

基于实例分割的视觉 SLAM 算法

何召兰, 何乃超⁺, 张庆洋, 姚 徐, 丁淑培

(哈尔滨理工大学 测控技术与通信工程学院, 黑龙江 哈尔滨 150080)

摘 要: 针对当前视觉同步定位与地图构建 (visual simultaneous localization and mapping, VSLAM) 算法鲁棒性较差, 无法生成适合自主导航的语义地图等问题, 提出一种基于实例分割的视觉 SLAM 算法。在对输入图像提取特征点的同时, 使用卷积神经网络对图像进行实例分割; 利用实例分割信息辅助定位, 剔除容易造成误匹配的特征点, 缩小特征匹配的区域; 使用实例分割的语义信息构建语义地图。使用 TUM 数据集对图像实例分割、视觉定位以及语义地图构建进行实验验证, 验证结果表明, 该算法提高了机器人定位的准确性, 生成了精确的语义地图, 满足机器人执行高级任务的需求。

关键词: 同步定位与地图构建; 卷积神经网络; 特征匹配; 实例分割; 语义地图

中图法分类号: TP391.41 **文献标识号:** A **文章编号:** 1000-7024 (2020) 10-2791-06

doi: 10.16208/j.issn1000-7024.2020.10.016

Visual SLAM algorithm based on instance segmentation

HE Zhao-lan, HE Nai-chao⁺, ZHANG Qing-yang, YAO Xu, DING Shu-pei

(School of Measurement and Communication, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: Aiming at the problem that the current visual simultaneous localization and mapping (VSLAM) algorithm has poor robustness and can not generate semantic maps suitable for autonomous navigation, a visual SLAM algorithm based on instance segmentation was proposed. Features from the input image were extracted and convolution neural network was used to segment the image. The information of instance segmentation was used to assist the localization, eliminating the feature which easily leaded to mismatching, and narrowing the feature matching area. A semantic map was constructed using the semantics information of instance segmentation. TUM data sets were used to validate image instance segmentation, visual localization and semantic map construction. The results show that the proposed algorithm can improve the accuracy of robot positioning and generate accurate semantic maps to meet the needs of robots to perform advanced tasks.

Key words: simultaneous localization and mapping; convolutional neural network; feature matching; instance segmentation; semantic map

0 引 言

移动机器人的智能化研究主要集中在对未知环境的探索认知、自我行为决策与规划、对未知信息的学习与表达等^[1]方面, 其中实现在未知环境中的自主定位导航, 是移动机器人智能化研究的基础与重点。要实现这一功能, 不仅需要精确估算出每一时刻移动机器人的位姿信息, 还要对每一时刻机器人所处位置的环境信息进行准确描述。1887 年, Smith 等提出了一种定位与构图同时进行的方

法^[2], 从此拉开对同步定位与地图构建 (simultaneously localization and mapping, SLAM) 研究的序幕。SLAM 也成为了实现移动机器人智能化导航的关键技术。

SLAM 算法即是移动机器人从未知环境的某一地点出发, 在运动过程中重复地读取传感器观测数据, 分析获取环境特征与自身位置姿态, 并以此实时构建周围环境的增量式地图^[3]。其中, 视觉传感器相比于其它传感器能够获得更丰富的图像信息, 同时视觉传感器轻巧廉价、易于安装^[4], 因此基于视觉传感器的 SLAM 研究成为了当下研究

收稿日期: 2019-07-04; **修订日期:** 2019-10-08

基金项目: 国家自然科学基金青年科学基金项目 (61803127); 黑龙江省普通本科高等学校青年创新人才培养计划基金项目 (UNPYSCT-2017093)

作者简介: 何召兰 (1969-), 女, 黑龙江哈尔滨人, 博士, 教授, 研究方向为图像处理与模式识别、机器人导航; ⁺通讯作者: 何乃超 (1993-), 男, 黑龙江哈尔滨人, 硕士研究生, 研究方向为机器人视觉导航; 张庆洋 (1994-), 男, 山东济南人, 硕士研究生, 研究方向为图像处理、语音识别; 姚徐 (1992-), 男, 安徽桐城人, 硕士研究生, 研究方向为图像处理; 丁淑培 (1996-), 女, 河南平顶山人, 硕士研究生, 研究方向为图像处理。E-mail: hezhaoan@sina.com

