

DOI: 10.19650/j.cnki.cjsi.J1904749

室内场景下实时地三维语义地图构建*

单吉超^{1,2}, 李秀智^{1,2}, 张祥银^{1,2}, 贾松敏^{1,2}

(1. 北京工业大学信息学部 北京 100124; 2. 计算智能与智能系统北京市重点实验室 北京 100124)

摘要: 移动机器人自主建图是完成智能行为的前提。为提高机器人智能水平和直观的用户交互, 地图需要扩展超出几何和外观信息的语义信息。研究了将基于深度残差网络(DRN)的像素级图像语义分割和三维同时定位与建图(SLAM)相融合的三维语义地图构建方法。首先, 采用一种联合中值滤波算法进行深度图像的修复, 使用改进的迭代最近点(ICP)算法得到相机估计位姿以及基于随机森林的闭环检测构建出三维环境地图; 其次, 采用优化的深度残差网络对输入的图像实现较精准的像素语义级别的预测与分割; 最后, 采用贝叶斯更新方法, 渐进式的将图像分割获取的语义分类标签迁移到重建的室内三维模型中, 获得完整的三维语义地图。实验表明, 所设计的方法可以在实际的、复杂环境下实时地构建语义地图。

关键词: 闭环优化; 同时定位与建图; 图像语义分割; 贝叶斯更新; 三维语义地图

中图分类号: TP242 TH72 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Real-time 3D semantic map building in indoor scene

Shan Jichao^{1,2}, Li Xiuzhi^{1,2}, Zhang Xiangyin^{1,2}, Jia Songmin^{1,2}

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China;

2. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China)

Abstract: Autonomous mapping for mobile robot is the premise of completing intelligent behavior. To improve the intelligence and intuitive user interaction of robot, maps are needed to achieve the semantics beyond geometry and appearance. This paper studies the 3D semantic map construction method, which fuses the pixel-level image semantic segmentation based on Deep Residual Networks (DRN) and Simultaneous Localization And Mapping (SLAM). Firstly, the combined median filter algorithm is used to restore the depth of the map. The improved Iterator Closest Point (ICP) algorithm is employed to estimate camera pose and loopback detection based on random ferns is proposed for 3D scene reconstruction. Then, the optimized DRN is utilized to achieve more accurate semantic prediction and segmentation. Finally, the predicted semantic classification labels are migrated to the 3D model by Bayesian based incremental transfer strategy to generate a globally consistent 3D semantic map. Experimental results show that the proposed method can build the real-time 3D semantic map in the real and complicated environment.

Keywords: closed-loop optimization; SLAM; image semantic segmentation; Bayesian update; 3D semantic map

0 引言

随着人工智能与机器人技术研究取得的卓越进展, 机器人正在逐渐走进人类生活之中。消费者对室内移动服务型机器人的需求呈现上升的趋势。智能机器人能够实现真正自主、智能的前提是自动感知、理解其所处的外部环境。移动机器人在移动过程中识别出物体的属性对

获得高层次的语义地图至关重要。

最初机器人构建的传统地图, 例如栅格地图^[1]与拓扑地图^[2]能够指导移动机器人完成定位、导航、路径规划等任务, 但缺乏机器人用于理解环境、人机交互等任务的高层次语义信息, 对周围环境的理解与人类相比相差甚远。而语义地图的构建可以使机器人在语义层次上理解环境信息, 这和人类对环境理解的方式更为接近。语义地图的研究是室内机器人一个有应用前景的重要方向,

收稿日期: 2019-02-20 Received Date: 2019-02-20

* 基金项目: 国家自然科学基金(61703012)、北京自然科学基金(4182010)、企业委托项目“人体目标跟踪系统开发”(2018223)资助

它将运动控制、图像处理、计算机视觉、信息融合等相结合,使机器人实现更高层次的智能任务成为了可能。

随着智能化机器人应用领域的发展,移动机器人的智能语义地图创建越发得到国内外学者的关注。Nieto-Granda 等^[3]将环境空间自动识别和分类为不同语义区域,并利用这些信息生成环境的拓扑图,基于人工增强映射的方法将语义标签与空间区域关联;Kostavelis 等^[4]提出了针对室内场景的双层机器人语义导航方法,深度相机的三维同时定位与建图(simultaneous localization and mapping, SLAM)系统获取低层数据,而高层则采用一种新的基于内容的表示算法用于空间抽象,两层之间通过包含环境认知属性的语义拓扑地图相互关联;Rituerto 等^[5]提出了一种基于全方位视觉的增强室内拓扑映射场景标记方法,然后将语义信息作为场景分割的标准,进行拓扑地图的划分;吴皓等^[6-7]提出了一种分层语义地图创建方法,地图分为 3 层:局部空间层、区域规划层和全局语义层。首先,机器人获取深度信息和 QR 码标签提供的信息构建物品功能图形成局部空间层。然后基于贝叶斯估计构建区域层的栅格地图形成区域规划层。最后,基于谱类算法建立具有房间分割功能的拓扑地图形成全局语义层。俄克拉荷马州立大学 Sheng 等^[8]为了有效地获取语义信息并将其链接到度量地图中,提出了在人机共存的环境中通过人类活动识别推断语义地图构建方法,但基于可穿戴设备略显繁琐。Shuran Song 等人^[9]利用单视角深度图的观测,生成一个完整的三维立体像素表示场景的空间占用和语义标签方法,引入端到端的语

