



计算机应用
Journal of Computer Applications
ISSN 1001-9081, CN 51-1307/TP

《计算机应用》网络首发论文

题目: 基于 RGB-D 图像的室内机器人同时定位与地图构建
作者: 赵宏, 刘向东, 杨永娟
收稿日期: 2020-04-23
网络首发日期: 2020-08-24
引用格式: 赵宏, 刘向东, 杨永娟. 基于 RGB-D 图像的室内机器人同时定位与地图构建. 计算机应用. <https://kns.cnki.net/kcms/detail/51.1307.TP.20200824.1204.006.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

基于 RGB-D 图像的室内机器人同时定位与地图构建

赵宏, 刘向东*, 杨永娟

(兰州理工大学 计算机与通信学院, 甘肃 兰州 730050)

(*通信作者电子邮箱 Liuxd1994@foxmail.com)

摘要：同时定位与地图构建 SLAM(Simultaneous Localization and Mapping)是机器人在未知环境实现自主导航的关键技术, 针对目前常用的 RGB-D SLAM(RGB-Depth SLAM)系统实时性差和精确度低的问题。提出一种 RGB-D SLAM 系统, 进一步提升实时性和精确度。首先, 采用 ORB(Oriented FAST and Rotated BRIEF)算法检测图像特征点, 并对提取的特征点采用基于四叉树的均匀化策略进行处理, 结合词袋模型 BoW(Bag of Words)进行特征匹配。然后, 在系统相机姿态初始值估计阶段, 结合 PnP(Perspective n Point)和非线性优化方法, 为后端优化提供一个更接近最优值的初始值; 在后端优化中, 使用光束平差法 BA(Bundle Adjustment)对相机姿态初始值进行迭代优化, 得到相机姿态的最优值。最后, 根据相机姿态和每帧点云地图的对应关系, 将所有的点云数据注册到同一个坐标系中, 得到场景的稠密点云地图, 并对点云地图利用八叉树进行递归式的压缩得到一种用于机器人导航的三维地图。在 TUM RGB-D 数据集上, 将构建的 RGB-D SLAM 同 RGB-D SLAMv2、ORB-SLAM2 系统进行了对比, 结果表明所构建的 RGB-D SLAM 系统在实时性和精确度上综合表现更优。

关键词：RGB-D 传感器; 同时定位与地图构建; 稠密点云地图; 八叉树地图

中图分类号：TP242.6

文献标识码：A

Indoor robot simultaneous localization and mapping based on RGB-D image

ZHAO Hong, LIU Xiangdong*, YANG Yongjuan

(School of Computer and Communication, Lanzhou University of Technology, Lanzhou Gansu 730050, China)

Abstract: Simultaneous Localization and Mapping (SLAM) is a key technology for robots to realize autonomous navigation in unknown environments. Aiming at the poor real-time performance and low accuracy of commonly RGB-D (RGB-Depth) SLAM system. A RGB-D SLAM system is proposed to further improve the real-time performance and accuracy. Firstly, the ORB (Oriented FAST and Rotated BRIEF) is used to detect the image feature points and the extracted feature points are processed by using the quadtree-based homogenization strategy, and the BoW (Bag of Words) is used to match the features. Then, in the stage of initial camera pose estimation, an initial value that is closer to the optimal value is provided for back-end optimization by combining the PnP and nonlinear optimization methods. In the back-end optimization, the BA (Bundle Adjustment) method is used to optimize the initial value of the camera pose iteratively to obtain the optimal value. Finally, according to the correspondence between the camera pose and point cloud of each frame image, all the point cloud data are registered in a coordinate system to obtain the dense point cloud map of the scene, and the octree is used to compress the point cloud map recursively. Therefore, a 3D map for robot navigation is obtained. On the TUM RGB-D data set, proposed the RGB-D SLAM system, RGB-D SLAMv2 system and ORB-SLAM2 system are compared. Experimental results show that the performance of the proposed RGB-D SLAM system in this paper has higher real-time and accuracy.

Key words: RGB-D sensor; Simultaneous Localization and Mapping (SLAM); dense point cloud map; octo-map

0 引言

2020 年的春天一场席卷全球的 COVID-19(Corona Virus Disease 2019)使得人人自危。医护人员每天和病患频繁的接触大大增加了交叉感染的几率。一款具有自主导航功能的送药、

送餐医疗机器人能减少医患人员之间接触的频次, 降低医护人员被感染的几率。同时定位与地图构建(Simultaneous Localization and Mapping, SLAM)技术能够让机器人定位自己在未知环境中的具体位置, 了解周围环境的结构。这些信息是机器人实现自主导航功能的基础。自 1986 年被提出之后, SLAM 技术一直是机器人领域的热点研究问题^[1-3]。RGB-

收稿日期: 2020-04-23; **修回日期:** 2020-08-10; **录用日期:** 2020-08-11。

基金项目: 国家自然科学基金(51668043, 61262016), 赛尔网络下一代互联网技术创新项目(NG1120160311, NG1120160112)。