热点。视觉 SLAM 的实现主要有特征点法和直接法^[5], 其中, 特征点法具有运行稳定、对光照及动态物体不敏感等优点^[6], 因此本文主要针对特征点法展开分析。近年来, 随着基于特征点法的视觉 SLAM 算法不断发展, 已经可以实现特定的场景下的实时跟踪定位。然而, 目前主流的视觉 SLAM 系统仍然存在以下几个问题: ①在特征匹配的过程中容易受到噪声和误匹配点的干扰, 鲁棒性较差。②无法从场景中提取语义信息, 仅能生成包含几何信息的地图, 无法满足机器人执行更高级任务的需求。因此本文将深度学习方法引入到视觉 SLAM 算法框架当中, 优化了图像特征匹配过程, 实现了位置信息和语义信息的融合。

1 视觉 SLAM 算法

1.1 SLAM 数学模型

移动机器人在一个未知环境中运动, 首先要建立固定不变的世界坐标系, 根据机器人自身传感器数据精准的建立时间与世界坐标系空间的对应关系, 即在世界坐标系下描述各个时刻移动机器人自身位置及其所处环境, 这是 SLAM 解决定位与构图问题的核心。视觉 SLAM 的自身传感器为相机, 因此移动机器人的运动可以抽象为相机在世界坐标系下的刚体运动。具体形式如图 1 所示。

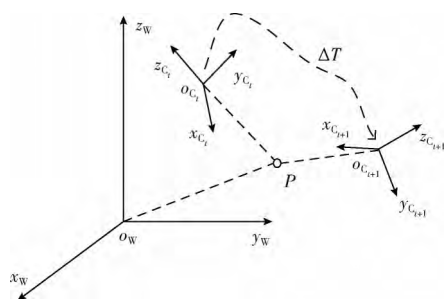


图 1 坐标系变换

图 1 中 (x_w, y_w, z_w) 为世界坐标系, (x_c, y_c, z_c) 为移动坐标系, 即相机所在坐标系, 点 P 为世界坐标系空间一点, 是路标 (landmark) 的抽象描述, 路标是场景中比较有辨识度的物体或者区域。在每个时刻, 传感器会观测到一部分路标, 得到它们的观测数据。设 t 时刻, 相机观测到空间点 P , 在移动坐标系下该观测点坐标为 $(^c_t p_x, ^c_t p_y, ^c_t p_z)$, 而 $t+1$ 时刻, 该空间点在移动坐标系下的坐标为 $(^{c_{t+1}} p_x, ^{c_{t+1}} p_y, ^{c_{t+1}} p_z)$, 那么相机的运动方程如式 (1) 所示

$$\begin{bmatrix} ^{c_{t+1}} p_x \\ ^{c_{t+1}} p_y \\ ^{c_{t+1}} p_z \\ 1 \end{bmatrix} = \Delta T \begin{bmatrix} ^c_t p_x \\ ^c_t p_y \\ ^c_t p_z \\ 1 \end{bmatrix} \quad (1)$$

其中, ΔT 为 t 时刻到 $t+1$ 时刻相机的位姿变换矩阵。

根据运动方程可知, 若已知移动机器人的初始位姿, 就可以通过各个时刻的观测数据, 估算出移动机器人在世界坐标系下的位姿, 同时利用位姿变换信息, 可以解算出各个路标点在世界坐标系下的坐标, 构建全局一致的地图。根据以上分析, 可以得到 SLAM 的数学模型, 如式 (2) 所示

$$\begin{cases} x_{t+1} = x_t \Delta T_t + w_t \\ z_{t,k} = h(^w P_k, x_t) + v_{t,k} \end{cases} \quad (2)$$

其中, x_t 为 t 时刻移动机器人的位姿, $z_{t,k}$ 表示移动机器人在 x_t 位置观察路标 $^w P_k$ 的观测数据。 w_t 与 $v_{t,k}$ 为噪声。

