

文章编号: 1001-0920(2005)05-0529-04

## 移动机器人智能寻线导航与策略控制

张淑军<sup>1</sup>, 孟庆春<sup>1,2</sup>, 吴 槟<sup>1</sup>, 费云瑞<sup>3</sup>

(1. 中国海洋大学 计算机科学系, 山东 青岛 266071; 2. 清华大学 智能技术与系统国家重点实验室, 北京 100084; 3. 中国海洋大学 电子工程系, 山东 青岛 266071)

**摘要:** 寻线导航是移动机器人寻航方式之一, 为使机器人沿地面标志线自主运动, 并在在适时离线执行任务后自动返航, 提出一种具有学习功能的智能寻线导航与策略控制方法以及系统模型, 并将其应用于自行开发研制的导医机器人中. 基于调制光检测原理, 通过状态评判网络对多传感器信息进行融合, 通过 $Q$ 学习方法获得离线走失后的最优控制策略, 实现了机器人在各种复杂未知环境下稳定、可靠地工作. 实验结果表明, 该方法可以有效地提高机器人的适应性和智能度.

**关键词:** 自主移动机器人; 寻线控制; BP网络;  $Q$ 学习

中图分类号: TP24

文献标识码: A

## Intelligent line-tracking navigation and policy control of mobile robot

ZHANG Shu-jun<sup>1</sup>, MENG Qing-chun<sup>1,2</sup>, WU Bin<sup>1</sup>, FEI Yun-ru<sup>3</sup>

(1. Department of Computer Science, Ocean University of China, Qingdao 266071, China; 2. State Key Laboratory of Intelligent Technology and Systems, Tsinghua University, Beijing 100084, China; 3. Department of Electronic Engineering, Ocean University of China, Qingdao 266071, China Correspondent: ZHANG Shu-jun, E-mail: Lindazsj@mail.ouc.edu.cn)

**Abstract:** An intelligent line-tracking navigation and policy control method is proposed and the system model is designed in order to assist a mobile robot walking along an indicated line on the ground, to ensure it automatically navigate home after leaving the line and finishing its tasks. The method and model are applied to self-developed hospital guiding robot. Multi-sensory information is fused using a state-evaluation network based on modulated light detection principle. The optimal control strategy after straying from the line is achieved through  $Q$ -learning and the robot's steady and reliable work in different complex unknown environments is realized. The experimental results show that the proposed method can effectively enhance the robot's adaptability and intelligence.

**Key words:** autonomous mobile robot (AMR); line-tracking and control; BP network;  $Q$ -learning

### 1 引言

自主移动机器人是目前机器人发展的主流, 随着其应用领域的不断扩展和自动化程度的进一步提高, 对机器人的功能提出了更高的要求<sup>[1]</sup>, 特别是需要各种具有不同智能程度的机器人, 能够设定自己的目标, 规划并执行自己的任务动作, 不断适应环境的变化. 在机器人系统中, 自主导航控制是一项核心

技术, 也是机器人研究领域的重点和难点问题. 导航过程中经常面临不可预测或动态变化的环境, 而机器人感知环境的手段通常是不完备的, 传感器给出的数据是不完全、不连续和不可靠的. 因此, 解决机器人的自主导航是一项迫切而又艰巨的任务<sup>[1]</sup>.

移动机器人有多种导航方式, 各有其特点, 适合不同的应用环境<sup>[2]</sup>. 目前对视觉导航的研究非常活

收稿日期: 2004-07-01; 修回日期: 2004-10-09

基金项目: 山东省科技攻关课题(031080124).

作者简介: 张淑军(1980—), 女, 山东泰安人, 博士生, 从事智能控制、机器学习的研究; 孟庆春(1958—), 男, 山东招远人, 教授, 博士生导师, 从事智能技术与系统、智能机器人等研究.

