

基于 SLAM 算法和深度神经网络的语义地图构建研究

白云汉

(复旦大学计算机科学与技术学院 上海 200433)

摘要 语义地图在移动机器人的导航等任务中有着关键的作用。目前基于视觉的机器人自动定位和制图 (SLAM) 系统已经能够达到较高的定位精度要求,但是基于多目几何的视觉 SLAM 算法并未充分利用空间中丰富的语义信息,地图中保留的地图点信息只是没有语义的空间几何点。由于基于卷积神经网络的算法在目标检测领域取得了突破性的成绩,利用目前最新的基于卷积神经网络的目标检测算法 YOLO,实现场景的实时目标检测,并结合 SLAM 算法构建语义地图。将视觉 SLAM 定位的精确性和深度神经网络在语义提取方面的优势相结合,既能够提高 SLAM 算法的准确性,同时采集到的丰富图像信息又能进一步训练更加深的神经网络。该算法能够应用于机器人的智能导航,同时也能作为数据采集器为其他视觉任务提供具备语义与几何信息的图像数据。

关键词 语义地图 机器人定位与导航 目标检测 深度学习

中图分类号 TP391

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2018.01.032

RESEARCH ON SEMANTIC MAPPING BASED ON SLAM ALGORITHM AND DEEP NERUAL NETWORK

Bai Yunhan

(School of Computer Science, Fudan University, Shanghai 200433, China)

Abstract Semantic mapping plays a key role in the task of mobile robot navigation. At present, the vision-based SLAM system has been able to achieve higher accuracy requirements, but the visual SLAM algorithm based on multi-view geometry cannot use the rich semantic information in space. The map point information reserved in the map is only spatial geometric point without semantic information. Since the algorithm based on convolution neural network has made a breakthrough in the field of object detection, the semantic mapping is constructed using the latest YOLO algorithm of object detection based on convolution neural network, which realized the real-time object detection in scene, combined with the SLAM algorithm. The proposed approach combines the accuracy of visual SLAM and the advantages of deep neural network in semantic extraction, which can improve the accuracy of SLAM algorithm, and the rich images collected can further help us to train deeper neural networks. This algorithm can be applied to the intelligent navigation of the robot, but also as a data collector for other visual tasks to collect the semantic and geometric information with the image data.

Keywords Semantic mapping SLAM Object detection Deep learning

0 引言

在机器人学的领域,一个重要的问题是如何使得机器人获得自身与环境的信息,因为只有机器人对自身和环境正确的建模后,机器人才可能完成其他任务,

如导航等。为了解决这个问题,首先要使得机器人具备感知环境的能力。在实际研究中,机器人常常配备有相机、GPS 设备、激光、声纳等传感器。这些传感器为机器人提供了原始的数据,但是还无法为机器人提供更加结构化和有意义的信息。

机器人自动定位和制图 (SLAM) 所要解决的是如

何利用传感器的原始信息来对机器人的位置与其所处环境的地图信息进行同时估计的问题。SLAM 算法的输入是连续的传感器信息,如视觉 SLAM 中,输入信息是配备相机的机器人在环境中连续采集到的图像帧序列。输出是机器人当前的位置以及机器人所处环境的地图点位置。但是在主流的 SLAM 算法中,机器人的位置以及地图点信息只是空间中密集或稀疏的几何点。通过对这些空间点的位置估计能够为我们提供相对精确的位置信息,但是无法提供更高层次的语义信息。机器人可以利用 SLAM 算法精确估计自己的位置,但是无法对空间中存在的物体进行识别和建模,这将导致在后续的任务中,机器人无法利用环境中丰富的语义信息。

目前深度学习方面的进展为解决这个问题提供了一个方向。深度神经网络强大的特征学习能力使得图像识别,目标检测领域取得了显著的进展,在图像识别数据集 ImageNet^[1] 的比赛中,基于 CNN^[2] 的深度学习模型已经取得了超过人类的成绩^[3],在目标检测方面以 RCNN^[4] 为代表的深度学习方法使得目标检测的准确性大幅度提升。视觉 SLAM 算法处理的图像序列常常包含了各类物体,结合深度神经网络的语义识别优势与 SLAM 算法提供的位置信息能够获得有关机器人自身和环境更加丰富的信息。

