

# 机场视觉移动服务机器人的设计与研究

赵小勇<sup>1</sup>, 章伦晖<sup>1</sup>, 王 勇<sup>1</sup>, 梁兵兵<sup>1</sup>, 张祖芳<sup>1,2</sup>

(1. 合肥工业大学 机械工程学院, 合肥 230009;

2. 合肥工业大学(马鞍山) 高新技术研究院, 安徽 马鞍山 243000)

**摘要:** 民航事业迅猛发展, 加上疫情突现, 机场中迫切需要一种无人的机器人移动服务平台, 以减少人力劳动和阻断病毒传播; 依据机场环境和功能要求, 对机场移动服务平台进行模块化设计与搭建; 平台集成 RGB-D 相机系统 Kinect 2.0, 采用张正友标定原理确定相机内外参数; 设计和实现移动平台的视觉 SLAM 整体架构, 利用 ORB 算法进行图像特征点的提取与匹配, 对邻帧图像进行估计与优化; 将平台样机投入实验场地, 设定粒子数为 50, 最小匹配得分为 30, 构建分辨率为每像素 0.02 m 的二维栅格地图和三维点云图; 该样机为机场提供一种以视觉传感器为基础的移动服务平台, 结合实际需求, 其他功能模块能快捷地集成于此, 发挥各自的服务作用。

**关键词:** 机场服务平台; 相机标定; 视觉 SLAM; ORB 算法

## Design and Research of Visual Airport Mobile Service Robot

Zhao Xiaoyong<sup>1</sup>, Zhang Lunhui<sup>1</sup>, Wang Yong<sup>1</sup>, Liang Bingbing<sup>1</sup>, Zhang Zufang<sup>1,2</sup>

(1. School of Mechanical Engineering, Hefei University of Technology, Hefei 230009, China;

2. High-tech Research Institute, Hefei University of Technology (Maanshan), Maanshan 243000, China)

**Abstract:** With the rapid development of civil aviation and the epidemic, there is an urgent need for an unmanned robot mobile service platform in the airport to reduce human labor and restrain the spread of virus. The airport mobile service platform is modular designed and built according to the airport environment and functional requirements. The platform is integrated with Kinect 2.0 camera system, which is calibrated by Zhang Zhengyou calibration principle to calculate the camera's internal and external parameters. The framework of visual SLAM (simultaneous localization and mapping) is designed and implemented. Then ORB (Oriented Fast and Rotated Brief) algorithm is used to extract and match the image feature points. Then, the images of adjacent frames are estimated and optimized. The platform prototype was put into the experimental site, by setting the number of particles as 50, the minimum matching score as 30, and the two-dimensional grid map with 0.02 m per pixel resolution can be obtained. The three-dimensional point cloud map was also constructed. The prototype provides a mobile service platform based on visual sensor for the airport. Combined with the actual demand, other functional modules can be integrated into this platform simply to play their respective service roles.

**Keywords:** airport service platform; camera calibration; visual SLAM; ORB

## 0 引言

民航运输作为现代五大交通运输方式之一, 占据较大的市场份额, 近年来民航业发展迅速, 中国逐渐从“民航大国”向“民航强国”转变<sup>[1]</sup>。随之产生的机场巨大客流量以及公众需求现状对机场服务装备提出了高要求<sup>[2]</sup>。结合当前新冠病毒具有人传人的特点, 疫情在全球肆虐, 而机场又是全球长途出行的重要场所, 迫切需要一种无人机场服务设备来代替人来提供基础服务, 保障机场工作人员的生命安全, 抑制直接传播的风险。

针对机场室内复杂广阔的环境, 服务移动平台在其中完成即时定位与地图构建是实现“自主”服务功能的前提, 而 SLAM (simultaneous localization and mapping) 技术能很好地解决这一问题<sup>[3]</sup>。该技术能为移动设备提供一种从未

