



基于 ORB-SLAM 的室内机器人定位和三维稠密地图构建

侯荣波, 魏武*, 黄婷, 邓超锋

(华南理工大学 自动化科学与工程学院, 广州 510640)

(* 通信作者电子邮箱 weiwu@scut.edu.cn)

摘要:针对在室内机器人定位和三维稠密地图构建系统中, 现有方法无法同时满足高精度定位、大范围 and 快速性要求的问题, 应用具有跟踪、地图构建和重定位三平行线程的 ORB-SLAM 算法估计机器人三维位姿; 然后拼接深度摄像头 KINECT 获得的三维稠密点云, 提出空间域上的关键帧提取方法剔除冗余的视频帧; 接着提出子地图法进一步减少地图构建的时间, 最终提高算法的整体速度。实验结果表明, 所提系统能够在 50 m 的大范围环境中准确定位机器人位置, 在运动轨迹为 50 m 的大范围中, 机器人的均方根误差为 1.04 m, 即误差为 2%, 同时整体速度为 11 帧/秒, 其中定位速度达到 17 帧/秒, 可以满足室内机器人定位和三维稠密地图构建的精度、大范围 and 快速性的要求。

关键词:同时定位和地图构建; 室内机器人; ORB-SLAM; 关键帧提取; KINECT; 图优化

中图分类号: TP242.6 **文献标志码:** A

Indoor robot localization and 3D dense mapping based on ORB-SLAM

HOU Rongbo, WEI Wu*, HUANG Ting, DENG Chaofeng

(School of Automation Science and Engineering, South China University of Technology, Guangzhou Guangdong 510640, China)

Abstract: In the indoor robot localization and 3D dense mapping, the existing methods can not satisfy the requirements of high-precision localization, large-scale and rapid mapping. The ORB-SLAM (Oriented FAST and Rotated BRIEF-Simultaneous Localization And Mapping) algorithm, which has three parallel threads including tracking, map building and relocation, was used to estimate the three-dimensional (3D) pose of the robot. And then 3D dense point cloud was obtained by using the depth camera KINECT. The key frame extraction method in spatial domain was introduced to eliminate redundant frames, and the sub-map method was proposed to reduce the cost of mapping, thereby the whole speed of the algorithm was improved. The experiment results show that the proposed method can locate the robot position accurately in a large range. In the range of 50 meters, the root-mean-square error of the robot is 1.04 m, namely the error is 2%, the overall speed is 11 frame/s, and the localization speed is up to 17 frame/s. The proposed method can meet the requirements of indoor robot localization and 3D dense mapping with high precision, large-scale and rapidity.

Key words: Simultaneous Localization And Mapping (SLAM); indoor robot; Oriented FAST and Rotated BRIEF-Simultaneous Localization And Mapping (ORB-SLAM); key frame extraction; KINECT; graph optimization

0 引言

在过去的十年里, 室外定位在以卫星定位为基础的技术上获得了快速的发展和广泛的应用, 如, 全球定位系统 (Global Positioning System, GPS)。但是我们超过 70% 的时间都是在室内, 因此室内定位技术具有巨大的研究价值和应用价值。由于室内环境复杂, 需要从一系列的测量数据中估计移动节点的实时位置, 所以到目前为止没有出现可以应用的解决方法^[1]。

在机器人的定位领域, 基于激光和视觉的同时定位和地图构建 (Simultaneous Localization And Mapping, SLAM) 系统得到了快速的发展。SLAM 是指搭载传感器的机器人在移动时利用传感器测量数据建立环境地图, 同时估计机器人自身的位姿^[2]。SLAM 同时包含定位与地图构建技术, 被认为是

实现机器人自主性的关键技术之一, 对机器人的控制、导航、任务规划等领域有重要的研究意义^[3]。根据 SLAM 的求解方法可以将其分为扩展卡尔曼滤波类 (Extended Kalman Filter, EKF) SLAM、粒子滤波类 (Particle Filtering, PF) SLAM 和图优化类 (Graph Optimization) SLAM。

EKF-SLAM 最早由 Smith 等^[4]提出, 在假设噪声为高斯噪声的基础上, 通过运动模型和观测模型来估计机器人自身及路标点的位置。之后许多专家学者在其方法上进一步改进, 针对 EKF-SLAM 的估计不一致性和非线性函数的线性化问题, Julier 等^[5]提出无迹卡尔曼滤波 (Unscented Kalman Filter, UKF) 方法, 利用样本加权求和直接逼近随机分布, 从而避免了对非线性函数雅各比矩阵的计算。张毅等^[6]提出了迭代扩展卡尔曼滤波 (Iterated Extended Kalman Filter, IEKF) 通过多次迭代更新获得更高的后验概率, 降低了估计