本文结合了基于单目视觉的 SLAM 系统和基于回归预测的深度卷积神经网络,设计并实现算法实现对语义地图的构建,将位置信息和语义信息融合。利用本文的算法,能够帮助机器人实现更为智能的导航任务。同时利用本文的算法机器人能够收集到有关物体的更加丰富的数据,这将进一步促进深度神经网络的性能提升。

1 相关工作

SLAM 算法是语义地图构建的基础。目前基于视觉的 SLAM 算法大致分为两个取向。一种是稀疏 SLAM。稀疏 SLAM 算法通过对视觉信息进行点特征提取,对提取的点特征进行匹配,从而实现对机器人位置的估计,同时构建环境中的稀疏地图点。稀疏 SLAM 的算法又可以大致分为基于滤波的方法,如 MONOSLAM^[6] 等。基于图优化的 SLAM 算法,如 PTAM^[7]、OKVIS^[8] 等。另外一个 SLAM 算法的研究取向是稠密 SLAM。稠密 SLAM 算法不对图像进行点特征提取,而是直接利用图像的梯度信息对机器人相机位置进行估计。稠密 SLAM 节约了特征提取过程的时间,并且利用了图像的所有像素信息,方便建立空间的

稠密地图,实现对空间的 3D 重构,但是由于要使用所有的像素信息,计算量常常过大,需要利用 GPU 并行加速。

以稀疏 SLAM^[9] 为例子,SLAM 算法可以分为前端和后端两部分。前端部分主要负责数据关联问题,后端部分负责对位置信息进行优化估计。本文采用了基于 ORB 特征的 ORBSLAM 算法,前端部分在图像提取 ORB 特征,与地图中的 3D 地图点进行匹配,后端部分利用匹配的结果构建一个因子图,在因子图上进行优化计算。具体流程如图 1 所示。

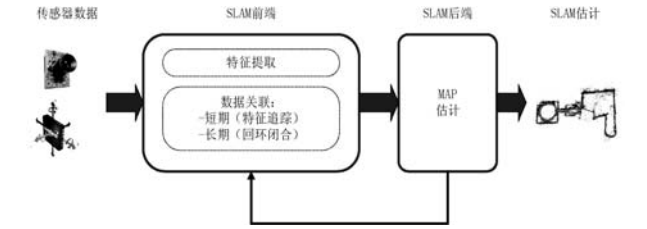


图 1 SLAM 算法的总体流程

语义地图的构建在 SLAM 算法之上,但是不同于 SLAM 算法,语义地图的构建需要对人类的概念如场景、物体、形状等进行抽象。语义地图构建的需求来自于机器人导航的需要,Kuipers^[10] 最早提出建立语义地图,他提出对空间知识进行建模的概念。随着近些年来 SLAM 算法越来越精确,将高精度的 SLAM 算法和语义地图结合成为研究的热点。Nielsen 等^[11] 将基于 SLAM 的语义地图构建看作是机器人和人类交互的界面。他提出将机器人的单帧图像信息与机器人的位置信息合并存储,将其作为几何地图的补充。

Galindo 等^[12] 将地图分为空间地图和概念地图,并采用分层的方式组织地图,并对空间地图节点和概念地图节点采用锚定,方便机器人根据语义信息进行导航。Civera 等^[13] 利用单目视觉 SLAM 算法建立起空间的栅格地图,在另外一个线程中实现物体识别算法,将识别结果作为地图数据库存储 Liu 等^[14] 则是使用 2D 的激光扫描建立起空间 occupancy 栅格地图,然后在此基础上建立起语义地图。斯坦福大学的 Carl 等^[15] 将 SLAM 地图与地图中的文字标签信息(如办公室门牌)结合,实现了机器人根据语义指示导航的功能。Rituerto 等^[16] 通过使用 catadioptric 视觉系统,实现了对拓扑地图进行语义标注。Fasola 等^[17] 的系统使用语义信息编码空间位置的相对关系,可以使得人类用自然语言和机器进行交互。

目前基于深度神经网络的视觉目标检测算法取得了巨大的突破,为我们构建语义地图提供了直接而准确的语义信息。本文将利用单目 SLAM 算法和目标检测算法,设计并实现语义地图构建算法,对地图点进行