义场景完成网络输出所有三维体素的占用率和语义标签;李等^[10]开发了一种轻量级目标检测模型,采用贝叶斯推理将检测的结果映射到栅格地图,完成了在嵌入式系统上的语义地图构建。近年来,基于 SLAM 的地图创建与基于卷积神经网络的图像分割方法的研究,有利于计算机视觉技术朝着实用化、工程化方向发展。

本文针对实际的室内环境,研究了一种基于 SLAM 的三维重建技术与图像语义分割相融合的实时三维语义地图创建方法。首先,为了更准确的构建出三维环境地图,采用一种联合中值滤波算法进行深度图像的修复,采用改进的迭代最近点(iterative closest point, ICP)算法计算旋转平移矩阵进而估计相机位姿以及基于随机藏类的闭环检测有效识别曾到达区域。其次,采用优化后的深度残差网络(deep residual network, DRN)对输入的图像实时进行像素语义级别的预测与分割;最后,经过贝叶斯更新的方法,渐进式的将图像分割获取的语义分类标签迁移到重建的室内三维模型中,获得完整的三维语义地图。

1 实时三维语义地图构建框架

本文所述语义地图构建框架如图 1 所示。三维地图创建系统通过 Kinect 相机对外部环境采集数据,基于 SLAM 算法对重建的场景建立一个完整的三维模型,采用优化的深度残差网络对输入的图像实现较精准的预测与分割。最后使用贝叶斯更新方法渐进式构建语义地图。

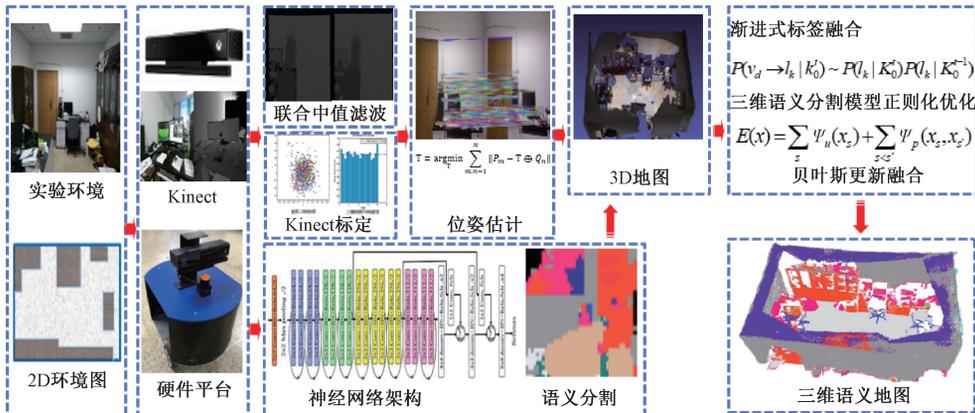


图 1 语义地图构建框架
Fig.1 Framework of the semantic mapping

2 基于 RGBD 相机的室内地图创建

为较好的完成室内环境下的三维地图创建任务,本研究引入基于图优化^[11-12]的 SLAM 方法,在图优化框架

下,采用联合中值滤波算法修复深度图像。采用改进的随机抽样一致性(random sample consensus, RANSAC)算法优化特征点对,然后通过 ICP 算法估计相机位姿,并通过随机藏类图形化编码识别曾到达过区域,得到室内环境的三维点云图。

2.1 基于联合中值滤波的深度图像修复

Kinect 采集到的深度图像能够得出三维环境的距离信息,如图 2 所示,由于目标遮挡、物体表面材质光滑发生镜面反射与光线剧烈变化,原始深度图像存在着像黑洞、黑边等形式的无效区域,信息缺失的深度图像势必对后面的算法造成很大的影响。



图 2 原始深度图像
Fig.2 Raw depth image

文献[13]采用高斯滤波器估计黑洞中的深度值,这是一种基于时间滤波器的方法。虽然此方法可以恢复黑洞和黑边处的像素值,但高斯滤波器忽略了图像的 RGB 颜色信息,填充效果的可信度效果不明显。本文受文献[14]的启发,提出一种联合中值滤波算法,在传统的中值滤波器之上,利用颜色信息来辅助填充失效像素的深度值。我们仅处理深度图像中的无效像素,对于每一个无效像素点的周围创建半径为 r 的窗口,在窗口中有两种像素点被舍弃:1) 无效像素点;2) 与窗口中邻接像素点的灰度差值大于设定阈值的中心像素点。最后,经过筛选,其他提取出的像素点按照深度值排序,选择这些点的中值作为无效点深度值的填充值。

2.2 基于投票式 RANSAC 算法与 ICP 算法相融合的相机位姿估计

本文使用联合中值滤波算法修复深度图像,然后利用 ICP 算法^[15]将深度图像转化成的点云与模型投影获得的点云进行配准,获取当前相机位姿。而 ICP 算法是局部优化器,如果场景复杂导致出现大量的错误特征点对,不能提供相对准确的特征点对集,就会出现错误的配准结果。为此,针对复杂环境中点集关联常常存在错误的点对,无法为 ICP 算法提供初始对齐的情况,提出了基于投票式 RANSAC 算法^[16]与 ICP 算法相融合的相机位姿估计算法。采用 ORB(oriented FAST and rotated BRIEF)特征匹配提取特征点,随即计算每个特征点对应的深度有效值,获得三维点集。在进行 ICP 配准之前,通过投票式 RANSAC 算法优化三维点集,将错误的特征点对剔除,然后通过 ICP 算法迭代的计算点集关联来优化求解。