收稿日期: 2016-10-14; 修回日期: 2016-12-21。

基金项目: 国家自然科学基金资助项目 (61573148); 广东省科技重大专项 (2015B010919007)。

作者简介: 侯荣波 (1993—), 男, 广东肇庆人, 硕士研究生, 主要研究方向: SLAM 算法、三维重构、计算机视觉; 魏武 (1970—), 男, 湖南益阳人, 副教授, 博士, 主要研究方向: 机器人控制、智能控制、模式识别、人工智能; 黄婷 (1991—), 女, 湖北孝感人, 硕士研究生, 主要研究方向: 计算机视觉、机器人嵌入式系统; 邓超锋 (1990—), 男, 江西鄱阳人, 硕士研究生, 主要研究方向: 机器人控制、智能控制。



误差。此外,EKF-SLAM 依赖于正确的数据关联,一旦数据关联出错则会导致构建地图失败;而且 EKF-SLAM 的计算成本高,计算量正比于地图规模的平方,因此只能对较小范围的环境构建地图。在 EKF-SLAM 框架上,一些研究人员在不同机器人系统上尝试了许多开发工作^[7]。

针对 EKF-SLAM 对正确数据关联十分敏感的问题,Sim 等^[8]引入粒子滤波器(PF)来解决,其思路是将机器人和路标点的位置用粒子的分布来表达,由于没有了 EKF-SLAM 对噪声的高斯假设,同时粒子数量较多,对机器人和路标点位置具有多个假设,因此对正确的数据关联不敏感。但是 PF-SLAM 仍面临着由于粒子数量大导致计算成本高,以及由于重采样导致的粒子耗尽问题。为了减少粒子数量,Blanco 等^[9]使用 KL 距离(Kullback-Leibler Distance, KLD)方法自适应确定粒子数量,从而剔除了冗余的粒子。

除了滤波器求解 SLAM 问题之外,最近基于图优化 SLAM 成为研究的热点。其起源于基于捆集调整(Bundle Adjustment, BA)的运动构建(Structure From Motion, SFM)领域,SFM 把位姿估计看作是一个全局优化问题,约束条件为相机的运动方程,目标函数为观测误差,最终通过最小化目标函数获得相机位姿。但是由于全局优化的计算量巨大,早期的 SFM 只能是离线的位姿估计和地图构建。后来研究者逐步认识到了 SLAM 求解问题的稀疏性,Sibley^[10]提出了把 SLAM 问题看作是一个图优化问题,图优化的节点描述了需要优化的变量,边则是与优化变量相关的约束条件,最终采用稀疏目标优化方法计算获得优化变量,即机器人的位姿和路标点位置。Klein 等^[11]提出了平行跟踪和地图构建方法(Parallel Tracking And Mapping, PTAM),把运动跟踪和地图构建分为两个任务,并在两个平行的线程同时进行处理,在线实时地实现了 SLAM 算法,虽然 PTAM 存在着缺少足够的回环检测、重定位不变性低、需要人工初始化地图等问题,但是其代表了视觉 SLAM 的一个重要突破。继 PTAM 之后,Mur-Artal 等^[12]提出了 ORB-SLAM(Oriented FAST and Rotated BRIEF-Simultaneous Localization And Mapping)算法,该算法具有跟踪、地图构建、重定位三个平行线程,有效地解决了回环检测、重定位和地图初始化等问题,能够在小范围和大范围的未知环境中在线实时地实现高精度定位,但算法最终只构建基于特征点的稀疏地图,无法应用于机器人导航、路径规划等实际领域。

针对 ORB-SLAM 算法的不足,引入了在线快速构建未知环境三维稠密地图的方法,算法的整体步骤如图 1 所示。

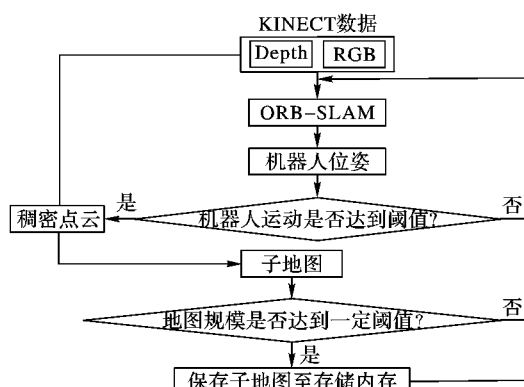


图 1 算法流程
Fig. 1 Algorithm flow

采用深度摄像头 KINECT 作为传感器设备,首先提出空间域上的方法提取关键帧,即当机器人的运动超过一定阈值时则把该帧选取为关键帧,并将它的点云拼接到已有的地图中;然后提出子地图法,将大规模的地图分解为一定规模的小地图,从而减少保存地图的运算内存和构建地图的时间,由此来提高算法的整体速度。实验表明,本文方法在标准 RGBD 数据集上能够取得较好的效果,在定位精度和快速性两方面都明显优于 RGBD-SLAM^[13]算法,定位误差仅为 RGBD-SLAM 的 40%,同时算法速度是 RGBD-SLAM 的 2.6 倍;在实际室内机器人实验中,本文方法能够快速准确地定位机器人位置,定位速度达到 17 帧/秒,能建立高精度的三维稠密地图。