1.2 传统视觉 SLAM 算法框架

随着国内外学者对视觉 SLAM 算法的不断研究改进, 视觉 SLAM 已经形成了一套较为完整的算法框架。如图 2 所示。

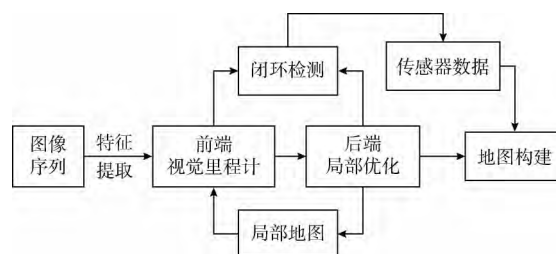


图 2 视觉 SLAM 算法框架

视觉 SLAM 主要由前端视觉里程计、闭环检测、后端局部优化以及全局地图构建 4 个部分构成^[7]。其中视觉里程计^[8]主要通过相邻两帧图像来估算移动机器人当前时刻的位姿变化, 根据 SLAM 数学模型分析可知, 机器人位姿是通过传感器测量的路标观测数据解算得到的。在视觉 SLAM 算法中, 传感器测量的路标观测数据为相机采集的图像帧的特征点集。因此计算机器人位姿首先要提取图像特征点。图像特征点是那些经过算法分析出来的, 含有丰富局部信息的点^[9]。在完成相邻的两帧图像 F_1, F_2 的特征点提取之后, 需要进行图像特征匹配^[10]。特征匹配解决的是相邻两帧图像特征点集数据的关联问题, 即确定当前所观察的路标点是否与上一时刻的路标点为同一个。经过特征匹配获得一组组匹配的特征点对, 如式 (3) 所示

$$\begin{cases} ^t P = \{^t P_1, ^t P_2, \dots, ^t P_N\} \in F_1 \\ ^{t-1} P = \{^{t-1} P_1, ^{t-1} P_2, \dots, ^{t-1} P_N\} \in F_2 \end{cases} \quad (3)$$

这些特征点对应的是世界坐标系下的路标点。根据 SLAM 数学模型, 这些特征匹配点对的变换关系如式 (4) 所示

$$\forall i, ^t P_i = \Delta T^{t-1} P_i \quad (4)$$

其中, ΔT 为机器人的位移变换矩阵。在特征匹配的过程中, 无法避免误差的产生, 为了尽量准确获得该时刻移动机器人的位姿数据, 本文采用最小化重投影误差的方法来

求解位姿。如式 (5) 所示

$$\{T\} = \operatorname{argmin} \sum_{i=1}^N \|P_i - (\Delta T^{-1} P_i)\|_2 \quad (5)$$

视觉里程计解算出的机器人位姿是短期的, 随着移动机器人的运行, 估算误差将不可避免的随时间积累, 因此视觉 SLAM 算法使用闭环检测与后端优化^[11]来避免误差积累。闭环检测通过读取传感器数据和视觉里程计信息来判断机器人是否在之前曾经到达过某一位置, 检测到正确的回环后, 将校正信息传递给后端进行优化和计算; 后端局部优化结合不同时刻视觉里程计估算出来的相机位姿数据以及闭环检测提供的回环信息进行全局优化, 从而得到精确的轨迹和地图。

根据上述分析可知, 传统的 SLAM 算法框架在估算机器人位姿时最重要的一步就是图像特征点匹配, 而一幅图像的特征点数目较多, 且图像特征点属于图像的局部特性, 如果机器人行驶到存在大量重复纹理的场景中, 会使得提取到的特征描述子非常相似, 如果不对图像的特征点进行筛选, 将导致较大特征点误匹配率。并且采用特征法的视觉 SLAM 建图为稀疏特征点地图, 这种地图只能满足机器人定位的要求, 而无法提供导航、避障与交互等智能化功能。

近些年来, 深度学习在图像领域取得了巨大进步, 为解决这些问题提供了新的思路。深度卷积神经网络结构具有强大的特征学习能力, 在图像识别、图像分割以及目标检测方面取得了十分显著的效果^[12]。尤其在目标检测领域, 由于卷积神经网络的引入, 物体检测的准确率与效率均得到了极大的提升^[13]。何凯明等在 Faster RCNN^[14]架构的基础上提出了 MASK R-CNN^[15], 并斩获了 ICCV 2017 年的最佳论文。MASK R-CNN 将目标检测与语义分割^[16]结合在一起实现了图像的实例分割, 相对目标检测的边界框, 实例分割可以获取精确到物体边缘轮廓的掩模信息; 而相对语义分割, 实例分割可以标注出图像上同一类物体的不同个体。移动机器人通过视觉传感器采集的图像序列包含不同类别的物体, 若结合 MASK R-CNN 的实例分割技术优势与视觉 SLAM 算法提供的解算数据, 不仅能够提高定位精度还能为移动机器人提供语义信息。