2.3 使用随机蕨类实现闭环匹配

在进行地图创建过程中,只使用图优化^[17]位姿优化方法,并不能识别曾到达区域,完成系统的全局闭环。在

长时间以及大场景地图创建过程中后端优化对于闭环的效果将逐渐减弱。在回到曾到达区域或反复进行局部建图时,会出现图 3 所示的重叠和错位现象。

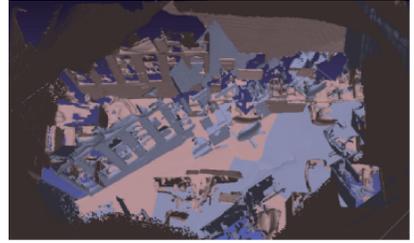


图 3 重叠及错位现象
Fig.3 The appearance of overlap and dislocation

为了解决上述问题,在优化相机位姿模型中使用随机蕨类模型。它使用图像编码的方式判断两帧之间的相似度,采用式(1)、(2)编码化关键帧。 m 个 *block* 的编码组成每幅图像 C 的编码,每个 *block* 由 n 个 *ferns* 组成。通过每个通道 x 处的像素值 θ 与阈值比较确定每个 *ferns*。

$$C = \{block_k\}_{k=1}^m \rightarrow block = \{ferns_i\}_{i=1}^n \quad (1)$$

$$f(I, \theta, \tau) = \begin{cases} 1, & I(\theta) \geq \tau \\ 0, & I(\theta) < \tau \end{cases}$$

$$\theta = \{c, x\} \quad (2)$$

如图 4 所示,将获取的关键帧进行编码化处理。关键帧 C 以及它对应的深度图像,共有 R、G、B、D 四个通道,分别比较每一通道像素值与选取的阈值的大小,来计算每一个 *block* 里的 *Fern* 列表中的值。

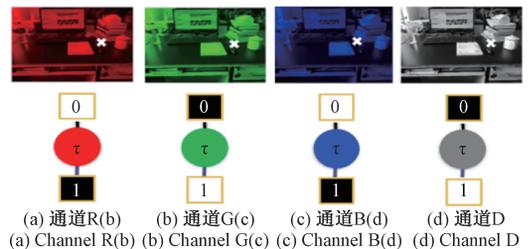


图 4 图像编码示意图

Fig.4 Illustration of image coding

以图 4 的图像编码为例,对获取到图像的 4 个通道的像素值与选取的阈值根据式(2)作比较,得出 *block* 的码值为 {1100}。图像的码值和关键帧 ID 的关系如表 1 所示,表 1 的左栏为通过式(2)计算得出的图像编码,右栏为图像编码所对应的关键帧编号。如果关键帧之间存在同样的图像编码,那么二者之间的相似度就高。根据 *block* 码值在表中索引,发现 {1100} 编码对应图像的 ID 为 2、7,那么图像 1 与图像 7 的相似度加 1,如果对图像所有 *block* 进行这样的操作,对比临近关键帧和历史关键帧,即可计算出此图像与所有图像之

间的相似度。判别关键帧间的相似度,决定当前关键帧是否加入回环,如果存在回环,那么和相似度大的关键帧配准,进行重定位。

表 1 关键帧编码表
Table 1 The encoding of key-frame

图像编码	关键帧 ID 编号
...	...
{ 1010 }	(0,3,5)
{ 1011 }	(1,4)
{ 1100 }	(2,7)
{ 1110 }	(6,8)
...	...

3 基于深度残差网络的图像语义分割

全卷积神经网络 (fully convolutional networks, FCN)^[18]是第一个实现端到端的基于卷积神经网络的语义分割模型。经典卷积神经网络 (convolutional neural network, CNN)是在卷积层后接上全连接层,得到固定长

度的特征向量,实现图像的分类,而 FCN 没有使用全连接层,网络模型全部由卷积层按照维度需求连接而成,采用反卷积层对特征图上采样实现分割。但 FCN 得到的分割结果不够精细,上采样的结果比较模糊和平滑,对图像中的细节信息不敏感。通过增加网络深度的方式可以有效地提取深层抽象特征使分割精度提高,但随着网络层数的加深,网络训练的难度显著增加,且识别率趋于饱和,甚至下降。针对于此,文献[19]提出了残差网络架构,在特定网络层的输入层加入前面网络层的输出结果,这种结构能够使得网络达到更深的层数,同时也易于训练。因此,本文融合了 DRN 与 FCN 网络的优点,设计了一种基于精简 DRN 网络的图像语义分割算法。首先,提出了一个简化的 DRN 网络,其参数量是 VGG-16 模型的 30%,在实时性能方面表现良好。同时,采用跳跃连接的方法,有效地融合各个卷积层学到的特征,尽可能地弥补网络前向提取特征时丢失的信息。

