

面向室内动态场景的多传感视觉 SLAM 方法

张建华, 张天晶, 赵 岩, 张 霖, 周 浩

河北工业大学机械工程学院, 天津 300132

基金项目: 国家重点研发计划(2018YFB1306500); 河北省自然科学基金(F2017202062, F2020202053); 天津市自然科学基金(19JCJQJC61600)

通信作者: 赵岩, yzhao@hebut.edu.cn 收稿/录用/修回: 2021-09-22/2021-12-31/2022-04-05

摘要

针对传统同时定位与地图构建(simultaneous localization and mapping, SLAM)框架面临动态场景时产生明显定位误差, 建立的场景稠密地图会包含动态对象及其运动叠影, 从而导致定位与建图鲁棒性不足的问题, 面向以人类为主要动态对象的室内动态场景, 从“温度”的角度出发, 提出基于热像仪与深度相机结合的多传感 SLAM 协同方案, 解决室内动态场景中的定位与建图难题。首先, 建立一套针对热像仪与深度相机的联合标定策略, 重新设计标定板与标定方案, 完成相机的内参标定、外参标定与图像配准, 得到一一对应的 RGB、深度、热(RDH)三模图像; 其次, 由热图像得到人体掩模图像, 进而在 ORB-SLAM2 系统框架下构建静态特征提取与数据关联策略, 实现基于三模图像的视觉里程计; 然后, 基于人体掩模图像更新深度图像, 滤除人体区域, 进而完成基于三模图像的静态环境稠密地图构建; 最后, 在室内动态场景下进行实验验证, 结果表明所提出算法在室内动态场景下可有效剔除动态对象的干扰特征, 相对传统 SLAM 算法具有明显优势。

关键词

动态场景
视觉 SLAM
热像仪
RDH 三模图像
多传感协同
中图法分类号: TP242.6
文献标识码: A

Multi-sensor Visual SLAM Method for Indoor Dynamic Scenes

ZHANG Jianhua, ZHANG Tianjing, ZHAO Yan, ZHANG Lin, ZHOU Hao

School of Mechanical Engineering, Hebei University of Technology, Tianjin 300132, China

Abstract

Owing to the localization error associated with traditional SLAM frameworks used in dynamic scenes, the established scene dense map contains dynamic objects with overlapping motions, leading to insufficient robustness of localization and mapping. For indoor dynamic scenes with humans as the main dynamic object, and from the perspective of “temperature”, we propose a collaborative scheme of multi-sensor SLAM using a combination of the thermal imaging system and depth camera to tackle the issue of positioning and mapping in indoor dynamic scenes. First, we establish a set of joint calibration strategies for the thermal imager and depth camera, redesign the calibration plate and scheme to complete the internal parameter calibration, external parameter calibration of the camera and image registration, and obtain the one-to-one corresponding three-mode images of RGB, depth, and heat (RDH). Second, we use the thermal image to obtain the mask image of the human body. Then, we construct the static feature extraction and data association strategy under

Keywords

dynamic environment;
visual SLAM;
thermal imaging system;
RDH three-mode images;
multisensing synergy

the framework of the ORB-SLAM2 system to realize the visual odometry using a three-mode image. The body region is filtered based on the updated depth image, and the static environment density map based on the three-mode image is constructed. Finally, the experiments are performed in a dynamic indoor scene, and the results indicate that the proposed algorithm effectively eliminates the interference features of dynamic objects in a dynamic indoor scene and has obvious advantages over traditional SLAM algorithms.

0 引言

SLAM 作为机器人学的一项关键技术,常被用于解决机器人在未知环境中的感知及定位问题^[1]。根据传感器不同,SLAM 主要分为激光 SLAM^[2]和视觉 SLAM^[3],其中视觉 SLAM 在环境感知方面优势明显,常常作为机器人环境感知的重要手段。动态环境是现实生活中最常见的情景,然而传统 SLAM 算法无法准确区分动态环境中的运动对象,从而影响环境感知的结果。因此,探索适用于动态环境的视觉 SLAM 算法对推进机器人的发展具有积极作用。

传统的视觉 SLAM 算法是在静态世界基本假设的前提下发展起来的,SLAM 的概念最早是由 Smith 等^[4]于 20 世纪 80 年代提出的。视觉 SLAM 表现出强大的环境感知能力。Klein 等^[5]首次将特征点法应用于 SLAM 系统并提出了 PTAM(parallel tracking and mapping),开创性地提出跟踪线程和建图线程,该系统第一次将非线性优化代替传统的滤波器应用于系统优化后端,为非线性优化作为视觉 SLAM 后端开创先河。针对 PTAM 存在的跟踪易丢失问题,萨拉戈萨大学的 Raul 等^[6]在 PTAM 的基础上于 2015 年提出了 ORB-SLAM,之后作者于 2017 年推出 ORB-SLAM 的迭代版本 ORB-SLAM2^[7],支持单目、双目、RGB-D(color-depth)等多种相机并且增加了地图复用功能,在 29 个公开数据集上得到的结果在当时具有最高的准确性。Engel 等^[8]提出了一种基于稀疏直接结构和运动公式的视觉里程计 DSO(direct sparse odometry)系统,以更整体、更优雅地方式处理了数据关联问题,但对场景光照变化敏感,适用于相机运动平缓、光照条件稳定的场景。

国内外针对动态环境下如何进行鲁棒定位与建图已经做了大量研究,目前的主流研究方案致力于消除动态对象的影响。中国人民解放军陆军工程大学的 Ai 等^[9]提出了一种能够在高动态场景下鲁棒运行的 DDL-SLAM(dynamic deep learning SLAM),得到仅包含静态区域的 RGB 图像及对应的深度图像,但实时性有待提高。萨拉戈萨大学的 Bescos 等

在 ORB-SLAM2 的基础上提出 DynaSLAM^[10],算法具备较好的动态目标检测和背景修复能力,但是不能达到实时性要求。Rünz 等^[11]提出以 RGB-D 数据流作为输入的 Co-Fusion 系统,使得该系统能生成一个可与环境交互的环境地图,前提需要模型的先验知识。随后,Rünz 等^[12]提出一个基于语义场景理解来建模和跟踪多个目标的 MaskFusion 系统,实现实例级对象重建,但是该系统对于小物体的跟踪和建模不够理想,在分割分类阶段没有考虑分类错误的情况。北京大学 Zhong 等^[13]提出 Detect-SLAM 系统,利用深度神经网络 SSD(single shot detector)构建目标检测器判断动态对象并移除其特征,改善 SLAM 的准确性和鲁棒性并以在线的方式构建环境的实例级语义地图。Henein 等^[14]提出了一种 Dynamic SLAM 算法,该算法利用语义分割来估计场景中刚性物体的运动,无需估计物体姿态或者 3D 模型先验知识。Yu 等^[15]提出了一种面向动态环境的鲁棒语义可视化的 DS-SLAM(dynamic semantic SLAM),将语义分割网络和运动一致性检查方法相结合,减小了动态对象的影响;Scona 等^[16]提出了一种 RGB-D SLAM 方法,在动态环境下检测运动目标的同时重构背景结构的鲁棒稠密建图方法。Rebecq 等^[17]提出了一个基于事件的视觉里程计(event-based visual odometry, EVO)算法,该方法利用了事件相机的突出特性来跟踪快速相机的运动,同时恢复半透明的三维环境地图;此外,还有 EM-Fusion^[18](expectation maximization fusion)、BaMVO^[19](backgroundmodel-based VO)、RO-SLAM^[20](range-only SLAM)、FlowFusion^[21]、ReFusion^[22]方法。现有 SLAM 算法面临动态场景时,存在着实时性和鲁棒性不足的问题。