语义关联,实现构建高精度语义地图。

2 SLAM 算法与目标检测算法

2.1 基于 ORB-SLAM 的 SLAM 算法

ORB-SLAM^[18]是目前基于关键帧的稀疏 SLAM 算法中性能较为出色的一个算法框架,ORB-SLAM 将 SLAM 算法分为跟踪、本地制图、回环修正三个线程。跟踪模块主要解决的是 SLAM 算法前端中的连续图像帧的数据关联问题,并且会对当前帧的位置进行优化估计。本地制图模块要解决的是地图点的创建与更新,回环检测的目标是消除 SLAM 算法中累计的误差。相比于其他基于关键帧的稀疏 SLAM 算法,ORB-SLAM 主要有这几个优势:

- 1) 采用 ORB^[19]特征作为视觉特征,具有提取速度快,视角光照不变性好的特点。
- 2) 采用 DBow2^[20]词带模型用于快速重定位和回环检测,提高了系统的鲁棒性。
- 3) 提出了共视图和本征图的概念,简化了地图构建中地图点和关键帧之间关联。
- 4) 在关键帧的选择机制,地图点的创建机制以及本地窗口的大小设定上更加合理,带来了性能上的提升。

在本文中,采用了 ORB-SLAM 作为我们的 SLAM 算法,并且修改地图点及关键帧的数据结构方便结合语义地图。

2.2 目标检测算法

传统的目标检测算法常常分为三个部分。

- 1) 选择检测的窗口。这一步对图像进行多位置和多尺度的窗口提取,采用 Selective Search^[21]等提取方法,并且基于颜色聚类、边缘聚类把无关区域去除。

- 2) 提取视觉特征。特征提取指的是在区域内提取视觉特征,常用的特征有 SIFT^[22]、HOG^[23]等。

- 3) 分类器分类。分类器分类指的是利用提取中的特征,使用机器学习模型对所得特征所属种类进行分类。常用的分类器有 SVM、随机森林、神经网络等。

传统的方法将检测分为这三个部分,本身不是一个端到端的算法,这将导致各个部分之间无法很好的学习和优化。并且在特征提取方面,传统方法提取的特征在分类任务中无法和基于深度卷积神经网络提取出的特征的性能相比。

在语义地图构建中,需要目标检测算法为我们提供原始的语义信息,并且因为 SLAM 算法的实时性,对算法的运行速度有着较高的要求。我们参照了基于深度

神经网络的 YOLO^[24]设计我们的网络,将输出层物体种类设为室内常见的 20 类物体,具体内容见实验部分。

YOLO 将物体检测作为一个回归问题求解,基于单一的端到端网络,完成对物体位置以及物体类别的同时预测。YOLO 与 RCNN^[29]、Fast RCNN^[28]、Faster RCNN^[27]不同之处在于它不需要显性地提取 region proposal,这使得端到端学习成为了可能。

YOLO 借鉴了 GoogleNet^[25]的分类网络结构,不同的是 YOLO 未使用 inception module,而是用了 1×1 的卷积层加上 3×3 的卷积层简单代替。YOLO 将输出图像分为 $S \times S$ 个格子,每个格子负责检测落入该格子的物体,若某个物体的中心位置坐标落在某个格子中,那么这个格子就负责检测该物体。每个格子负责输出 B 个 bounding box 信息,以及 C 个物体属于某种类别的概率信息。每个 bounding box 有 5 个数值,分别代表当前格子预测得到的 bounding box 的中心坐标,宽度和高度,以及当前 bounding box 包含物体的置信度。因此最后全连接层输出的维度为 $(S \times S) \times (B \times 5 + C)$ 。

在本文的网络中,我们的分类种类 $C = 5$,并设置 $B = 2, S = 13$ 。本文网络的损失函数由三部分构成,分别为预测数据与标定数据之间的坐标误差、IOU 误差和分类误差:

$$Loss = \sum_{i=0}^{S^2} coorError + iouError + classError \quad (1)$$

具体的损失函数如下:

$$Loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} I_{ij}^{noobj} (C_i - \hat{C}_i)^2 + \sum_{i=0}^{S^2} \sum_{j=0}^B I_{ij}^{obj} \sum_{c \in classes} (p_i(c) - \hat{p}_i(c))^2 \quad (2)$$

式中: x, y, w, h 为网络预测是坐标信息,其中 x, y 指的是预测中心位置, w, h 为 bounding box 的宽和高。对应的是标签信息。 I_{ij}^{obj} 代表物体落入格子 i 的第 j 个 bounding box 里, I_{ij}^{noobj} 代表未落入。 C 代表是预测的置信度, \hat{C} 为标注值。