1 ORB-SLAM 概述

ORB-SLAM 是一种基于图像特征和非线性优化的单目视觉 SLAM 系统。它包括基于 ORB(Oriented FAST and Rotated BRIEF)图像特征的词袋(Bag of Words, BoW)用于位置识别及回环检测^[14]、信息关联视图^[15]、G2o 图优化通用框架^[16]。对于大范围的地图构建,应用了尺度感知的回环检测。该算法在全部的处理当中只用到了 ORB 来作特征检测和描述,这样一来提高了其在位置识别和回环检测的效果。文献[12]完整地介绍了 ORB-SLAM 算法,在此主要总结了它的三个主要步骤,包括跟踪、构建特征地图、回环检测,如图 2 所示。

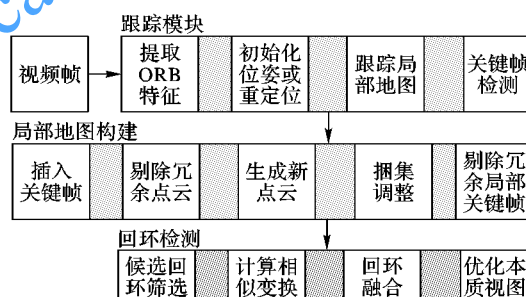


图 2 ORB-SLAM 系统流程

Fig. 2 System overview of ORB-SLAM

1.1 跟踪

这一部分的主要任务是通过 KINECT 获得视频帧,并用光流法^[17]来跟踪其位置。光流法是利用图像序列中像素在时间域上的变化,求解出空间运动物体在观察成像平面上的像素运动的瞬时速度 $s = (u, v)$,从而计算出相邻帧之间物体的运动信息,估计上一帧跟当前帧之间存在的对应关系。假设像素的光流运动微小和亮度恒定,可以得到 $I(x, y, t) = I(x + dx, y + dy, t + dt)$,由一阶泰勒展开得:

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt \quad (1)$$

即 $I_x dx + I_y dy + I_t dt = 0$,并令 $u = \frac{dx}{dt}$, $v = \frac{dy}{dt}$,那么则有:

$$\begin{bmatrix} I_x & I_y \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = -I_t \quad (2)$$

对于 N 个像素点的光流:



$$\begin{bmatrix} I_{x1} & I_{y1} \\ I_{x2} & I_{y2} \\ \vdots & \vdots \\ I_{xN} & I_{yN} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_{d1} \\ I_{d2} \\ \vdots \\ I_{dN} \end{bmatrix} \quad (3)$$

将式(3)记为 $A \cdot s = b$, 那么:

$$s = \arg \min_s \|A \cdot s - b\| \quad (4)$$

假设地图点的三维位置可用,并且用 ORB 特征来描述。如果在当前图像帧中跟踪成功,则通过之前图像帧的运动模型平均值来估计当前 KINECT 的位姿,然后重映射地图点估计其在当前帧上的图像。用 ORB 特征描述的地图点与在预测点附近范围检测到的特征进行匹配,图像上最小汉明距离的特征点被选为匹配点。然后利用非线性优化最小化重投影误差来计算获得相机位姿。优化之后,匹配点根据合适的阈值被分为内点和外点。如果跟踪失败,首先把当前图像帧转换成图像词袋,检索图像数据库,为全局重定位查找关键帧。然后计算 ORB 特征和每个关键帧的地图云点的对应关系,最后对每个关键帧执行随机一致性检验(RANdom Sample And Consensus, RANSAC)迭代计算,用 PnP(Perspective-n-Point)算法估计 KINECT 位姿。一旦估计获得 KINECT 的位姿,就可以恢复跟踪模块。

1.2 构建局部地图

为了构建环境的三维地图,系统在 KINECT 获得的视频帧中提取合适的帧。利用跟踪模块获得的匹配特征点,估计关键帧之间的匹配关系。一旦匹配有效,就采用 BA 来优化估计三维地图点位置和关键帧位姿。算法在一个和跟踪线程平行的线程中以较低的频率连续计算匹配和迭代提高地图点的精度。BA 最小化关键帧匹配点的重映射误差, T_{w,c_i} 为第 i 个关键帧的位姿, $X_{w,j}$ 为第 j 个三维地图点:

$$\arg \min_{T_{w,c_i}, X_{w,j}} \sum \rho_h(e_{i,j}^T \Omega_{i,j}^{-1} e_{i,j}) \quad (5)$$