知室内环境出发, 在没有环境先验信息的条件下, 通过自身携带的传感器, 在移动过程中获取周围环境信息的能力, 具体表现为移动设备的实时定位和环境地图构建。

早期 SLAM 技术基本采用滤波器方法, 依赖激光传感器实现, 经过多年发展成为比较成熟的方案。激光 SLAM 拥有高精度的优势, 但其设备成本过高, 应用场景小, 地图缺乏语义信息。随着芯片技术和视觉传感器的发展, VSLAM (visual simultaneous localization and mapping) 技术突飞猛进, 使其成为未来机器人研究领域的主流方向。按照传感器解决方案, 能细分出不同 SLAM 技术: 从 2007 年被 Davison<sup>[4]</sup>提出的第一个 Mono-SLAM 单目算法, 到 Stereo-SLAM 双目视觉算法, 再到基于深度相机的 RGB-D SLAM。

本文应对机场复杂环境, 模块化地设计和搭建一款基于视觉 SLAM 的自主移动机器人样机<sup>[5]</sup>。它采用一台标定

收稿日期: 2020-08-04; 修回日期: 2020-09-02。

作者简介: 赵小勇(1962-), 男, 山东海阳人, 硕士, 主要从事测试技术方向的研究。

通讯作者: 王 勇(1969-), 男, 安徽无为, 博士, 主要从事零件成形制造方向的研究。

引用格式: 赵小勇, 章伦晖, 王 勇, 等. 机场视觉移动服务机器人的设计与研究[J]. 计算机测量与控制, 2021, 29(3): 209-213, 219.

过内外部参数的 Kinect 2.0 深度相机作为传感器来进行环境数据采集, 在 ROS 系统中完成彩色图像与深度图像的特征点识别与匹配处理, 再利用 Gmapping 建图和 RViz 平台可视化显示。样机被投入到实验室场地进行试验, 构建出测试环境的二维地图和三维点云图。

### 1 机场服务机器人平台

为了满足机场复杂环境的要求, 保障避障和语音服务等功能的实现, 整个移动平台系统采用模块化设计, 包括: 电机驱动模块、电源模块、通讯模块、红外避障模块、视觉模块、语音模块、循迹模块等。这些模块各司其职, 在驱动程序和视觉算法的控制下实现完整的系统功能, 设备的整体系统原理如图 1 所示。

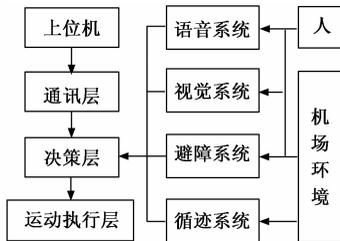


图 1 视觉移动平台系统运行原理框图

为保证原理样机移动底盘能在多变的机场环境中, 向任意方向平移与旋转, 执行层的轮胎选用三英寸麦克纳姆轮, 它能使机器人在移动过程中保持稳定。采集麦克纳姆轮的运动信息来估计移动机器人的位姿, 增加机器人运动模型里程计数据的采集, 保证移动平台后续能为 SLAM 里程计计算提供位移补偿的数据支持, 从而能在没有可视化数据的前提下预测摄像机位姿, 这比通过概率模型得到的预测值更加准确。装配各系统部件和结构框架平台, 最后的移动平台样机实物如图 2 所示。

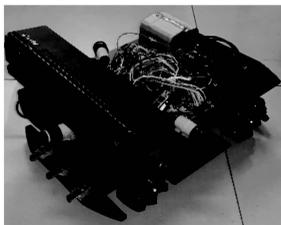


图 2 机场视觉 SLAM 移动平台原理样机

### 2 相机数据获取及标定

与电磁导引相比, 视觉相机传感器可适用于未知的动态环境, 不局限于已知环境的循迹导航; 与单目视觉传感器相比, RGB-D 相机可以同时获取目标的彩色图像信息和深度图像信息; 与双目视觉传感器相比, 双目系统获取深度信息是通过间接解析的方式, 导致计算量大, RGB-D 相机能通过红外结构光或 TOF 原理, 直接发射和接收光信号, 测量出相机和物体的距离, 从而大大减少获取深度信息的计算量, 测距精度更高, 但容易受到阳光的影响, 更