2 基于实例分割改进的视觉 SLAM

2.1 MASK R-CNN 模块构建

MASK R-CNN 是一个开源的基于深度学习的图像实例分割框架, 它可以有效检测出图像中的物体, 并精确分割每一个实例, 同时为每个实例生成高质量的语义掩模。

如图 3 所示, MASK R-CNN 框架可以并行完成目标分类、目标定位、语义分割 3 项任务。不同于传统的语义分割, 基于 MASK R-CNN 框架的实例分割不仅仅可以标注出图像中物体的语义信息, 还能够区分同一类物体的不同

个体, 并可以用边界框 (bounding box) 精确定位目标物体在图像中的位置, 这对于 SLAM 这种要求检测精度较高的任务来说十分重要。

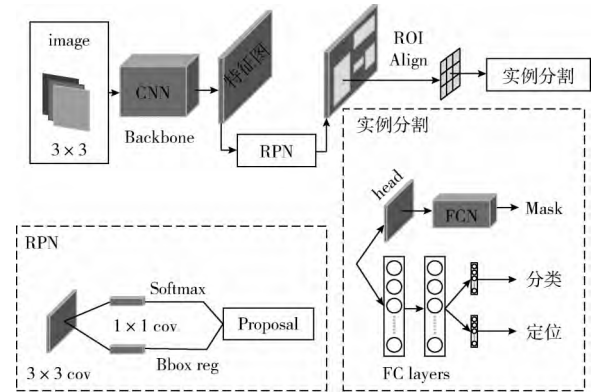


图 3 MASK R-CNN 框架

MASK R-CNN 主要由卷积神经网络构成, 通过多分支子网络并行实现端到端的实例分割。MASK R-CNN 实现实例分割的具体过程如下:

(1) 将图像输入主干网络 (backbone network), 生成特征图 (feature maps)。主干网络的卷积层由 MASK R-CNN 整体共享, 主要的功能是学习并提取图像特征。本文使用深度残差网络 (ResNet)^[17] 以及特征金字塔网络 (feature pyramid networks, FPN)^[18] 相互结合构建主干网络;

(2) 将特征图输入区域提议网络 (region proposal network, RPN), RPN 是一个轻量级的卷积神经网络结构。该网络首先利用 3×3 的卷积层对原图进行滑动扫描, 为图像的每个像素生成锚框 (anchor box), 然后将锚框送入分类分支以及边框回归分支 (Bbox reg), 分类分支输出每个锚框存在物体的概率; 而边框回归分支将锚框进行伸缩平移, 使它更接近地回归于标记好的真实边界框 (ground truth box), 输出初步的候选框坐标及其高与宽。最后在 RPN 末端综合两分支的结果, 并利用非极大值抑制 (non-maximum suppression, NMS)^[19] 对锚框进行初步筛选, 最终获得基于原图的候选区域 (region proposal);

(3) 利用 ROI Align 对 RPN 输出的候选区域的图像进行处理。ROI Align 使用双线性内插的方法将基于原图的矩形候选框坐标映射到特征图上, 并使所有映射到特征图的候选区域图像大小一致, 使之符合全连接层 (fully connected layers, FC layers) 的输入要求;

(4) 将经过 ROI Align 层处理的候选区域图像送入目标检测分支及掩模预测分支, 完成分类、定位以及掩模预测任务。其中掩模预测是利用全卷积神经网络 (fully convolutional networks, FCN)^[20] 对候选区域图像的每一个像素点进行分类, 生成物体的掩模, 实现对候选区域图像的语义分割。

