

动态环境下的语义地图构建



齐少华 徐和根 万友文 付豪

同济大学电子与信息工程学院 上海 201804

(2316187741@qq.com)

摘要 三维语义地图在移动机器人的导航、路径规划、智能抓取、人机交互等任务中有着关键的作用,因此如何实时地构建三维语义地图尤为重要。当前同时定位和地图构建(Simultaneous Localization And Mapping,SLAM)算法已经可以达到较高的定位和制图精度,但是在动态环境下如何通过剔除动态物体来获得较高的定位精度,以及理解周围场景中存在的物体及其位置信息等问题没有得到很好的解决。在此,文中提出了一种可在动态环境下构建语义地图的算法。该算法在ORB-SLAM2上进行改进,在跟踪线程中加入动静点检测算法来剔除检测为动点的特征点,提高了动态环境下的定位精度;添加目标检测线程对关键图像进行目标检测,在地图构建线程中构建Octo-Map地图,同时根据检测结果构建3D目标数据库。为了证明该算法的可行性,以实验室为测试环境,分别进行了目标检测、动态点检测、三维目标信息获取和动态环境下语义地图构建的实验。在目标检测实验中,训练了速度和精度较高的目标检测网络——mobilenet-v2-ssdlite,检测速度可以达到7帧/秒,基本可以实现实时检测。在动态点检测中,采用光流法剔除动态点,处理速度为16.5帧/秒。文中创建了数据集来评测算法性能,相比原版ORB-SLAM2算法,结合光流法后的算法的定位精度提高了5倍;在三维目标信息获取上,采用了基于深度滤波和基于点云分割两种方法,结果表明后者的3D目标获取更为精确。最后,对整个实验室进行动态环境下的语义地图构建,构建Octo-Map稠密地图,根据检测结果构建3D目标数据库,并将目标尺寸和位置的检测值与真实值进行对比,误差均在5厘米以内。实验结果表明所提算法具有较高的精度和实时性。

关键词: 语义地图构建;动态点检测;目标检测;视觉SLAM

中图法分类号 TP249

Construction of Semantic Mapping in Dynamic Environments

QI Shao-hua, XU He-gen, WAN You-wen and FU Hao

College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China

Abstract Three-dimensional semantic maps play a key role in tasks such as robot navigation, path planning, intelligent grasping and human-computer interaction. So how to construct 3D semantic maps in real time is especially important. The current SLAM (simultaneous localization and mapping) algorithm can achieve higher positioning and mapping accuracy. However, how to eliminate dynamic objects to obtain higher positioning accuracy in a dynamic environment, and to understand the existence of objects and their location information in the surrounding scenes are still not well solved. This paper presents an algorithm for constructing semantic maps in a dynamic environment. This algorithm is improved on ORB-SLAM2. The dynamic and static point detection algorithm is added to the tracking thread to eliminate the feature points detected as dynamic feature points, which improves the positioning accuracy in dynamic environment. Object detection threads are added to detect key images. The mapping threads are added with the Octo-Map dense map construction. At the same time, the 3D object database is constructed according to the detection results. In order to prove the feasibility of the algorithm, the laboratory is used as the test environment, and the object detection, dynamic point detection, 3D target information acquisition, and semantic map construction experiments in the dynamic environment are performed. In the object detection experiment, a high-speed and high-precision object detection network, mobilenet-v2-ssdlite, is trained, which can reach a detection speed of 7 frames/s, which can basically achieve real-time detection. In dynamic point detection, the optical flow method was used to eliminate dynamic point, processing speed is 16.5 frames/s. And this paper creates a data set to evaluate the performance of the algorithm. Compared with the original ORB-SLAM2 algorithm, the positioning accuracy is improved by 5 times after combining with the optical flow method. For the acquisition of three-dimensional object information, two methods based on depth filtering and point cloud segmentation are adopted. The results show that the latter's 3D object acquisition is more accurate. Finally, the entire laboratory is constructed with a semantic map in a dynamic environment, an Octo-Map dense map is constructed, and a 3D object database is constructed based on the detection results. The detected values of the object size and position are compared with the true values, and the errors are within 5 cm. The results show that the proposed algorithm has high accuracy and real-time performance.

