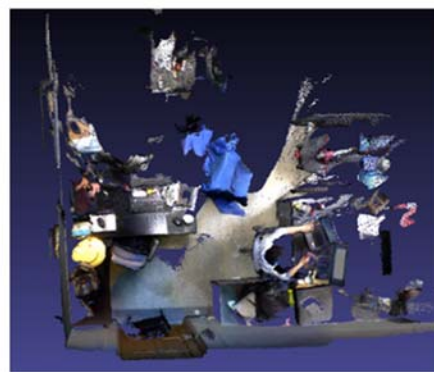
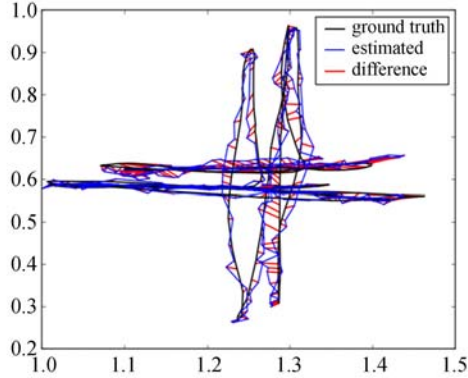


图 13 旋转一周后的彩色渲染图

Fig. 13 Color rendering after rotate a circle

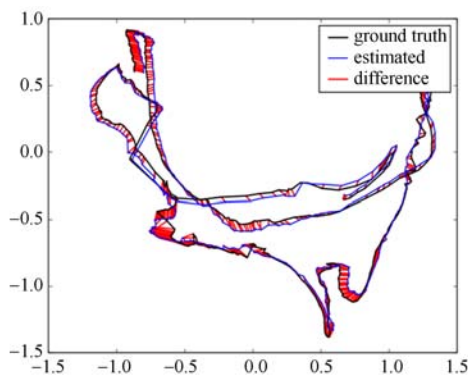
5.2 利用 TUM 数据集测试相机轨迹准确性

TUM 数据集^[18]中的 RGB-D 数据包主要包含 Kinect 采集的 RGB 图片、深度图片和使用高精度物理运动捕捉系统采集的相机真值轨迹。其中相机真值轨迹包含表示相机位姿的四元数和空间坐标。



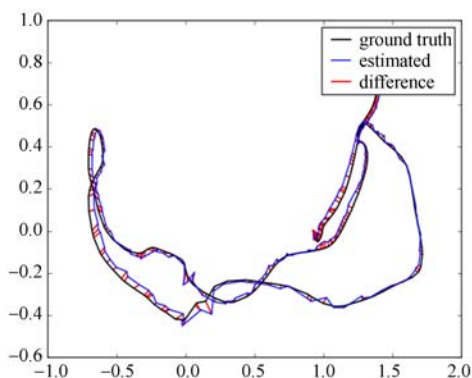
(a) fr1/xyz 轨迹图

(a) Track diagram of fr1/xyz



(b) fr1/room 轨迹图

(b) Track diagram of fr1/room



(c) fr1/desk 轨迹图

(c) Track diagram of fr1/desk

图 14 RGB-D SLAM 数据集轨迹图

Fig. 14 Track diagram of RGB-D SLAM datasets

为了验证本文提出的混合位姿模型算法的先进有效性,实验使用 fr1/xyz、fr1/room、fr1/desk3 个数据集进行实验,其轨迹图如图所示。

本文提出的混合位姿模型所创建的相机位姿轨迹与数据集所提供的真值相比较,由图 14 的三张图可以看出,创建的轨迹误差较小,没有表现出累积误差并且每一个区域都没有出现较大的误差,验证了该模型的准确性和鲁棒性。

在硬件平台上,对 r1/xyz、fr1/room、fr1/desk3 个数据集分别用 DVO、RGB-D SLAM、KinectFusion3 个开源算法进行对比试验,实验结果如表 4 所示。均方跟误差能够较好的反映测量值与真值之间的差别^[18]。本文提出的混合位姿模型在 fr1/xyz、fr1/room、fr1/desk3 个数据集上对比与同样使用 TSDF 模型进行地图创建的 KinectFusion 系统,均方根误差分别减少 0.59 cm、3.14 cm、0.94 cm。特别在代表较大环境的 fr1/room 数据集下,均方根误差减少了 31.2%。

表 4 与开源算法对比

Tab. 4 Comparison of open source algorithm (m)

