

基于 Kinect 传感器的移动机器人环境检测及行为学习^①

段 勇^② 盛栋梁 于 霞

(沈阳工业大学信息科学与工程学院 沈阳 110870)

摘 要 研究了一种基于深度图像和强化学习算法的移动机器人导航行为学习方法。该方法利用机器人装配的 Kinect 传感器检测工作环境信息,然后对获取的深度图像数据和视频图像进行处理、融合和识别,并由此构建机器人任务学习的状态空间,最终利用强化学习方法实现移动机器人的导航任务的自学习。该方法的有效性通过实验得到验证。实验表明,该方法能够使机器人具有较强的环境感知能力,并能够通过自学习的方式掌握行为能力。

关键词 移动机器人导航, Kinect 传感器, 深度图像, 强化学习(RL)

0 引 言

机器人学习是指机器人利用自身配置的传感器,通过与环境的不断交互来获取知识并掌握完成任务的能力,它是使机器人具有生物智能的一种有效方式,对于提高机器人的智能水平、自动化程度以及对环境的适应性具有重要意义。

在未知环境中机器人进行的任务学习包括环境感知、自学习方法以及如何应用学习方法实现机器人的任务执行等关键问题。其中机器人装配的传感器的感知能力是机器人进行一切工作的基础和前提条件,目前在移动机器人导航领域应用的传感器主要包括:超声测距传感器、激光测距传感器以及单目视觉、全景视觉、双目视觉传感器和 3D 激光摄像机等^[1]。而近年来适用于移动机器人应用的新类型传感器不断出现,其中 Kinect 传感器的出现和应用为机器人的研究提供了新的方法并受到了广泛的关注^[1-3]。Kinect 是微软公司于 2010 年推出的基于体感交互的传感器设备,能够实现视频图像、深度数据、语音信号的获取和处理,在人机交互、虚拟现实、语音识别、智能机器人等领域体现出了巨大的潜

力^[4-6]。Kinect 传感器能够直接获取彩色图像数据和覆盖区域的深度距离信息,因此与超声、激光、单目视觉等常用传感器相比能够提供更为丰富的感知信息,同时也具有较好的使用 and 开发便利性,因此其研究和开发具有重要实用价值。有效的学习方法是机器人完成学习任务的关键。强化学习(reinforcement learning, RL)是一种重要的机器学习方法^[7],其学习目标是实现状态空间到动作空间的最优映射策略,这与反应式移动机器人导航方式有着天然的关联,因此已成为一种有效的机器人自学习方法^[8-10]。基于此,本研究给出了一种基于 Kinect 传感器环境信息感知的移动机器人导航任务自学习方法。该方法首先对 Kinect 获取的环境信息进行处理、识别和理解,再根据导航任务需求,应用强化学习方法来实现机器人的行为动作学习。本研究通过实验测试了机器人传感器对环境的感知能力,并验证了机器人执行导航任务的有效性。

1 Kinect 系统原理及感知信息

1.1 Kinect 传感器介绍

Kinect 传感器如图 1 所示,数据获取器件从左

① 国家自然科学基金(60905054),辽宁省高等学校优秀科技人才支持计划(LR2015045)和辽宁省自然科学基金(2015020010)资助项目。
② 男,1978 年生,博士,副教授;研究方向:智能机器人,计算机视觉,机器学习等;联系人,E-mail: duanyong0607@163.com
(收稿日期:2015-09-20)

至右分别是红外激光发射器、彩色摄像头和红外接收器,分别用来获取普通视频图像和深度数据^[11,12]。



图 1 自主移动机器人及 Kinect 传感器

1.2 环境信息的获取与处理

Kinect 传感器获取同一室内场景的视频图像和深度数据映射图像如图 2 所示。为了方便观察和利用深度数据,将其进行可视化处理,每个像素的深度数据映射到颜色空间,显示结果见图 2(b),图中不

同灰度对应不同的深度距离数值。图中黑色像素用来标识传感器不能正确检测的深度值,出现该问题的原因来自于 Kinect 传感器本身的局限性。造成这些无效深度数据出现的情况主要包括环境中的凹形物体、物体边缘、特殊材质表面物体以及光照条件等。