作者简介: 赵宏(1971—), 男, 甘肃西和人, 教授, 博士生导师, 博士, CCF会员, 主要研究方向: 并行与分布式处理、三维重建、深度学习; 刘向东(1994—), 男, 甘肃陇西人, 硕士研究生, CCF会员, 主要研究方向: 三维重建、视觉SLAM; 杨永娟(1996—), 女, 甘肃靖远人, 硕士研究生, CCF会员, 主要研究方向: 三维重建、视觉惯性SLAM。

D(RGB-Depth)传感器能有效避免单目和双目相机恢复的三维点云尺度不确定与精度低的问题。基于 RGB-D 传感器的 SLAM 方案,对室内机器人自主导航具有重要的意义^[4,5]。

基于特征点的视觉 SLAM(Visual SLAM)^[6]根据相机的种类可以分为三类:1)单目 SLAM^[7]。单目相机结构简单、成本低廉,单目 SLAM 通过相机的运动恢复自身的位置和周围场景的结构,但存在尺度不确定性的问题,且恢复的是稀疏结构。2)双目 SLAM^[2]通过视差恢复场景的稠密结构,但双目相机恢复的空间三维点的坐标误差较大。3)RGB-D SLAM^[8]则是借助 RGB-D 相机^[9]实现的,RGB-D 相机由彩色摄像头和深度摄像头组成,可以同时采集场景的彩色图像(如图 1(a))和深度图像(如图 1(b))。深度图像是一种特殊的“灰度图”,图中每个像素的值代表相机和场景之间的距离。当环境中不存在大量的红外光源时,RGB-D 相机可以准确的恢复出场景的三维点云(如图 1(c))。



图 1 RGB-D 相机采集的图像及生成的三维点云

Fig.1 Images captures by RGB-D camera and generates 3D point clouds

Hentry 等搭建了一套基于特征点的 RGB-D SLAM^[10],可以完成稠密点云的重建,但实时性不高。Endres 在迭代最近点(Iterative Closest Point, ICP)的基础上提出了 RGB-D SLAMv2 系统^[11],建立了一套完整的 RGB-D SLAM 系统,该系统提供了 ORB(Oriented FAST and Rotated BRIEF)^[12]、SIFT(Scale Invariant Feature Transform)^[13]、SURF(Speeded up Robust Features)^[14]、SIFTGPU^[15]四种特征提取算法的接口,使用之前只需要在配置文件中进行配置。并提供了 ROS(Robot Operating System)^[16]系统的接口,方便使用者进行实地建模。但系统在运行时有卡顿现象,且生成的点云文件占用存储空间过大,机器人在导航时难以应用。文献[17]采用非线性优化的方法求解相机姿态,准确度较 ICP 方法有所提升。文献[18]提出的 ORB-SLAM2 是 ORB-SLAM^[17]的升级版,提供了单目、双目、RGB-D 相机的接口,整个系统围绕 ORB 特征检测算法建立,可以在 CPU 上实时完成稀疏三维点云的重建,并使用三个线程完成 SLAM。在目前基于特征点的 SLAM 方案中,ORB-SLAM2 拥有较高的精度和实时性,但 ORB-SLAM2 得到的是稀疏的点云地图,无法满足机器人导航和避障的需求。文献[19]在 ORB-SLAM2 的基础上,增加了构建稠密地图的功能,并针对点云地图难以用于机器人导航的问题,提出了一种基于八叉树的导航地图构建方法,完成机器人在未知空间的导航。

为弥补现有 RGB-D SLAM 系统实时性差和精度低的问题,本文搭建一套 RGB-D SLAM 系统,主要内容如下:

(1)对特征检测算法进行改进,在 ORB 特征的基础上引入一种基于四叉树的特征点均匀化策略,保证特征点的均匀分布。并在特征匹配阶段,结合词袋模型(Bag of Words, BoW)^[20]提升特征匹配的速度和精度;

(2)为准确求解相机姿态初始值,采用 PnP(Perspective n Point)^[21]+RANSAC(Random Sample Consensus)^[22]+非线性优化的方法估计每帧图像的相机姿态;

(3)为避免误差的累积,结合回环检测和 BA(Bundle Adjustment)^[2]对相机姿态初始值进行优化,得到相机姿态的最优值。并在此基础上,根据点云数据和相机坐标系的对应性,将所有点云数据转换到同一个坐标系中,得到场景的稠密点云地图;