人类与周围环境相比具有明显不同的特性,即温度不同。基于此,面向以人类为主要动态对象的室内场景,本文提出了基于热像仪与深度相机的多传感协同方案,以解决动态场景中的定位与建图问题。算法采用热像仪与 RGB-D 深度相机构建多传感视觉系统。首先,进行多传感视觉系统联合标定与图像配准;然后,基于热像图实现人体分割提取人体

掩模, 深度图像经人体掩模图像处理后便可同 RGB 图像根据各帧对应的位姿生成点云; 最后, 对点云数据进行拼接、滤波、转换等处理, 生成最终的地图形态。

1 多传感觉系统标定与图像配准

由于热像仪和 RGB-D 深度相机属于异源传感器, 在用于动态场景中的定位与建图前, 需要对两个传感器构成的多传感觉系统进行严格的内参标定、外参标定与图像配准工作。

本文使用的 RGB-D 深度相机为 Kinect V2 深度相机, 包括彩色摄像头、红外发射器和红外接收器, 其中深度信息是通过投射的红外线返回时间来取得深度信息, 即可认为深度图像来源于红外图像, 两者显示的内容一致, 但由于深度图像无法显示标定板角点, 故在进行相机标定中使用红外图像进行标定。

1.1 标定方法

多传感相机标定算法采用的标定工具为 Matlab Camera Calibrator^[23], 其核心操作是需要对若干张棋盘格图像进行角点检测, 计算出相机的内参矩阵和畸变系数。

热像仪根据周围环境的热辐射情况进行成像, 选择“黑热”模式成像。由于普通棋盘格在热像仪镜头下看不到角点, 故制作表面哑光喷漆的不锈钢镂空棋盘格标定板, 格子尺寸为 30 mm × 30 mm, 数量为 7 × 4, 如图 1 所示, 其中黑色格子为镂空区域。

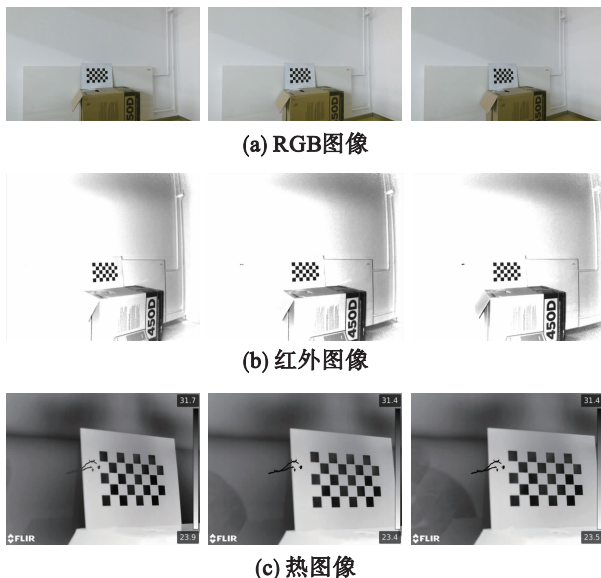


图 1 标定图像

Fig.1 Calibration images

将不锈钢标定板置于热源电暖气前方, 其间加

装一张黑色纸, 由于不锈钢和黑纸的厚度和传热系数不同, 短时间内标定板的格子区域与标定板之间产生较大的温差, 因此可以在热像仪镜头下产生棋盘状的图像, 如图 1(c) 所示。固定标定板, 多次改变传感器位置进行标定图像采集, 同步采集了 30 张标定板图像, 其中部分图像如图 1 所示。

1.2 热像仪与 RGB-D 相机内参标定

热像仪使用 FLIR 热像仪, 成像过程符合针孔模型, 相机内参矩阵为

$$\mathbf{K} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix} \quad (1)$$

式中, f_x 、 f_y 为使用像素来描述 x 轴、 y 轴方向焦距的长度, c_x 、 c_y 为主点的实际位置。

根据标定结果, 得到 RGB 镜头、红外镜头和热镜头的内参矩阵 $\mathbf{K}_{\text{rgb}}$ 、 $\mathbf{K}_{\text{d}}$ 、 $\mathbf{K}_{\text{t}}$ 为

$$\mathbf{K}_{\text{rgb}} = \begin{pmatrix} 1\ 047.232\ 2 & 0 & 951.520\ 4 \\ 0 & 1\ 048.059\ 7 & 541.544\ 7 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{K}_{\text{d}} = \begin{pmatrix} 369.930\ 8 & 0 & 252.920\ 2 \\ 0 & 369.901\ 8 & 200.105\ 7 \\ 0 & 0 & 1 \end{pmatrix} \quad (2)$$

$$\mathbf{K}_{\text{t}} = \begin{pmatrix} 2\ 205.483\ 1 & 0 & 483.346\ 6 \\ 0 & 2\ 208.447\ 5 & 350.071\ 7 \\ 0 & 0 & 1 \end{pmatrix}$$

由标定结果得到各传感器的径向畸变系数 k_1 、 k_2 , 具数值如表 1 所示。

表 1 各传感器的径向畸变系数
Tab.1 Radial distortion coefficient of each sensor

传感器	k_1	k_2
RGB 相机	0.013 5	0.839 9
深度/红外相机	-0.061 9	1.140 7
热像仪	-0.253 7	0.677 0

1.3 热像仪与 RGB-D 相机外参标定

热像仪、RGB-D 深度相机的相对位置如图 2 所示, 为了能够进行图像配准, 需要将它们置于同一坐标系下进行表示。用 $\mathbf{R}_{\text{d} \rightarrow \text{rgb}}$ 、 $\mathbf{t}_{\text{d} \rightarrow \text{rgb}}$ 分别表示 RGB 镜头相对于红外镜头的空间旋转矩阵及中心点间的平移向量, $\mathbf{R}_{\text{d} \rightarrow \text{t}}$ 、 $\mathbf{t}_{\text{d} \rightarrow \text{t}}$ 分别表示热镜头相对于红外镜头的空间旋转矩阵及中心点间的平移向量。