跃,视觉传感器获取的信息量大,而且比较接近人的导航机理,但存在算法复杂性与实时性的矛盾<sup>[1]</sup>。为了提高系统的实时性,降低系统的复杂性,非视觉传感器导航方法已经开始引起人们的兴趣。地面标志线导航(即寻线导航)作为一种非视觉传感器组合导航方式,铺设简单,灵活方便,对周围环境的依赖性小,已获得广泛应用。寻线导航方式以光电传感器为硬件基础,利用调制光检测原理,实现环境识别与定位。目前国内外对此均有不同程度的研究,文献[3,4]利用调制光寻线系统控制自动行走机器人,取得了较好的效果,但由于检测点过多,光电检测装置体积大,电路比较复杂。本文利用较少的检测点,通过设计相应的算法实现对机器人导航的优化控制。

导医机器人是全自主式服务型机器人。根据其一般工作在相对规整环境中的特点,选择地面标志线导航方式,并设计了一种新的移动机器人智能寻线与策略控制方法。它根据传感器信息的融合结果,使用BP网络评判当前状态,并根据当前状态,采用Q学习方法,选择最优动作,使机器人不仅能准确地沿标志线运动,而且可在一定环境内离线行走,还能通过学习调整自动返航,灵活自主地完成各种预定的任务。

## 2 寻线系统硬件结构及原理

机器人采用圆形的底盘结构,独立双轮步进电机驱动模式,利用安装在底盘中央的非接触式光电传感器为机器人导航,探测与地面有一定颜色对比度的标志线。光电传感器由比较电路板和光发射接收板两部分组成。比较电路板将光发射接收板输入的模拟信号经比较器转换为数字信号,输入到单片机中。光发射接收板完成光的发射、接收,将光信号转换为模拟信号。每一块光发射接收板就是机器人的一个检测点,其结构如图1所示<sup>[3]</sup>。它由3个白发红的亮度发光二极管(LED)和1个光电三极管组成。3个LED发出的红色光照射到地面上,产生漫反射,反射光进入光电三极管后,光电三极管产生光电流,其产生光电流的能力随地面反射光的强弱而显著变化。每个光电三极管的输出信号都通过比较器进行处理,得到一个模数转换后的AD值。由于线和地面的AD值可通过阈值上下分开,通过输出的AD值即可确定光电三极管是否在线上,即检测点是否在线。

本寻线系统设置了6个光电检测点,分为3排安装在机器人小车的底盘上,布局如图2所示。6个检测点逆时针依次编码为0~5,各检测点在安装板上成非线性的排列结构,相对于导航线的中轴成对称分布。前后4个检测点分布在导航线中轴的左右两

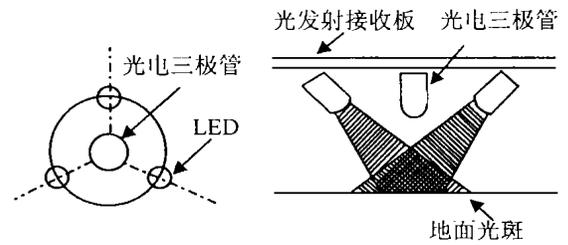


图1 检测点结构和原理

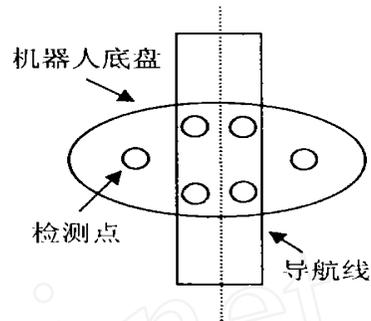


图2 检测点布局

边,用于跟踪导航直线,使检测点相对于线的位置变化敏感。分布在两侧的两个检测点是为了在机器人转弯或偏差过大时“捕捉”航线<sup>[4]</sup>。

利用光寻线系统可识别浅色地板上的黑色导航线或深色地板上的白色导航线<sup>[1]</sup>(根据先验知识设置不同的区分阈值)。在环境信息完全未知的情况下,机器人在行走的过程中利用传感器不断地对导航线进行探测,并调整行进路线与导航线的偏差。当遇到障碍时,或停下等待或绕开障碍,避障后再根据导航线的指引回到原来的路线上,最终在导航线的指引下到达指定的目的地<sup>[1]</sup>。