适合在机场室内这种环境中使用。

Kinect 2.0 相机能调用红外发射器主动发射近红外电磁波, 当投射到粗糙物体或是穿透毛玻璃后, 会发生漫反射, 形成随机的反射斑点, 深度摄像头则通过读取散斑的信息, 创建探测范围内的深度图像, 而后融合在彩色图像中, 丰富场景中像素点的深度信息。

#### 2.1 相机的标定

相机的彩色摄像头和深度摄像头在位置结构上有偏移, 为确保相机获取的彩色图像和深度图像像素点一一对应, 需要通过标定得知相机的内外部参数。

本文采用张正友<sup>[6]</sup>教授提出的单平面棋盘格相机标定法, 首先利用图像之间的特定关系来对相机参数进行初步估计, 然后通过优化函数对所构造参数进行迭代优化处理。该方法的原理性质介于传统标定法与自标定法之间。设定棋盘格所在平面为世界坐标系  $Ox_w Y_w Z_w$  中,  $Z_w=0$  处, 各点坐标表达为  $(X_w, Y_w, 0, 1)$ , 通过相机坐标系与世界坐标系的坐标变换:

$$Z_c \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A[r_1 \ r_2 \ r_3 \ T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = A[r_1 \ r_2 \ T] \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \quad (1)$$

令  $P = [p_1 \ p_2 \ p_3] = \lambda A [r_1 \ r_2 \ T]$ , 其中,  $\lambda$  为比例因子,  $p_1, p_2$  来自  $P = [p_1 \ p_2 \ p_3]$  的  $3 \times 3$  单应性矩阵, 每幅图像存在如下关系:

$$\begin{cases} p_1^T A^{-T} A^{-1} p_2 = 0 \\ p_1^T A^{-T} A^{-1} p_1 = p_2^T A^{-T} A^{-1} p_2 = I \end{cases} \quad (2)$$

其中:  $A$  为相机内部参数矩阵:

$$A = \begin{bmatrix} \alpha & \gamma & u \\ 0 & \beta & v \\ 0 & 0 & 1 \end{bmatrix} \quad (3)$$

不妨令:

$$H = A^{-T} A^{-1} - I = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (4)$$

根据 (2) 式求解相机的内部参数  $A$  如下:

$$\begin{cases} v = \frac{h_{12}h_{13} - h_{23}h_{11}}{h_{11}h_{22} - h_{12}^2} \\ \lambda = h_{33} - \frac{[h_{13}^2 + v(h_{12}h_{13} - h_{23}h_{11})]}{h_{11}} \\ f_x = \sqrt{\lambda/h_{11}} \\ f_x = \sqrt{\frac{\lambda/h_{11}}{h_{11}h_{22} - h_{12}^2}} \\ Z_c = -h_{12}f_x^2 f_y / \lambda \\ u = sv/f_y - h_{13}f_x^2 / \lambda \end{cases} \quad (5)$$

其中:  $\lambda$  为比例因子;  $Z_c$  为像素点对应的深度值。联立公式 (5) 及单应性矩阵, 求得相机的外部参数如下:

$$\begin{cases} r_1 = \lambda A^{-1} p_1 \\ r_2 = \lambda A^{-1} p_2 \\ r_3 = r_1 \cdot r_2 \\ T = \lambda A^{-1} p_3 \end{cases} \quad (6)$$

联立红外摄像头相对于彩色摄像头外部参数的旋转矩阵  $R$  和平移向量  $T$  为:

$$\begin{cases} R = R_{ir} R_{rgb}^{-1} \\ T = T_{ir} - R_{ir} R_{rgb}^{-1} T_{rgb} \end{cases} \quad (7)$$