2.2 MASK R-CNN 改进的视觉 SLAM 算法

为提高视觉 SLAM 位姿估计的准确性并生成语义地图, 本文利用 MASK R-CNN 对视觉 SLAM 进行了改进, 如图 4 所示。

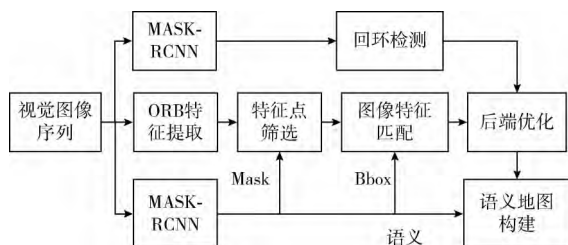


图 4 基于实例分割的视觉 SLAM 框架

本文在传统的视觉 SLAM 框架的基础上引入了 MASK R-CNN 分支, 利用 MASK R-CNN 的信息辅助定位并构建语义地图。具体改进如下:

(1) 利用 MASK R-CNN 生成的掩模信息对特征点进行筛选; 本文采用图像的 ORB (oriented FAST and rotated BRIEF) 特征作为视觉 SLAM 的路标特征, ORB 特征是目前最具代表性的实时图像特征^[21]。它改进了 FAST 检测子不具有方向性的问题, 并采用速度极快的二进制描述子 BRIEF (binary robust independent elementary features), 使图像特征提取整个环节大大提速。对视觉图像进行特征提取时, 将生成大量的 ORB 特征点, 这些特征点所包含的信息是冗余的, 大量的特征点不仅会加大系统的运算量, 更会导致误匹配概率加大。MASK R-CNN 可以对视觉图像进行目标检测, 将图像进行了目标与背景的分割, 本文借助 MASK R-CNN 的掩模信息, 剔除图像背景的特征点, 保留前景物体的特征点, 筛选后的特征点归属不同的物体, 因此 ORB 特征点描述子的相似性将大大降低;

(2) 利用边界框信息对视觉图像进行分区, 提高特征匹配效率; 在传统的 SLAM 算法中要完成一组特征点匹配, 需要遍历整幅图像的特征点, 运算效率极低, 而 MASK R-CNN 利用边界框将视觉图像划分成有语义标记的不同实例区域, 在进行特征匹配时, 只需要在同一实例区域进行匹配, 这不仅极大提高了特征匹配的效率, 更进一步减少了误匹配现象的发生, 提高了位姿匹配的估算精度;

(3) 利用语义信息进行回环检测; 传统的 SLAM 算法使用对特征描述进行聚类的方法来建立词袋 (bag of words, BOW) 模型, 而 MASK R-CNN 可以利用图像分类的高级语义信息更直接而准确地建立词袋模型, 并实现回环检测;

(4) 利用语义信息构建语义地图; MASK R-CNN 可以对机器人在运动过程中检测到的物体打上语义标签, 并结合 SLAM 框架生成带有高级语义标签的语义地图。

3 实验分析

为验证本文算法的有效性和正确性, 本节对基于实例分

割的 SLAM 算法进行实验验证。主要的验证方法为对比实验。本文采用时下应用最为广泛的 ORB-SLAM2 算法作为本文算法的对照组。ORB-SLAM2 是 ORB-SLAM 的改进版本, 支持单目, 双目以及 RGB-D 相机, 该算法具备传统 SLAM 所有的模块功能, 在处理速度、追踪效果和地图精度上都取得了不错的效果。同时, ORB-SLAM2 算法同样使用 ORB 图像特征, 因此本文可以通过对比本文算法与 ORB-SLAM2 算法的结果来验证本文算法的效果。主要验证的指标为: 特征匹配效率与精度、定位的准确性, 以及构图效果。

3.1 实验数据集

实验采用两种数据集, 一种是 MSCOCO 数据集, 另一个是 TUM 数据集。MSCOCO 数据集是微软专门为图像识别、检测以及语义分割所设计的大型图像数据集, 包含超过 30 万张图片, 200 多万个实例以及 80 个对象类别。但该数据集的图像并不具备时间连续性, 只使用该数据集无法满足实验要求。因此本文使用 TUM 的 freiburg1_room 数据集进行扩充。TUM 数据集是德国慕尼黑工业大学利用深度相机采集的一组连续的室内图像集, 提供了深度相机参数以及相机在室内运动的真实轨迹 (ground-truth), 可以满足本文实验的需求。