3.1 模型结构

本文设计的 DRN 网络结构如图 5 所示,结构为 22 层。它由 3 个用于上采样的反卷积层和 5 种类型的卷积层组成。具有特定颜色的每种类型的卷积层包括若干个残差网络所需要的卷积层。

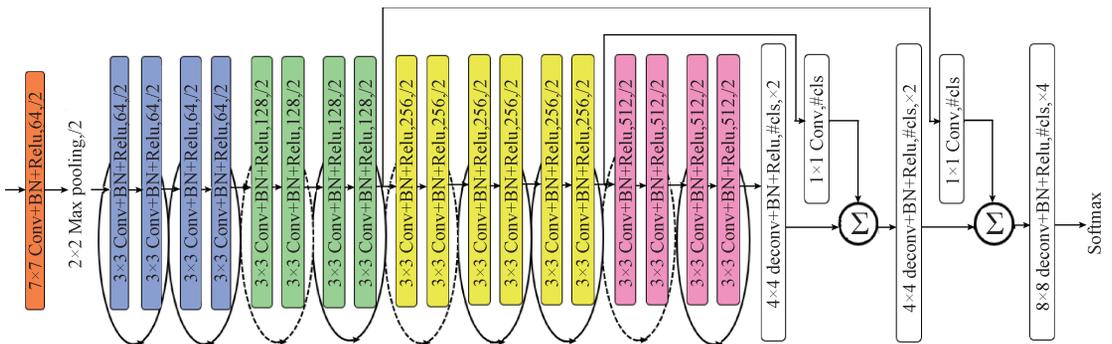


图 5 CNN 架构:结合跳跃连接、反卷积的残差神经网络

Fig.5 CNN architecture: Residual network with skip connection and deconvolution

将输入的图片统一成 224×224 的分辨率,第 1 种类型的卷积层使用 64 个 7×7 的卷积核获得 64 个 112×112 的特征图,然后将 64 个特征图经过 3×3 卷积核大小的池化层获得 64 个 56×56 的特征图(M1)。第 2 种类型的卷积层将 M1 作为输入,使用四个相同结构的 64 个 3×3 的卷积核获得 64 个 56×56 的特征图(M2)。以此类推,经过第五种类型的卷积层,得到 512 个 7×7 的特征图(M5),经过多次的卷积和池化,特征图变小,分辨率降低。为了恢复到图像的原始分辨率,本文采用跳跃连接的方法来减小浅层上采样的步长,将第 3 层和第 4 层的输出特征图与反卷积上采样之后的特征图进行融合,从而丰富了上采样的部分损失信息,最后使用 8 倍上采样

来实现图像的像素级分割。同时,残差学习在每一种类型的卷积层中被使用到,这种结构能够使得网络达到更深的层数,同时也易于训练优化。

3.2 网络训练

ImageNet、VOC 等图像数据库提供了大量的带有真值标签的图像数据。在配有 Intel i7-6700 CPU @ 3.40 GHz、16 G 内存的 GTX1070 GPU 和操作系统为 Ubuntu16.04 LTS 的电脑上用于训练和测试,开发语言是 python。本文在 PASCAL VOC-2012 分割数据集上对本文网络进行训练,分割精度与 FCN 中的 VGG 模型进行对比。同时,为了对实验室等实际场景的特定物品进行图像语义分割实验,在 NYUv2 数据集重新训练了网络,使用标准随机梯

度下降的方法,学习率为 0.01,动量为 0.9,权重衰减为 0.0005 进行优化。经过 10 000 次迭代后,将学习速率降低到 0.001。使用 64 大小 mini-batch,在 Nvidia GTX 1070 的 2 d 内对网络进行总计 20 000 次的迭代训练。

4 基于贝叶斯更新的语义地图融合

本文研究了将基于深度残差网络的像素级图像语义分割和三维同时定位与建图相融合的三维室内语义地图构建方法。将上述论述的三维场景重建、图像语义分割通过基于贝叶斯更新的融合算法有机地整合为一个整体,将神经网络识别的结果与三维点云地图整合进统一的语义地图中,实现面向移动机器人的三维语义地图构建。

通过传感器采集数据,由于其不稳定性,在未知的环境下,在连续相邻的关键帧中语义分割通常缺乏一致性。为此,通过渐进式语义标签融合^[58]关联多个关键帧中的语义标签,图像语义标签随场景的关联而及时更新。若当前关键帧为 K_t , K_t 的三维体素 V_d 上的语义标签分布为 l_k 。需要得到每一个三维体素在其语义标签集合上的独立概率分布 $P(v_d \rightarrow l_k | k'_0)$,而目前关键帧集合为 $K'_0 = \{K_0, K_1, \dots, K_t\}$,递归式贝叶斯如式(3)所示。

$$P(v_d \rightarrow l_k | k'_0) = P(K_t | K_0^{t-1}, l_k) P(l_k | K_0^{t-1}) \frac{1}{Z_i} \quad (3)$$