依照两个摄像头之间的坐标关系变换, 完成 Kinect 相机的配准。

### 2.2 标定实验和参数获取

本实验使用的标定模板为  $7 \times 10$  的棋盘格, 每个棋盘格边长为 30 mm。棋盘标定模板在世界坐标系中为固定状态, 采用 A3 图纸, 采集图像时, 将棋盘平整的贴在固定的墙面上。通过 Windows 系统下的“Color Basics-WPF”模块和“Infrared Basics-WPF”模块采集彩色和深度图像。

操作时, 通过改变 Kinect 摄像机与棋盘之间的相对位置来采集不同的彩色图像和红外图像。每次采集图像时, 要保证 Kinect 相机与固定棋盘格之间的相对位置尽可能不同, 两者角度尽量控制在  $45^\circ$  以下, 且彩色图像和深度图像要一一对应。分别采集彩色图像与红外图像各 49 张, 再将两组照片分别导入到程序中标定计算。如图 3 和图 4 所示, 分别为对应的 8 张彩色图像与红外图像的角点提取结果, 方形为标记的棋盘坐标原点, (高亮处理) 圆圈为棋盘上被检测到的特征点, 加号代表特征点的重投影位置, 提取角点之后可分别对两个图像集进行标定操作。

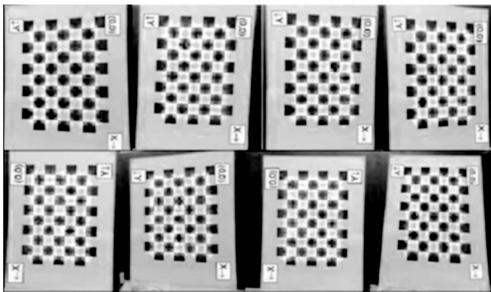


图 3 彩色图像中角点提取结果

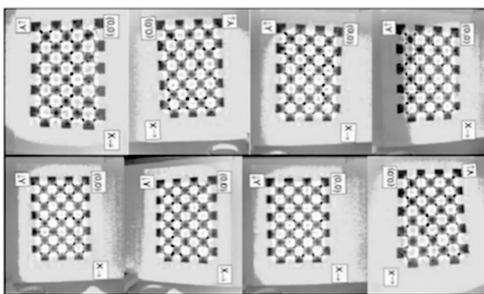


图 4 红外图像中角点提取结果

将摄像机与 49 幅标定图像的相对位姿表达在坐标系中, 以相机为静止参照物, 如图 5 所示, 样本图像的采集

容量足够大。

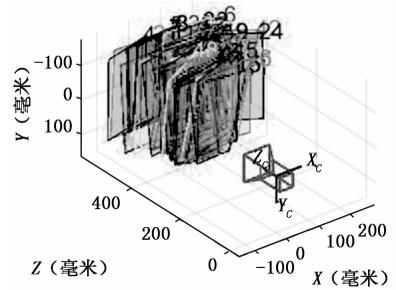


图 5 Kinect 2.0 与标定图像相对位姿

经过标定所得摄像头内部参数如表 1 所示。彩色图像中只有一张图像的误差超过 1, 平均投影误差为 0.485 0, 红外图像的平均投影误差仅为 0.213 3, 平均投影误差均低于 0.5, 由于样本容量足够大, 说明该标定实验的内部参数是精确的。

表 1 Kinect 2.0 内部参数

参数	彩色摄像头	红外摄像头
焦距 $(f_x, f_y)$	(107.25, 107.36)	(368.786 6, 368.616 3)
中心点 $(c_x, c_y)$	(969.864 4, 528.235 6)	(261.907 2, 201.593 0)
不对称度	0	0
径向畸变 $(k_1, k_2)$	(0.080 1, -0.075 1)	(0.129 1, -0.290 2)
切向畸变 $(k_3, k_4)$	(0, 0)	(0, 0)
平均投影误差	0.485 0	0.213 3