在式(2)中,第 1、2 行代表是对坐标预测的惩罚。在本文网络中,对坐标预测的惩罚加了一个参数 λ_{coord} 进行增强。这是因为如果不加这个参数,那么在损失函数中,预测位置的重要性就和预测是否包含物体的重要性相等,这不符合最大化平均精度的想法。同时

在 3,4 行计算 IOU 误差的时候,包含物体的格子和不包含物体的格子两者的 IOU 误差对 Loss 的贡献应该是不同的。若采用相同的权值,那么不包含物体的格子的 confidence 近似为 0,变相地放大了包含物体的格子的 confidence 误差在计算梯度时的影响。因此将不包含物体的格子的 IOU 误差加上参数 λ_{noobj} 进行削弱。最后一行为分类的惩罚。我们将 λ_{coord} 设为 5.0,而将 λ_{noobj} 设为 0.5。

3 语义地图构建

在 SLAM 算法中,图像特征与地图点直接对应。而本文的目标检测算法检测出的目标框内有图像特征,因此可以将存储的地图点与目标物体对应。如图 2 所示,每个地图点与一个目标物体对应,而每个关键帧通过地图点间接地和一个或多个目标相连接。由此建立图像数据和目标物体之间的连接关系。

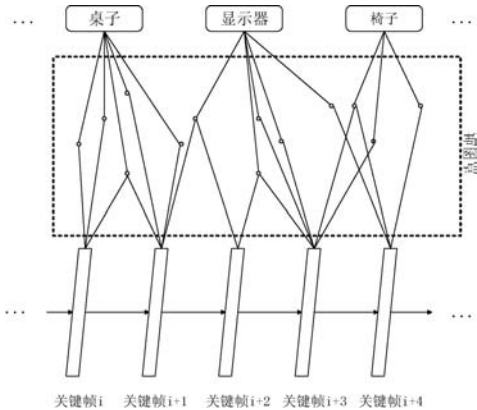


图 2 关键帧、空间地图点和语义信息的关系图
语义地图构建算法流程如图 3 所示。

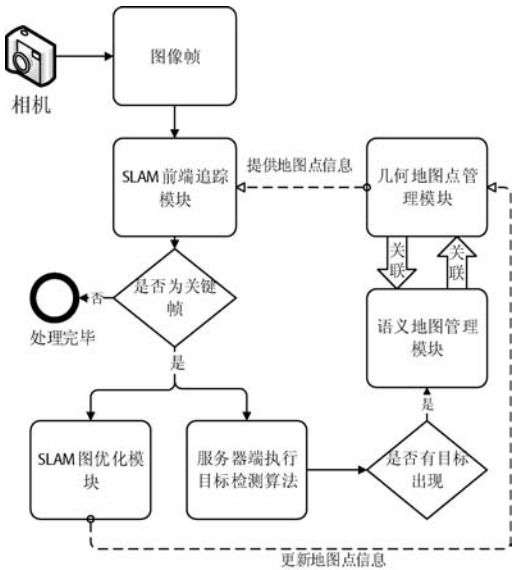


图 3 语义地图构建算法流程图

首先,SLAM 算法前端追踪模块决定是否将当前

图像帧当做关键帧,关键帧选择规则如下:

- 1) 至少离上一次重定位距离 20 帧。
- 2) 本地制图模块空闲或离上一次关键帧插入相差已有 20 帧。
- 3) 当前帧中至少有 50 个特征点。

其次,当前帧被选择为关键帧之后,关键帧将被两部分算法同时处理:第一部分是目标检测算法,第二部分是 SLAM 后端图优化算法。在进行语义地图模块之前,需要等待 SLAM 后端图优化模块完成,因为后端图优化模块将可能产生新地图点。SLAM 的后端图优化模块主要包括两部分:

- 1) 当前关键帧位置与本地地图点的优化更新。
- 2) 通过三角法创建新的空间地图点。

最后启动语义地图管理模块。语义地图管理模块将进行两方面的工作:

- 一是对地图中不可用点进行区分和标记。例如将检测为 people、cat、dog、bicycle、toy 这几个类别目标框中的特征点标记为不可用。这是由于这些类别的物体常常发生移动,因此不应该将其上面的点作为地图点。
- 二是建立地图点和语义信息的联系。通过查找目标物体框中的特征点对应的地图点所连接的语义信息。