Keywords Semantic mapping, Dynamic point detection, Object detection, Visual SLAM

1 引言

机器人技术代表了当今世界高新技术的前进方向。机器人对周围环境的感知、建模和理解的能力是决定其智能水平的关键因素。在实际研究中,机器人常常装载着相机、GPS、激光雷达、超声波、IMU 等传感器。这些装置为机器人提供最原始的数据,但是还不能为机器人提供更加结构化、层次化的信息,需要经过多种算法的处理才能获取较好的结果。

在移动机器人技术领域,地图构建技术是感知、建模、规划与理解的基础与核心。同时定位与地图构建技术(SLAM)能利用传感器的数据进行自定位,与此同时构建出周围环境的地图。视觉 SLAM 是以相机为主要外部传感器的 SLAM 系统,有时还会结合其他传感器的信息,如 IMU 信息,可应用于机器人、无人驾驶和增强现实(Augmented Reality, AR)等领域,是近年来热门的研究方向。目前大多数的视觉 SLAM 算法是在静态场景这一假设下运行的,获得的信息中不包含高层的语义信息。通过 SLAM 算法,机器人可以精确定位,但无法对环境中的物体进行识别和建模,这导致机器人无法根据环境中的高层语义信息进行智能操作。随着室内外移动机器人技术、自动驾驶和增强现实技术等研究及应用的发展,室内外场景中语义理解能力的重要性愈加显著。

近几年,深度学习技术快速发展,为解决上述问题提供了新的思路。考虑快速性与实时性而提出的 SSD^[1] 目标检测框架在保证检测精度的同时提高了检测速度,由于其使用了多尺度预测组件,相比 YOLO 框架^[2] 提高了检测精度,并且由于其结构紧凑,相比 RCNN^[3] 框架提高了检测速度。为了在嵌入式平台上运行神经网络,近期研究者又提出了以 Mobile Net^[4-5] 为代表的小型化网络,这些网络通过巧妙设计网络结构、简化卷积核等方法使得网络的运算量降低。ORB-SLAM2^[6-7] 算法是近年来基于特征点法的视觉 SLAM 算法中较为出色的一个算法框架。输入视觉 SLAM 算法的图像信息流中常常包含各种物体,该算法结合了目标检测网络的语义信息提取优势与 SLAM 算法获取的精确的几何信息,使得机器人可以从周围环境中获取更加结构化、语义化、层次化的地图信息。近年来,一些研究根据目标检测结果的先验信息和动态点检测算法的测量信息来剔除相机视野中的动态点,以提高算法的定位精度,如 DynaSLAM^[8] 使用多视角几何和语义分割来剔除动态点,DS-SLAM^[9] 使用光流一致性来剔除动态点。

在语义构图方面,Salas-Moreno 等^[10] 提出的 SLAM++ 系统将点云特征与目标数据库进行对比,并在地图中插入目标的对应点云,但 SLAM++ 只能对预定义的物体进行建图。Hermans 等^[11] 基于贝叶斯更新和稠密条件随机场进行 2D-3D 的标签转换,生成三维空间中的语义地图。Concha 等^[12] 将语义分割与半稠密的大规模直接法(Large Scale Direct Method, LSD)进行单目 SLAM 融合,通过超像素分割得到图像平面,利用大规模直接法得到边缘等显著特征,并将两者进行融合得到较稠密的实时建图结果,但该方法的平面精确度不够理想。Tateno 等^[13] 提出将 CNN 预测与 SLAM 得到的深度图进行融合,从而得到单一视角下的语义地图,该方法主要侧重深度预测与融合。McCormac 等^[14] 利用稠密的 ElasticFusion 构建三维地图,以反卷积神经网络进行图像分割,该方法相对耗时,而且资源消耗很大。Jiang 等^[15] 实现了基于全局的条件随机场下的大规模道路场景致密语义地图构建

方法。Yu 等^[16] 在云平台上获取环境语义并完成了室内环境本体构建与推理,然后在标识归属物品划分的基础上实现了语义地图构建。Xing^[17] 首先进行未知环境的地图构建,然后通过人机对话的方式构建语义地图,最后进行路径规划。Tian 等^[18] 选择人工语义标签来提供环境功能信息和上下文等语义信息,并通过射频识别技术实现了环境语义信息的组织和机器人导航。Zhao 等^[9] 基于 RGB-D 相机,使用密集的条件随机场在全局范围内对空间进行语义分割来构建语义地图。