分别用彩色摄像头和深度摄像头对相同位姿的标定板求外参数, 结合两者结果, 带入公式 (7), 求得红外摄像头与深度摄像头之间坐标变化的旋转矩阵和平移向量分别为  $R$  和  $T$ :

$$R = \begin{bmatrix} -1.0595 & 0.0124 & -0.0002 \\ 0.0178 & 1.0445 & -0.0264 \\ 0.0093 & 0.0019 & 0.9971 \end{bmatrix} \quad (8)$$

$$T = [24.995 \quad 23.331 \quad -2.346]^T \quad (9)$$

### 3 视觉 SLAM 设计

设计搭建视觉 SLAM 的整体框架, 完成视觉 SLAM 五大主要步骤: 采集图像、视觉里程计、回环检测、位姿图优化和地图构建<sup>[7]</sup>。

#### 3.1 SLAM 整体架构

准确的建图来自精确的定位, 精确的定位来自准确的建图, 二者互相依存, 不可分割<sup>[8]</sup>。如图 6 所示为本文所构建的机场服务机器人视觉 SLAM 算法整体架构。

服务机器人平台处在机场动态环境中, 从一个陌生的地点出发, 在移动过程中通过 Kinect 2.0 相机获取环境的彩色图像和深度图像。在完全未知的环境下得到地图数据, 这涉及到环境特征的提取问题, 在视觉 SLAM 的前端线程中, 首先利用 RGB-D 传感器数据, 采用 ORB 算法完成图

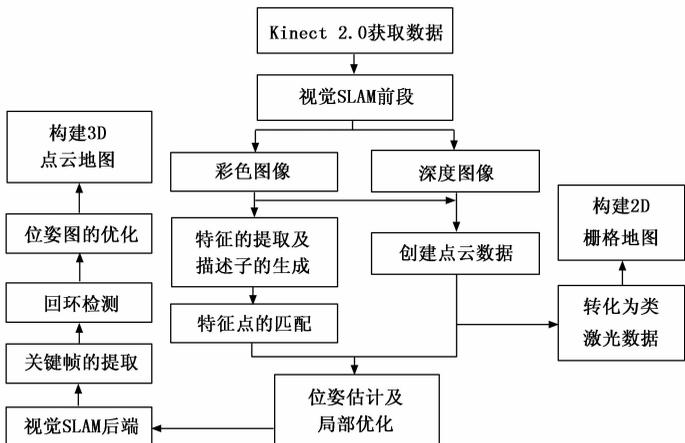


图 6 机场服务机器人视觉 SLAM 流程图

像特征点的提取，描述子的生成以及特征点的匹配等；然后采用 P3P 运动估计方法对移动机器人的运动位姿进行估计；最后采用非线性优化方法对运动估计值进行局部优化。在后端线程中，首先提取一些特殊的图像帧作为关键帧，对关键帧进行回环检测以实现对位姿图的优化，构建三维点云图。回环检测可有效减少机场服务机器人在移动过程中的累积误差，从而在视觉 SLAM 的实现过程中得到精确度更高的运动轨迹。而深度数据点云经计算能得到类激光数据，构建二维地图。

### 3.2 SLAM 前端设计

前端部分采用 ORB 算法，主要对观察对象进行特征点的提取和实现特征点的匹配，建立起相邻两帧之间对应的三维点云集合，进而对机器人的运动状态进行估计和优化。

ORB 算法性能强大，计算速度是 SIFE 算法的 100 倍，是 SURF 算法的 10 倍，其中的 FAST 算法负责完成图像特征点的检测和提取。设定算法中特征点判定阈值  $T$  为扫描像素点亮度的 30%，利用 FAST-12 筛选图片中符合条件的像素角点，得出合适的特征点，而后利用灰度质心法为 FAST 算法检测出的特征点赋予主方向。

