(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112130570 B
(45)授权公告日 2023.03.28

(21)申请号 202011033850.3

(22)申请日 2020.09.27

(65)同一申请的已公布的文献号
申请公布号 CN 112130570 A

(43)申请公布日 2020.12.25

(73)专利权人 重庆大学
地址 400044 重庆市沙坪坝区沙坪坝正街174号

(72)发明人 陈刚 林卓龙

(74)专利代理机构 北京同恒源知识产权代理有限公司 11275
专利代理师 赵荣之

(51)Int.Cl.
*G05D 1/02*(2020.01)

(56)对比文件
CN 111142536 A,2020.05.12
CN 111142536 A,2020.05.12
CN 205251976 U,2016.05.25
CN 111609851 A,2020.09.01
CN 105403222 A,2016.03.16
CN 111367282 A,2020.07.03
US 2013262353 A1,2013.10.03
方勇纯,等.基于路径积分强化学习方法的蛇形机器人目标导向运动.《模式识别与人工智能》.2019,第32卷(第1期),
严涛.改进的强化学习算法研究及其在机械臂控制中的应用.《中国优秀博硕士学位论文全文数据库信息科技辑》.2020,(第8期),
孙彧,等.多智能体深度强化学习研究综述.《计算机工程与应用》.2020,第56卷(第5期),
黄志峰.深度逆向强化学习在机器人视觉伺服控制中的应用.《中国优秀博硕士学位论文全文数据库信息科技辑》.2020,(第1期),
Kao-shing. Hwang,等.An unified approach to inverse reinforcement learning by oppositive demonstrations.《2016 IEEE International Conference on Industrial Technology (ICIT)》.2016,

审查员 张艺

权利要求书6页 说明书16页 附图6页

(54)发明名称

# 一种基于强化学习的最优输出反馈控制器的导盲机器人

(57)摘要

本发明涉及一种基于强化学习的最优输出反馈控制器的导盲机器人，属于机器人技术领域。通过采用realsense D435i深度摄像机作为视觉传感器，能够准确且高效的获取导盲机器人在前进引导过程中的实时环境信息。为解决导盲机器人在移动过程中所面临诸多不稳定因素的问题，设计了一种基于ADP方法的无模型同步积分强化学习控制器，通过构建基于强化学习的导盲机器人系统的代价函数，建立所构建代价函数的HJB方程，通过基于同步强化学习的方法求解HJB方程，最后通过迭代的方法得到最优解，实现导盲机器人系统的最优控制。

CN 112130570 B

1.一种基于强化学习的最优输出反馈控制器的导盲机器人,其特征在于:包括底层硬件层、感知层和策略层;

采用分层控制,基于ROS机器人操作系统,采用4个伺服电机配套4个万向轮的轮式机器人;

底层用于完成机器人本体的硬件平台搭建;

其中DSP作为底层的控制器,用于采集陀螺仪和里程计信息,并且控制伺服电机的运动;

感知层和策略层的PC用于感知层和策略层的信息采集与计算;

所述导盲机器人的动态模型为:

$$\begin{cases}\dot{x}(t)=f(x(t))+g(x(t))u(t),\\ y(t)=Cx(t).\end{cases} \tag{1}$$

其中$x(t)\in\Re^{n}$是不可测系统状态向量,$u(t)\in\Re^{n*m}$是系统的控制输入,y(t)是系统唯一输出;

现假定f(0)=0,f(x)是未知的且满足||f(x)||≤$b_f$||x||,$b_f$是一个常量;g(x)是已知且有界的,0<||g(x)||≤$b_g$,$b_g$是一个常量;

定义导盲机器人系统的代价函数:

$$J(y(t),\ u(t))=\int_{t}^{\infty}[Q(y(\tau))+U(u(\tau))]d\tau \tag{2}$$

其中,$\tau=[\tau_1,\ \tau_2,\cdots,\ \tau_m]^T\in\Re^m$,Q(y(τ))=$y^T$(τ)Qy(τ)是正定且连续可微的;U(u(τ))是被积函数;考虑系统的输入受限,定义以下一个非二次性能函数:

$$U(u)=2\int_{0}^{u}(\lambda\beta^{-1}(\vartheta/\lambda))^{T}Rd\vartheta \tag{3}$$

其中,$\vartheta\in\Re^m$,β(•)=tanh(•),λ是饱和有界的;R=diag($r_1$,$r_2$,…,$r_m$)>0是对角型;

通过设置基于输出反馈的神经网络观测器,导盲机器人运行时,将实时状态传给设计的控制器进行处理后使系统稳定;

系统状态x(t)不可测,基于输出反馈的状态观测器的动态模型如下:

$$\begin{cases}\dot{\tilde{x}}(t)=f(\tilde{x}(t))+g(\tilde{x}(t))u(t),\\ \tilde{y}(t)=C\tilde{x}(t).\end{cases} \tag{4}$$