式中: $Z_i = P(K_t | K_0^{t-1})$,同时在 $P(K_t | K_0^{t-1}, l_k)$ 上采用马尔科夫假设,得到式(4)。

$$P(v_d \rightarrow l_k | k'_0) = P(K_t | K_0^{t-1}, l_k) P(l_k | K_0^{t-1}) \frac{1}{Z_i} = \frac{P(K_t) P(l_k | K_t)}{P(l_k)} P(l_k | K_0^{t-1}) \frac{1}{Z_i} \quad (4)$$

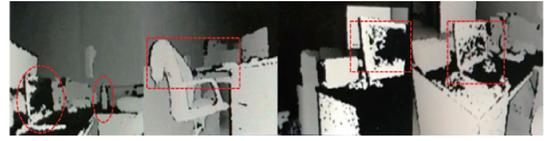
假设 $P(l_k)$ 与时间变量无关,当获取下一关键帧图像时,可以用式(5)更新三维体素的语义标签类别:

$$P(v_d \rightarrow l_k | k'_0) \sim P(l_k | K'_0) P(l_k | K_0^{t-1}) \quad (5)$$

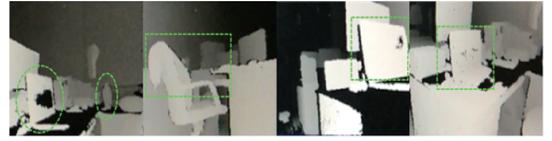
5 实验结果与分析

5.1 深度图像修复实验

本文研究采用联合中值滤波算法修复原始深度图像存在着黑边、黑洞等大量无效区域。本文比较了在真实实验室环境下获取的深度图像以及经过算法处理后的图像。图 6(a) 所示为 Kinect 采集的原始深度图,图 6(b) 所示为对应修复后的深度图。实验结果表明所采用算法很好的填充了深度图像中的无效距离值。修复后的深度图像由于填充了部分无效值,有利于构建较好的致密三维模型。



(a) 深度图像修复前
(a) Depth image before restoration



(b) 深度图像修复后
(b) Depth image after restoration

图 6 深度图像修复算法结果

Fig.6 Restoration results on depth maps

5.2 基于 RGBD 相机的三维环境建图实验

在图 7 所示的实验室环境下,手持 Kinect 实现采图和三维建图。移动速度约为 0.15 m/s,采集帧率为 30 Hz。



(a) 视角 1
(a) Visual angle 1
(b) 视角 2
(b) Visual angle 2

图 7 实验环境

Fig.7 The laboratory scene

为验证改进 RANSAC-ICP 算法的有效性,使用标准数据集^[20]对比不同方法得到相对旋转与平移误差,实验结果如表 2 所示。标准数据集是由 3 种不同程度复杂度的环境被相机以 2 种不同的速度及运动方式拍摄获得。表 2 对比了使用 DVO (dense visual odometry)^[21], GICP (generalized-ICP)^[22], NICP (normal-ICP)^[23] 及所提方法处理其中 6 种数据集时所得实验结果。表 2 标出了使用不同方法处理当前数据集时所得的最好结果(加粗字体)。实验表明 RANSAC-ICP 算法无论在图像清晰还是模糊时都能减小估计位姿的旋转与平移误差,得到较好的位姿估计结果。

利用 Kinect 视觉传感器得到室内环境信息,加入随机类检测模型,在某一局部实验室场景中验证其有效性。实验调用 Kinect 摄像头,运行建图系统,如图 8 所示,对办公桌区域进行建图,反复创建这一局部环境,没有出现重影和误差。表明系统能够有效识别曾经到达过的位置,并且和相邻区域建立了较准确的拼接。

表 2 相机位姿旋转与平移误差表
Table 2 Rotation and translation error of camera pose (m)

方法	评价标准	高复杂度	中速高复杂度	低速低复杂度	中速低复杂度	低速中复杂度	中速中复杂度	低速
DVO	平移误差	0.247	0.143	0.634 1	0.499 1	0.472 1	0.256	
	旋转误差	9.914	4.875	3.793	4.791	2.692	6.218	
GICP	平移误差	0.089	0.062	0.073	0.062	0.073	0.057	
	旋转误差	5.018	3.246	3.870	2.850	4.316	2.769	
NICP	平移误差	0.072	0.059	0.079	0.057	0.063	0.044	
	旋转误差	5.694	3.524	4.349	2.699	4.136	2.601	
本文方法	平移误差	0.060	0.061	0.076	0.049	0.051	0.033	
	旋转误差	5.011	3.249	3.973	2.612	4.123	2.541	

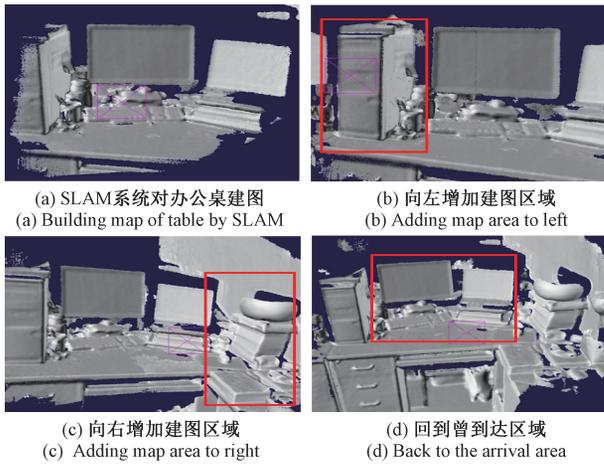


图 8 随机蕨类效果

Fig.8 The effectiveness of random ferns

进一步扩大实验场景,手持 Kinect 对整个实验室扫描一周并三维建图,如图 9 所示,运动角速度约为 0.15 rad/s。图 9 比较了使用基于随机蕨类的闭环检测前后的效果差异,软件环境为 meshlab。从图 9 可以看出,使用随机蕨类可以有效消除系统的累积误差,实现闭环建图。