1) 若所有地图点均无语义信息说明这是第一次检测出该目标,将结果暂存在缓存中。当连续 5 帧检测出相同目标则将所有缓存地图点与当前目标检测结果对应。

2) 若已经有语义信息,说明当前目标物体已经存在于语义地图中,此时更新目标框内所有地图点语义信息为当前目标。

4 实验

4.1 目标检测网络训练

为了防止无关语义信息对地图构建的干扰,本文对 YOLO v2 网络结构进行调整。以 COCO^[26]和 PASCAL^[5]数据集为基础,筛选出室内场景中常见的 20 类物体。再对这些图片数据做随机拉伸剪切等数据增广处理,最后得到 20 000 张图像,物体种类见表 1。

表 1 选取的 20 类物体的具体类别

Chair	Clock	Refrigerator	Toy
Sofa	Toilet	Washing Machine	Window
Bed	People	Air condition	Fan
Tvmonitor	Book	Dog	Floor lamp
Laptop	Door	Cat	Biycle

在训练目标检测网络之前首先设置该深度神经网络的超参数。考虑到 YOLO v2 的网络特性,实验通过反复对比不同参数下的检测精度,在原始超参数的基础上进行调整,最终得到的超参数如表 2 所示。

表 2 深度神经网络的超参数设置

参数名	参数值
batch	64
subdivisions	8
height	416
width	416
channels	3
momentum	0.9
decay	0.000 5
angle	0
saturation	1.5
exposure	1.5
hue	0.1
learning_rate	0.000 1
max_batches	45 000
policy	steps
steps	100,25 000,35 000
scales	10,0.1,0.1

在 Ubuntu 14.04 的系统环境下训练和测试,处理器型号为 Intel i7-5960X,内存为 64 GB。为了得到更高的训练和测试速度,本文使用两张 GTX 1080 显卡加速训练。共训练 45 000 个 batch。

4.1.1 目标检测精度分析

采用留出法,将训练数据的 10% 留出作为测试数据。为了衡量算法的精确程度,本文与 Faster RCNN^[27]、YOLOv2^[24] 做对比。Faster RCNN 采用基于 tensorflow 实现的版本,YOLOv2 采用作者提供的原始版本。实验结果如表 3 所示。

表 3 目标检测精度对比实验结果

网络 类别	本文的网络	YOLOv2	Faster RCNN
Chair	52.3	51.8	45.6
Sofa	62.3	58.4	65.7
Bed	67.5	67.3	66.5
Tvmonitor	65.0	65.5	61.5
Laptop	84.7	82.2	85.3
Clock	83.1	83.0	82.8
Toilet	64.7	63.8	64.9
People	87.9	87.8	89.4
Book	63.3	63.3	63.9
Door	70.1	70.1	73.2
Efrigerator	62.1	63.0	63.2
Washing Machie	60.9	59.2	59.4

续表 3

网络 类别	本文的网络	YOLOv2	Faster RCNN
Air condition	67.9	66.3	66.3
Dog	92.7	92.2	86.9
Cat	90.9	91.5	88.5
Toy	50.6	49.8	48.3
Window	63.2	63.3	62.2
Fan	51.9	51.9	52.2
Floor lamp	64.8	62.3	57.3
Biycle	84.7	84.7	80.2
平均值(mAP)	69.53	68.87	68.17

AP 即平均精确度(Average precision),是目标检测算法中衡量精度的指标。每个类别根据精度(Precision)和召回率(Recall)得到 ROC(Receiver Operating Characteristic)曲线,AP 即是这条曲线之下的面积,mAP 是所有类别的平均 AP。

可以看到通过利用室内场景的图片对网络进行微调,本文的网络目标检测精确程度相比于 YOLOv2 和 Faster RCNN 有了明显的提高。

4.1.2 目标检测速度分析

为了达到实时构建语意地图的要求,要平衡目标检测的处理时间和处理精度。为了衡量本文算法的运行时间效率,本文在数据集上测试目标检测算法的运行效率,以处理帧率作为衡量标准,并与高精度的 Fast RCNN^[28]、Faster RCNN 等检测算法,以及实时的 100 Hz DPM、30Hz DPM 检测算法进行对比。实验的平台为 Ubuntu 14.04 的系统环境,处理器型号为 Intel i7-5960X,内存为 64 GB,显卡型号为 GTX 1080。实验结果如表 4 所示。

表 4 目标检测速度对比实验结果

检测算法	mAP/%	FPS
100 Hz DPM	16.0	100
30 Hz DPM	26.1	30
Fast RCNN	67.32	0.5
Faster RCNN	68.17	7
本文的网络	69.53	45