(4)对稠密点云地图进行压缩和转换,得到用于机器人导航的八叉树地图^[19]。

1 系统设计

系统实现主要分为:特征点检测与匹配、相机姿态估计及优化、回环检测、地图构建等部分。系统整体框架如图 2 所示。

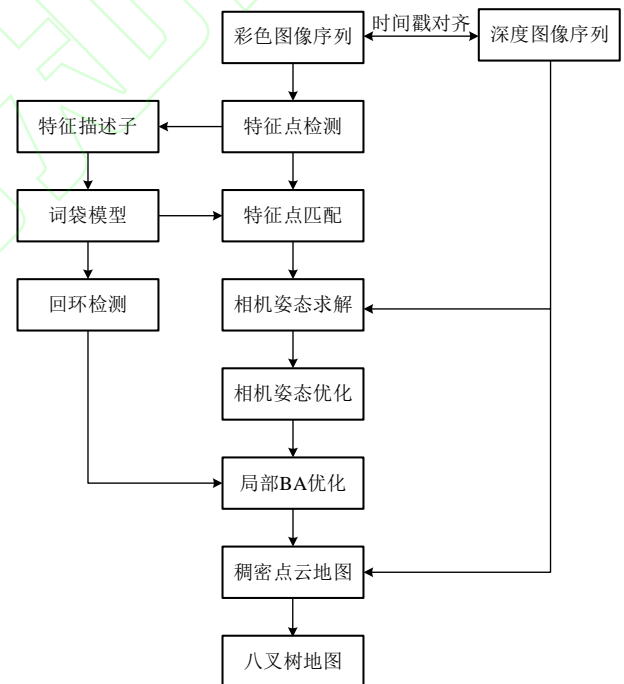


图 2 系统框架

Fig.2 System framework

1.1 特征检测与匹配

文献[12]提出的 ORB 特征检测算法由 FAST(Features from Accelerated Segment Test)关键点和 BRIEF(Binary Robust Independent Elementary Feature)描述子构成。相比 SIFT^[13]和 SURF^[14]算法,ORB 具有更快的速度(三者的比较如表 1,检测

的特征点数量为 1000，数据来源于文献[12])。因此本文采用 ORB 算法检测特征点。

表 1 特征检测算法时间对比

Tab.1 Comparison of feature detection algorithm time

算法	耗时/ms
ORB	15.3
SURF	217.3
SIFT	5228.7

ORB 特征检测算法由四个部分构成：1) 图像金字塔的构造。2) FAST 特征点检测。3) 计算特征点的旋转角度。4) 生成特征点的描述子。本文在此基础上对 ORB 算法进行改进，引入一种基于四叉树的特征点均匀分布策略，从而使 ORB 提取的特征点分布更加均匀。特征点的均匀分布步骤如下：

Step1: 对输入的图像构建图像金字塔。

Step2: 在金字塔的每一层图像中，按照大小为 30×30 的像素，进行分格处理。

Step3: 在每个格子中，提取 FAST 角点。

Step4: 对所有提取到的角点，采用四叉树进行划分，初始节点为 1。

Step5: 计算当前节点的特征点数量，如果大于 1，在当前节点继续按照四叉树进行划分，否则，当前节点停止划分；

Step6: 计算节点的数量，如果小于设定的需求特征点数量 N ，继续划分。否则，停止所有划分。

Step7: 遍历满足条件的节点，对每个节点中按照特征响应值 c 排序， c 的计算如式(1)所示，选取特征响应值最大的特征点作为代表。

$$c = \sum_i^{16} |I_p - I_i| \quad (1)$$

其中， I_p 是特征点对应的灰度值， I_i 是以特征点为圆心半径为 3 的圆上标号为 1~16 的像素的灰度值。

如图 3 所示，OpenCV(Open Source Computer Vision Library)[23]实现的 ORB 关键点有扎堆现象，本文实现的 ORB 关键点分布均匀。



(a) OpenCV 实现的结果 (b) 本文实现的结果

图 3 ORB 特征的均匀分布

Fig.3 Uniform distribution of ORB features

特征点检测与匹配的速度显著影响系统的实时性。本文在使用改进的 ORB 算法检测特征点的基础上，在特征匹配阶段，引入一种基于词袋模型[20]的特征匹配方法。首先，对输入的图像使用 ORB 算法提取特征。然后，运用 K-means++将 ORB 算法提取的图像描述子进行聚类，构建 k 叉树，生成图像的词典。最后，当进行特征匹配时，从 k 叉树的根节点出发，

查找特征对应的单词，将图像特征转换成单词向量的描述。当对应的单词向量一致时，对特征进行匹配。在该过程中，有效减少了特征点匹配的范围，因此特征匹配的准确率得到提升。同时， k 叉树便于查找的特性，保证了在匹配时有较快的速度。词袋模型的生成原理如图 4 所示，最终得到一颗深度为 d ($d = \log_k N$ ， N 为特征点的个数)的 k 叉树。

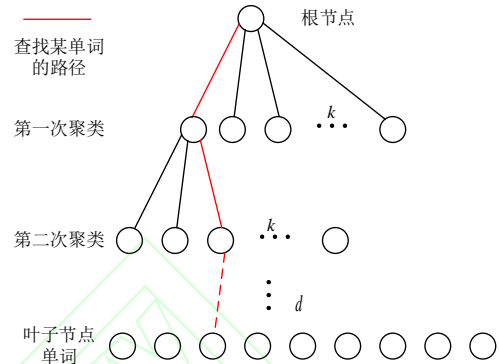


图 4 词袋模型的生成原理示意图

Fig.4 The generating principle of BoW

1.2 相机姿态的求解与优化

视觉 SLAM 将整个系统分为前端视觉里程计 (Visual Odometry, VO)[21]和后端优化两个部分。后端对前端输入的相机姿态进行优化，从而得到相机姿态的最优值。因此，前端估计的相机姿态初始值对于整个系统的精度具有显著的影响。本文采用 PnP+RANSAC+非线性优化的方法估计相机姿态的初始值。