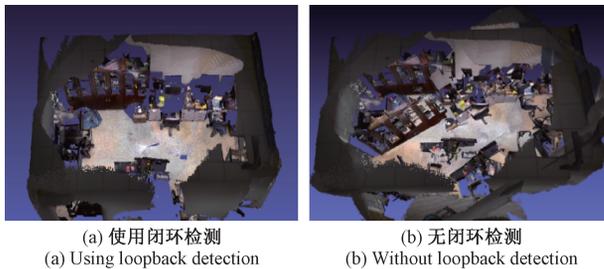


图 9 随机蕨类模型全局效果

Fig.9 The global effectiveness of random ferns

5.3 图像语义分割实验

本文图像语义分割模型的网络结构,由图 5 的网络

作为训练网络。将 PASCAL VOC 的子集包括 9 963 个训练图像和相应的标签对象用于模型训练。有 21 类数据集,如汽车、人、卡车和自行车等。这些标签由两部分组成:类别分割,标注出每一个像素的类别;实例分割,标记每个像素属于哪个对象。本文选择类别分割标签用于评估所用方法的有效性并与 FCN 方法中最精细的 FCN-8 s 作对比。

本文采用平均交并比作为与其他方法对比的评价指标。它定义为计算真实值和预测值交集和并集之比。平均交并比(mean IoU)公式如下:

$$(1/n_c) \sum_i n_{ii} / (t_i + \sum_j n_{ji} - n_{ii}) \quad (6)$$

式中: n_{ij} 是类别 i 对应的像素被预测属于类别 j 的总数目; n_c 代表所有类别的数目; t_i 代表类别 i 对应的像素的总数目。

在 PASCAL VOC-2012 测试集上的分割结果如表 3 所示,本文的算法在分割精度以及网络传播速度上均优于 FCN。产生的可视化效果如图 10 所示,图像分割效果更加精细,对于小物体的识别分割有明显改善。

表 3 分割精度结果

Table 3 The result of segmentation accuracy

分割方法	精确度(平均交并比)时间/ms
FCN-32 s	59.2~160
FCN-8	62.2~175
Ourmethod	70.6~98

为了验证本文的网络模型在实际环境中的效果,对实验室实际场景进行图像语义分割实验。同样由图 5 的网络在 NYUv2 数据集重新训练。将待识别的三维场景中的物体划分为 10 个颜色类别对应 10 个编号,即代表 10 种语义信息,通过 DRN 网络预测得到待分割图像的几何轮廓图,实现较为精准的图像语义分割。