可以看到 Faster RCNN 与 Fast RCNN 的帧率远远低于本文的目标检测算法,无法达到实时的要求。而 DPM 算法虽然运行 FPS 指标与本文接近,但算法精度较差。

通过实验可以看出本文的算法在精度和时间效率上取得了较好的平衡,在实时运行的基础之上能够达到很好的 mAP。

4.2 语义地图构建

由于目标检测算法在没有 GPU 加速的情况下耗时较长,而移动机器人上目前还很难搭载高性能的 GPU,因此本文将目标检测算法放到服务器端进行。服务器端和移动机器人端之间采用 C++ 编写的基于 gRPC 的 RPC 框架进行通信,服务端的目标检测算法使用 C 语言编写的 darknet 作为框架,并搭载了 GTX1080 显卡加速运算,每幅图像的处理时间平均约 0.04 s。得到关键帧的目标检测结果之后,我们将结果通过 gRPC 传到移动机器人的语义地图管理模块中。

在 SLAM 系统中地图点与图像关键帧有着关联关系,每个地图点可以从多个角度观察到,因此将地图点和多个关键帧进行关联,而一个地图点理论上只能存在于一个物体之上。最终得到的地图点提取结果如图 4 所示。



图 4 地图点提取示意图

本文在室内场景下测试语义地图构建系统,软件的运行效果如图 5 所示。

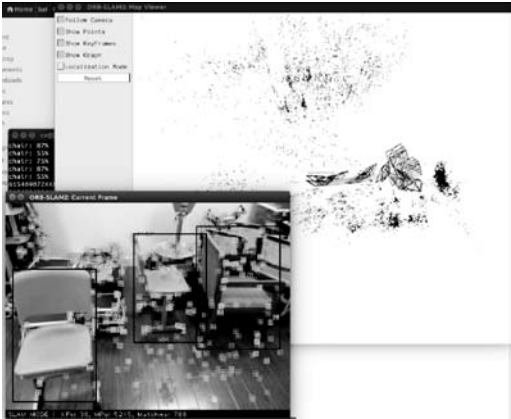


图 5 软件运行效果图

软件的前端部分展示了 SLAM 算法的定位效果,并且同时目标检测的结果也在图像中显示。语义管理模块对原始关键帧图片进行保存。在系统关闭时该模块将所有关键帧图片、对应的地图点、语义地图信息,以及目标物体在关键帧中的位置信息以 Protocol Buffer

的格式存储。其中图片数据以图片路径的形式保存在 Protocol Buffer 格式的数据里。通过将目标检测算法放到云端服务器中运行,语义地图构建能够实时地在机器人上运行。

5 语义地图的应用

本文的语义地图构建结果主要可以应用于两个方面。

第一,智能导航。机器人的导航需要机器人知道当前位置和目标位置,而在人类与机器人交互的过程中,人类常常无法提供给机器人目标在空间坐标系下的准确位置。语义地图提供了一种新的人机交互方式,人们可以以语义的形式给机器人提供指令,机器人将查询其语义地图,得到与语义地图相关联的地图点信息。而地图点包含了坐标信息,机器人根据 SLAM 模块得到的自身位置信息以及地图点的坐标信息,选择合适的路径规划算法。

第二,数据采集。传统的图像识别图像分割数据集,如 ImageNet 等,包含了数以百万计的图片。然而观察这些图片将会发现这些图片大部分是从正面拍摄的物体的照片,这是由于这些数据集图片的来源多是摄影作品。另外一个问题是这些图片中没有包括图像采集者与目标物体之间的相对位置信息。通过本文的算法构建的语义地图很好地弥补了这个问题。利用我们的语义地图构建算法,将得到关于一个目标物体在各个角度各个距离上的图像。同时,这些图像中的目标物体都是有标注的,这是因为我们的语义信息与 SLAM 中的地图点是对应的,我们可以将地图点投影到图像帧中,获得目标物体在图像帧中的位置信息。这将大大增加我们可以得到的有标注数据的数量,并且所得数据分布在更多角度和尺度之上,这将对训练更深更复杂的深度学习模型提供数据上的支持。