PnP 是一种相机位姿求解方法。P3P 算法[24](图 5)是 PnP 中的一种，只需四对 3D-2D 匹配点。其中一对匹配点作为验证点。其余三对匹配点中，2D 点 a 、 b 、 c 是 3D 点 A 、 B 、 C 在相机成像平面上的投影。 O 表示相机的光心。

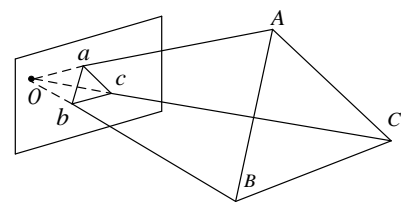


图 5 P3P 问题示意图

Fig.5 The principle of P3P problem

三角形之间的对应关系如式(2)所示。

$$\begin{cases} \triangle Oab \sim \triangle OAB \\ \triangle Obc \sim \triangle OBC \\ \triangle Oac \sim \triangle OAC \end{cases} \quad (2)$$

根据余弦定理，三角形之间的关系如式(3)所示。

$$\begin{cases} x^2 + y^2 - 2xy \cos \langle a, b \rangle - \varphi = 0 \\ y^2 + 1 - 2y \cos \langle b, c \rangle - \omega \varphi = 0 \\ x^2 + 1 - 2x \cos \langle a, c \rangle - \rho \varphi = 0 \end{cases} \quad (3)$$

其中， $x = OA/OC$ ， $y = OB/OC$ ， $\varphi = AB^2/OC^2$ ， $\omega \varphi = BC^2/OC^2$ ， $\rho \varphi = AC^2/OC^2$ 。

根据 2D 点的像素坐标、 φ 、 ρ 以及三个余弦角 $\cos \langle a, b \rangle$ 、 $\cos \langle b, c \rangle$ 、 $\cos \langle a, c \rangle$ 的值, 可得式(3)是一个关于 x 和 y 的二元二次方程, 使用吴消元法^[24]求解。

经过改进的特征匹配关系中仍然可能存在误匹配。误匹配会造成 P3P 算法的失效。因此, 本文采用 RANSAC 模型来增强 P3P 算法的鲁棒性。

$$\left\| \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} - T \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} \right\|_2 \leq \beta \quad (4)$$

其中 (X_w, Y_w, Z_w) 表示空间点 P 在世界坐标系下的三维坐标, (X_c, Y_c, Z_c) 表示 P 点在相机坐标系下的三维坐标。 T 表示变换矩阵, 由旋转矩阵 R 和平移向量 t 构成。 β 表示距离阈值。运用 RANSAC 算法的具体步骤如下:

Step1: 随机抽取 4 对匹配点, 作为初始的内点集合, 通过该内点集合计算出变换矩阵 T 。

Step2: 依次判断剩余的点, 计算三维坐标点之间的距离, 如果距离小于设定的阈值(本文设定的阈值为 $4.0^{[2]}$), 则将当前点加入到内点集合中。

Step3: 重复迭代步骤 M (M 的值为 $100^{[2]}$)次, 选取内点样本数量最多的一组作为最终的匹配点集。

为对 P3P 算法求解的相机姿态进行优化, 将相机的姿态根据 BA 定义成一个最小二乘问题, 采用小孔相机模型描述成像原理。小孔相机模型的定义如下:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_w} K \begin{bmatrix} R & t \\ 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (5)$$

其中, (X_w, Y_w, Z_w) 是空间点 P 在世界坐标系中的坐标, K 是摄像机的内参, R 是旋转矩阵, t 是平移向量, (u, v) 是 P 点投影在成像平面上的像素坐标。

待优化的误差项 e 的计算如式(6)所示:

$$e = P' - \frac{1}{s} K \exp(\xi^{\wedge}) P \quad (6)$$

其中, P' 是投影点 P 的真实像素坐标, s 是 P 点的深度, K 是摄像机的内参, ξ 是相机姿态的李代数^[2], 最小化误差项的计算如式(7)所示。

$$\xi^* = \arg \min_{\xi} \frac{1}{2} \sum_{i=1}^n \left\| P'_i - \frac{1}{s_i} K \exp(\xi^{\wedge}) P_i \right\|_2^2 \quad (7)$$

使用列文伯格马夸尔特(Levenberg Marquardt, LM)方法不断调整 ξ 的值, 最终得到与当前匹配关系最优的相机姿态。在 LM 方法中, 使用 $J^T J$ 代替海塞矩阵(Hessian Matrix), 避免计算量大的问题。相机姿态的雅可比矩阵 J_c 的计算方法如式(8)所示。

$$\frac{\partial e}{\partial \delta \xi} = - \begin{bmatrix} -\frac{f_x X_c Y_c}{Z_c^2} & f_x + \frac{f_x X_c^2}{Z_c^2} & -\frac{f_x Y_c}{Z_c} \\ -f_y - \frac{f_y Y_c^2}{Z_c^2} & -\frac{f_y X_c Y_c}{Z_c^2} & \frac{f_y X_c}{Z_c} \end{bmatrix} \quad (8)$$