无效深度像素通常会成块出现,因此常规的图像去噪方法难以解决该问题。基于此,Silbeman 等提出利用条件随机场模型深度表示区域来实现室内场景分割,以此消除 Kinect 采集的无效数据^[13]。文献[14]也给出了基于双边联合滤波方法来填补深度图像中的无效数据空洞。但以上方法需要较为复杂的运算,将会影响机器人视觉等应用领域对处理实时性的要求。因此我们在满足需求的情况下折衷考虑性能和效率,使用无效数据相邻区域同一行数据的均值来添补无效数据的方法,同时为了防止无效数据出现在边缘的情况,使用水平双方向进行扫描,添补无效数据的深度映射图像,如图 2(c)所示。

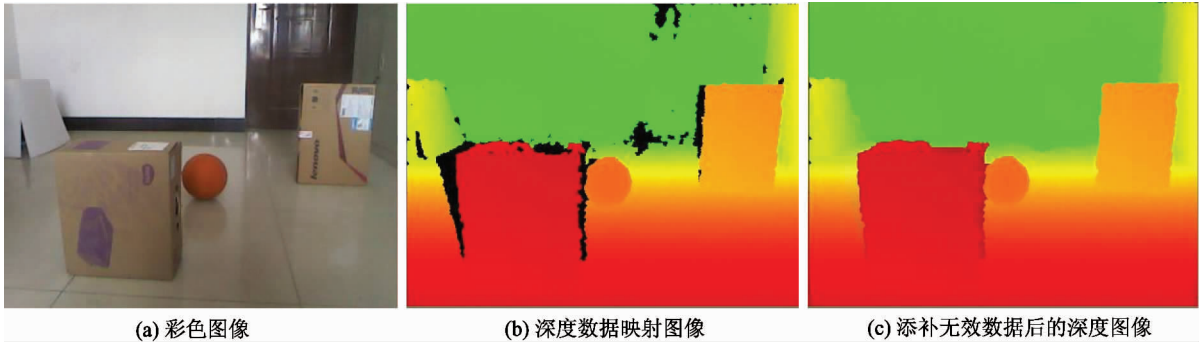


图 2 Kinect 传感器数据采集和处理

2 基于 Kinect 传感器的机器人系统结构

研究实验平台使用博创公司的 Voyager II 型移动机器人,如图 1 所示。移动机器人导航问题可以认为是机器人能够根据传感器获取的环境信息,完成在有障碍物的环境中的特定任务。将自主移动机器人系统分为传感器、主控系统、运动机构三层,Kinect 传感器层用于感知环境信息,将获取的视频图像和深度数据传递给机器人主控系统;机器人主控

系统进行感知信息融合处理以实现障碍物的检测以及目标物的定位,然后基于感知信息进行导航任务学习,学习后建立导航策略,再将决策指令发送给底层运动结构;运动机构控制机器人执行避障、接近目标等行为动作。

机器人检测的环境信息主要包括完全未知的障碍物和具有已知颜色特征的目标物。未知障碍物是机器人室内工作环境的家具、墙壁等物体,已知特征目标为机器人需要跟踪的目标物,

基于 Kinect 感知信息进行多障碍物检测方法如下:(1)建立环境地面模型,检测多障碍物。由于

Kinect 获取的深度数据反映了物体表面与传感器的距离信息,因此当传感器俯仰角度不变的情况下,地面距离也保持不变。事先采集无障碍深度数据图像作为地面模型,然后将检测深度数据与地面模型进行帧差,可初步检测多障碍物。(2)根据 Kinect 测距原理和深度数据特性,利用障碍物深度图像的边缘特征和距离特征进行障碍物的分割;(3)对 Kinect

深度传感器进行标定(类似与针孔成像模型),利用深度数据图像的像素与空间位置关系确定分割障碍物的位置信息。(4)根据实际要求设定障碍物距离阈值,确定障碍物所在机器人视野中的区域,实现障碍物的检测。检测过程及结果如图 3 所示,检测后可获得障碍物的距离信息和与机器人的相对方位信息。