本文以复杂的室内实验室环境为背景,探索动态环境下构建语义地图的方法,结合基于 RGBD 相机的视觉 SLAM 系统和基于回归预测的多尺度预测的深度卷积神经网络 SSD 框架来设计算法,以剔除动态物体点、融合位姿信息和语义信息以及构建语义目标数据库,实时构建高层次语义地图。本文方法在动态环境下的语义地图构建等方面具有一定的参考价值,可以帮助机器人实现更为智能的导航任务。

2 动态信息检测和 3D 语义信息获取

本文提出 4 种算法来解决动态环境和语义地图构建问题。本文针对动态环境,提出了一种基于光流法的动态点检测算法来提高系统的定位精度;针对语义地图构建问题,提出了一种结合 2D 目标检测和 3D 点云信息的算法;考虑到目标检测算法的实时性要求,采用基于 Mobile Net 网络和 SSD 检测框架的算法;考虑到目标 3D 定位精度,推荐采用基于点云聚类分割的目标 3D 信息获取算法。

2.1 动态信息检测

当前的 SLAM 技术大多假设场景是静态的,如果有运动物体进入相机的视野,则会影响相机自身位姿的估计,运动物体的轨迹也将出现在点云图中,并且在将点云图转换为其他类型时动态轨迹信息将被保留。包含动态对象信息的地图不能直接用于导航,因此在图像序列中寻找动态对象的区域,并消除这些区域的影响,是一项具有挑战性的工作。

为消除相机运动带来的影响,本文首先设置一个光流幅值较大的阈值,来滤除噪声和相机运动带来的光流,但这样有一个缺点,即无法检测到相机视野中运动量过小的物体,但是处理的速度较快,同时可以排除一些运动量较大的物体;其次,对相机的运动进行建模,利用相邻两帧的特征点对求解出相机的单应变换矩阵 H , 然后对当前图像进行反变换后,再与上一帧图像一起求解光流场,这样可以有效剔除一些相机的运动,但在相机视角变化太大时,效果较差。矩阵 H 的求解可以使用 4 对点对联立线性方程组进行奇异值分解,由于匹配点对数量较多,需要使用 RANSAC (RANDOM SAMPLE Consensus) 方法得到更加鲁棒的结果,如图 1 所示。

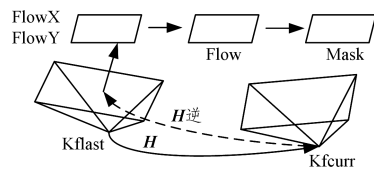


图 1 基于光流的动态点检测算法

Fig. 1 Dynamic point detection based on optical flow

实验中,先对两幅图像进行下采样,然后计算光流场,之后对光流图像进行上采样得到原始图像尺寸的光流场图像,这样可以大幅度提高计算速度,但是光流场的精度稍微有所降低。得到光流场之后,可以通过水平和垂直光流值计算光

流幅值 f_m , 然后使用阈值对光流场进行二值化, 再对二值形式的光流场 f_{mask} 进行腐蚀和膨胀操作, 以降低噪声。阈值是从实验中获得的数据, 相机静止不动时, 阈值的取值范围为 3~4, 相机运动时其取值范围为 6~9; 也可以根据相机的运动量大小设置自适应阈值。若阈值过大, 则算法会对一些运动量小的物体不灵敏, 若阈值过小, 则算法会被过多的噪声所影响。

2.2 2D 目标检测

考虑到直接对点云进行 3D 语义分割的复杂性以及低实时性, 本文采用 2D 目标检测结合点云的方式获取 3D 目标信息。我们用 Mobile Net 网络作为 SSD 目标检测网络的前端特征提取网络, 其局部结构如图 2 所示。