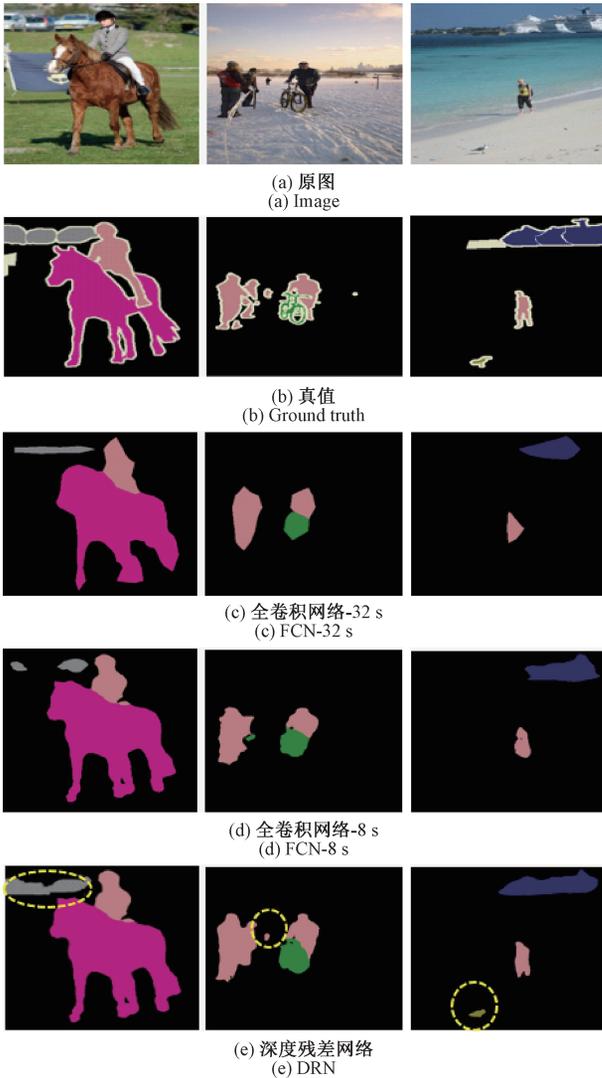


图 10 语义分割结果对比

Fig.10 The comparison of semantic segmentation results

本文提出的物体识别分割算法的结果如图 11 所示,可以看出即使是在相对复杂、狭窄的实验室环境,本文使用的语义分割算法仍然表现良好。

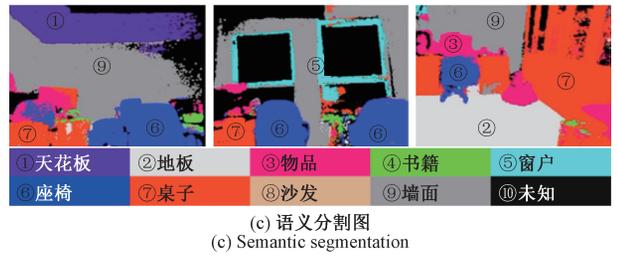
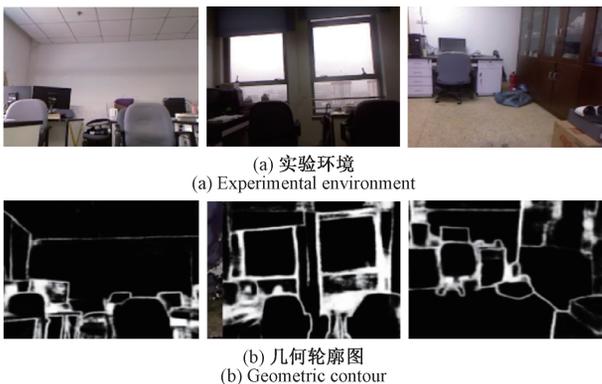


图 11 图像语义分割结果

Fig.11 Result of image semantic segmentation

5.4 语义地图融合实验

采用视觉 SLAM 方法对环境进行三维重建,采用优化的深度残差网络对输入的图像实现较精准的预测与分割。然后经过贝叶斯更新的方法,渐进式的将图像分割获取的语义分类标签迁移到重建的室内三维模型中,获得完整的三维语义地图。

为了验证本文提出的语义地图构建方法,在 8 m × 4 m 的实验室中进行三维语义地图创建。图 12 所示为实时创建语义地图时部分关键帧的语义分割图,图 13 所示为整个实验室场景下构建的三维语义地图。能够看出得到了完整的三维语义地图,验证了算法的效果。

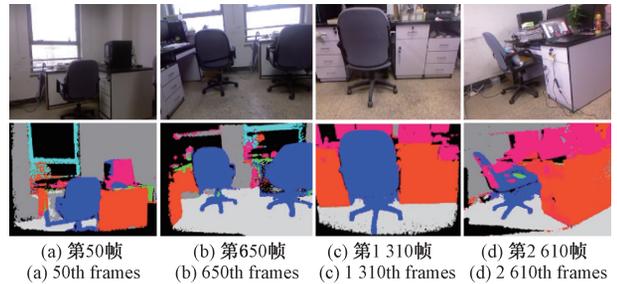


图 12 关键帧语义分割结果

Fig.12 The results of semantic segmentation about key-frame

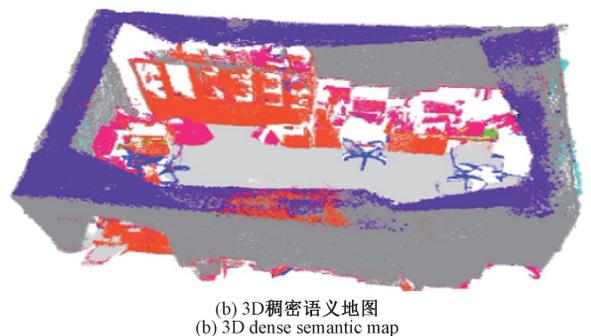


图 13 实验场景与 3D 稠密语义地图

Fig.13 The laboratory scene and 3D dense semantic map

6 结 论

本文提出了一个室内场景实时三维语义地图构建方法。首先,在室内场景三维重建方面,使用联合中值滤波算法实现实时地深度图像修复,采用投票式 RANSAC 算法与 ICP 算法相融合得到较好的相机估计位姿,关键帧图像使用随机藤类进行编码化处理实现闭环检测,识别曾到达区域。同时,采用优化的 DRN 模型对输入的图像进行像素语义级别的预测与分割。最后,采用基于贝叶斯更新的概率分布模型将图像中的语义标签迁移到三维模型中,获得完整的三维语义地图。通过对真实场景的语义地图构建实验,表明了本文设计的三维语义地图构建方法可以在实际的、复杂环境下实时地构建语义地图。由于部分结果存在零散分割区域,因此如何进一步提高图像分割精度,提高平均交并比是下一步研究的重点。在语义分割的过程中增加图像的空间信息、全局信息,引入条件随机场是下一步研究的重点。