6 结 语

本文的主要工作是通过利用高精度的单目 SLAM 算法,结合基于深度神经网络的快速目标检测算法,对机器人所处的空间构建语义地图。本文的算法将 SLAM 算法提供的精确的几何信息和深度学习提供的丰富的语义信息有效地结合起来,建立语义信息和空间地图点之间的映射关系,并且利用深度神经网络的输出结果纠正 SLAM 算法中可能存在的错误。本文实现的算法将需要 GPU 加速的部分放到服务器端,通过高效的 RPC 框架进行实时通信。最终得到的语

义地图能够应用于机器人的智能导航系统,也可以应用于数据采集。

本文的算法将 SLAM 的地图点与语义信息之间建立起联系,把精确的地图点和语义信息之间建立映射。

未来的研究方向有:

1) 如何有效解决遮挡问题。当物体被另一物体遮挡时,目标检测算法检测出的目标框将发生重合,这时候需要有效消除重合的影响。

2) 根据探索情况更新原有地图。当机器人发现原来某地的物体消失,应能及时更新地图。

参 考 文 献

- [1] Deng J, Dong W, Socher R, et al. ImageNet: A large-scale hierarchical image database [C] // Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009: 248 - 255.
- [2] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C] // Advances in neural information processing systems, 2012. 2012: 1097 - 1105.
- [3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. 2016: 770 - 778.
- [4] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation [J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 38 (1) : 142 - 158.
- [5] Everingham M, Van Gool L, Williams C K, et al. The pascal visual object classes (voc) challenge [J]. International journal of computer vision, 2010, 88 (2) : 303 - 338.
- [6] Davison A J, Reid I D, Molton N D, et al. MonoSLAM: real-time single camera SLAM [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, 29 (6) : 1052 - 1067.
- [7] Klein G, Murray D. Parallel Tracking and Mapping for Small AR Workspaces [C] // IEEE and ACM International Symposium on Mixed and Augmented Reality. IEEE Computer Society, 2007: 1 - 10.
- [8] Leutenegger S, Furgale P, Rabaud V, et al. Keyframe-Based Visual-Inertial SLAM using Nonlinear Optimization [C] // Robotics, Science and Systems, 2013. 2013: 789 - 795.
- [9] 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.
- [10] Kuipers B. Modeling spatial knowledge [J]. Cognitive science, 1978, 2 (2) : 129 - 153.
- [11] Nielsen C W, Ricks B, Goodrich M A, et al. Snapshots for semantic maps [C] // Systems, Man and Cybernetics, 2004 IEEE International Conference on, 2004. IEEE, 2004: 2853 - 2858.
- [12] Galindo C, Saffiotti A, Coradeschi S, et al. Multi-hierarchical semantic maps for mobile robotics [C] // Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on, 2005. IEEE, 2005: 2278 - 2283.
- [13] Civera J, Gálvez-López D, Riazuelo L, et al. Towards semantic SLAM using a monocular camera [C] // Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, 2011. IEEE, 2011: 1277 - 1284.
- [14] Liu Z, von Wichert G. Extracting semantic indoor maps from occupancy grids [J]. Robotics and Autonomous Systems, 2014, 62 (5) : 663 - 674.
- [15] Case C, Suresh B, Coates A, et al. Autonomous sign reading for semantic mapping [C] // Robotics and Automation (ICRA), 2011 IEEE International Conference on, 2011. IEEE, 2011: 3297 - 3303.
- [16] 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.
- [17] Fasola J, Mataric M J. Using semantic fields to model dynamic spatial relations in a robot architecture for natural language instruction of service robots [C] // Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, 2013. IEEE, 2013: 143 - 150.
- [18] 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.
- [19] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF [C] // Computer Vision (ICCV), 2011 IEEE International Conference on, 2011. IEEE, 2011: 2564 - 2571.
- [20] Gálvez-López D, Tardos J D. Bags of binary words for fast place recognition in image sequences [J]. IEEE Transactions on Robotics, 2012, 28 (5) : 1188 - 1197.
- [21] Uijlings J R, Van De Sande K E, Gevers T, et al. Selective search for object recognition [J]. International journal of computer vision, 2013, 104 (2) : 154 - 171.
- [22] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International journal of computer vision, 2004, 60 (2) : 91 - 110.
- [23] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005. IEEE, 2005: 886 - 893.
- [24] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection [C] // Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition,2016. 2016:779 – 788.

[25] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2015. 2015:1 – 9.

[26] Lin T, Maire M, Belongie S, et al. Microsoft coco: Common objects in context [C]//European Conference on Computer Vision,2014. Springer,2014:740 – 755.