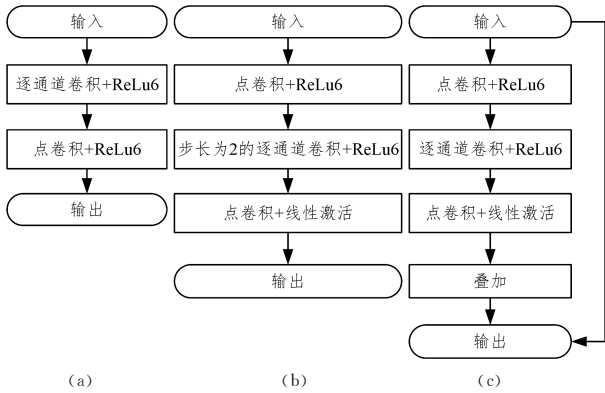


图 2 Mobile Net 的结构

Fig. 2 Structures of Mobile Net

图 2(a) 给出了 Mobile Net-v1 结构, 图 2(b) 和图 2(c) 分别给出了 Mobile Net-v2 的带下采样和无下采样的两个版本。Mobile Net 的核心思想是使用逐通道卷积代替普通卷积降低运算量, 提高运算速度。本文结合 Mobile Net 和 SSD 形成的 Mobile Net-SSD 目标检测网络可以在普通 PC 机上使用 NC-NN 平台近似实时运行。整体网络结构如图 3 所示。

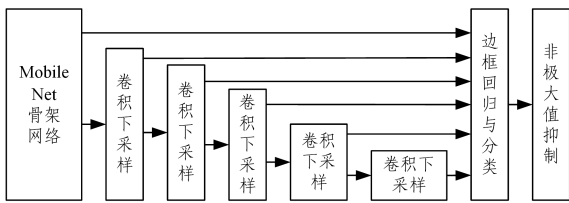


图 3 Mobile Net-SSD 目标检测网络

Fig. 3 Mobile Net-SSD object detection network

2.3 3D 语义信息获取

本文使用 2.2 节中目标检测的 2D 检测结果和 RGB-D 相机捕获的点云来获取 3D 语义目标信息。点云获取有两种思路: 1) 以 2D 检测信息为主在有序点云上框取对应的点云, 经过滤波等处理获取点云团的位置和空间尺寸信息; 2) 以点云为主, 对点云进行分割后获取有序点云团, 并将其投影在图像平面上的 2D 框, 再与目标检测的 2D 框进行对比, 根据交并比计算相似性进行匹配。前者根据 RANSAC 方法在对应检测框内选取目标物体点云的平均深度, 根据平均深度对 2D 框内的点云进行滤波后计算中心点、3D 包围框尺寸等信息, 具体方法如图 4(a) 所示。后者, 先对点云进行滤波、分割聚类, 然后将点云分解成独立的点云团, 再将点云团反投影到 RGB 图像平面, 计算点云团对应投影点的 2D 矩形框, 最后计

算与目标检测的 2D 矩形框的相似度 (IOU 交并比等), 每一个检测框匹配一个点云团进而计算该点云团的 3D 信息。该方法的具体流程如图 4(b) 所示。

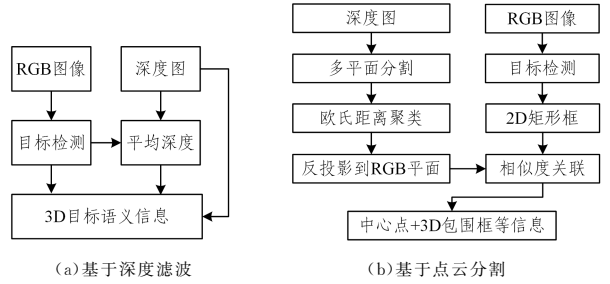


图 4 基于点云分割 3D 语义信息获取

Fig. 4 3D semantic information acquisition

3 动态语义地图构建

在 SLAM 建图过程中, 动态目标会影响相机位姿的估计, 而且动态目标的轨迹将出现在点云图中, 影响地图的构建。我们在 ORB-SLAM2 框架的基础上进行改进, 加入动态点检测算法, 以实现动态环境下的语义地图构建过程, 完成动态环境下的目标识别定位实验, 算法流程如图 5 所示。基于 ORB-SLAM2, 图 5 添加了动静点检测、SSD 目标检测、Octo-map 动态建图和目标数据库构建等模块。在跟踪线程的特征提取后加入动静点检测算法, 可使用基于光流的算法以及基于多视角几何的算法, 之后剔除落入动态点 mask 内部的特征点, 再使用过滤后的特征点进行数据关联、位姿优化, 以获取当前帧的位姿, 之后再根据一些条件选取关键帧, 在此引入一个约束, 避免当前图像与地图的相似性太大。然后, 新建一个线程对关键图像进行目标检测, 在原局部地图线程中加入 Octo-Map^[20] 三维占有格概率地图构建任务, 对地图进行动态更新, 同时结合 2D 目标检测结果, 使用 2.3 节中的算法获取目标的 3D 信息, 构建一个目标数据库。