假设在红外相机归一化坐标系的空间点 $\mathbf{P}_{\text{d}}$, 它在红外相机像素平面上的投影点为 $\mathbf{u}_{\text{d}}$, 设该点在热像仪归一化坐标系下对应的空间点为 $\mathbf{P}_{\text{t}}$, 经过投影

得到热像仪像素平面上的投影点 u_t , 根据内参矩阵 K_t , 通过空间位置变换关系, 可得:

$$P_d = K_d^{-1} u_d \quad (3)$$

$$P_t = P_{d \rightarrow t} P_d + t_{d \rightarrow t} \quad (4)$$

$$u_t = K_t P_t \quad (5)$$



图2 多传感视觉系统

Fig.2 Multi-sensor vision system

假设世界坐标系下有空间点 P_w , 根据空间位置变换关系, 可以得到该点在红外相机坐标系下的空间点 P_d , 在热像仪坐标系下的空间点 P_t , 即:

$$P_t = R_t R_d^{-1} P_d - R_t R_d^{-1} t_d + t_t \quad (6)$$

式中, R_d 、 t_d 表示红外相机的当前位姿; R_t 、 t_t 表示热像仪的当前位姿。

对比式(4)、式(6), 可以得到:

$$R_{d \rightarrow t} = R_t R_d^{-1} \quad (7)$$

$$t_{d \rightarrow t} = t_t - R_t R_d^{-1} t_d = t_t - R_{d \rightarrow t} t_d$$

根据式(7)可知, 只要获取同一时刻下红外相机的位姿和热像仪的位姿, 就可以得到两者之间的空间位置关系 $R_{d \rightarrow t}$ 、 $t_{d \rightarrow t}$ 。同理可获得红外相机和 RGB 相机之间的空间位置关系 $R_{d \rightarrow \text{rgb}}$ 、 $t_{d \rightarrow \text{rgb}}$:

$$R_{d \rightarrow \text{rgb}} = R_{\text{rgb}} R_d^{-1} \quad (8)$$

$$t_{d \rightarrow \text{rgb}} = t_{\text{rgb}} - R_{\text{rgb}} R_d^{-1} t_d = t_{\text{rgb}} - R_{d \rightarrow \text{rgb}} t_d$$

现需要通过标定, 获得同一场景下 RGB 图像、红外图像和热图像分别对于标定板的位姿, 将得到的位姿通过式(7)、式(8)进行计算得到各传感器之间的相对位置。同一场景下各传感器获得的标定图像, 如图3所示。

对每帧标定图像, Matlab Camera Calibrator 可以给出其对应的相机位姿, 计算得到结果为

$$R_{d \rightarrow \text{rgb}} = \begin{pmatrix} 1.000 & 0.070 & -0.001 & 0.038 & -0.004 & 0.254 \\ 0.000 & 0.959 & 0.999 & 0.334 & -0.036 & 0.710 \\ 0.004 & 0.182 & 0.036 & 0.583 & 0.999 & 0.320 \end{pmatrix} \quad (9)$$

$$t_{d \rightarrow \text{rgb}} = (-52.705 \ 098 \ 6.049 \ 173 \ -23.092 \ 634)^T$$

$$R_{d \rightarrow t} = \begin{pmatrix} 1.000 & 0.034 & -0.006 & 0.201 & 0.006 & 0.771 \\ 0.006 & 0.207 & 0.999 & 0.987 & 0.007 & 0.662 \\ -0.006 & 0.879 & -0.007 & 0.658 & 0.999 & 0.939 \end{pmatrix} \quad (10)$$

$$t_{d \rightarrow t} = (102.255 \ 793 \ -122.786 \ 627 \ 17.521 \ 777)^T$$



图3 同一场景的标定图像

Fig.3 Calibrated images of the same scene

1.4 基于深度数据的三模图像配准

为了保证多传感器可以获得完全一致的图像, 找到3种模态图像之间的相同区域, 需要对来自不同传感器的原始图像进行配准处理, 得到像素级对应的具有相同图像内容的 RGB 灰度图像、红外图像和热图像。

假设深度图像中有一个像素点 u_d , 其深度值为 z_d , 可以得到该点在相机坐标系下的3维坐标 P_d :

$$P_d = z_d K_d^{-1} u_d \quad (11)$$

已知红外镜头和 RGB 镜头的空间关系 $R_{d \rightarrow \text{rgb}}$ 、 $t_{d \rightarrow \text{rgb}}$, 整理得到 RGB 图像与红外图像的像素点 u_{rgb} 、 u_d 之间的空间位置关系:

$$z_{\text{rgb}} u_{\text{rgb}} = z_d K_{\text{rgb}} R_{d \rightarrow \text{rgb}} K_d^{-1} u_d + K_{\text{rgb}} t_{d \rightarrow \text{rgb}} \quad (12)$$

同理得到热图像与红外图像的像素点之间的空间位置关系:

$$z_t u_t = z_d K_t R_{d \rightarrow t} K_d^{-1} u_d + K_t t_{d \rightarrow t} \quad (13)$$

根据式(13), 可以找到深度图像中某一像素点 u_d 对应的 RGB 图像像素点 u_{rgb} 和热图像像素点 u_t , 遍历深度图像中每一个像素点, 即可找到该深度图像对应的 RGB 图像像素群和热图像像素群, 通过重组该像素群得到与深度图像内容一致的全新 RGB 灰度图像、深度图像和热图像, 如图4所示。

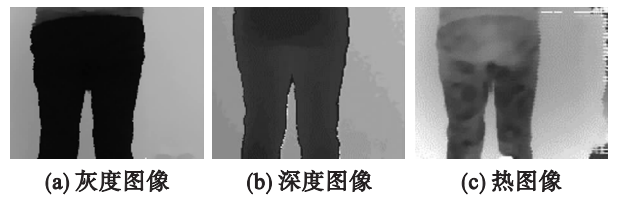


图4 配准后的 GDH 三模图像

Fig.4 RDH three-mode images after registration

2 基于 RDH 三模图像的视觉里程计

针对室内主要的动态对象——人类, 利用热图

像来定位分割人类, 获得人体掩模图像, 从而确定场景中的静态区域, 然后通过静态区域的信息开展室内动态场景下的鲁棒定位, 完成视觉里程计的实现。

2.1 基于热图像获取人体掩模图像

热像仪采用“黑热”模式成像, 故通过确定人体区域内像素灰度值所在区间, 就可以明确图像内任一像素点的归属域。在数据集[24]中, 经过多组实验选择像素灰度值区间[22, 95]为人体区域, 首先对热图像进行双边滤波处理, 然后二值化分割得到初始人体分割图像, 如图 5(b)所示, 对其进行腐蚀与膨胀处理, 去除伪人体区域, 得到图 5(c)所示的图像, 称为人体掩模图。