其中: ρ_h 为 Huber 函数, $\Omega_{i,j}$ 为协方差矩阵, $e_{i,j}$ 由式(6)、(7)计算获得:

$$e_{i,j} = u_{i,j} - \text{CamProj}(T_{w,c_i}, X_{w,j}) \quad (6)$$

$$\text{CamProj}(T_{w,c_i}, X_{w,j}) =$$

$$\frac{1}{Z_{X_{w,j}}} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} f_x & 0 & c_x & 0 \\ 0 & f_y & c_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} T_{w,c_i} X_{w,j} \quad (7)$$

其中: $u_{i,j}$ 为第 j 个地图点在第 i 个关键帧上的观测; CamProj 计算了地图点在已知相机位姿上的投影,其定义为式(7),其中 f_x, f_y, c_x, c_y 为相机内参。当 KINECT 到达一个新的区域,那么新的关键帧就会添加到地图中。添加关键帧之后,当前关键帧就会和之前的关键帧建立关联并初始化新的地图点。

最初,地图点和关键帧是冗余的,那么需要进一步更为严格地筛选它们。出现以下情况的地图点就会被剔除:1)这些点在接下来的图像帧中无法跟踪和匹配;2)投影光线通过三角化计算点处于低视差;3)三角测量点会产生的重投影误差较大。这个严格地图点筛选过程保证了构建地图的鲁棒性。为了保证重构地图的简洁性,降低 BA 过程中的复杂度,检测冗余关键帧并删除它们,在关键帧集中,如果一个关键帧的 90% 的地图点至少在其他的三个关键帧中被检测到,那么则剔除该关键帧。

1.3 回环检测

为了降低视觉里程计过程中的积累误差,利用各关键帧的相互关联性进行有效的闭合回环,在本质视图上优化位姿。这样就可以将累计的误差分散到姿态图中,并通过相似变换矫正尺度偏移。通过回环检测优化位姿之后,根据优化后的关键帧矫正地图点云。首先计算关键帧 K_i 的词袋和其数据关联视图附近关键帧的相似度,并提出相似度较低的关键帧,同时删除和 K_i 直接连接的关键帧,最终获得闭合回环。然后当闭合回环达到一定程度时,利用通用图优化框架(General Graph optimization, G2o)优化回环位姿图,如图3所示。最后更新地图点,融合重复的地图点。位姿图优化公式如下:

$$\arg \min_{X_{w,i}, X_{w,j}} \sum (e_{i,j}^T \Delta_{i,j} e_{i,j}) \quad (8)$$

其中 $e_{i,j}$ 由式(9)计算获得:

$$e_{i,j} = \log_{\text{sim}(3)}(X_{i,j} X_{w,j} X_{w,i}^{-1}) \quad (9)$$

其中: $X_{i,j}$ 是位姿 $X_{w,j}$ 到位姿 $X_{w,i}$ 的相似变换, $\log_{\text{sim}(3)}$ 将转移矩阵的位姿误差映射到 7 维欧氏空间 \mathbf{R}^7 , $\Delta_{i,j}$ 为边的信息矩阵。

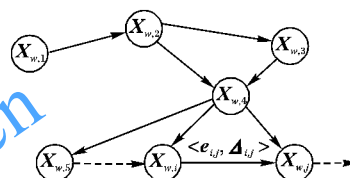


图3 回环检测位姿图

Fig. 3 Pose graph of loop detection

2 三维稠密地图快速构建

在 ORB-SLAM 系统中,其构建的是三维稀疏特征地图,分辨率极低,无法在实际的机器人领域中应用,如机器人导航、路径规划等。为了克服这一不足之处,本文提出了利用深度摄像头 KINECT 作为传感器设备来构建三维稠密地图。如果将每一帧的点云都融合到地图中,那么地图的容量会很大,从而降低了系统实时性能。由于机器人在运动过程中相邻图像帧具有连续性,即相邻帧的位姿在空间上变化较小,提出了空间域上的关键帧提取方法来筛选合适的图像帧,并引入子地图法进一步减少构建地图的时间,由此来提高算法的速度。

KINECT 摄像头由彩色相机和深度传感器组成,其中深度传感器由红外发射器和红外接收器两部分构成,如图4所示。彩色相机可以获得每个像素点的 RGB 值即彩色图像,而深度传感器则可以测量到像素点的距离信息即深度图像,根据 KINECT 成像原理,如图5所示,融合彩色图和深度图像计算获得三维点云。

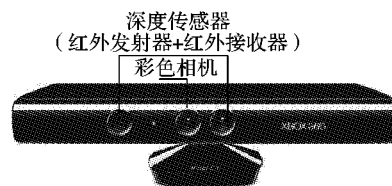


图4 KINECT 结构

Fig. 4 Structure of KINECT

从二维彩色图像和深度图像计算三维点云的公式如下:



$$\begin{cases} z = d/s \\ x = (u - c_x) \cdot (z/f_x) \\ y = (v - c_y) \cdot (z/f_y) \end{cases} \quad (10)$$