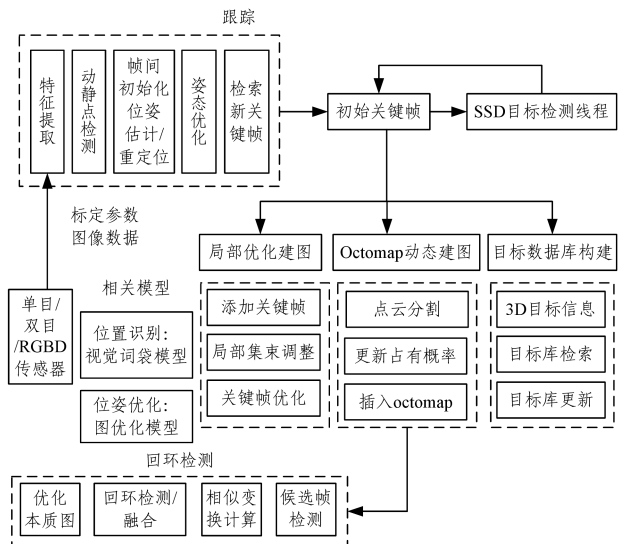


图 5 动态环境下语义地图构建算法 (电子版为彩色)

Fig. 5 Semantic map construction algorithm in dynamic environment

Octo-Map 将地图按八叉树结构进行存储, 最小的叶子结

点尺寸由分辨率参数决定,由于环境中存在动态物体,因此一个叶子的占有概率应由多次观测的联合概率决定,我们对联合概率进行 logit 变换(见式(1)),得到概率更新的简化方式。

$$\alpha = \log(p/1-p) \quad (1)$$

上述 logit 变换的反变换为 sigmoid 函数形式,如式(2)所示:

$$p = 1/(1 + \exp(-\alpha)) \quad (2)$$

概率更新的简化方式如式(3)所示:

$$L(n|Z_{1:T}) = L(n|Z_{1:T-1}) + L(n|Z_T) \quad (3)$$

其中, $L(n|Z_{1:T-1})$ 为前面时刻占有格的 α 值,其加上最近一次的 α 值 $L(n|Z_T)$ 可以得到当前综合的 α 值,再经过式(2)的反变换可以得到占有格的占有概率值。 $L(n|Z_T)$ 由式(4)决定。

$$L(n|Z_T) = \begin{cases} l_{occ}, & \text{节点 } n \text{ 被观测到} \\ 0, & \text{节点 } n \text{ 未被观测到} \\ l_{free}, & \text{地面等不考虑} \end{cases} \quad (4)$$

其中, l_{occ} 与 l_{free} 是预定义的值,这里分别取 0.85 和 0.4,分别对应占有概率值 0.7 和 0.4,而 0 对应的概率值为 0.5。如果节点 n 经常被观测到,其占有概率会上升,而未被观测到的以及不被考虑的点(地面点以及动态点)的占有概率会相对减小。

在目标数据库的构建过程中,首先生成每一帧观测中存在的 3D 目标物体,然后将其加入数据库。当下一次检测到相同类别时,首先根据位置信息判断是否是同一个物体,如果是则更新该物体的空间坐标等信息,否则记录为新物体。语义目标数据库的构建流程如图 6 所示。

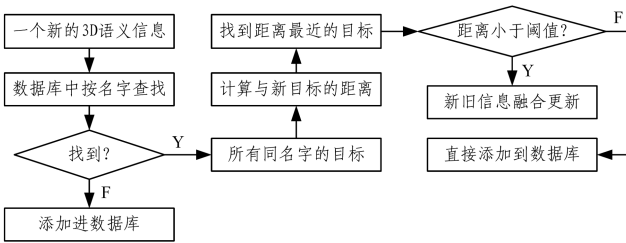


图 6 语义目标数据库的构建流程

Fig. 6 Construction process of semantic object database

4 实验

4.1 目标检测网络的构建

为了使系统具有实时性,本文训练了速度和精度较高的目标检测网络——mobilenet-v2-ssdlite,网络先在 Image net 数据集上进行预训练,然后迁移到自制数据集上进行目标检测方面的训练,自制数据集的物体类别有:椅子、显示器、柜子等 10 类实验室常见目标。对于网络的前向推理过程,我们选择了嵌入式优化等级较高的库——NCNN。实验的平台为 Ubuntu 16.04 的系统环境,处理器型号为 Intel i7-8750H,6 核,内存为 16 GB,使用的 RGB-D 摄像头型号为图漾 FM810-HD,其原理是双目结构光。对各尺寸图像进行 1000 次检测的平均检测时间如表 1 所列。