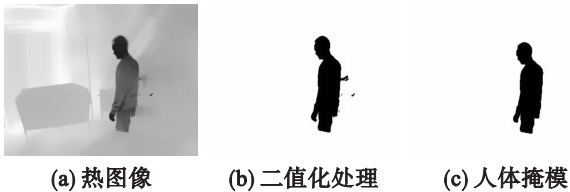


图 5 人体掩模图像

Fig.5 Human mask images

2.2 基于三模图像进行位姿估计

ORB-SLAM2 算法作为基本框架, 是特征点法的集大成者, 提取的特征点为 ORB 特征^[25]。在 2.1 节中通过热图像得到的人体掩模图像, 其作用是确定对应的 RGB 图像上的感兴趣区域(region of interest, ROI), 即不含人类的区域, 对 ROI 区域进行 ORB 特征提取, 即可实现仅提取静态特征的目的, 如图 6 所示。

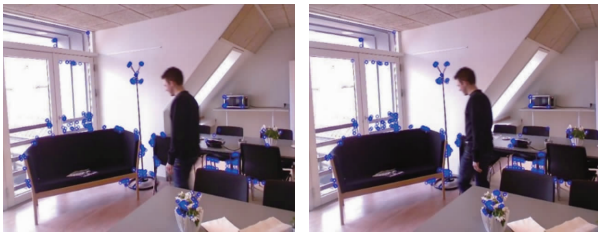


图 6 静态 ORB 特征提取结果

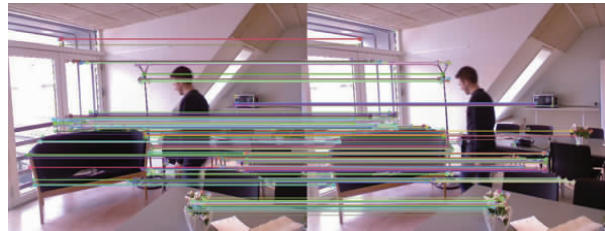
Fig.6 The features extraction result of ORB

ORB 特征的信息利用 Steer BRIEF 描述子进行表达, 考虑到本系统面临的环境中特征点数不多, 故采用暴力搜索方法来完成两幅图像之间的特征匹配, 结果如图 7(a)所示。可以看到, 由于室内物品相似度较高, 不同地点的特征较为相似等一些原因会导致大量的误匹配, 需要对误匹配进行剔除。采用的剔除策略为: 匹配过程中计算出所有匹配特征

的最小汉明距离, 当某个匹配特征的汉明距离大于最小距离的 2 倍时, 则认为是误匹配予以剔除, 结果如图 7(b)所示。



(a) 原始特征匹配



(b) 改进特征匹配

图 7 特征匹配结果

Fig.7 The result of feature matching

对匹配到的特征点选择 PnP (perspective-n-point) 算法进行建模, 使用非线性优化算法对其求解, 构建最小二乘问题并迭代求解, 完成优化, 从而计算出当前的相机运动。

3 静态环境稠密地图构建

提出的基于 RDH 三模图像构建静态环境地图的方案。首先利用热图像分割生成人体掩模图像, 然后用其进行深度图像更新, 从源头上避免动态像素的建模。最后根据相机位姿将 RGB-D 数据转换为点云并完成拼接, 再通过滤波、转换等后处理, 实现动态场景下 3 维静态地图的构建。

3.1 系统框架

基于 ORB-SLAM2 算法框架, 增加一个稠密建图 (dense mapping) 线程, 提出一种多传感视觉 SLAM 算法。算法同时运行着跟踪 (tracking)、局部建图 (local mapping)、闭环检测 (loop closing) 及稠密建图 4 个线程, 系统框架如图 8 所示。其中, 前 3 个线程实时运行, 主要基于输入的 RDH 三模图像构建视觉里程计, 得到局部优化的相机位姿、特征点地图以及相应的关键帧序列; 所增加的稠密建图线程主要负责将关键帧序列中的深度图像进行人体滤除, 并根据相应的位姿进行点云构建与拼接, 再经过后处理得到最终的场景静态地图, 由于其较大的计算量且不需要实时预览, 故在后台运行。

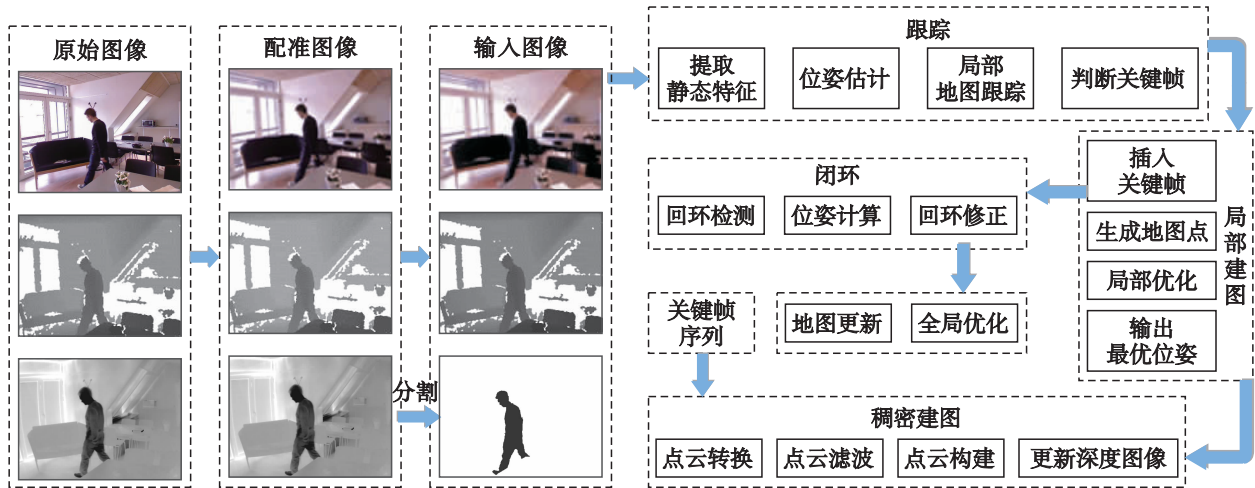


图8 系统框架

Fig.8 System framework

3.2 基于热图像更新深度图像

已知某像素点的像素坐标和深度值，根据当前相机位姿和内参矩阵可以计算出该像素点的世界坐标；反之当其深度值 $z=0$ 时，则无法得到其3维空间点。利用这一原理，依赖人体掩模图像，将每一帧深度图像人体区域的深度值修改为0。

具体过程如下：构造两个指针，分别指向人体掩模图像与深度图像的第一个像素，若该指针所指向的像素属于人体区域（即在人体掩模图中灰度值为0），则将深度图中指针所指向像素的深度值修改为0，否则不做任何修改；移动指针指向下一个像素，重复上述过程，直到处理完所有像素点，深度图像更新效果如图9所示。