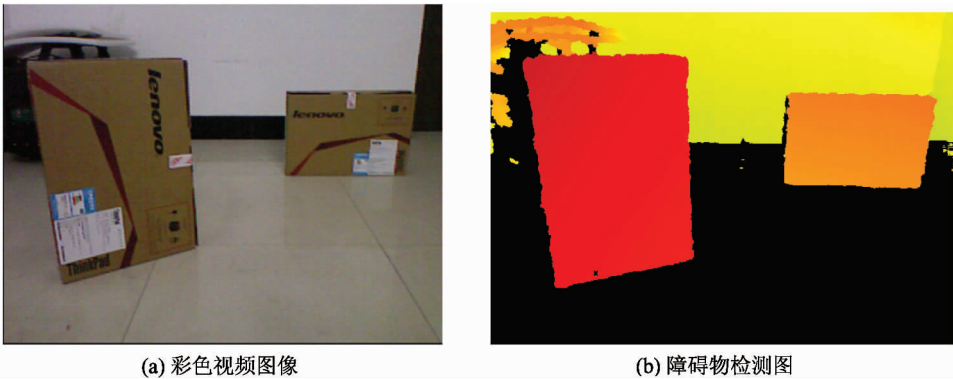


图 3 基于 Kinect 传感器的障碍物检测

此外,在研究中选择小球作为机器人跟踪的可移动目标,基于 Kinect 传感器的目标物检测方法如下:(1)首先注册视频图像和深度数据图像,使得两者在像素级上对应。注册后可以实现视频数据和深度数据的信息融合,从而处理复杂机器视觉任务;(2)根据目标物体的颜色特征进行搜索和分割,包括颜色采集、确定颜色分割阈值、区域连通、像素搜索等操作,具体实现方法可参考之前的研究工

作^[15];(3)基于二维图像实现目标的分割,但要确定目标的三维空间定位,需要进行摄像机标定和几何关系计算等复杂操作^[15]。而 Kinect 传感器为目标空间定位提供了方便且精确的方法,只需根据视频图像获取的目标物像素坐标位置,即可对应深度数据图像直接得到目标位的距离信息。其检测过程及结果如图 4 所示,检测后可以获得目标物的距离信息和与机器人的相对方位信息。

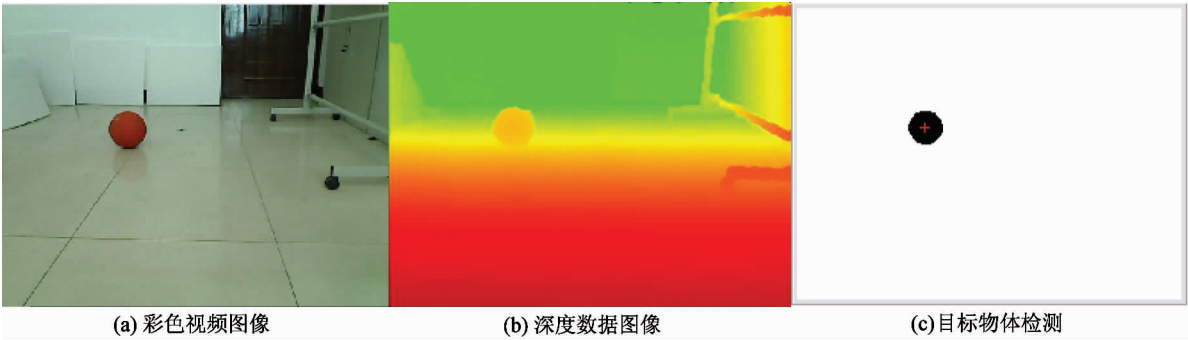


图 4 基于 Kinect 传感器的目标物检测

3 移动机器人导航行为自学习

强化学习是一种基于感知环境信息获取机器人

导航行为能力的有效方法。它只需定义机器人的学习任务,再将机器人置于未知环境中通过自身的传感器来获取环境状态,通过机器人与环境的交互来

积累成功和失败经验,以试错的方式不断提高机器人的行为能力。本节将讨论使用强化学习方法来学习移动机器人的躲避障碍物和接近目标行为。

3.1 强化学习基本理论

定义智能体的状态集合为 $S = \{s_i | s_i \in S\}$, 它执行的动作集合表示为 $A = \{a_i | a_i \in A\}$ 。在 t 时刻,根据智能体的状态 s_t , 选择并执行动作 a_t , 导致智能体的状态转移到 s_{t+1} , 同时从环境得到强化信号 r_t , 它对动作 a_t 的效能进行评价。强化学习的任务是得到一个映射策略 $\pi: S \rightarrow A$ 。