表 1 目标检测网络的运行时间

Table 1 Running time of objection detection network

Image Size	Time/ms
320 * 240	140
640 * 480	141
1280 * 720	143
1280 * 960	144

图像均被缩放到 300 * 300 后输入网络中,因此各尺寸图像的运行时间大致相同。由表 1 可以看出,该算法可以在 1 秒内处理 7 帧图像,因此能够实现近似实时的目标检测,具有较强的实际应用价值。测试结果如图 7 所示。

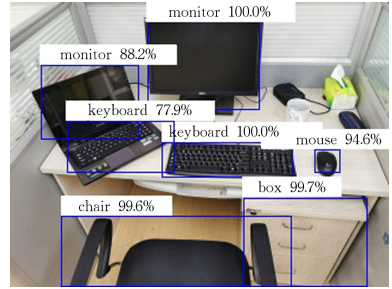


图 7 目标检测的网络测试结果

Fig. 7 Network test results of object detection

4.2 动态点检测

TUM RGB-D^[21] 数据集是使用 Microsoft Kinetic 传感器在不同室内场景下录制的包含 39 个序列的数据集,包含 RGB 图像和深度图像以及由高精度运动捕捉系统得到的真实的相机轨迹。我们在 walking(w) 的序列中进行实验,该序列包含 833 张彩色图和深度图,其场景是有两个人从后方走到前方并坐到椅子上。基于光流算法的动态点检测算法的检测结果如图 8 所示。该算法耗时 56 秒,平均每秒处理 16.5 帧,虽然相比原 ORB-SLAM2 算法(耗时 31 秒)速度有所降低,但也能满足实时的定位需求。



图 8 基于光流的动态点检测算法

Fig. 8 Dynamic point detection algorithm based on optical flow

为了评测算法的定位性能,我们创建了一个包含 1918 张彩色图和深度图的数据集,该数据集是用相机在 3.6 m * 2.7 m 的矩形轨迹上水平运动得到的。将原 ORB-SLAM2 算法和加入光流动态点检测的相机定位结果进行对比,结果如图 9 所示。

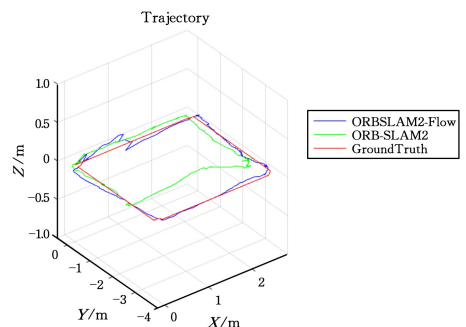


图 9 各算法结果的对比

Fig. 9 Comparisons of each algorithm results

其中,groundTruth 代表运动捕捉设备记录的相机真实轨迹,ORB-SLAM2 代表原 ORB-SLAM2 算法获得的相机轨迹,ORB-SLAM2-Flow 代表使用光流算法剔除动态点的 ORB-SLAM2 算法得到的相机轨迹。算法结果的误差对比分析如表 2 所列。

表 2 算法误差分析

Table 2 Algorithm error analysis

(单位:m)					
Algorithm	RMS	average	Standard Deviation	MIN	MAX
ORB-SLAM2	0.49	0.37	0.32	0.02	1.22
Slam2+flow	0.10	0.07	0.07	0.00	0.45

由于真实轨迹录制时的坐标系和算法一开始的坐标系存在差异,因此算法估计的相机轨迹和真实轨迹之间存在一个欧氏变换,故对估计值和真实值进行配准后,需要求解真实值和匹配的估计值之间的欧氏变换。表 2 列出了对估计值进行欧氏变换后再与真实值计算三维点距离的差值,对比了误差的均方根、均值、标准差值等信息。可以看出,光流法在保持实时性的同时相对原 ORB-SLAM2 算法精度提高了 5 倍。