为了提高 MASK R-CNN 在视觉 SLAM 算法框架中的表现力, 本文在使用 MSCOCO 数据集训练的基础上, 又使用 TUM 数据集对 MASK R-CNN 的网络参数进行微调。TUM 数据集没有对图像进行标注且包含的场景单一, 因此本文首先利用 labelme 工具对选取的图像进行标注并生成掩模, 然后对图像进行裁剪、缩放以及旋转等处理, 以扩充数据。经过训练后, MASK R-CNN 进行实例分割最终结果如图 5 所示。

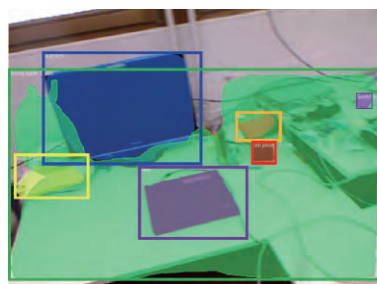


图 5 实例分割

可以看出, MASK R-CNN 能够较为准确检测出图中的物体, 并对检测物体进行像素级别的分类。

3.2 实验对比

首先本文对特征匹配效率与精度这一指标进行验证, 将 TUM 数据集图片分为完全相同的两组, 其中一组输入传统的 ORB-SLAM2 算法直接进行特征匹配, 特征匹配效果如图 6 所示。另一组输入本文的算法框架, 在进行特征提取的同时, 利用实例分割进行特征点筛选并优化特征匹

配过程, 匹配效果如图 7 所示。

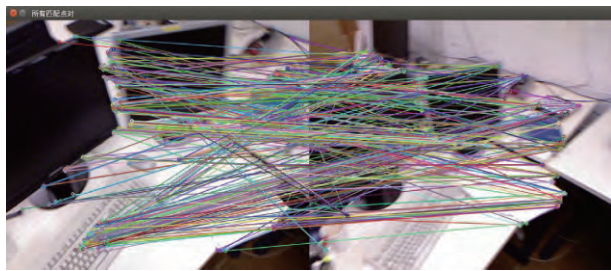


图 6 传统算法的特征匹配效果

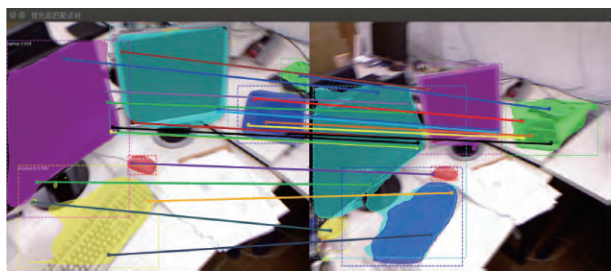


图 7 实例分割优化的特征匹配效果

通过对比, 可以看出 ORB-SLAM 算法提取得到的特征点数目是巨大的, 且在特征匹配时存在很多误匹配点, 大量的特征点匹配需要进行复杂的运算, 匹配效率低下, 同时大量的误匹配点还会为后续的位姿估计带来较大的误差。而本文算法利用实例分割对特征匹配过程进行优化, 从图中可以看出, 本文算法进行特征匹配的特征点对较少, 且匹配点均位于已分割出的语义区域, 因此误匹配现象较少。

接下来, 本文对定位精度这一指标进行验证。视觉 SLAM 能够实时估算出移动机器人位姿, 若将各个时刻的机器人位姿连接起来, 就形成了机器人运行轨迹。TUM 数据集提供了采集图像时机器人运动的真实轨迹, 如图 8 所示。