图9 深度图像更新

Fig.9 Depth image update

深度图像经人体掩模图像处理后便同RGB图像根据各帧对应的位姿生成点云(point cloud, PC)，最后将点云数据进行拼接、滤波、转换处理，生成最终的地图形态。

本文采用RGB-D相机作为重要的图像采集传感器，可以同时获取同一场景的RGB图像和深度图像，即可以同时得到像素的坐标值 $p = (u, v)^T$ 和深度值 z ，根据相机成像模型和当前的相机位姿，得到深度值已知的像素点所对应的世界坐标为

$$P_w = zT^{-1}K^{-1}p \quad (14)$$

$$\text{式中, } T = \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix}。$$

遍历深度图，对每一个像素点进行上述反投影过程，得到点云。再将RGB图像对应像素的颜色信息附着在点云上，依次处理关键帧序列中的其他帧，最终得到全局静态稠密点云。

点云的生成是不断向空间添加点的过程，对于经过深度滤除的某一帧，即便没有获得人体区域遮挡的场景，但随着人体移动，被遮挡的场景仍会被传感器视野捕捉，这些区域的数据仍会被投影到空间生成点云。

对于生成的点云中存在的离群异常点和冗余数据，首先采用统计滤波器^[26]剔除离群异常点；然后采用体素滤波器^[27]，进一步剔除冗余数据。

对河北工业大学北辰智能装备研究院5号楼4单元1层进行3维重建验证滤波算法，选择参数 $k=50$, $n=1$ ，统计滤波后的点云如图10(b)所示。原始点云共包含3 402 359个数据点，经过统计滤波器处理后剩余2 261 548个数据点，数据量减少33.5%，占用空间从54.5 Mb减小至36.2 Mb。体素滤波将体素的体积设置为 $1.5 \text{ cm} \times 1.5 \text{ cm} \times 1.5 \text{ cm}$ ，滤波后的点云如图10(c)所示，此时点云数据点数量降至1 108 887，相比原始点云减少67.4%。

为适用于避障和导航，将滤波后的点云格式转换为OctoMap^[28]地图形式。首先以1.5 cm为最小分辨率创建一个八叉树模型，读取点云文件，遍历点云数据，将每一个点的位置和颜色信息插入到对应位置的OctoMap中，形成带颜色的“体数据”，最

后由 OctoMap 中所有的体数据构成最终的八叉树地图模型, 如图 11 所示, 此时文件占用存储空间仅为 8.4 Mb, 相比于原始点云减少 84.6%。

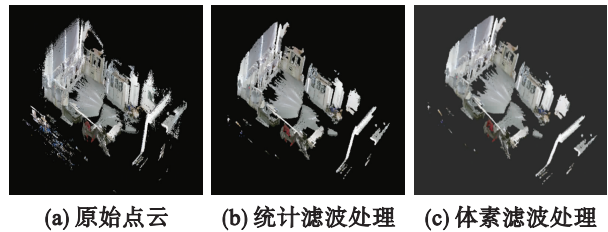


图 10 滤波处理

Fig.10 Filtering processing

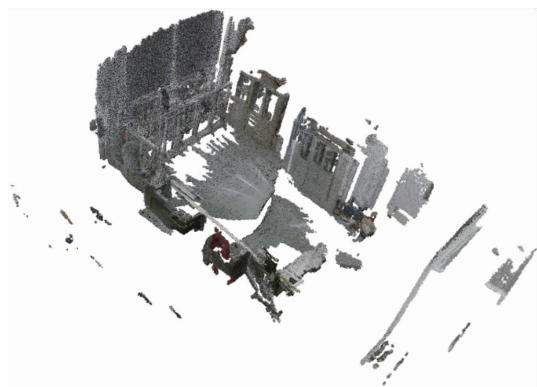


图 11 八叉树地图模型

Fig.11 The model of OctoMap

4 实验验证与分析

为验证本文提出的算法, 搭建一套如图 12 所示的多传感视觉 SLAM 系统实验平台, 平台主体为 HandsFree 轮式移动机器人, 底盘由 1 个万向轮和 2 个驱动轮构成; 多传感系统由 Kinect V2 深度相机和 FLIR T1040 热像仪组成, 其详细参数如表 2 所示; 计算平台为 ThinkPad E470, 其处理器为 Inter Core i7-7500U, 运行内存 8 G, 主频 2.70 GHz, 最高 2.9 GHz, 运行的软件系统为 Windows 10 + Ubuntu 16.04 LTS 双系统组合。

选择河北工业大学北辰智能装备研究院 5 号楼 4 单元 1 层车间作为实际场地进行数据采集, 其中固定热成像传感器的测温范围为 20 °C ~ 45 °C, 黑热模式。获得的图像序列分别包含 RGB 图像、深度图像和热图像, 数据集部分序列如图 13(a) ~ 图 13(c) 所示, 图像经处理后得到如图 13(d) 所示的人体掩模图像。以一组图像为例, 配准后得到如图 14 所示的 RGB 图像、深度图像和热图像, 根据热图像处理后得到如图 14(d) 所示的人体掩模图像。

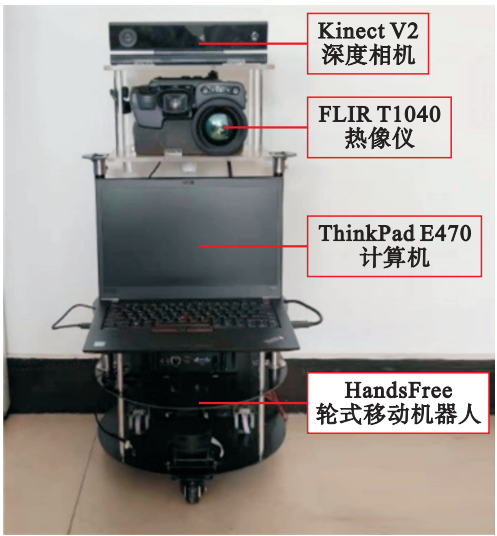


图 12 实验平台

Fig.12 Experiment platform

表 2 多传感器视觉系统参数

参数	RGB-D 相机	热像仪
型号	Kinect V2	FLIR T1040
分辨率	RGB: 1 920 × 1 080 Depth: 512 × 414	1 024 × 768
视场角 (FOV)	RGB: 84.1 × 53.8 Depth: 70 × 60	28 × 21
帧率	30 帧/s	30 帧/s

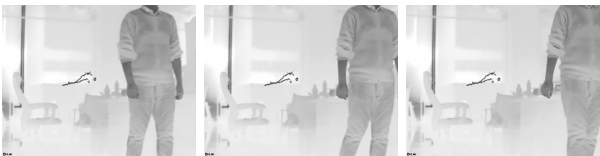


图 13 数据集序列(部分)