其中: f_x, f_y, c_x, c_y 为相机内参, (u, v) 为图像坐标, (x, y, z) 为图像坐标系坐标, d 深度相机测得像素点的距离, 单位毫米 (mm), s 为实际距离和测得距离 d 的比例系数, 这里取 1000。

从相机坐标系点到全局坐标系点云的变换公式如下:

$$X_{w,j} = T_{w,c_i} X_{c_i,j} \quad (11)$$

其中: T_{w,c_i} 为第 i 个关键帧的位姿, $X_{c_i,j}$ 为在第 i 个关键帧坐标系上的点云, $X_{w,j}$ 是变换后获得在全局坐标系上的点云。

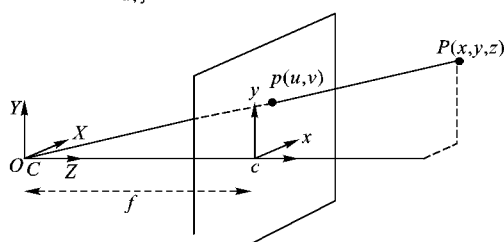


图5 针孔模型成像原理

Fig. 5 Imaging principle of pinhole model

为了避免三维点云的冗余,造成不必要的计算量,本文采用基于空间域上的方法来提取关键帧,其思想是只有当相机在相邻帧运动了一定大小时才把该帧视为关键帧,并把它点云叠加到现有的地图中去。其计算公式如下:

$$\min_norm \leq \|\Delta t\| + \min(2\pi - \|r\|, \|r\|) \leq \max_norm \quad (12)$$

其中: Δt 为相机在相邻帧之间的位移矢量; r 为相机在相邻帧之间的旋转角度,用它们的范数和来描述相机运动大小; \min_norm 为相机的最小运动,即当相机在相邻帧运动大于 \min_norm 时则把该帧提取为关键帧; \max_norm 为相机的最大运动,即当其运动大于 \max_norm 时,可认为是相机位姿估计错误,剔除该帧。在本文中根据实验经验, \min_norm 和 \max_norm 分别取值为 0.4, 5。

在大范围环境中建立三维稠密地图时,整个地图会相当大,占用大量的计算机内存,从而降低了系统的整体速度。本文提出了子地图法,将整个大地图分成多个包含一定关键帧的子地图。当地图规模达到了一定的程度即一定关键帧数量时,将该子地图从计算内存中释放出来,保存到计算机硬盘中。待系统将要关闭,则将多个子地图整合为完整的地图。本文的子地图规模为 10 个关键帧。

3 实验结果和分析

为了验证本文算法的有效性,进行了两组实验,包括:

1) 对比实验。利用 RGBD-SLAM 提供的标准数据集进行相机定位和地图构建实验,并和 RGBD-SLAM 进行对比。

2) 实际环境实验。利用 Turtlebot2 机器人在实验室楼道进行机器人定位和稠密地图构建实验。

下面分别是这两组实验的过程和结果分析。

3.1 对比 RGBD-SLAM

本节主要从算法的准确性和快速性两方面,将本文算法和 RGBD-SLAM 算法进行对比。

文献[13]提供了标准的 RGB-D 数据集,包括由 KINECT 获取的彩色图和深度图,由高精度动作捕捉系统获得的

KINECT 运动的真实位姿以及获取各个数据集的相机内参。除此之外,该文献还提供了计算估计位姿和真实位姿的均方根误差 (Root Mean Square Error, RMSE) 工具,这样可以有效地评估算法的定位精度。

本文主要利用 RGB-D 数据集中 FR1 组数据进行实验,估计 KINECT 相机运行轨迹并构建三维稠密地图。实验对比结果如表 1 所示,其中 RGBD-SLAM 实验结果来自文献[13],图 6 和图 7 分别是本文算法在 FR1 数据集上三维地图局部图、估计轨迹和实际轨迹对比图。