其中, f_x 和 f_y 分别是摄像机在水平和竖直方向的焦距, (X_c, Y_c, Z_c) 表示 P 点在相机坐标系下的三维坐标。

上述内容介绍了系统前端相机姿态的求解方法。由于噪声的存在, 上一时刻的误差会传递到下一时刻, 随着系统的运行, 误差会不断的积累, 最终造成系统的奔溃。为减小误差累积, 系统在后端将前端输入的所有相机姿态和三维点进行更大范围的优化。后端局部 BA 优化的计算如式(5)所示, 空间三维点的雅可比矩阵 J_p 的计算方法如式(9)所示。

$$\frac{\partial e}{\partial P} = - \begin{bmatrix} \frac{f_x}{Z_c} & 0 & -\frac{f_x X_c}{Z_c^2} \\ 0 & \frac{f_y}{Z_c} & -\frac{f_y Y_c}{Z_c^2} \end{bmatrix} R \quad (9)$$

1.3 回环检测

回环检测的目的是让机器人在运动的过程中识别出曾经到过的地方, 从而在采集的图像数据中添加一个有间隔时间的约束, 进而减小误差的积累, 提高重建系统的精度。本文采用基于词袋模型^[20]的回环检测方法, 判断图像之间的相似性。该方法, 首先将图像用词袋模型转换成单词向量, 然后通过对单词向量的比较判断两帧图像之间的相似性。两幅图像 f_A 和 f_B 之间的相似性计算如式(10)所示。

$$s(f_A, f_B) = \frac{1}{2} \sum_{i=1}^n \{ |f_{A_i}| + |f_{B_i}| - |f_{A_i} - f_{B_i}| \} \quad (10)$$

回环检测的具体步骤如下:

Step0: 输入图像。

Step1: 计算特征点和描述子。

Step2: 从 k 叉树的根节点出发, 查找描述子对应的单词, 生成单词向量。

Step3: 和关键帧序列中的图像进行对比, 计算图像之间的相似性。

Step4: 如果满足设定的阈值(本文设定的阈值为 0.32), 则将当前帧设置为回环帧候选帧, 否则, 返回 Step0。

Step5: 计算当前帧和回环候选帧的特征匹配数量(本文设定的阈值为 45)。如果满足阈值, 则判定当前帧和当前回环候选帧产生回环。否则, 返回 Step0。

Step6: 在当前帧和回环帧中计算相机姿态, 更新当前帧的状态。

1.4 导航地图

受深度相机量程和视角的限制, 使用传感器每次仅仅能获取一部分待重建的场景, 因此需要通过不同角度来扫描, 最

终根据深度图和相机坐标系的对应关系,使用优化后的相机姿态 R_i 和 t_i 将各个角度下的深度图恢复的点云数据转换到同一个坐标系中,得到场景的稠密点云地图。

但稠密点云地图的文件存储量过大,包含了大量对于机器人导航无用的信息,且无法处理动态的物体。因此难以应用于机器人导航。八叉树地图是一种灵活的、压缩的、能实时更新的地图形式,相比点云地图更有利于机器人的导航。八叉树的生成原理如图6所示。

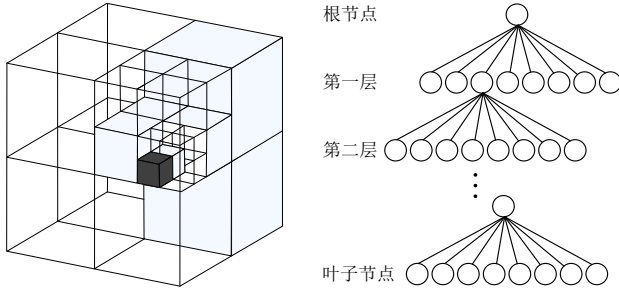


图6 八叉树示意图

Fig.6 The principle of Octree

八叉树的每个叶子节点中存储是否被点云占据的概率值,即表示机器人在某个地方能不能通过。对于第 n 个叶子节点,设系统在 T_i 状态下观测到的数据为 z_i , 则第 n 个叶子节点存储的概率值 P_{oct} 为:

$$P_{oct}(n | z_{1:i}) = \left[1 + \frac{1 - P_{oct}(n | z_i) \cdot 1 - P_{oct}(n | z_{1:i-1}) \cdot 1 - P_{oct}(n)}{P_{oct}(n | z_i) \cdot P_{oct}(n | z_{1:i-1}) \cdot 1 - P_{oct}(n)} \right]^{-1} \quad (11)$$

引入 logit 变换。

$$Q = \text{logit}(P_{oct}) = \left\{ \frac{P_{oct}}{1 - P_{oct}} \right\} \quad (12)$$

则式(11)变换为:

$$L(n | z_{1:i+1}) = L(n | z_{1:i-1}) + L(n | z_i) l(n) = \frac{1}{8} \sum_{i=1}^8 L(n_i) \quad (13)$$