Fig.13 Part of dataset sequence

如图 14 所示,根据第 2 节中提出的配准方法,基于 RGB 图像、热图像分别与深度图像的像素点之间的空间位置关系,找到和深度图像对应的 RGB 图像像素群和热图像像素群,通过重组像素群得到与深度图像内容一致的全新 RGB 图像和热图像如图 14(a)、图 14(c)所示。然后基于配准后的热图像,二值化分割获得的人体掩模图像如图 14(d)、图 14(e)所示,其中图 14(e)中红色线框为深度图像 14(b)的人体区域轮廓,从图中可以看出掩模图像中人体轮廓与深度图像中人体轮廓吻合较好。进一步地,图 13 所示序列中 3 帧 RGB 图像、深度图像、热图像及处理的到的人体掩模图像横向对比表明,所提出算法可稳定提取动态场景中动态对象的人体掩模图像。

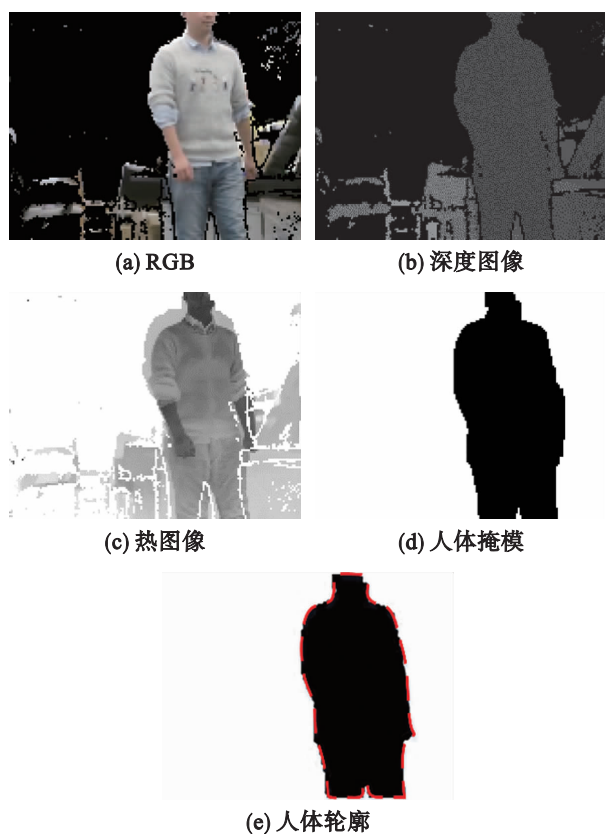


图 14 图像数据
Fig.14 The image data

将图像送入多传感视觉 SLAM 系统,对其运行状况进行截图,如图 15(a)所示。作为对比,图 15(b)为将数据集集中的 RGB 图像序列和深度图像序列送入常规的视觉 SLAM 系统,对其运行状况进行截图所示,经过对比,可以发现动态对象上的特征点明显变少。

在构建的动态数据集上运行多传感协同算法以及 ORB-SLAM2 算法,得到的轨迹如图 16 所示。从

图中可以发现,两种算法形成了差异较大的轨迹:两个算法在运行初期静态环境中地图轨迹基本相同;运行 3 s 后,由于动态对象的出现,传统算法的定位产生较大干扰,而多传感视觉算法具备一定的抗干扰能力。

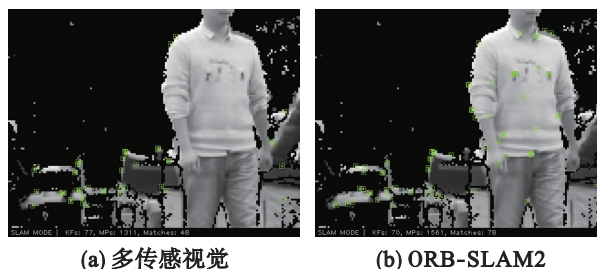


图 15 算法运行截图

Fig.15 Screenshot of algorithm running

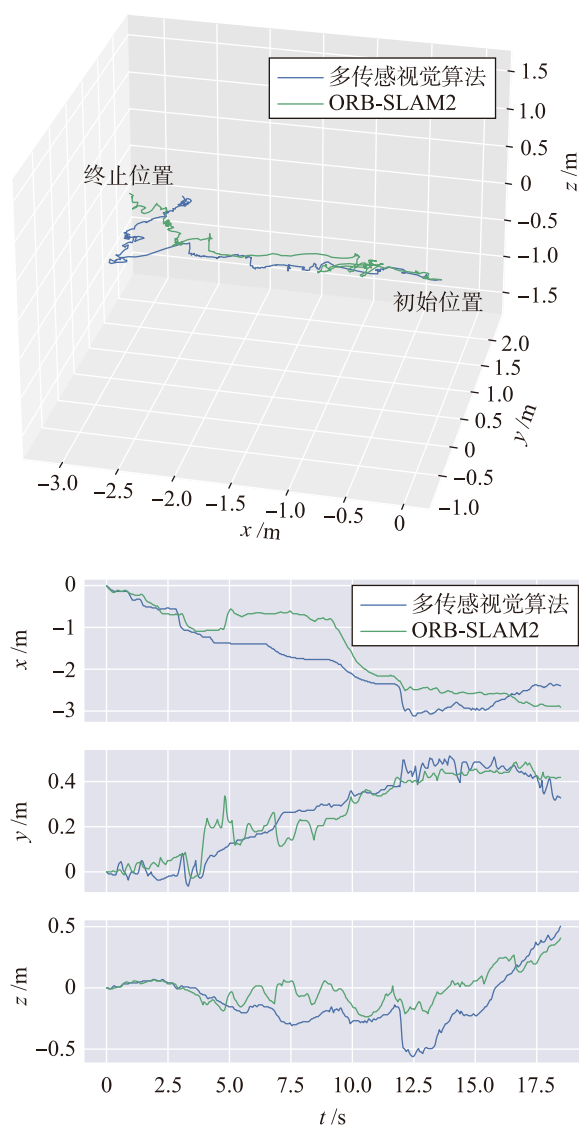


图 16 轨迹对比实验结果

Fig.16 The experiment results of trajectory comparison

从轨迹波动及整体形状来看, 多传感视觉算法的运行轨迹更符合当时采集的路线, 且对于动态特征具有抑制作用。多传感视觉算法运行轨迹与 ORB-SLAM2 算法得到的轨迹之间的差值最大为 1.134 m, 均方根差为 0.560 m, 多传感视觉算法的数据集对应的总距离为 10.65 m。

多传感视觉算法得到的稠密点云如图 17(a) 所示, 作为对比, ORB-SLAM2 算法得到的稠密点云如图 17(b) 所示。从图 17 可以看出, 多传感视觉算法取得了良好的实验效果, 动态对象几乎被剔除干净。