米	数据集	均方根	平均值	标准差	最大值
混合优化位姿模型	fr1/xyz	0.018 4	0.017 8	0.008 1	0.059 3
	fr1/room	0.069 1	0.077 2	0.025 1	0.361 2
	fr1/desk	0.031 8	0.023 1	0.011 4	0.120 8
DVO	fr1/xyz	0.029 7	0.017 2	0.011 2	0.109 2
	fr1/room	0.094 5	0.091 2	0.041 2	0.410 2
RGB-D SLAM	fr1/desk	0.031 4	0.031 2	0.017 8	0.142 8
	fr1/xyz	0.019 2	0.016 8	0.009 3	0.061 7
Kinect Fusion	fr1/room	0.081 0	0.073 6	0.033 7	0.350 9
	fr1/desk	0.029 1	0.026 1	0.012 9	0.139 8
Kinect Fusion	fr1/xyz	0.024 3	0.027 5	0.013 5	0.069 3
	fr1/room	0.100 5	0.082 5	0.058 2	0.601 8
	fr1/desk	0.041 2	0.038 1	0.021 75	0.210 9

为了验证混合位姿模型在实际在线场景中的表现,使用本章介绍的硬件平台在 8 m×4 m 的实验室中进行地图创建。从图中可以明显看出,由于算法使用了模型对模型的优化回环匹配与基于随机森林的闭环检测,创建的地图几乎没有出现重影和误差,验证了该算法可以创建出准确的较大场景三维地图。

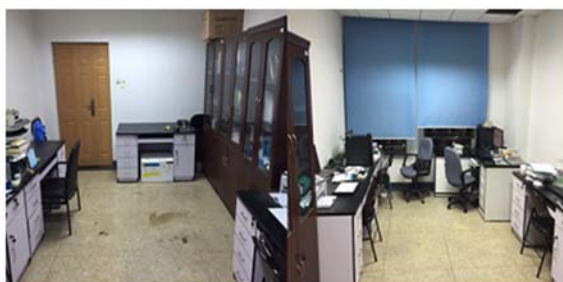


图 15 实验室场景

Fig. 15 Laboratory scene



图 16 3D 稠密地图

Fig. 16 3D dense map

6 结 论

本文提出一种采用混合位姿模型优化的 TSDF V-SLAM 系统。使用 5.1 部分介绍的硬件平

台对外界环境进行数据采集。通过 g2o 构造迭代函数求得误差最小的观测状态,使用 GPU 并行处理 ICP 匹配投影算法,构造模型对模型的匹配策略。关键帧图像使用随机蕨类进行编码化处理。最后构造回环检测模型实现全局 SLAM 匹配并在 GPU 上实时使用随机蕨类对采集图像进行检测,实现闭环检测,识别曾到达区域。

对比本文提出的算法与经典 SLAM 算法,混合位姿模型可以增加当前机器人位姿与历史机器人位姿之间的联系,增加局部回环和大回环。通过这种方式可以有效消除系统的累积误差,充分利用相机得到的数据信息,校正机器人姿态。对于关键帧图像使用随机蕨类进行编码化处理,进行相似度打分,比较当前帧与历史帧之间的相似程度,可以识别曾到达区域,消除点云模型的重影与不匹配现象,进而创建精度更高的 V-SLAM 系统,使用 GPU 并行计算进一步优化用户体验,提高系统鲁棒性。故而本文算法在 TUM 数据集中的 fr1/xyz、fr1/room、fr1/desk 上均取得较好的效果。对比同样使用 TSDF 模型的 KinectFusion,3 个数据集的均方根误差分别减少 0.59 cm、3.14 cm、0.94 cm。特别在较大环境的 fr1/room 数据集下,均方根误差减少了 31.2%,由此可见本文提出的混合位姿模型优化的 TSDF 系统在准确度和鲁棒性上均取得更好的效果。

参考文献:

- [1] FUENTES-PACHECO J, RUIZ-ASCENCIO J, RENDÓN-MANCHA J M. Visual simultaneous localization and mapping: a survey[J]. *Artificial Intelligence Review*, 2015, 43(1):55-81.
- [2] 曹天扬,蔡浩原,方东明,等. 结合图像内容匹配的机器人视觉导航定位与全局地图构建系统[J]. *光学精密工程*, 2017, 25(8): 2221-2232.
CAO T Y, CAI H Y, FANG D M, *et al.*. Robot vision system for keyframe global map establishment and robot localization based on graphic content matching[J]. *Opt. Precision Eng.*, 2017, 25(8): 2221-2232. (in Chinese)
- [3] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: A Versatile and Accurate Monocular SLAM System[J]. *IEEE Transactions on Robotics*, 2017, 31(5):1147-1163.
- [4] GAO X, ZHANG T. Unsupervised learning to detect loops using deep neural networks for visual SLAM system[J]. *Autonomous Robots*, 2017, 41(1):1-18.
- [5] KERL C, STURM J, CREMERS D. Robust odometry estimation for RGB-D cameras[C]// *IEEE International Conference on Robotics and Automation*. IEEE, 2013:3748-3754.
- [6] ENDRES F, HESS J, ENGELHARD N, *et al.*. An evaluation of the RGB-D SLAM system[C]// *IEEE International Conference on Robotics and Automation*. IEEE, 2012:1691-1696.
- [7] ENDRES F, HESS J, STURM J, *et al.*. 3-D Mapping With an RGB-D Camera[J]. *IEEE Transactions on Robotics*, 2017, 30(1):177-187.

- [8] KÜMMERLE R, GRISETTI G, STRASDAT H, *et al.*. G2o: A general framework for graph optimization[C]// *IEEE International Conference on Robotics and Automation*. IEEE, 2011:3607-3613.
- [9] 杜航原, 郝燕玲, 赵玉新, 等. 用概率假设密度滤波实现同步定位与地图创建[J]. *光学精密工程*, 2011, 19(12):3064-3073.
- DU H Y, HAO Y L, ZHAO Y X, *et al.*. Implementation of SLAM by probability hypothesis density filter[J]. *Opt. Precision Eng.*, 2011, 19(12): 3064-3073. (in Chinese)
- [10] 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*. IEEE, 2012:127-136.
- [11] ANDUJAR D, DORADO J, RIBEIRO A. An Approach to the Use of Depth Cameras for Weed Volume Estimation[J]. *Sensors*, 2016, 16(7).
- [12] TATENO K, TOMBARI F, LAINA I, *et al.*. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction[C]// *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, Hawaii, USA, June, 2017.
- [13] YAN Z, YE M, REN L. Dense Visual SLAM with Probabilistic Surfel Map. [J]. *IEEE Transactions on Visualization & Computer Graphics*, 2017, PP(99):1-1.
- [14] 刘宗明, 张宇, 卢山, 等. 非合作旋转目标闭环检测与姿态优化[J]. *光学精密工程*, 2017, 25(4): 1036-1043.
- LIU Z M, ZHANG Y, LU SH, *et al.*. Closed-loop detection and pose optimization of non-cooperation rotating targets[J]. *Opt. Precision Eng.*, 2017, 25(4): 1036-1043. (in Chinese)
- [15] GLOCKER B, SHOTTON J, CRIMINISI A, *et al.*. Real-Time RGB-D Camera Relocalization via Randomized Ferns for Keyframe Encoding [J]. *IEEE Trans Vis Comput Graph*, 2015, 21(5): 571-583.
- [16] 张百强, 储海荣, 孙婷婷, 等. 应用 RB 无迹卡尔曼滤波组合导航提高 GPS 重获信号后的导航精度[J]. *光学精密工程*, 2016, 24(4):835-843.
- ZHANG B Q, CHU H R, SUN T T, *et al.*. Precision improvement methodology for INS/GPS after GPS outage using RB-UKF[J]. *Opt. Precision Eng.*, 2016, 24(4): 835-843. (in Chinese)
- [17] 祝继华, 郑南宁, 袁泽剑, 等. 基于 ICP 算法和粒子滤波的未知环境地图创建[J]. *自动化学报*, 2009, 35(8):1107-1113.
- ZHU J H, ZHENG N N, YUAN Z J, *et al.*. A SLAM Approach by Combining ICP Algorithm and Particle Filter[J]. *Acta Automatica Sinica*, 2009, 35(8):1107-1113. (in Chinese)
- [18] STURM J, ENGELHARD N, ENDRES F, *et al.*. A benchmark for the evaluation of RGB-D SLAM systems [C]// *Ieee/rsj International Conference on Intelligent Robots and Systems*. IEEE, 2012:573-580.

作者简介:



贾松敏(1964—),女,北京人,博士,教授,2002年于日本国立电气通信大学获得博士学位,主要从事智能服务机器人,计算机视觉等方面的研究。E-mail: jsm@bjut.edu.cn

通讯作者:



李柏杨(1992—),男,内蒙古包头人,硕士研究生,2015年于北京工业大学获得学士学位,主要从事机器人实时定位与导航,计算机视觉等方面的研究。E-mail: liboyang92@163.com