在强化学习算法中, Q 学习是最重要也是使用最广泛的一种算法,它通过定义函数 $Q(s, a)$ 来评估状态相对应的各个动作。Q 学习算法的基本形式为^[2,5]

$$\hat{Q}(s_t, a_t) = Q(s_t, a_t) + \eta_t \cdot (r_t + \gamma \cdot \max_a Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)) \cdot e(s_t, a_t) \quad (1)$$

其中, η_t 为学习率, γ 为折扣因子。 $e(s_t, a_t)$ 定义为资格迹^[7], 它主要用于解决强化学习的信度分配问题。

3.2 基于 Kinect 感知信息的强化学习方法

本文应用表格型 Q 学习方法来使机器人学习避障和接近目标行为的学习。表格型 Q 学习是将状态空间和动作空间进行离散化,再由离散的状态变量和动作变量构成查找表格,表格中的值为对应的状态-动作对的评估值 Q 。根据式(1)实现学习过程中 Q 值的更新,在学习结束后选择具有最大 Q 值的动作和对应的状态变量形成映射策略。

强化学习的状态信息来自于机器人 Kinect 传感器采集数据的处理结果。将感知范围划分成 9 个区域(示意图见图 5), 它基于摄像机模型由 Kinect 获取的图像数据投射到机器人实际工作环境来得到。其中垂直 3 个划分分别代表障碍物与机器人的距离为近(near)、中(mean)、远(far);水平方向 3 个划分表示障碍物在机器人的左、中、右三个方向。障碍物所在区域需要根据上节的处理结果,根据机器人实际工作任务要求定义距离阈值(threshold): D_{th}^{Near} 和 D_{th}^{Far} , 则表示垂直离散划分的强化学习状态分量定义为

$$s_i = \begin{cases} Near, & d < D_{th}^{Near} \\ Mean, & D_{th}^{Near} < d < D_{th}^{Far} \\ Far, & d > D_{th}^{Far} \end{cases} \quad (2)$$

式中 d 为障碍物与机器人的距离,该值可以通过第 3 节分割的障碍物所对应的深度信息得到。在距离垂直划分中,对于没有障碍和距离划分为“远”的机器人避障策略是相同的。

通过对特征目标的识别和定位可以得到目标相对于机器人的角度信息和距离信息,获取后机器人可形成记忆位置信息,将其定义为状态变量 s_{tr} 和 s_{id} 。再根据机器人实际工作的需求将其进行适当的离散化。

结合避障和接近目标两种行为,强化学习每个状态变量可以定义为: $s_t = \{s_l, s_c, s_r, s_{tr}, s_{id}\}$, 分别用来表示机器人 Kinect 感知范围内左、中、右三个方向障碍物的距离状态以及目标的方位和距离信息,将其可能出现的各分量组合构成状态对应序号编码,并构成表格型 Q 学习的离散状态。

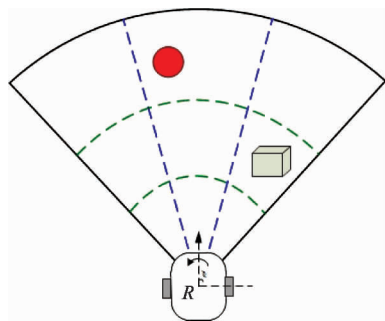


图 5 机器人行为学习状态划分

Q 学习的动作空间由移动机器人运行信息给出。移动机器人采用左右轮独立驱动方式,其旋转半径范围可从零(原地打转)到无限大(直线运动)。机器人的运动包括两个部分:车体中心的平动和绕小车中心的转动,机器人的基本运动控制单元为左右轮速(左右驱动电机转速),表示为 v_l 和 v_r 。定义的 Q 学习动作变量为“宏动作”,具体分为直行前进、左/右小转、左/右大转、原地左/右转向,如图 6 所示。每个宏动作通过给定的不同的左右轮速来实现。