因此通过比较 SLAM 估算得到的运行轨迹与真实轨迹

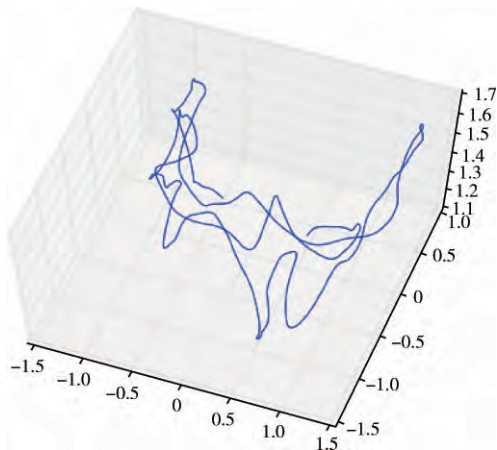


图 8 TUM 数据集真实轨迹

的差值, 如式 (6) 所示

$$\begin{aligned} & \text{MIN}(\sqrt{(x_i - x_i^s)^2 + (y_i - y_i^s)^2 + (z_i - z_i^s)^2}) \\ & \text{MAX}(\sqrt{(x_i - x_i^s)^2 + (y_i - y_i^s)^2 + (z_i - z_i^s)^2}) \\ & \frac{1}{n} \sum_{i=1}^n \sqrt{(x_i - x_i^s)^2 + (y_i - y_i^s)^2 + (z_i - z_i^s)^2} \end{aligned} \quad (6)$$

式中: (x_i^s, y_i^s, z_i^s) 为真实轨迹的空间坐标点, 而 (x_i, y_i, z_i) 为算法估算的运行轨迹, 空间坐标点, 通过计算运行轨迹与真实轨迹的最大, 最小以及平均误差, 就可以验证 SLAM 算法定位的准确程度, 运行轨迹的差值数据见表 1。

表 1 运行轨迹差值

SLAM 算法对比	最大误差 /m	最小误差 /m	平均误差 /m
ORB-SLAM2	0.211	0.02	0.135
MASK-RCNN SLAM	0.096	0.0001	0.047

根据表 1 中数据显示, 本文算法的运行轨迹更加接近于真实轨迹, 故基于实例分割的视觉 SLAM 算法的定位精度更高。

最后对本文算法的构图模块进行验证。基于特征点法的视觉 SLAM 算法通常构建的为稀疏特征地图, 这种地图无法具体的描述机器人当前所处的环境。随着计算机硬件的发展, 构建稠密的三维点云地图成为了当前的主流, 如图 9 所示。这种地图虽然可以直观反应环境信息, 但对于计算机来说, 该地图只不过由三维点云堆积而成, 机器人无法凭借该地图执行更高级的智能任务。本文利用 MASK R-CNN 的语义信息构建了具有语义信息的点云地图。该点云地图的点云像素被赋予了高级语义信息, 为直观描述, 本文利用不同颜色来标注语义, 如图 10 所示。通过语义信息, 移动机器人可以完成如自主导航、人机交互等高级任务。

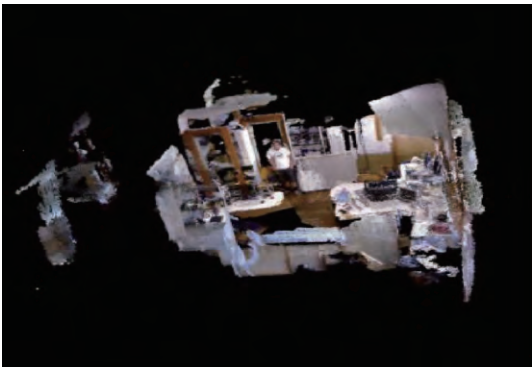


图 9 三维点云地图

4 结束语

基于实例分割的视觉同步定位与地图构建算法将 MASK R-CNN 作为一个独立的分支引入到传统的视觉

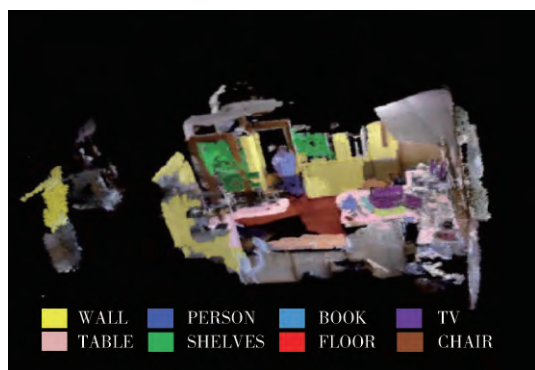


图 10 语义地图

SLAM 框架当中。通过理论分析以及实验验证,可以得到以下结论:①利用实例分割可以辅助完成 SLAM 的定位过程,提高了定位的精确性、鲁棒性以及定位速度;② MASK R-CNN 能够对视觉图像进行目标检测,并对检测到的物体进行像素级别的分类,赋予三维点云地图语义信息,构建具有语义信息的三维点云地图。满足了移动机器人执行更高级的智能化任务的需求。下一步,本文将进一步的将深度学习引入到视觉 SLAM 框架中,利用深度学习在图像特征提取方面的优势,真正实现端到端的机器人定位与语义地图构建。