## 3 具有学习能力的寻线导航与策略控制

移动机器人要在未知环境中安全、可靠地完成指定任务,必须能够处理各种突发情况,逐渐适应环境,提高工作效率,这就要求其导航控制系统具有灵活性和适应性<sup>[5]</sup>。为此提出具有学习能力的寻线导航与策略控制方法。

### 3.1 状态评判网络

由于不同的光照条件以及地面存在各种影响因素,如地面颜色与导航线的颜色较接近,地面带有自铺设的金属线,有裂痕、斑迹等干扰因素,导致光电传感器的检测可能会产生误差,仅根据某些检测点的输出无法断定机器人是否正常在线。为此,采用BP网络进行状态评判,以各个检测点的输出AD值作为网络输入向量,通过训练样本调节权重,得到机器人的当前状态。神经网络是一个高度并行的分布式系统,故可用于完成对输入信息的快速处理<sup>[1]</sup>,充分利用其强大的非线性映射能力达到环境状态辨识

的目的 它还具有一定的容错能力, 并且对学习过程中未遇到的情况也能进行一定的处理<sup>[1,2]</sup>, 使机器人能够适应各种不同的地面环境

### 3.2 基于Q 学习的寻线导航与策略控制方法

机器人离线行走过程中, 由于传感器提供的信息是不完整、不精确的, 存在一定的冗余和不可靠因素, 因而可能出现走失现象; 另外, 如果地面过于光滑, 也会造成机器人在运动中的打滑或丢步。实验中发现机器人若使用没有学习能力的路径规划方法, 会意外地出现迷失而行动混乱甚至死机的现象, 无法返回导航线。为解决此类问题, 使机器人在复杂未知环境中仍然能准确到达目的地, 提出基于Q 学习的寻线与策略控制算法, 用于机器人走失后的寻线返航

Q 学习<sup>[6,7]</sup> 是一种典型的强化学习算法, 通过试错与环境交互获得策略的改进, 学习的目的是获得一个能最快到达目标状态的控制策略。设机器人可区分的环境状态集为  $S$ , 可选择的动作集为  $A$ , 机器人的学习任务是学习一个策略  $\pi: S \rightarrow A$ , 它基于当前观察到的状态  $s_t$ , 选择下一步动作  $a_t$ , 即  $\pi(s_t) = a_t$ , 使得从任意初始状态  $s_t$  获得的累积回报值

$$V^\pi(s_t) = \sum_{i=0}^{\infty} \gamma^i r_{t+i} \quad (1)$$

最大 其中:  $r_{t+i}$  为  $t+i$  时刻环境给予的回报值;  $0 < \gamma < 1$  为折扣因子, 确定了延迟回报与立即回报的相对比例

Q 算法中学习器存储一个Q 二维表, 每个状态-动作都有一表项  $Q(s, a)$ 。Q 值定义为执行当前所选动作之后按照某一策略执行, 将得到的折扣累积回报。最优Q 值可表示为  $Q^*$ , 其定义为: 执行当前动作并按最优策略执行, 将得到的折扣累积回报, 即

$$Q^*(s, a) = r(s, a) + \mathcal{W}^*(s), \quad (2)$$

其中  $s$  表示在状态  $s$  执行动作  $a$  以后转移到的新状态

Q 表可初始化为 0 或随机值, 机器人重复地观察当前状态  $s$ , 选择某动作  $a$ , 执行此动作; 然后观察回报  $r(s, a)$  以及新状态  $s'$ , 按照下式更新 Q 表:

$$Q(s, a) = r(s, a) + \gamma \max_a Q(s', a). \quad (3)$$

研究表明, 随着学习次数的增加, 不断地探索状态空间, Q 值将逐步趋近于  $Q^*$ , 即 Q 学习最终将收敛到最优策略<sup>[6,7]</sup>。

机器人采用基于 Q 学习的寻线与策略控制算法, 流程见图 3。机器人根据传感器信息确定当前状态, 通过与环境的交互进行学习并选择合适的动作执行。学习的目标是找到线恢复正常运行, 完成导航、避障、路径规划等任务。