即第 n 个叶子节点的值当前观测结果与上一次观测结果的和。在系统运行的过程中,根据观测到某个节点被占据的情况,在初始值(P_{oct} 的初始值为0.5^[19])的基础上不断增加或减小 P_{oct} 的值。当 Q 的值从 $-\infty$ 变到 $+\infty$, P_{oct} 的值相应地从 0 变到 1。父亲节点的概率值为孩子节点的平均值,父亲节点 $l(n)$ 的计算如式(14)所示。

$$l(n) = \frac{1}{8} \sum_{i=1}^8 L(n_i) \quad (14)$$

2 实验评估

2.1 实验环境及数据集

为验证本文构建的系统的有效性,使用 TUM RGB-D 数据集^[25],从系统运行时间和绝对位姿误差(Absolute Pose Error, APE)^[25]两个方面与现有的系统进行对比实验。实验环

境如表2所示。选用的三个数据集序列分别是: fr1_desk2、fr1_xyz、fr1_room。

表2 实验环境

Tab.2 Experimental environment

参数	数值
CPU	Intel(R) Core(TM) i5-4258U @ 2.40GHz
内存	8GB
系统	Windows10/Ubuntu16.04
语言	C/C++

2.2 运行时间对比

本文系统、ORB-SLAM2^[18]、RGB-D SLAMv2^[11]在运行时间方面的对比如表3所示。为保证公平性,RGB-D SLAMv2中同样使用ORB算法检测特征点。

表3 系统运行时间对比

Tab.3 Comparison of system running time

图像序列	运行时间/s		
	本文系统	ORB-SLAM2	RGB-D SLAM v2
fr1_desk2	53.418	47.4528	87.1943
fr1_xyz	57.5159	51.9736	96.8286
fr1_room	129.517	86.1355	195.505

由表3可知,本文构建的系统平均运行时间约为ORB-SLAM2^[18]的1.25倍,这是由于本文进行稠密点云地图重建造成的。文献[18]的方法虽然快,但得到稀疏点云地图不能应用于机器人导航。另外,本文系统在文献[11]的基础上平均运行时间降低了37.697%,这得益于基于词袋模型的特征匹配方法具有较快的速度。最后,在运行期间,场景的点云不断更新的,没有卡顿现象,证明本文构建的系统可以实时运行。

2.3 系统精度对比

绝对位姿误差 APE 描述的是每一帧相机的真实姿态和估计姿态之间的差值,可以直观的反应系统的精度。对于第 i 帧图像, APE 的计算如式(15)所示。

$$E_i = Q_i^{-1} S T_i \quad (15)$$

其中, Q_i 是真实的相机姿态, T_i 是系统估计的相机姿态, S 是两者之间的转换矩阵。使用均方根误差(Root Mean Square Error, RMSE)统计总体值,计算方法如式(16)所示:

$$RMSE(E_{i:n}, \Delta) = \left(\frac{1}{m} \sum_{i=1}^m \| \text{trans}(E_i) \|^2 \right)^{\frac{1}{2}} \quad (16)$$

本文系统、ORB-SLAM2^[18]、RGB-D SLAMv2^[11]在精度方面的对比如表4所示。

由表4可知,在 fr1_desk2 和 fr1_xyz 序列上,本文系统与文献[18]的系统精度几乎相同。在 fr1_room 序列上,本文的精度约为文献[18]的57.89%。与文献[11]的比较中,在三个序列上,本文实现的结果精度均优于文献[11],在文献[11]的基础上平均提升了约56.104%。

表 4 系统 APE 精度对比

Tab.4 Comparison of system APE accuracy

图像序列	RMSE		
	本文系统	ORB-SLAM2	RGB-D SLAM v2
fr1_desk2	0.022186	0.023211	0.058751
fr1_xyz	0.009708	0.010084	0.019186
fr1_room	0.077027	0.044592	0.177780

三个图像序列上,APE 的实时变化情况分别如图 7 所示。

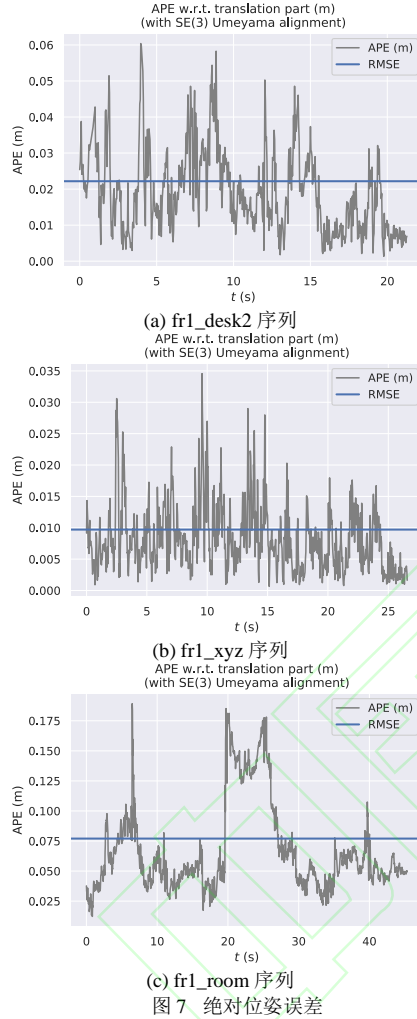
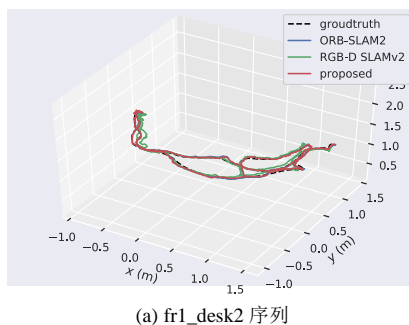