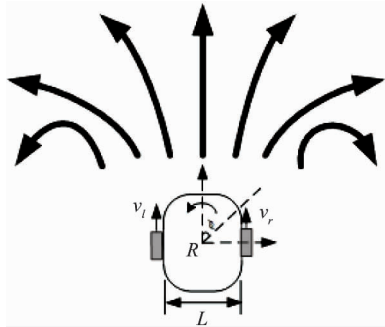


图 6 强化学习动作变量

Q 学习最重要的学习依据是强化信号函数(回报),它代表在一定状态下选择并执行动作获得的立即评估回报。对于机器人所要学习任务需要在避障的同时接近目标,对于避障行为,希望机器人远离障碍物,因此机器人越接近障碍物,越应该得到惩罚(负回报);相反,机器人应该获得奖励(正回报)。对于接近目标行为,机器人接近目标物应获得奖励,而远离目标则应得到惩罚。综合两种行为需求,定义 Q 学习的强化信号(回报)函数如下:

$$r_t = \begin{cases} -1.0, & d_t \leq D_c \\ -\kappa(D_{A0} - d_t), & D_c < d_t \leq D_{A0} \\ +0.2, & d_t^T < d_{t-1}^T \\ 0, & \text{其他} \end{cases} \quad (3)$$

其中, r_t 表示机器人 t 时刻的强化信号函数, d_t 为机器人与障碍物的距离, κ 为比例系数,用来计算回报; D_c 代表碰撞距离阈值,当机器人与障碍物的距离小于该值时,认为机器人发生碰撞; D_{A0} 为机器人避障阈值,在该阈值范围内需要机器人进行避障; d_t^T 和 d_{t-1}^T 分别表示机器人 t 时刻和 $t-1$ 时刻机器人与目标点的距离值,其含义是当机器人更接近目标时,应获得正回报。

4 实验结果及分析

基于以上研究工作,对 Kinect 传感器环境感知和利用强化学习实现移动机器人导航任务学习进行实验。实验平台采用 Voyager II 自主机器人,系统开发使用 Windows 操作系统环境下 C++ 语言的开发环境。系统操作界面如图 7 所示,可以显示传感器采集的信息、处理结果、定位信息等,能够实现各

种处理的操作,机器人的行为学习和导航控制等。

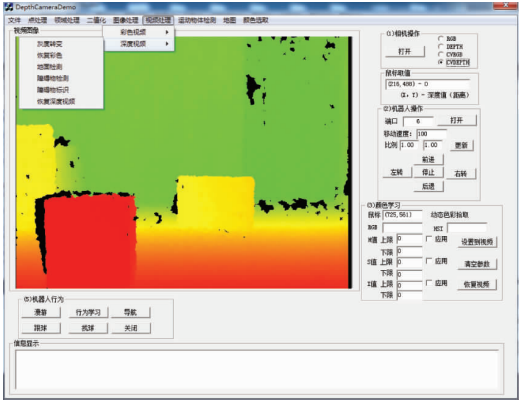


图 7 系统操作界面

实验环境为实验室的室内场地,室内环境的墙壁,障碍物和具有颜色特征的目标球构成基本的实验场景,通过改变障碍物和目标物的位置来得到多个变化的实验环境。

4.1 移动机器人环境信息检测实验

首先利用机器人装配的 Kinect 传感器进行环境检测实验并同常用超声传感器(声纳)检测结果进行比较。机器人通常配置多个声纳传感器,每个传感器按照一定的距离间隔分别安装在机器人同一水平的圆周上,见图 1。首先使用超声进行环境检测,环境中的障碍物如图 8 所示,该障碍物正面接触面积狭窄,同时下部还有一较低高度的横梁,这类物体在室内环境中十分常见。图 9 为机器人超声传感器的检测结果,图中以标靶图的方式标识出了机器人前方声纳的测距数据,显示了传感器并没有完整地检测出环境中的障碍物。说明声纳虽然检测出图中靠左侧的立柱,但由于声纳测距角度分辨率较低,无法确定宽度较窄障碍物的具体方位信息。