完成图像特征点的提取后，还需要通过某种形式来描述特征点，特征点属性的输出称之为该特征点的描述子<sup>[9]</sup>，ORB 采用 BRIEF 算法来计算特征点的描述子，本文设定最大描述子提取数为 500，金字塔抽取率 1.2，edgeThreshold 和 patchSize 参数同为 31，BRIEF 描述符生成的随机像素数量为 3，以此调教出适用于本次实验环境的计算参数。此参数程序在实验场地的实现效果如图 7 所示，图中高亮圆为通过 ORB 算法生成图像特征点的描述子。

特征点的匹配能确定出 Kinect 相机当前观察到的陆标与之前观察到的陆标的对应关系。特征点的匹配问题可通过描述子的差异性进行比较判断，两副图像描述子通过汉明距离公式来进行比较：

$$d(a_1, a_2) = a_1 \oplus a_2 \quad (10)$$

汉明距离公式把两个描述子的二进制串的位置做异或运算，异或值越小，两个特征点之间的相似度就越高，若

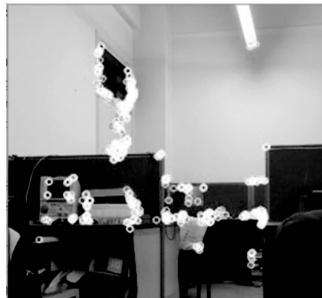


图 7 图像描述子的生成

异或值  $d$  不超过设定的汉明距离值，则认为两个特征点匹配成功，继而得到两幅图像对应特征点的坐标。但计算量为两幅图像特征点数乘积，为减少匹配计算量，对其进行随机抽样，采用双向匹配交叉过滤方法进行第一次筛选，把明显错误的匹配剔除，减少迭代次数，再采用 RANSAC 算法迭代剔除此匹配算法的外点，最后排序输出实现改进后的 ORB 特征匹配算法，流程图如图 8 所示。

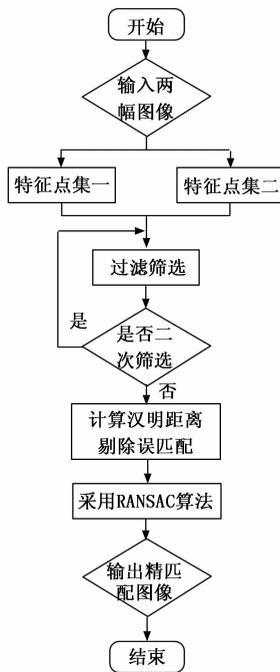


图 8 改进后的 ORB 特征匹配算法流程图

使用改进后的 ORB 特征点匹配算法对两幅图像进行特征匹配，匹配结果如图 9 所示，没有出现误匹配点对，为移动机器人求取运动估计值提供一个良好的初始值。

最后采用 P3P 算法对相邻帧间 Kinect 的六自由度位姿进行估计，并使用非线性优化方法对相机运动的估计值进行优化。

### 3.3 后端设计

随着图像帧数的增加，单纯计算相邻帧的匹配关系来估计相机位姿，将导致位姿估计过程中的误差累积。运用图优化理论对移动机器人的估计位姿进行优化，可有效减小累积误差，回环检测即通过某种手段让 Kinect 相机检测



图 9 改进后的 ORB 算法特征点匹配

到曾经到达过的场景, 该方法能够显著减少累积误差, 获得全局一致的估计值。

回环检测的方法主要是采用视觉词袋<sup>[10]</sup>模型 (Bag-of-Word, BoW) 进行处理。视觉词袋的构建可以看作是描述子聚类的过程, 将提取到的关键帧全部进行特征的检测和提取, 从而完成对视觉词袋的构建。回环检测过程中, Kinect 2.0 相机采集到的图像被处理后, 图像特征将转化成词袋向量, 通过词袋向量来比较当前帧与候选帧之间的相似度。然后引入归一化的先验相似评分进行相似度的评判, 排除错误干扰, 保留真实的回环。最后利用回环检测两个标准: 准确率 (Precision) 和召回率 (Recall) 对回环检测的效果进行判断。