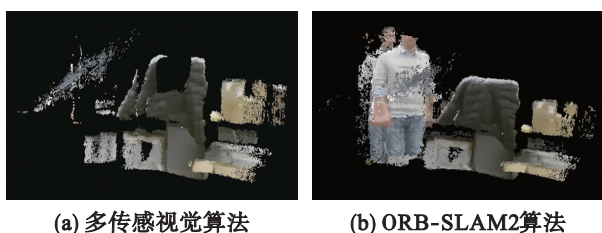


图 17 稠密 3 维点云

Fig.17 Dense 3D point clouds

对多传感视觉算法生成的点云进行滤波处理, 如 4.2 节内容所述, 统计滤波参数取 $k = 50$, $n = 1$, 得到滤波后的效果如图 18(a) 所示, 对比图 17(a) 滤波前图片可以得知, 存在的离群异常点被滤除; 而后设置体素空间为 $1.5 \text{ cm} \times 1.5 \text{ cm} \times 1.5 \text{ cm}$, 经过体素滤波后的点云如图 18(b) 所示, 在剔除冗余数据的情况下, 使得点云地图更加精确美观。

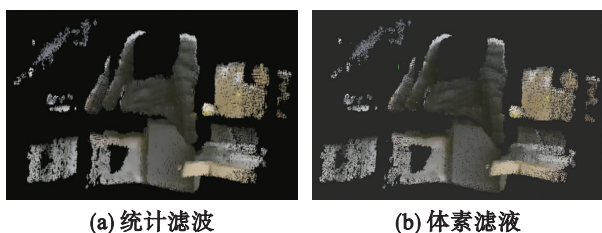


图 18 滤波后的点云数据

Fig.18 Filtered point cloud data

对多传感视觉算法生成的点云滤波处理后进行格式转换, 以 1.5 cm 为最小分辨率转换生成的 OctoMap

如图 19 所示。



图 19 八叉树地图

Fig.19 OctoMap

本文还可以从以下几方面不足进行改进:

1) 使用效果更好的 RGB-D 相机和热成像传感器。由于设备有限, 采用的 Kinect V2 RGB-D 相机深度图像的质量不高, 导致在图像配准时生成的 RGB 图像和热图像不够精确; 热成像传感器视场角过小, 间接导致图像配准完成之后图像分辨率仅有 160×120 。

2) 由于环境中可能存在与人体体温相近且体积较大的特殊物体, 这部分物体在环境建模时会被过滤掉。在今后的研究中可以采用深度学习的方法处理热图像, 更好地提取人体掩模图像。

5 结论

针对传统视觉 SLAM 在面临室内动态场景时产生的实时性、鲁棒性不足及建模冗余问题, 本文从“温度”的角度出发, 提出基于热像仪与深度相机相结合的多传感 SLAM 协同方案。首先, 实现深度相机与热成像组成的多传感视觉系统的联合标定与多模态图像配准, 然后基于 RDH 三模图像构建视觉里程计, 创建静态地图, 最后对该视觉系统应用于 SLAM 系统, 提出多传感视觉 SLAM 算法。经实验验证, 在面临动态场景时, 本文所提算法相比于传统视觉 SLAM 算法能够有效去除动态对象特征点干扰, 提升了建图效果。在后续研究工作中, 将针对不同动态场景种类、复杂度和稠密度等参数进行研究, 探索其对算法的影响。

参考文献

- [1] Durrant-Whyte H, Bailey T. Simultaneous localization and mapping: Part I[J]. IEEE Robotics & Automation Magazine, 2006, 13(2): 99–110.
- [2] 危双丰, 庞帆, 刘振彬, 等. 基于激光雷达的同时定位与地图构建方法综述[J]. 计算机应用研究, 2020, 37(2): 327–332. Wei S F, Pang F, Liu Z B, et al. Survey of Li DAR-based SLAM algorithm[J]. Application Research of Computers, 2020, 37(2): 327–332.
- [3] 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.
- [4] Smith R C, Cheeseman P. On the representation and estimation of spatial uncertainty[J]. The International Journal of Robotics Research, 1986, 5(4): 56 – 68.
- [5] Klein G, Murray D. Parallel tracking and mapping for small AR workspaces[C]//6th IEEE & ACM International Symposium on Mixed & Augmented Reality. Piscataway, USA: IEEE, 2007: 225 – 234.
- [6] 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.
- [7] Mur-Artal R, Tardós J D. ORB-SLAM2: An open-source SLAM system for monocular, stereo and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5): 1255 – 1262.
- [8] Engel J, Koltun V, Cremers D. Direct sparse odometry[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(3): 611 – 625.
- [9] Ai Y, Rui T, Lu M, et al. DDL-SLAM: Arobust RGB-D SLAM in dynamic environments combined with deep learning[J]. IEEE Access, 2020, 8: 162335 – 162342.
- [10] Bescos B, Fàcil J M, Civera J, et al. DynaSLAM: Tracking, mapping and inpainting in dynamic scenes[J]. IEEE Robotics & Automation Letters, 2018, 3(4): 4076 – 4083.
- [11] Rünz M, Agapito L. Co-Fusion: Real-time segmentation, tracking and fusion of multiple objects[C]//IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2017: 4471 – 4478.
- [12] Rünz M, Buffier M, Agapito L. MaskFusion: Real-time recognition, tracking and reconstruction of multiple moving objects [C]//IEEE International Symposium on Mixed and Augmented Reality. Piscataway, USA: IEEE, 2018: 10 – 20.
- [13] Zhong F, Wang S, Zhang Z, et al. Detect-SLAM: Making object detection and SLAM mutually beneficial[C]//2018 IEEE Winter Conference on Applications of Computer Vision. Piscataway, USA: IEEE, 2018: 1001 – 1010.
- [14] Henein M, Zhang J, Mahony R, et al. Dynamic SLAM: The need for speed[C]//2020 IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2020: 2123 – 2129.
- [15] Yu C, Liu Z, Liu X J, et al. DS-SLAM: Asemantic visual SLAM towards dynamic environments[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2018: 1168 – 1174.
- [16] Scona R, Jaimes M, Petillot Y R, et al. StaticFusion: Background reconstruction for dense RGB-D SLAM in dynamic environments[C]//2018 IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2018: 3849 – 3856.
- [17] Rebecq H, Horstschaefer T, Gallego G, et al. EVO: Ageometric approach to event-based 6-DOF parallel tracking and mapping in real time[J]. IEEE Robotics & Automation Letters, 2017, 2(2): 593 – 600.
- [18] Strecke M, Stuckler J. EM-fusion: Dynamic object-level SLAM with probabilistic data association[C]//2019 IEEE/CVF International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 5864 – 5873.
- [19] Kim D H, Kim J H. Effective background model-based RGB-D dense visual odometry in a dynamic environment[J]. IEEE Transactions on Robotics, 2016, 32(6): 1565 – 1573.
- [20] Kim J H, Kim D. Cooperative range-only SLAM based on sum of gaussian filter in dynamic environments[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2019: 2139 – 2144.
- [21] Zhang T, Zhang H, Nakamura Y, et al. FlowFusion: Dynamic dense RGB-D SLAM based on optical flow[C]//2020 IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2020: 7322 – 7328.
- [22] Palazzolo E, Behley J, Lottes P, et al. Refusion: 3D reconstruction in dynamic environments for RGB-D cameras exploiting residuals[C]//2019 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2019: 7855 – 7862.
- [23] 陈昕, 谢志宏, 宗艳桃. 基于 Matlab Camera Calibrator 的摄像机标定方法研究[J]. 电脑编程技巧与维护, 2015(16): 81 – 82, 92.
- Chen X, Xie Z H, Zong Y T. Camera calibration method based on Matlab camera calibrator[J]. Programming Skills & Maintenance, 2015(16): 81 – 82, 92.
- [24] Palmero C, Clapés A, Bahnsen C, et al. Multi-modal RGB-depth-thermal human body segmentation[J]. International Journal of Computer Vision, 2016, 118(2): 217 – 239.
- [25] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]//IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2011: 2564 – 2571.
- [26] 李鹏飞, 吴海娥, 景军锋, 等. 点云模型的噪声分类去噪算法[J]. 计算机工程与应用, 2016, 52(20): 188 – 192.
- Li P F, Wu H E, Jing J F, et al. Noise classification denoising algorithm for point cloud model[J]. Computer Engineering and Applications, 2016, 52(20): 188 – 192.