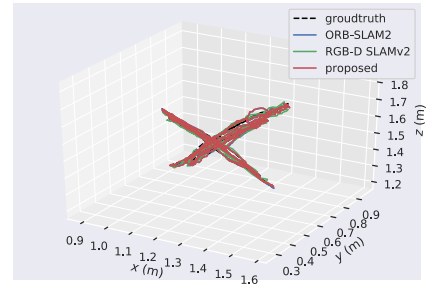
图 7 绝对位姿误差

Fig.7 Absolute pose error

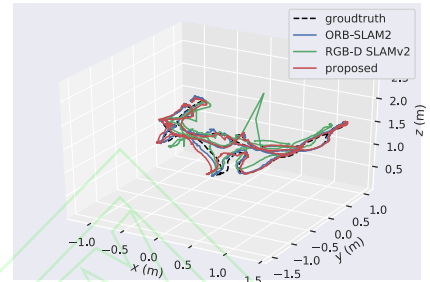
三个系统和真实轨迹的对比结果分别如图 8 所示(其中 groundtruth 是真实轨迹, proposed 为本文系统)。和真实轨迹的对比也印证了系统的精度,本文和文献[18]生成的轨迹和真实轨迹的贴合程度较高,而文献[11]生成的轨迹偏离程度较大。



(a) fr1_desk2 序列



(b) fr1_xyz 序列



(c) fr1_room 序列

图 8 相机轨迹对比

Fig.8 Comparison of camera trajectory

2.4 八叉树地图

本文系统在 fr1_room 序列上得到的稠密点云地图如图 9 所示,稠密点云地图经过八叉树压缩得到的 3D 导航地图(分辨率为 0.05 米)如图 10 所示。虽然,本文选择的系统精度衡量标准不能直接衡量生成地图的质量。但是,相机坐标系和单帧点云数据具有对应关系,相机姿态的偏差程度直接决定了点云地图的偏差程度。



图 9 稠密点云地图

Fig.9 Map of dense point cloud



图 10 3D 导航地图

Fig.10 Map of 3D navigation

三个图像序列上稠密点云地图和八叉树地图文件的存储量对比如表 5 所示。

表 5 地图文件存储量对比
Tab.5 Comparison of map file size

图像序列	地图文件存储量/MB	
	稠密点云地图	八叉树地图
fr1_desk2	31.3	0.818
fr1_xyz	7.7	0.216
fr1_room	84.2	2.932

由表 5 可知, 经过八叉树压缩后的稠密点云地图的文件大小显著减小, 平均大小约为原来稠密点云地图大小的 3.217%。在机器人自主导航时, 较小的地图文件, 便于机器人携带和加载。

3 结语

本文基于 RGB-D 图像, 根据视觉 SLAM 的相关技术原理, 建立了一套便于机器人导航的八叉树地图。在 TUM RGB-D 数据集上的实验结果证明, 本文构建的 RGB-D SLAM 系统在运行速度和精度上要优于 RGB-D SLAMv2^[11]系统, 在 RGB-D SLAMv2 的基础上, 平均运行速度降低约 37.697%, 平均精度提升约 56.104%。与目前具有代表性的 ORB-SLAM2^[18]系统相比, ORB-SLAM2 的平均运行速度约为本文系统速度的 1.25 倍, 这是本文进行稠密点云地图重建造成的。就系统的精度而言, 在 fr1_desk2 和 fr1_xyz 序列上两个系统几乎相同。在 fr1_room 序列上, 本文构建的 RGB-D SLAM 系统约为 ORB-SLAM2 的 57.89%。综合而言, 本文构建的 RGB-D SLAM 系统具有较高的实时性和精度。另外, 建立的基于八叉树的导航地图, 平均约为稠密点云文件存储量的 3.217%, 满足了机器人对导航地图的需要。在接下来的工作中将探索机器人导航时, 如何在八叉树地图上进行路径规划与避障。