4.3 三维目标信息的获取

对 2.3 节 3D 语义信息获取的两种算法进行实验。基于深度信息滤波的方法的处理结果如图 10 所示。



图 10 基于深度滤波的算法结果

Fig. 10 Result of algorithm based on depth filtering

可以看出,由于目标检测算法获取的 2D 框包含了较多其他非目标的区域,容易得到很多噪声点云,获取的 3D 框和位置信息不精确。而对于基于点云分割的算法,只要目标物体放置得不太紧密都能获得较好的分割结果,如图 11 所示。



图 11 基于点云分割的算法结果

Fig. 11 Result of algorithm based on point cloud segmentation

图 11 中,一个人坐在椅子上,与椅子有重叠,该算法在人与椅子靠背有间隙的情况下也能获得较好的分割结果,可以看出只要物体不是紧密接触,该方法就能获取较为精确的识别和定位精度。

4.4 动态环境下的语义地图构建

本节在日常实验室场景下进行动态语义地图的地图构建实验,实验中有部分人在走动,同时有椅子被移动的情况发生,地图构建的结果如图 12 所示。

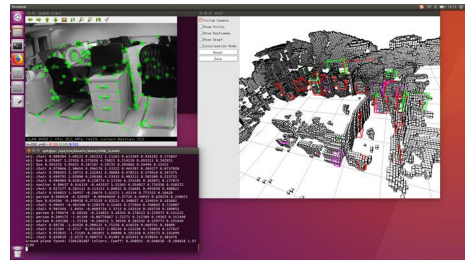


图 12 动态环境下语义地图构建结果(电子版为彩色)

Fig. 12 Semantic map construction result in dynamic environment

本文使用基于光流法的动静点检测算法提高位姿估计精度,同时对关键图像使用目标检测算法进行 2D 目标检测,使用基于点云分割的 3D 目标检测框生成算法生成 3D 目标信息,使用 3D 目标信息构建目标数据库,然后根据点云分割信息更新 Octo-Map 地图(地面和动态点云概率降低,非地面点云概率升高,因此可结合目标检测信息),在节点占有概率为 0.8 时,更新 Octo-Map 地图的相应节点。

图 12 中,检测到的目标物体以相应颜色(红色表示椅子,绿色表示显示器等)的立体框标出,并显示出目标数据库信息。数据库包含的信息有物体类别、类别概率值位置坐标以及 3D 包围框尺寸。从建立的语义目标数据库中截取部分有代表性的值,测量目标对应的真实三维坐标及尺寸,并计算相应的误差,结果如表 3 所列。可以看出,目标的估计位置及尺寸与真实位置及尺寸的误差均控制在 5 厘米以内。

表 3 目标数据库结果

Table 3 Object database results

(单位:m)

object	value	X	Y	Z	长	宽	高
monitor	result	0.19	-0.02	1.01	0.26	0.13	0.06
	truth	0.20	-0.03	1.00	0.22	0.15	0.05
	error	0.01	0.01	0.01	0.04	0.02	0.01
chair	result	1.17	0.02	1.01	0.48	0.55	0.13
	truth	1.19	0.20	1.00	0.44	0.57	0.15
	error	0.02	0.18	0.01	0.04	0.02	0.02
box	result	2.39	0.09	0.96	0.44	0.46	0.04
	truth	2.32	0.08	0.95	0.40	0.50	0.03
	error	0.03	0.01	0.01	0.04	0.04	0.01

结束语 本文基于高精度的 ORB-SLAM2 算法、MobileNet-SSD 目标检测算法、动态点检测算法、3D 目标信息提取算法以及动态地图更新算法,在动态环境下实时构建语义地图。本文的主要贡献为提出了一种简单的基于光流算法的动态点检测算法与两种 3D 目标信息获取算法。本文将视觉 SLAM 算法获取的精确定位信息和目标检测网络得到的语义信息有效地结合起来,实现了语义信息和空间点云团之间的一一映射,构建了一个有效的语义目标数据库,可以为机器人智能导航提供有效的语义地图信息,所提算法能够实现近似实时运行,具有较强的实际应用价值。

算法生成的语义地图可应用于两个方面。1)用于机器人智能导航抓取等。语义地图使得机器人具有语义理解的能力,可以用语义的形式给机器人发送指令,如语音命令。机器人解析语义命令后,在语义数据库中查询指定的目标,得到与语义地图相关联的目标信息。机器人根据 SLAM 的定位信息以及语义目标坐标信息,选择合适的路径规划算法到达指