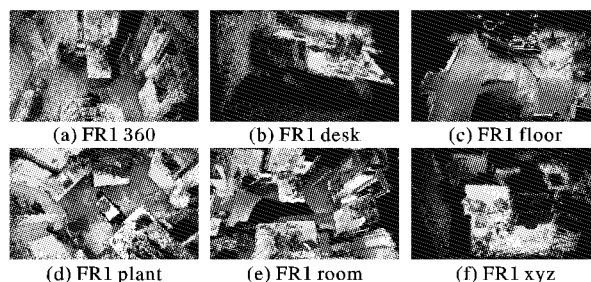


图6 部分 FR1 数据集三维地图

Fig. 6 3D maps of part FR1 dataset

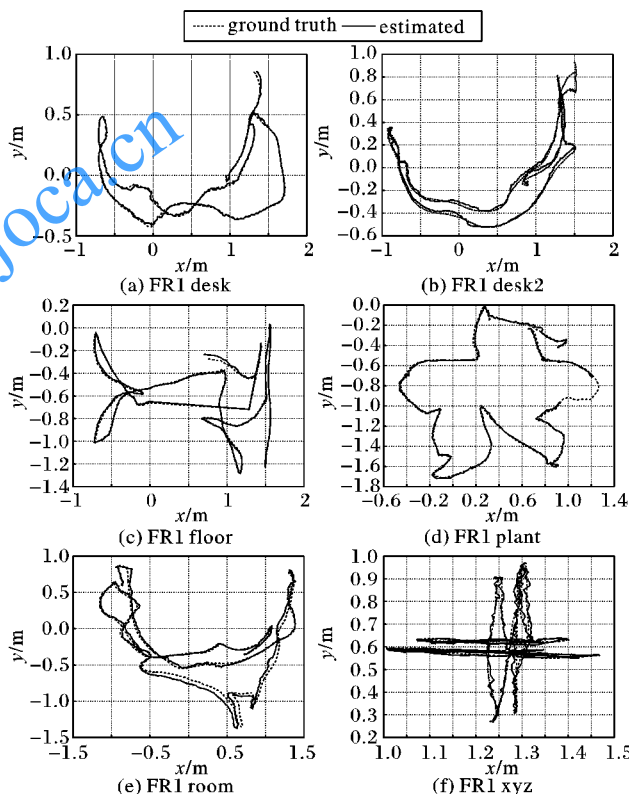


图7 FR1 数据集估计轨迹和实际轨迹对比

Fig. 7 Comparison between ground truth and estimated trajectory of FR1 dataset

从表 1 可以知道:1) 在定位精度方面,本文方法明显高于 RGBD-SLAM 算法。RGBD-SLAM 在 FR1 room 数据包上的定位误差最大,达到了 0.219 m,但是本文方法在这个数据包上的定位误差只有 0.045 m。RGBD-SLAM 在 FR1 数据集上平均定位误差是 0.097 m,而本文方法的平均定位误差为 0.038 m,仅为 RGBD-SLAM 的 40%;2) 在快速性方面,本文方法也明显优于 RGBD-SLAM 算法。RGBD-SLAM 在 FR1 数据集上平均每帧的处理时间是 0.349 s,而本文方法仅用了 0.132 s,速度约是 RGBD-SLAM 的 2.6 倍。



表1 本文算法与 RGBD-SLAM 性能对比

Tab. 1 Performance comparison between the proposed method and RGBD-SLAM

数据包	长度/m	帧数	每帧处理时间/s		RMSE/m	
			RGBD-SLAM	本文方法	RGBD-SLAM	本文方法
FR1 360	5.82	745	0.195	0.172	0.103	0.127
FR1 desk2	10.16	614	0.287	0.118	0.102	0.043
FR1 desk	9.26	575	0.346	0.141	0.049	0.015
FR1 floor	12.57	1214	0.402	0.069	0.055	0.015
FR1 plant	14.80	1112	0.381	0.155	0.142	0.018
FR1 room	15.99	1332	0.318	0.147	0.219	0.045
FR1 rpy	1.66	687	0.354	0.131	0.042	0.023
FR1 teddy	15.71	1395	0.399	0.142	0.138	0.045
FR1 xyz	7.11	788	0.463	0.114	0.021	0.009
平均			0.349	0.132	0.097	0.038

3.2 实际环境实验

为了进一步分析本文方法在实际室内机器人定位和三维稠密地图构建的效果,利用开源机器人平台 Turtlebot2 在实验室楼道进行了分析实验。Turtlebot2 机器人主要由 KOBUKI 移动底座、KINECT 深度摄像头和笔记本组成,其结构如图 8 所示。本文的 SLAM 算法是在 i5-3210M, 2.5 GHz CPU, 8 GB RAM, ubuntu14.04 操作系统的笔记本上用 C++ 实现的。



图8 Turtlebot2 机器人
Fig. 8 Turtlebot2 robot

实验室楼道由四段直线道组成,如图 9 所示,楼道存在大量的相似物体,如门、瓷砖等,并且瓷砖的角点数量较少,这些给特征匹配和回环检测带来了巨大的挑战。Turtlebot2 机器人上配备了 KINETCT 摄像头,并使得机器人坐标系和 KINECT 摄像头坐标系重合。实验过程中,用 XBOX 遥控手柄控制机器人以 0.4 m/s 速度水平运动,本文系统在线定位机器人位姿并构建三维稠密地图,最终获得机器人水平二维轨迹图(图 10)和三维地图(图 11、图 12)。