[27] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [C]//Advances in neural information processing systems, 2015. 2015:91 – 99.

[28] Girshick R. Fast r-cnn [C]//Proceedings of the IEEE International Conference on Computer Vision, 2015. 2015: 1440 – 1448.

[29] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of the IEEE conference on computer vision and pattern recognition,2014. 2014:580 – 587.

(上接第 131 页)

表 2 传统模板匹配与改进模板匹配检测算法对比

性能测试	传统模板匹配	改进模板匹配
误差范围/ μm	- 16. 39 ~ 9. 86	- 4. 25 ~ 3. 97
平均绝对误差/ μm	13. 5	2. 18
样本检测平均 耗费时间/ms	232. 5	196. 5

表 2 中根据误差范围、平均绝对误差、样品检测平均耗费时间这三个指标作为评判标准,显然改进模板匹配后的算法对系统的识别性能有所提高。

4 结 语

本文提出了符合光纤 PCB 批量识别的模板匹配算法。通过构建特征点,采用形状匹配的方式识别黄圈光学中心。通过改进 SAD,用基于灰度值的匹配方式识别黑圈的光学中心。并在此基础上构建了单块 PCB 识别的整体检测算法模型。为了迎合批量检测的需求,用 C#搭建了用户友好型界面,操作者只需要一个按钮就能够得到识别后的所有参数。对于 300 块样本 PCB 的检测表明,每个样本的绝对误差都能控制在 5 微米以内,检测平均耗时少于 200 ms,符合大规模检测要求。与传统模板匹配相比,改进模板匹配识别性能更佳。对于光纤 PCB 版目标检测算法设计,能很大程度上克服传统通过人眼识别导致误差大、效率低等

缺点,也为其他高精度产品自动化加工检测系统提供重要借鉴作用。

参 考 文 献

[1] 王强,宋京民,胡建平,等. 一种快速模板匹配目标识别算法[J]. 计算机工程与应用,2000,36(6):42 – 43,69.

[2] 赵梓森. 光纤通信的过去、现在和未来[J]. 光学学报, 2011,31(9):91 – 93.

[3] 李侠,李文举,孙娟红,等. 基于模板匹配和边缘检测的车标定位方法[J]. 计算机工程与设计,2011,32(10):3453 – 3456.

[4] 薛峰,丁晓青. 基于形状匹配变形模型的三维人脸重构[J]. 电子学报,2006,34(10):1896 – 1899.

[5] 郑江云,江巨浪,黄忠. 基于 RGB 灰度值缩放的彩色图像增强[J]. 计算机工程,2012,38(2):226 – 228.

[6] 梁路宏,艾海舟,何克忠,等. 基于多关联模板匹配的人脸检测[J]. 软件学报,2001,12(1):94 – 102.

[7] 吴成茂,郭锐. 三种典型模板匹配算法性能比较[J]. 西安邮电大学学报,2014(3):15 – 21.

[8] Wang C Y, Qu M J, Bi H Z, et al. Facial feature detection algorithm based on main characteristic of eyes [J]. Applied Mechanics and Materials,2013,303 – 306:1402 – 1405.

[9] Ouyang W, Tombari F, Mattoccia S, et al. Performance Evaluation of Full Search Equivalent Pattern Matching Algorithms [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence,2011,34(1):127 – 143.

[10] David W, Palieroni, Siddharth M. Signal to noise behavior for matches to gradient direction models of corners in images [R]. Computer Science Technical Report,2008.

[11] Kim J S, Kim H W, Kweon I S. A camera calibration method using concentric circles for vision applications [C]//The 5th Asian Conference on Computer Vision,2002:515 – 520.

[12] 王慧. 基于模板匹配的手写体字符识别算法研究[D]. 北京交通大学,2012.

[13] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection [C]//Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. IEEE,2005:886 – 893.

[14] 姜映映,田丰,王绪刚,等. 基于模板匹配和 SVM 的草图符号自适应识别方法[J]. 计算机学报,2009,32(2):252 – 260.

[15] 周瑜,刘俊涛,白翔. 形状匹配方法研究与展望[J]. 自动化学报,2012,38(6):889 – 910.

[16] 郑丹晨. 基于轮廓点空间结构特征的形状匹配方法研究 [D]. 大连理工大学,2014.

[17] 朱铮涛,张宏. 基于特征点的形状匹配技术及其算法实现 [J]. 计算机工程与设计,2010,31(5):1096 – 1099.