参考文献

- [1] 唐宏伟,孙伟,杨凯,等.基于 SURF 特征的多机器人栅格地图拼接方法[J].电子测量与仪器学报,2017,31(6):859-868.
TANG H W, SUN W, YANG K, et al. Grid map merging approach of multi-robot based on SURF feature [J]. Journal of Electronic Measurement and Instrumentation, 2017, 31(6) : 859-868.
- [2] SCHWERTFEGER S, BIRK A. Map evaluation using matched topology graphs [J]. Autonomous Robots, 2016, 40(5) : 761-787.
- [3] NIETO-GRANDA C, ROGERS J G, TREVOR A J B, et al. Semantic map partitioning in indoor environments using regional analysis [C]. International Conference on Intelligent Robots and Systems, 2010, 1451-1456.
- [4] KOSTAVELIS I, GASTERATOSA. Learning spatially semantic representations for cognitive robot navigation [J]. Robotics and Autonomous Systems, 2013, 61(12) : 1460-1475.
- [5] RITUERTO A, MURILLO A C, GUERRERO J J. Semantic labeling for indoor topological mapping using a wearable catadioptric system [J]. Robotics and Autonomous Systems, 2014, 62(5) : 685-695.
- [6] 吴皓,田国会,薛英花,等.基于 QR code 技术的家庭半未知环境语义地图构建[J].模式识别与人工智能,2010,23(4):464-470.
WU H, TIAN G, XUE Y, et al. QR code based semantic map building in domestic semi-unknown environment [J]. Pattern Recognition and Artificial Intelligence, 2010, 23(4) : 464-470.
- [7] 吴皓,田国会,王家超,等.室内非结构化环境三维栅格语义地图的构建[J].模式识别与人工智能,2012(4):564-572.
WU H, TIAN G H, WANG J CH, et al. Three-dimensional grid semantic map building in unstructured indoor environment [J]. Pattern Recognition and Artificial Intelligence, 2012(4) : 564-572.
- [8] SHENG W, DU J, CHENG Q, et al. Robot semantic mapping through human activity recognition: A wearable sensing and computing approach [J]. Robotics and Autonomous Systems, 2015(68) : 47-58.
- [9] SONG S R, YU F, ZENG A. Semantic scene completion from a single depth image [C]. International Conference on Computer Vision and Pattern Recognition, 2017: 190-198.
- [10] 李秀智,李尚宇.实时的移动机器人语义地图构建系统[J].仪器仪表学报,2017,38(11):2769-2778.
LI X ZH, LI SH Y. A system of real time mobile robot semantic map building [J]. Chinese Journal of Scientific Instrument, 2017, 38(11) : 2769-2778.
- [11] LU F, MILIOS E. Globally consistent range scan alignment for environment mapping [J]. Autonomous Robots. 1997, 4(4) : 333-349.
- [12] GUTMANN J S, KONOLIGE K. Incremental mapping of large cyclic environments [C]. IEEE International Symposium on Computational Intelligence in Robotics and Automation, 1999.
- [13] VIJAYANAGAR K R, LOGHMAN M, KIM J, et al. Refinement of depth maps generated by low-cost depth sensors [C]. International SoC Design Conference, 2012: 355-358.
- [14] 李应彬,冯杰,张华熊,等.基于改进双边滤波的 Kinect 深度图像空洞修复算法研究[J].工业控制计算机,2013(11):105-106,109.
LI Y B, FENG J, ZHANG H X, et al. New algorithm of depth hole filling based on intensive bilateral filter [J]. Industrial Control Computer, 2013(11) : 105-106, 109.
- [15] BESL P J, MCKAY N D. A method for registration of 3-D shapes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1992, 14(2) : 239-256.
- [16] SHANJ C, LIX Z, JIA S M, LU X Y. Variational optical flow based velocity estimation for omni-directional intelligent wheelchair [C]. International Symposium on Computational Intelligence and Design, 2017: 418-422.
- [17] KUMMERLE R, GRISETTI G, STRASDAT H. G2o: A general framework for graph optimization [C]. International Conference on Robotics and Automation,

2011:3607-3613.

- [18] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [C]. International Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [19] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]. International Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
- [20] POMERLEAU F, MAGNENAT S, COLAS F, et al. Tracking a depth camera: Parameter exploration for fast ICP [C]. International Conference on Intelligent Robots and Systems, 2011: 3824-3829.
- [21] STEINBRUCKER F, STURM J, CREMERS D. Real-time visual odometry from dense RGB-D images [C]. International Conference on Computer Vision Workshops, 2011: 719-722.
- [22] SEGAL A, HAHNEL D, THRUN S. Generalized-ICP [C]. DBLP, 2009.
- [23] CARLONE L, KIRA Z. Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors [C]. IEEE International Conference on Robotics and Automation, 2014: 4290-4297.

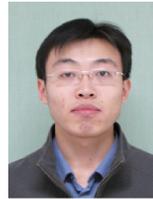
作者简介



单吉超(通信作者), 2015年于菏泽学院获得学士学位, 现为北京工业大学硕士研究生, 主要研究方向是机器视觉、计算机视觉。

Email: 1028406182@qq.com

Shan Jichao (Corresponding author) received his B. Sc. degree from Heze University in 2015. He is currently a M. Sc. candidate at Beijing University of Technology. His main research interests include machine vision and computer vision.



李秀智, 2008年于北京航空航天大学获得博士学位, 现为北京工业大学副教授、硕士生导师, 主要研究方向是智能机器人导航、机器视觉。

Email: xiuzhi.lee@163.com

Li Xiuzhi received his Ph. D. degree from Beihang University in 2008. He is currently an associate professor and M. Sc. supervisor at Beijing University of Technology. His research interests include intelligent robot navigation and computer vision.