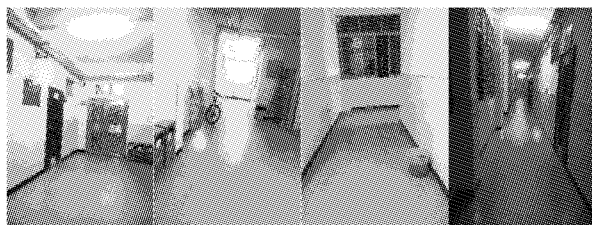


图9 四段楼道局部图
Fig. 9 Four sections of the corridor

由于缺少运动捕抓系统,无法确定机器人的真实位置,为了分析机器人的定位精度,把每段道路中的一点和终点选为测量点。由于机器人是水平运动,因此测量点的真实位置可以用二维坐标点 (x, z) 来描述。本文采用卷尺测量获得 5 个测量点的真实位置,最后用测量点的真实位置和算法估计位置的 RMSE 来度量定位精度。均方根误差计算方法 $RMSE = \left[\left(\sum_{i=1}^n (t_i - \bar{t}_i)^2 \right) / n \right]^{1/2}$, 其中 \bar{t}_i 、 t_i 分别为测量点的真实位置

和估计位置,在本文中 $n = 5$ 。五个测量点的真实位置、估计位置以及 RMSE 见表 2。机器人在楼道运动过程中位姿估计存在误差累计,5 个测量点的均方根误差为 1.04 m,机器人整个运动轨迹的距离为 50 m,即存在 2% 的误差。

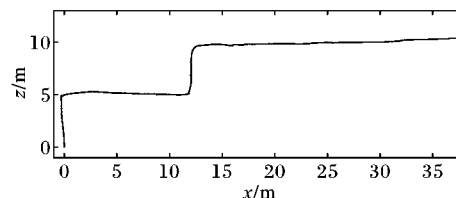


图10 机器人水平二维轨迹
Fig. 10 Horizontal 2D trajectory of the robot



图11 重构后四段路局部图
Fig. 11 Four sections of the corridor after reconstruction

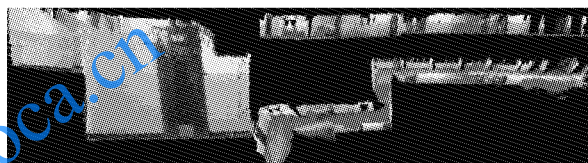


图12 构造的三维稠密地图
Fig. 12 Constructed 3D dense map

表2 测量点位置估计的结果

Tab. 2 Results on position estimation of the measured points

测量点序号	真实值/m	估计值/m	误差/m
1	(-0.3, 4.96)	(-0.3, 4.90)	0.0036
2	(8.13, 5.86)	(7.6, 5.1)	0.8500
3	(12.29, 9.34)	(12.1, 8.5)	0.6000
4	(30.7, 11.2)	(29.8, 10.5)	1.3000
5	(38.7, 11)	(37.1, 10.6)	2.7000

从图 11 和图 12 中可以看到,本文系统即使在大范围的环境下依然能够取得较好的构图效果,三维地图和实际三维环境相符,没有出现地图重叠和变形等情况,并且地图中楼道的反光瓷砖、地面和杂物都清晰可见。

在整个实验中,系统共处理了 1548 张视频帧,关键帧为 130 帧,整体速度为 11 帧/s,其中定位速度为 17 帧/s。结合上文对定位精度和三维稠密地图的效果分析,系统可以满足机器人同时定位和地图构建的精度和速度上的要求。

4 结语

为了建立高精度定位、大范围和快速的室内机器人定位和三维稠密地图构建系统,本文利用 ORB-SLAM 算法估计机器人三维位姿,然后采用深度摄像头 KINECT 获得每一帧的稠密点云,为了剔除冗余的视频帧引入了空间域上的关键帧提取方法,并采用子地图法进一步减少地图构建的时间以提高整体速度。利用机器人 Turtlebot2 在实验室楼道进行了分析实验,结果表明,本文的方法能够在大范围环境中准确地定



位机器人位置,在运动轨迹为 50 m 的大范围中,机器人的均方根误差为 1.04 m,即误差为 2%;同时系统的整体速度为 11 帧/秒,其中定位速度达到 17 帧/秒,可以满足室内机器人在精度、大范围和快速性的要求。但系统仍存在不足,由于 ORB-SLAM 算法是基于 ORB 角点特征来实现位置识别和回环检测的,所以在角点特征较少的环境中机器人的定位效果较差,容易定位失败。今后将会结合基于图像像素的直接 SLAM 进一步提高 ORB-SLAM 方法的定位效果。

参考文献 (References)