在图优化方面, 本文采用只考虑 Kinect 相机运动位姿边约束的图优化模型, 位姿图中添加的回环约束边越多, 图优化模型就会变得越健壮。待位姿图构建完毕, 将所有位姿节点与边约束一起进行优化, 将得到的位姿误差  $e_{ij}$  进行最小化处理, 从而使得整体的位姿估计更符合实际。

至此, 完成本课题视觉 SLAM 算法整体框架的设计, 前端线程保证机场服务机器人视觉检测可以稳定的实现, 后端线程则保证了 Kinect 2.0 构建的地图更精确。

## 4 实验结果与分析

利用该原理样机的嵌入式 ROS 系统驱动 Kinect 2.0 相机, 在实验室等场景中进行机场服务机器人的视觉 SLAM 的运行, 创建二维栅格地图和三维点云地图, 验证方案可行性。

### 4.1 2D SLAM 实验

通过 iai-Kinect 2.0 接口保证工具库 Libfreenect2 在 Ubuntu16.04 系统下驱动 Kinect 2.0。利用 Gmapping 功能包创建实验室和实验室外部走廊的二维栅格地图, 并使用 RViz 可视化平台显示。

Gmapping 使用的是粒子滤波算法, 粒子不断迭代更新, 选取一个合适粒子数, 在保证算法准确性的前提下, 同时可以提高算法的运算效。设定计算参数: 粒子数为 50, 最小匹配得分为 30, 地图分辨率为 0.02 m, 保证运算效率和算法的准确性。将样机投入室内实验室环境采集图像数据, 得到其 2D 栅格地图如图 10 所示。

如图 11 所示, 将移动机器人平台移至室内走廊环境进行实验。设置相同的计算参数, 得到走廊环境 2D 栅格地图如右图所示。

对比实验地图结果与实际采集环境, 在结果方面, 实

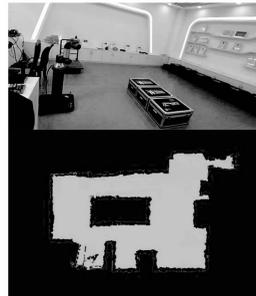


图 10 实验室环境与其栅格地图

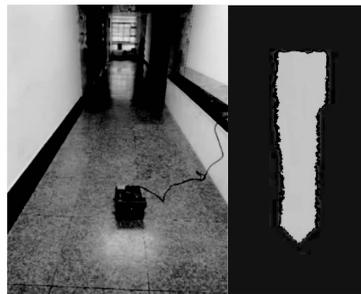


图 11 走廊环境与其栅格地图

验程序能鲜明地分割墙体与道路区间, 给出环境稠密地图的通过信息, 回环检测发挥能效, 地图可行域边界轮廓实现充分闭合; 在误差方面, 地图边界会出现轻微弯曲, 墙面存在噪点误差, 墙角等特征也与真实环境存在差异。分析实验可知误差来源于移动平台本身的运动误差, 导致视觉里程计误差的存在; 环境反射材质的多变, 导致像素点深度信息采集误差的存在。但构图实验仍能高度还原场景的环境信息, 为地图导航提供的信息基础。

### 4.2 3D SLAM 实验

该实验采用手持相机绕实验场地运动的方式采集数据, 通过 g2o 图优化器解决图优化 BA (Bundle Adjustment) 问题, 得到全局一致的运动轨迹, 进而得到误差较小的三维点云地图, 如图 12 为所构建的实验室三维点云图。



图 12 实验室点云图

手持移动轨迹的抖动必然存在, 轨迹算法会将其平滑化, 如图 13 所示, 为视觉算法反求的 Kinect 2.0 相机坐标系移动轨迹。图 12 的多帧三维点云数据实现拼接, 地图效果反应实验室环境中的物体信息, 但点云密度有待提高。

本实验验证了, 在室内环境中, 利用 Kinect 2.0 相机实现 2D SLAM 和 3D SLAM 建图的有效性。

(下转第 219 页)