- [19] 张晓路, 李斌, 常健, 等. 水下滑翔蛇形机器人滑翔控制的强化学习方法[J]. 机器人, 2019, 41(3): 334–342.
Zhang X L, Li B, Chang J, et al. A reinforcement learning method for gliding control of underwater gliding snake-like robot[J]. Robot, 2019, 41(3): 334–342.
- [20] Ignacio C, Mariano D P, Corina B, et al. A reinforcement learning control approach for underwater manipulation under position and torque constraints[C]//Global Oceans 2020. Piscataway, USA: IEEE, 2020: 1–7.
- [21] Tobin J, Fong R, Ray A, et al. Domain randomization for transferring deep neural networks from simulation to the real world [C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2017: 23–30.
- [22] Peng X B, Andrychowicz M, Zaremba W, et al. Sim-to-real transfer of Robotic Control with dynamics randomization[C]//2018 IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2018: 3803–3810.
- [23] Chebotar Y, Handa A, Makoviychuk V, et al. Closing the sim-to-real loop: Adapting simulation randomization with real world experience[C]//2019 International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2019: 8973–8979.
- [24] Mehta B, Diaz M, Golemo F, et al. Active domain randomization[C]//3rd Annual Conference on Robot Learning. Osaka, Japan: PMLR, 2019: 1162–1176.
- [25] Zhang Q F, Zhang Y X, Huo L Q, et al. Design and pressure experiments of a deep-sea hydraulic manipulator system[C]// 2014 International Conference on Intelligent Robotics and Applications. Berlin Germany: Springer, 2014: 117–128.
- [26] Sutton R S, Barto A G. Reinforcement learning: An introduction[M]. Cambridge, USA: MIT Press, 1998.
- [27] Schulman J, Wolski F, Dhariwal P, et al. Proximal policy optimization algorithms[DB/OL]. (2017–08–28)[2021–11–01] <https://arxiv.org/abs/1707.06347>. pdf.
- [28] Morison J R, Johnson J W, Schaaf S A. The force exerted by surface waves on piles[J]. Journal of Petroleum Technology, 1950, 2(5): 149–154.
- [29] Todorov E, Erez T, Tassa Y. MuJoCo: A physics engine for model-based control[C]//2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2012: 5026–5033.
- [30] Wang C, Zhang Q F, Zhang Q C, et al. Construction and research of an underwater autonomous dual manipulator platform [C]// 2018 OCEANS-MTS/IEEE Kobe Techno-Oceans. Piscataway, USA: IEEE, 2018: 1–5.
- [31] Kingma D P, Ba J, Bengio Y, et al. Adam: A method for stochastic optimization[C]//3rd International Conference on Learning Representations. Berlin, Germany: Springer, 2015.
- [32] Koenig N, Howard A. . Design and use paradigms for Gazebo, an open-source multi-robot simulator[C]//2004 IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2004: 2149–2154.
- [33] Quigley M, Conley K, Gerkey B, et al. ROS: An open-source robot operating system[C/OL]//ICRA Workshop on Open Source Software. Piscataway, USA: IEEE, 2009 [2021–08–01]. <http://www.robotics.stanford.edu/~ang/papers/icraossos-ROS.pdf>.
- [34] Manhoes M M M, Scherer S A, Voss M, et al. UUV Simulator: A Gazebo-based package for underwater intervention and multi-robot simulation[C]//2016 OCEANS, Piscataway, USA: IEEE, 2016: 1–8.
- [35] Garrido-Jurado S, Muñoz-Salinas R, Madrid-Cuevas F, et al. Automatic generation and detection of highly reliable fiducial markers under occlusion[J]. Pattern Recognition, 2014, 47(6): 2280–2292.
- [36] Zhang Z, Wang C, Zhang Q, et al. Research on autonomous grasping control of underwater manipulator based on visual servo [C]// 2019 Chinese Automation Congress. Piscataway, USA: IEEE, 2019: 2904–2910.

作者简介

王 聪(1991–), 男, 博士生。研究领域为水下机器人自主作业, 强化学习。

张子扬(1994–), 男, 硕士生。研究领域为水下机械手控制。

陈言壮(1988–), 男, 博士生。研究领域为水下液压系统。

(上接第 650 页)

- [27] 李瑞雪, 邹纪伟. 基于 PCL 库的点云滤波算法研究[J]. 卫星电视与宽带多媒体, 2020(13): 237–238.
Li R X, Zou J W. Point cloud filtering algorithm based on PCL library[J]. Satellite TV & IP Multimedia, 2020(13): 237–238.
- [28] Hornung A, Kai M W, Bennewitz M, et al. OctoMap: An efficient probabilistic 3D mapping framework based on octrees[J]. Autonomous Robots, 2013, 34(3): 189–206.

作者简介

张建华(1979–), 男, 博士, 教授。研究领域为智能机器人, 医疗机器人, 协作机器人。

张天晶(1997–), 女, 硕士生。研究领域为机器视觉, SLAM。

赵 岩(1988–), 男, 博士, 讲师。研究领域为医疗机器人, 机器视觉, SLAM。