参考文献

- [1] LIU H M, ZHANG G F, Bao H J. A survey of monocular simultaneous localization and mapping[J]. Journal of computer Aided Design and Compute Graphics, 2016, 28(6): 855-868.
- [2] 高翔, 张涛, 颜沁睿, 等. 视觉 SLAM 十四讲: 从理论到实践[M]. 北京: 电子工业出版社, 2017: 132-180. (GAO X, ZHANG T, YAN Q R, et al. The 14 lectures on visual SLAM: from theory to practice[M]. Beijing: Publishing House of Electronics Industry, 2017: 132-180.)
- [3] CESAR C, CARLONE L, LUCA. Past, present, and future of simultaneous localization and mapping: toward the robust-perception age [J]. IEEE Transactions on Robotics, 2016, 32(6): 1309-1332.
- [4] 胡凌燕, 曹祿, 熊鹏文, 等. 基于 RGB-D 图像的三维同步定位与建图研究[J]. 系统仿真学报, 2017, 29(11): 243-249. (3D simultaneous localization and mapping based on RGB-D Images[J]. Journal of System Simulation, 2017, 29(11): 243-249.)
- [5] WANG J, HUANG S, ZHAO L, et al. High quality 3D reconstruction of indoor environments using RGB-D sensors[C]// Proceedings of the 12th IEEE Conference on Industrial Electronics and Applications. 2017.
- [6] 权美香, 朴松昊, 李国. 视觉 SLAM 综述[J]. 智能系统学报, 2016, 11(6): 768-776. (QUAN M X, PIAO S H, LI G. An overview of visual SLAM[J]. CAAI Transactions on Intelligent Systems, 2016, 11(6): 768-776.)
- [7] DAVISON A, REID I, MOLTON N. Mono Slam: real-time single camera SLAM[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6):1052-1067.
- [8] 陈世浪, 吴俊君. 基于 RGB-D 相机的 SLAM 技术研究综述[J]. 计算机工程与应用, 2019, 55(7):30-39. (CHEN S L, WU J J. RGB-D SLAM: a survey[J]. Computer Engineering and Applications, 2019, 55(7):30-39.)
- [9] 余涛. Kinect 应用开发实战: 用最自然的方式与机器对话[M]. 北京: 机械工业出版社, 2013: 48-62. (YU T. Kinect application development practice: The most natural way to talk to the machine[M]. Beijing: China Machine Press, 2013:48-62.)
- [10] HENRY, P, KRAININ, M, HERBST, E, et al. RGB-D mapping: Using kinect-style depth cameras for dense 3D modeling of indoor environments[J]. International Journal of Robotics Research, 2012, 31(5):647-663.
- [11] ENDRES F. RGB-D SLAMv2[EB/OL].[2020-1-16]. <https://github.com/felixendres>.
- [12] RUBLEE E, RABAUDE V, KONOLIGE K. ORB: an efficient alternative to SIFT or SURF[C]. Proceedings of the 2011 IEEE International Conference on Computer Vision. 2011, 2564-2571.
- [13] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 62(2): 91-110.
- [14] BAY H, TUVTELARRS T, GOOL L J V. SURF: Speeded up robust features[J]. Computer Vision and Image Understanding, 2008, 110(3): 404-417.
- [15] SIFTGPU[EB/OL]. [2019-9-18]. <https://github.com/pitzer/SiftGPU>.
- [16] ROS[EB/OL]. [2019-10-19]. <http://wiki.ros.org/ROS/>.
- [17] 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.
- [18] MUR-ARTAL R, TARDOS J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D camera[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255-1262.
- [19] 林辉灿, 吕强, 王国胜, et al. 基于 VSLAM 的自主移动机器人三维同时定位与地图构建[J]. 计算机应用, 2017, 37(10): 2884-2887. (LIN H C, LYU Q, WANG G S, et al. 3D simultaneous localization and mapping for mobile robot based on VSLAM[J]. Journal of Computer Applications, 2017, 37(10): 2884-2887.)
- [20] 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.
- [21] 陈鹏, 王晨晓. IEPnP: 一种基于 EPnP 的相机位姿迭代估计算法[J]. 光学学报, 2018, 38(4): 130-136. (CHEN P, WANG CH X. IEPnP: an iterative camera pose estimation algorithm based on EPnP[J]. Acta Optica Sinica. 2018, 38(4): 130-136.)
- [22] 樊彦国, 柴江龙, 许明明, 等. 基于 ORB 与 RANSAC 融合改进的图像配准[J]. 光学精密工程, 2017, 27(3): 702-717. (FAN Y G, CHAI J L, XU M M, et al. Improved fast image registration algorithm based on ORB and RANSAC fusion[J]. Optics and Precision Engineering. 2017, 27(3): 702-717.)
- [23] OpenCV[EB/OL]. [2019-8-19]. <https://opencv.org/>.
- [24] GAO X S, HOU X R, TANG J, et al. Complete solution classification for the perspective-three-point problem[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(8):930-943.
- [25] JURGEN S, ENGELHARD N, ENDRES F. A benchmark for the evaluation of RGB-D SLAM systems[C]// Proceedings of the 2012 IEEE International Conference on Intelligent Robots & Systems. 2012: 580-593.

This work is partially supported by the National Natural Science Foundation of China (51668043, 61262016), the next generation Internet technology innovation project of the SEIR network (NG1120160311, NG1120160112).

ZHAO Hong, born in 1971, Ph. D., professor. His research interests include parallel and distributed processing, 3D Reconstruction, deep learning.

LIU Xiangdong, born in 1994, M. S. candidate. His research interests include 3D Reconstruction, Visual SLAM.

YANG Yongjuan, born in 1996, M. S. candidate. Her research interests include 3D Reconstruction, Visual-Inertia SLAM.