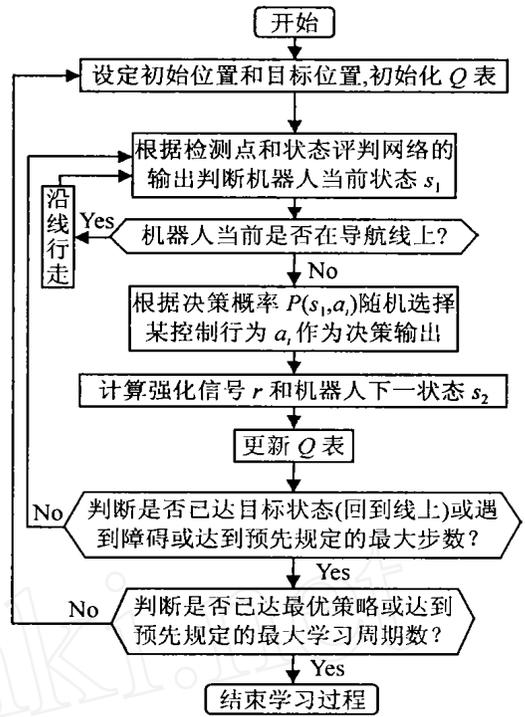


图 3 基于Q 学习的寻线与策略控制算法流程

### 3.3 智能寻线及策略控制系统模型

图 4 为所提出的智能寻线和策略控制系统模型。机器人通过状态评判网络对多个传感器的信息进行融合, 以确定自身状态和相对于导航线的姿态, 根据不同的状态选择动作规划模块或学习/自适应模块进行决策, 进而做出相应控制。将该模型用于自主移动导医机器人, 机器人借助光电传感器自动跟踪地面上成矩形铺设的导航线(宽度约为 35 mm), 沿线行走过程中随机地离线, 利用红外和超声传感器检测障碍, 遇障碍停止运动, 自动判别服务体进行服务, 达到设定的时间阈值后返航; 返回过程中不断检测标志线, 不到线则继续前进, 出现走失状态则通过在线学习调整, 直到返回标志线; 若到线端点, 则转身 180°, 继续沿线行走。

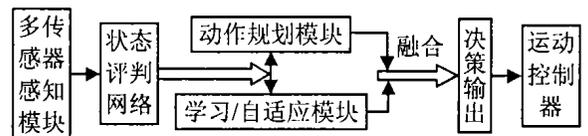


图 4 智能寻线及策略控制系统模型

## 4 实验结果及分析

机器人通过上位机与单片机的通讯实现对底层硬件的实时控制, 上位机采用 VC++ 6.0 编程, 在 Windows 98 环境下运行, 单片机采用 80C51 芯片。针对各种不同的地面环境, 寻线系统的检测点将给出不同的 AD 值。采集了 100 组检测点输出值和机器

人状态训练样本,对状态评判网络进行训练.采用单隐层BP网络,输入维数为6,输入变量为6个检测点的AD值;输出层有5个神经元,理想输出向量分别为10000~00001,代表机器人的状态向量.状态空间 $S = \{s_0, s_1, s_2, s_3, s_4\} = \{\text{正常在线, 线上左偏, 线上右偏, 到线端点, 不在线}\} = \{\text{online, left-off, right-off, at-the-end, out-of-line}\}$ .采用自适应学习率的改进BP算法,利用Matlab仿真得到网络训练误差变化曲线如图5所示.图中实线和虚线分别表示训练误差变化曲线和目标输出线.由于加入了动量因子,训练曲线在第200步左右跳出局部极小,当训练至第357步时,网络性能达标,训练结果比较满意.