图 8 机器人工作环境中的障碍物

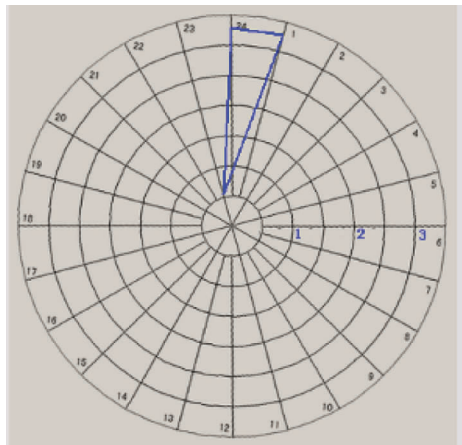


图9 基于超声的障碍物检测结果

同时,使用 Kinect 传感器对于同一环境进行检测, Kinect 传感器获取的深度数据进行处理,分割出该障碍物,如图 10 所示。可以 Kinect 传感器能够较为完整准确地实现复杂障碍物的检测,相对于机器人常用的超声传感器具有更强的感知能力和适用性。



图10 基于 Kinect 传感器障碍物检测结果

此外,本研究利用 Kinect 传感器的目标定位方法,即通过融合自身获取的彩色图像和深度数据来实现,首先根据目标颜色特征进行分割,然后在像素级对应深度数据图像来获得其距离信息。在完成目标分割后就能够直接从深度数据中得到目标的相对距离值。实验中,分别进行 10 次测量,每次改变目标物(小球)和机器人的相对位置,求得多次定位相对误差的平均值为 1.1%,定位距离最大相对误差为 1.8%。

由于各传感器的精度与工作距离有关,因此本

文采用相对误差来反映测距精度,实验中作为标准的精确数据通过单点激光测距仪(SW-40,精度 $\pm 2\text{mm}$)获得。

4.2 移动机器人导航任务学习实验

使用 Q 学习方法实现移动机器人导航任务学习。学习过程分为两个阶段,首先在仿真环境进行,仿真设置机器人各区域障碍和目标位置。当仿真学习进行一定阶段后,将机器人置于真实环境进行学习来完善学习结果。这样做一方面可以提高学习效率节省训练时间,另一方面能够减少机器人的消耗并保障机器人的安全。将机器人置于放置障碍物和目标物的实验环境中同时学习避障和接近目标行为,机器人获取环境状态后通过 EEP 随机策略选择动作,再控制机器人到达新的状态,同时得到即时强化信号,最后更新 Q 函数查找表。当机器人发生碰撞、接触目标或完成预定的学习迭代步数时,机器人重新开始新一轮的学习。强化学习算法中参数选择如下:学习率 $0.3 \leq \eta \leq 0.8$, 折扣算子 $0.8 \leq \gamma \leq 0.95$, EEP 动作选择策略采用 Epsilon 贪婪策略^[7], 其中最优动作选择概率 $0.2 \leq \varepsilon \leq 0.5$ 。其中学习率的值体现了对学习速度的控制。折扣算子则用于控制立即回报对学习过程的影响。最优动作选择概率在学习的初始阶段选择较小的值,其目的是使学习量有充分的机会参与学习;而在后面阶段取较大值,其目标是为了算法尽快收敛,获得最优策略。根据机器人的实际体积、传感器的检测范围、预留机器人避障安全区域和障碍物体积等因素,令分割状态空间的距离阈值为 $D_{th}^{Near} = 1.2\text{m}$, $D_{th}^{Far} = 2.2\text{m}$ 。在强化信号函数中,定义碰撞距离阈值 $D_c = 0.3\text{m}$,避障阈值 $D_{A0} = 1.7\text{m}$,以保障机器人的安全性和学习的有效性。

强化学习后,基于最终的 Q 表确定状态变量和动作变量的最优映射策略,并用其来控制机器人执行导航任务。图 11 为在学习过程中,某一状态一动作对应的根据最大 Q 值进行归一化后的 Q 值变化曲线,表明了回报函数作用于 Q 值的更新,并最终使 Q 值到达最大值,实现了该状态到动作的最优映射策略。