- [1] DARDARI D, CLOSAS P, DJURIC P M. Indoor tracking theory, methods, and technologies [J]. *IEEE Transactions on Vehicular Technology*, 2015, 64(4): 1263–1278.
- [2] CHOI H, YANG K W, KIM E. Simultaneous global localization and mapping [J]. *IEEE/ASME Transactions on Mechatronics*, 2014, 19(4): 1160–1170.
- [3] CHEN Z, SAMARABANDU J, RODRIGO R. Recent advances in simultaneous localization and map-building using computer vision [J]. *Advanced Robotics*, 2007, 21(3/4): 233–265.
- [4] SMITH R, SELF M, CHEESEMAN P. Estimating uncertain spatial relationships in robotics [C]// *Autonomous Robot Vehicles*. New York: Springer, 1990: 167–193.
- [5] JULIER S J, UHLMANN J K. Unscented filtering and nonlinear estimation [J]. *Proceedings of the IEEE*, 2004, 92(3): 401–422.
- [6] 张毅, 汪龙峰, 余佳航. 基于深度信息的移动机器人室内环境三维地图创建 [J]. *计算机应用*, 2014, 34(12): 3438–3445. (ZHANG Y, WANG L F, YU J H. Depth-image based 3D map reconstruction of indoor environment for mobile robots [J]. *Journal of Computer Applications*, 2014, 34(12): 3438–3445.)
- [7] SCHMIDT A. The EKF-based visual SLAM system with relative map orientation measurements [C]// *Proceedings of the 2014 International Conference on Computer Vision and Graphics*. Berlin: Springer, 2014: 570–577.
- [8] SIM R, ELINAS P, LITTLE J. A study of the Rao-Blackwellised particle filter for efficient and accurate vision-based SLAM [J]. *International Journal of Computer Vision*, 2007, 74(3): 303–318.
- [9] BLANCO J L, GONZALEZ J, FERNANDEZ-MADRIGAL J A. An optimal filtering algorithm for non-parametric observation models in robot localization [C]// *Proceedings of the 2008 IEEE International Conference on Robotics and Automation*. Piscataway, NJ: IEEE, 2008: 461–466.
- [10] SIBLEY G. Relative bundle adjustment [J]. *Electronic Notes in Theoretical Computer Science*, 2009, 220(3): 976–982.
- [11] KLEIN G, MURRAY D. Parallel tracking and mapping for small AR workspaces [C]// *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality*. Washington, DC: IEEE Computer Society, 2007: 225–234.
- [12] MUR-ARTAL R, MONTIEL J M M, TARDÓS J D. ORB-SLAM: a versatile and accurate monocular SLAM system [J]. *IEEE Transactions on Robotics*, 2015, 31(5): 1147–1163.
- [13] STURM J, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems [C]// *Proceedings of the 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Piscataway, NJ: IEEE, 2012: 573–580.
- [14] GALVEZ-LÓPEZ 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.
- [15] STRASDAT H, DAVISON A J, MONTIEL J M M, et al. Double window optimization for constant time visual SLAM [C]// *Proceedings of the 2011 International Conference on Computer Vision*. Washington, DC: IEEE Computer Society, 2011: 2352–2359.
- [16] KUEMMERLE R, GRISETTI G, STRASDAT H, et al. G2o: a general framework for graph optimization [C]// *Proceedings of the 2011 IEEE International Conference on Robotics and Automation*. Piscataway, NJ: IEEE, 2011: 3607–3613.
- [17] LUCAS B, KANADE T. An iterative image registration technique with an application to stereo vision [C]// *Proceedings of 7th International Joint Conference on Artificial Intelligence*. San Francisco: Morgan Kaufmann, 1981: 674–679.

This work is partially supported by the National Natural Science Foundation of China (61573148), the Guangdong Provincial Science and Technology Major Project (2015B010919007).

HOU Rongbo, born in 1993, M. S. candidate. His research interests include simultaneous localization and mapping, three-dimensional reconstruction, computer version.

WEI Wu, born in 1970, Ph. D., associate professor. His research interests include robot control, intelligent control, pattern recognition, artificial intelligence.

HUANG Ting, born in 1991, M. S. candidate. Her research interests include computer version, robotic embedded system.

DENG Chaofeng, born in 1990, M. S. candidate. His research interests include robot control, intelligent control.

Piscataway, NJ: IEEE, 1993: 40–44.

This work is supported by the National Defense Key Laboratory Fund Program of China (9140C120402120C1208).

YANG Xuefeng, born in 1976, Ph. D., lecturer. His research interests include image restoration and super-resolution reconstruction, pattern recognition, computer vision.

CHENG Yaoyu, born in 1966, Ph. D., professor. His research interests include information acquisition and processing, image processing, measurement and control system.

WANG Gao, born in 1973, Ph. D., professor. His research interests include optoelectronic information processing.

(上接第 1433 页)

- [19] EMMANUEL J, MICHAEL B. An introduction to compressive sampling [J]. *IEEE Signal Processing Magazine*, 2008, 25(2): 21–30.
- [20] MICHAL A, MICHAEL E, ALFRED B. K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on Signal Process*, 2006, 54(11): 4311–4322.
- [21] YANG J Y, PENG Y G, XU W L, et al. Ways to sparse representation: an overview [J]. *Science in China Series F: Information Sciences*, 2009, 52(4): 695–703.
- [22] PATI Y C, REZAIFAR R, KRISHNAPRAS P S. Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition [C]// *Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers*.