定的目标位置进行操作。2)用于数据采集扩充训练数据集。传统的目标检测等数据集中的大部分图片是从正面拍摄的,缺乏各个角度的信息。利用本文算法可以得到目标在各个角度、位置上的图像,所得数据分布在更多角度和尺度之上,对训练复杂、高精度的模型提供数据上的支持。

针对算法出现的误差,未来的研究方向有:1)如何更有效地解决遮挡问题。本文基于点云分割的3D目标信息获取算法在物体放置不紧密的情况下能够获得很高的精度,但是在物体紧密放置情况下,由于欧氏距离分割算法不能将物体分割开,导致分割错误,因此可以采用颜色凸凹性的超体聚类分割算法来实现更精细的分割。2)如何有效地检测出动态点。在相机定位过程中,动态关键点的存在会干扰相机的定位,如何在相机定位中剔除动态点是提高相机在动态环境中定位精度的关键。本文提出的基于光流算法的动点检测算法速度快,但是精度低,而且容易产生很多误检测,因此可以考虑采用基于多视角几何的算法来减少误检问题。

参 考 文 献

- [1] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]// European Conference on Computer Vision. Berlin, German: Springer, 2016: 21-37.
- [2] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[J]. arXiv: 1506. 02640 v4, 2015.
- [3] 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 & Machine Intelligence, 2015, 39(6): 1137-1149.
- [4] HOWARD A, ZHU M, CHEN B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv: 1704. 04861, 2017.
- [5] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks [J]. arXiv: 1801. 04381 [cs. CV], 2018.
- [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, TARDOS 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] BERTA B, FACIL J M, JAVIER C, et al. DynaSLAM: Tracking, Mapping and inpainting in Dynamic Scenes[J]. arXiv: 1806. 05620v1, 2018.
- [9] YU C, LIU Z, LIU X, et al. DS-SLAM: A Semantic Visual SLAM towards Dynamic Environments [J]. arXiv: 1809. 08379v2, 2018.
- [10] SALAS-MORENO R F, NEWCOMBE R A, STRASDAT H, et al. SLAM++: Simultaneous localisation and mapping at the level of objects[C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2013: 1352-1359.
- [11] HERMANS A, FLOROS G, LEIBE B. Dense 3D semantic mapping of indoor scenes from RGB-D images[C]// IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2014: 2631-2638.
- [12] CONCHA A, CIVERA J. DPPTAM: Dense piecewise planar tracking and mapping from a monocular sequence[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2015: 5686-5693.
- [13] TATENO K, TOMBARI F, LAINA I, et al. CNN-SLAM: Real-time dense monocular SLAM with learned depth prediction [C]// IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 6565-6574.
- [14] MCCORMAC J, HANDA A, DAVISON A, et al. Semantic Fusion Dense 3D semantic mapping with convolutional neural networks[C]// IEEE International Conference on Robotics and Automation. Piscataway, USA: IEEE, 2017: 4628-4635.
- [15] JIANG W T. Construction of dense semantic map of large-scale road[D]. Hangzhou: Zhejiang University, 2016.
- [16] YU J S, WU H, TIAN G H, et al. Semantic database design and semantic map construction of robots based on the cloud[J]. Robot, 2016, 38(4): 410-419.
- [17] XING Q W. Construction of robot semantic map under unknown environment [D]. Changchun: Northeast Normal University, 2017.
- [18] HAO W, GUOHUI T, PENG D, et al. Characterization of large-scale unknown environmental information based on RFID technology[J]. Journal of Central South University(Natural Science Edition), 2013(S1): 166-170.
- [19] ZHAO Z, CHEN X. Building 3D semantic maps for mobile robots using RGB-D camera [J]. Intelligent Service Robotics, 2016, 9(4): 1-13.
- [20] HORNUNG A, WURM K M, BENNEWITZ M, et al. OctoMap: an efficient probabilistic 3D mapping framework based on octrees[J]. Autonomous Robots, 2013, 34(3): 189-206.
- [21] JÜRGEN S, ENGELHARD N, ENDRES F, et al. A benchmark for the evaluation of RGB-D SLAM systems[C]// 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2012.



Qi Shao-hua, born in 1995, postgraduate. His main research interests include visual SLAM, deep learning and augmented reality.



Xu He-gen, born in 1972, Ph.D, associate professor. His main research interests include image processing, machine vision, pattern recognition and robot.