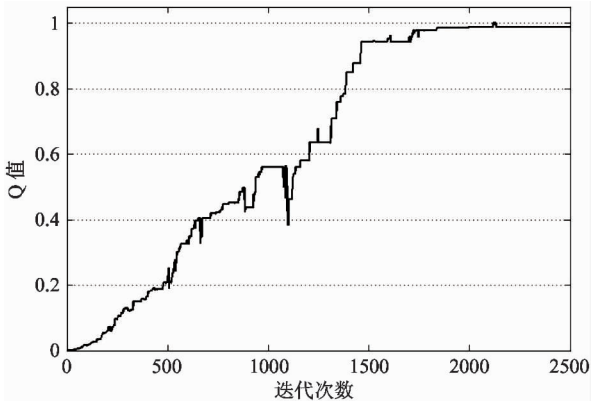


图 11 强化学习中某一 Q 值变化曲线

图 12 为在应用学习策略控制机器人在仿真环境中的导航运行轨迹,其中对机器人运动和障碍物位置进行模拟,目标位置设定为已知。可见机器人能够在较为复杂的环境中完成避障且接近目标。

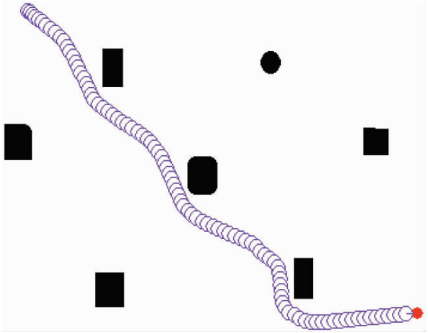
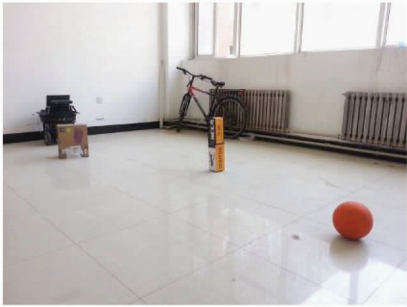
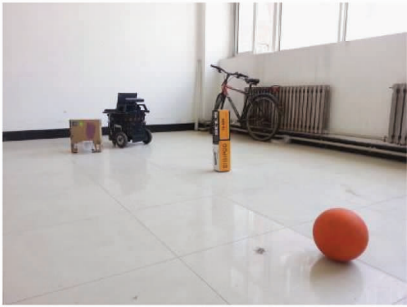


图 12 机器人导航轨迹

此外,我们使用实体移动机器人进行实验,通过改变目标物的位置来得到多组目标物的位置信息,其位置信息来自于上节的识别结果。其位置信息坐标分别为: $(3.55\text{m}, 7.10\text{m})$, $(1.53\text{m}, 6.76\text{m})$, $(5.32\text{m}, 6.81\text{m})$ 。然后分别进行机器人导航实验,图13显示了其中一组的实体移动机器人在工作环



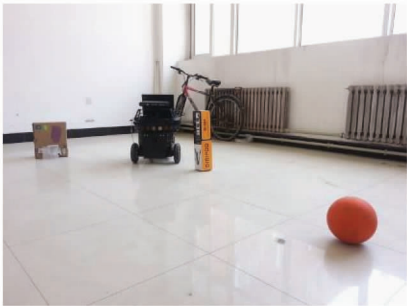
(a) 位置 1



(b) 位置 2



(c) 位置 3



(d) 位置 4



(e) 位置 5



(f) 位置 6

图 13 移动机器人导航运行过程

境下的运行情况,可见机器人可以按照较为优化的路径无碰撞地到达目标。

本文对于由障碍物和目标物定义的强化学习离散状态空间划分没有过于精细,主要原因是由于机器人车体、障碍物和目标均具有一定的体积,特别是机

器人接近时物体在图像中会有较大的比例,因此会得到“远粗近细”的机器人控制效果,能够满足机器人工作的要求。此外,若划分过细也会使得对应的强化学习状态空间过大,影响学习效率,引起维数灾难问题。