参考文献:

- [1] Yunhui L, Fan Z, Ruibin G, et al. Robot intelligence for real world applications? [J]. Chinese Journal of Electronics, 2018, 27 (3): 446-458.
- [2] Cadena C, Carlone L, Carrillo H, et al. Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age [J]. IEEE Transactions on Robotics, 2016, 32 (6): 1309-1332.
- [3] Li Ruihao, Wang Sen, Gu Dongbing, et al. Ongoing evolution of visual SLAM from geometry to deep learning: Challenges and opportunities [J]. Cognitive Computation, 2018, 10 (6): 875-889.
- [4] Zaffar M, Ehsan S, Stolkin R, et al. Sensors, SLAM and long-term autonomy: A review [C] //NASA/ESA Conference on Adaptive Hardware and Systems. NEW YORK: IEEE Press, 2018: 285-290.
- [5] Yousif K, Bab-Hadiashar A, Hoseinnezhad R. An overview to visual odometry and visual SLAM: Applications to mobile robotics [J]. Intelligent Industrial Systems, 2015, 1 (4): 289-311.
- [6] Taketomi T, Uchiyama H, Ikeda S. Visual SLAM algorithms: A survey from 2010 to 2016 [J]. IPSJ Transactions on Computer Vision and Applications, 2017, 9 (1): 16-21.
- [7] Huletski A, Kartashov D, Krinkin K. Evaluation of the modern visual SLAM methods [C] //Artificial Intelligence and Natural Language and Information Extraction, Social Media and Web Search FRUCT Conference. NEW YORK: IEEE, 2015: 19-25.
- [8] Zhao B, Hu T, Shen L. Visual odometry-A review of approaches [C] //IEEE International Conference on Information & Automation. NEW YORK: IEEE, 2015: 2569-2573.
- [9] LI Shimiao. A review of feature detection and match algorithms for localization and mapping [J]. IOP Conference Series: Materials Science and Engineering, 2017, 231 (1): 012003.
- [10] Gao L, Yuan P, Wang T, et al. New research on SLAM algorithm based on feature matching [C] //2nd International Conference on Advanced Robotics and Mechatronics. NEW YORK: IEEE Press, 2017: 1-6.
- [11] Liu G, Hu Zhaozheng. Fast loop closure detection based on holistic features from SURF and ORB [J]. Jiqiren/Robot, 2017, 39 (1): 36-45.
- [12] Shin H C, Roth H R, Gao M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning [J]. IEEE Transactions on Medical Imaging, 2016, 35 (5): 1285-1298.
- [13] Zhao Z Q, Zheng P, Xu S T, et al. Object detection with deep learning: A review [J]. IEEE Transactions on Neural Networks and Learning Systems, 2019, PP (99): 1-21.
- [14] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39 (6): 1137-1149.
- [15] He Kaiming, Gkioxari Georgia, Dollar Piotr, et al. Mask R-CNN [C] //IEEE International Conference on Computer Vision. NEW YORK: IEEE, 2017: 2980-2988.
- [16] Abdelnasser Heba, Mohamed Reham, Elgohary Ahmed, et al. Semantic SLAM: Using environment landmarks for unsupervised indoor localization [J]. IEEE Transactions on Mobile Computing, 2016, 15 (7): 1770-1782.
- [17] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] //IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2016: 770-778.
- [18] Lin T Y, Dollar P, Girshick R, et al. Feature pyramid networks for object detection [C] //IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2017: 936-944.
- [19] Hosang Jan, Benenson Rodrigo, Schiele Bernt. Learning non-maximum suppression [C] //IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society, 2017: 6469-6477.
- [20] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39 (4): 640-651.
- [21] Li L, Wu L, Gao Y. Improved image matching method based on ORB [C] //17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing. NEW YORK: IEEE, 2016: 465-468.