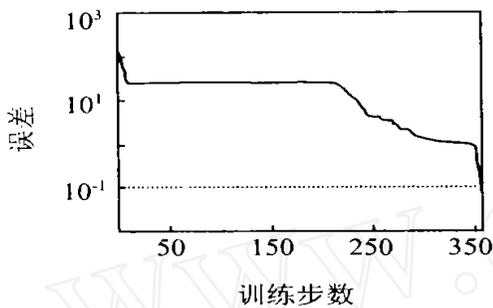


图5 状态评判网络训练误差变化曲线

基于Q学习的寻线与策略控制算法中,状态空间即状态评判网络的输出 $S$ ,动作空间 $A = \{\text{前进, 后退, 停止, 左转 } 15^\circ; \text{右转 } 15^\circ; \text{左转 } 30^\circ; \text{右转 } 30^\circ; \text{左转 } 90^\circ; \text{右转 } 90^\circ; \text{转身 } 180^\circ\}$ .为考察采用Q学习算法后机器人的运动效果,取每100次学习中到达目标(导航线)所用的步数进行统计平均,这个平均值越小,说明控制策略越好.根据实验数据得到图6.图中横坐标表示进行统计平均的时间序号,纵坐标表示达到目标所用步数的平均值.可以看出,随着学习过程的进行,到达目标所用的平均步数呈下降趋势,并收敛于一个最小值.这个值是最优策略所决定的平均步数.实验结果表明,机器人在瞬间的走失状态出现后能够在线学习,经过寻线策略寻优,最终将沿着一

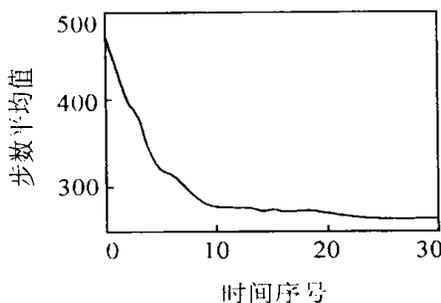


图6 成功返航所用的平均步数与学习时间的关系

条比较优化的路径返回航线.随着学习步数和实验次数的增多,学习效果明显改善.

## 5 结 论

未知环境中的移动机器人导航与控制方法的研究是机器人学和智能控制的一个重要研究领域<sup>[5]</sup>.实际应用中的机器人往往受到传感范围有限的制约,没有全局视野,规划只能依靠实时测得的局部信息.本文提出带有学习功能的智能寻线导航与策略控制方法,并给出了相应的系统模型,使机器人不仅可以在复杂环境下自动跟踪导航线运动,而且能够在偏离或丢失状态下通过学习调整自动返航,提高了控制系统的实时性、鲁棒性和适应性.仿真实验及实际运行结果表明,本文所提出的方法能很好地满足移动机器人对运动控制的需求.

## 参考文献 (References)

- [1] 陆新华,张桂林.室内服务机器人导航方法研究[J].*机器人*, 2003, 25(1): 80-87.  
(Lu X H, Zhang G L. Summarization on indoor service robot navigation[J]. *Robot*, 2003, 25(1): 80-87.)
- [2] 卢韶芳,刘大维.自主式移动机器人导航研究现状及其相关技术[J].*农业机械学报*, 2002, 33(2): 112-116.  
(Lu S F, Liu D W. A survey of research situation on navigation by autonomous mobile robot and its related techniques [J]. *Trans of the Chinese Society for Agricultural Machinery*, 2002, 33(2): 112-116.)
- [3] 万永伦,丁杰雄.一种机器人寻线控制系统[J].*电子科技大学学报*, 2003, 32(1): 48-51.  
(Wan Y L, Ding J X. A Line-tracking control system of robot[J]. *J of UEST of China*, 2003, 32(1): 48-51.)
- [4] 许欢,唐竞新.调制光寻线系统及其在自动行走机器人中的应用[J].*清华大学学报*, 2002, 42(1): 115-117.  
(Xu H, Tang J X. Light modulation line-tracking system and its application in auto-running robot[J]. *J of Tsinghua University*, 2002, 42(1): 115-117.)
- [5] 蔡自兴,贺汉根,陈虹.未知环境中移动机器人导航控制研究的若干问题[J].*控制与决策*, 2002, 17(4): 385-390.  
(Cai Z X, He H G, Chen H. Some issues for mobile robots navigation under unknown environments [J]. *Control and Decision*, 2002, 17(4): 385-390.)
- [6] Tom M. Mitchell. *机器学习*[M].北京:机械工业出版社, 2003.
- [7] 李晓萌,杨煜普,许晓鸣.基于独立学习的多智能体协作决策[J].*控制与决策*, 2002, 17(1): 29-32.  
(Li X M, Yang Y P, Xu X M. Multiragent cooperative decision making based on independent learning [J]. *Control and Decision*, 2002, 17(1): 29-32.)