5 结 论

感知系统是机器人执行一切工作任务的基础,也是决定自主机器人智能程度的关键,Kinect 传感器的产生和应用为移动机器人了解环境信息提供了更为有效、便捷的途径。基于此,本文研究了 Kinect 传感器在移动机器人领域的应用问题,给出了障碍物和目标物的检测及定位方法,弥补了传统超声传感器和视觉传感器在机器人环境感知方面的局限和不足。此外,本文提出了利用 Kinect 感知信息并基于强化学习的移动机器人复杂任务学习方法,该方法能够在没有专家知识和精确环境模型条件下,利用传感器感知环境信息,通过机器人自学习来掌握行为动作能力。因此,研究工作能够有效地提高机器人的智能水平、学习能力和环境的适应性,同时也能够进一步拓展到其他复杂任务的学习中。

参考文献

- [1] Stoyanov T, Mojtahedzadeh R, Andreasson H, et al. Comparative evaluation of range sensor accuracy for indoor mobile robotics and automated logistics applications. *Robotics and Autonomous Systems*, 2013, 61: 1094-1105
- [2] 杨东方,王仕成,刘华平等. 基于 Kinect 系统的场景建模与机器人自主导航. *机器人*, 2012, 34(5): 281-289
- [3] Tolgyessy M, Hubinsky P. The Kinect sensor in robotics education. In: *Proceedings of the 2nd Robotics in Education*, Vienna, Austria, 2011. 143-146
- [4] 刘鑫,许华荣,胡占义. 基于 GPU 和 Kinect 的快速物体重建. *自动化学报*, 2012, 38(8): 1288-1297
- [5] Han J G, Shao L, Xu D, et al. Enhanced computer vision with Microsoft Kinect sensor: a review. *IEEE Transactions on Cybernetics*, 2013, 43(5): 1318-1334
- [6] 罗元,谢彧,张毅. 基于 Kinect 传感器的智能轮椅手势控制系统的设计与实现. *机器人*, 2012, 34(1): 110-113, 119
- [7] Sutton R S, Barto A G. *Reinforcement learning: an introduction*. Cambridge: MIT Press, 1998
- [8] McPartland M, Gallagher M. Reinforcement learning in first person shooter games. *IEEE Transactions on Computational Intelligence and AI in Games*, 2011, 3(1): 43-56
- [9] Chen C L, Li H X, Dong D Y. Hybrid control for robot navigation—a hierarchical Q-learning algorithm. *IEEE Robotics & Automation Magazine*, 2008, 15(2): 37-47
- [10] Yamaguchi A, Takamatsu, Ogasawara T. OCOB: action space for reinforcement learning of high DoF robots. *Autonomous Robots*, 2013, 34(4): 327-346
- [11] 余涛. *Kinect 应用开发:用最自然的方式与机器对话*. 北京:机械工业出版社, 2013
- [12] Khoshelham K, Elberink S O. Accuracy and resolution of Kinect depth data for indoor mapping application. *Sensors*, 2012, (12): 1437-1454
- [13] Silberman N, Fergus R. Indoor scene segmentation using a structured light sensor. In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*, Barcelona, Spain, 2011. 601-608
- [14] Camplani M, Salgado L. Efficient spatio-temporal hole filling strategy for Kinect depth maps. *Is&spie Electronic Imaging*, 2012, 8290(9): 841-845
- [15] 段勇,徐心和. 自主足球机器人视觉系统结构及关键技术研究. *东北大学学报(自然科学版)*, 2006, 27(1): 9-12

Mobile robot environment detection and behavior learning based on Kinect sensor

Duan Yong, Sheng Dongliang, Yu Xia

(School of Information Science and Engineering, Shenyang University of Technology, Shenyang 110870)

Abstract

A behavior learning algorithm for mobile robot navigation based on depth images and reinforcement learning is proposed. The algorithm uses the Kinect sensor on a mobile robot to capture the environmental data of the robot, then, processes, fuses and identifies the data of depth images and video images among them to establish the state space for robot learning, and finally, uses the reinforcement learning method to implement the mobile robot's self-learning of navigation tasks. The effectiveness of proposed algorithm was verified by experiment. The experimental results show that the method can make a mobile robot possess the stronger ability of perceiving environments and capacity of mastering behaviors by self-learning.

Key words: mobile robot navigation, Kinect sensor, depth image, reinforcement learning (RL)