由于系统动态模型$f_i(x_i)$的内部函数未知,用神经网络来估计$f_i(x_i)$:

$$f_i(x_i)=A_{oi}x_i+\omega_{oi}^{T}\varphi_{oi}(x_i)+\varepsilon_{oi}(x_i) \tag{5}$$

其中$A_{oi}\in\mathbb{R}^{n\times n}$是赫尔维茨矩阵;$\omega_{oi}\in\mathbb{R}^{l_0\times n}$为神经网络权重,且||$\omega_{oi}$||≤||$\omega_{omi}$||;$\varphi_{oi}(x_i)\in\mathbb{R}^{l_0}$为神经网络激活函数,且$\varphi_{oi}(x_i)\le\varphi_{omi}$;

不妨将系统的动态模型写成:

$$\begin{cases}\dot{x}(t)=A_{0i}(x(t))+n(x(t))+g(x(t))u(t),\\ y(t)=Cx(t).\end{cases}\tag{6}$$

其中，$n(x(t))=f(x)-A_{0i}x(t)$，$A_{oi}\in\Re^{n*n}$是赫尔维茨矩阵；

则观测器的动态模型为：

$$\begin{cases}\dot{\tilde{x}}(t)=F\left(y(t),\tilde{x}(t)\right)+g\left(\tilde{x}(t)\right)u(t),\\ \tilde{y}(t)=C\tilde{x}(t).\end{cases}\tag{7}$$

其中，$F\left(y(t),\tilde{x}(t)\right)=A_{0i}\tilde{x}(t)+\tilde{\omega}_0{}^{T}\varphi_0\left(\tilde{x}(t)\right)+K\left(y(t)-\tilde{y}(t)\right)$，$\tilde{x}(t)$和$\tilde{y}(t)$是所设计观测器的状态；K是观测器的增益，$(A_{0i}-KC)$是赫尔维茨矩阵；系统满足：

$$(A_{0i}-KC)^{T}P+P(A_{0i}-KC)=-qI\tag{8}$$

其中，q是正常量，I是一个单位矩阵，P是一个对称正定矩阵；

定义观测器误差为$\hat{x}_i=x_i-\tilde{x}_i$，则：

$$\dot{\hat{x}}(t)=(A_{0i}-KC)\hat{x}+\hat{\omega}_0{}^{T}\varphi_0\left(\tilde{x}(t)\right)+\xi(x(t))\tag{9}$$

其中，$\hat{\omega}_0=\omega_0-\tilde{\omega}_0$是构造的神经网络观测器的估计误差；

选择下面李雅普诺夫函数：

$$L_{oi}(t)=\frac{1}{2}\tilde{x}_i^{T}P_i\tilde{x}_i+\frac{1}{2}tr\left\{\tilde{\omega}_{oi}^{T}\eta_i^{-1}\tilde{\omega}_{oi}\right\}\tag{10}$$

将上面李雅普诺夫函数进行求导：

$$\dot{L}_{oi}(t)=\frac{1}{2}\dot{\tilde{x}}_i^{T}P_i\tilde{x}_i+\frac{1}{2}\tilde{x}_i^{T}P_i\dot{\tilde{x}}_i+tr\left\{\tilde{\omega}_{oi}^{T}\eta_i^{-1}\dot{\tilde{\omega}}_{oi}\right\}\tag{11}$$

根据观测器测得的误差$\hat{\omega}_0=\omega_0-\tilde{\omega}_0$，知：

$$\dot{\hat{\omega}}=\beta_i\varphi_0\left(\tilde{x}_i\right)\hat{y}_iC_i(A_0-KC)^{-1}+\eta_i\|\hat{y}\|\left(\omega_0-\hat{\omega}_0\right)\tag{12}$$

将(8)(9)(12)带入(11)得：

$$\begin{aligned}\dot{L}_{oi}(t)=&\frac{1}{2}\left[(A_{oi}-K_iC_i)\tilde{x}_i+\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right]^{T}P_i\tilde{x}_i\\&+\frac{1}{2}\tilde{x}_i^{T}P_i\left[(A_{oi}-K_iC_i)\tilde{x}_i+\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right]\\&+tr\left\{\tilde{\omega}_{oi}^{T}\eta_i^{-1}\left[\beta_i\varphi_{oi}(\hat{x}_i)\tilde{y}_i^{T}C_i(A_{oi}-K_iC_i)^{-1}+\eta_i\|\tilde{y}_i\|(\omega_{oi}-\tilde{\omega}_{oi})\right]\right\}\\=&-\frac{1}{2}q_i\tilde{x}_i^{T}\tilde{x}_i+\tilde{x}_i^{T}P_i\left(\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right)\\&+tr\left\{\eta_i^{-1}\beta_i\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)\tilde{y}_i^{T}C_i(A_{oi}-K_iC_i)^{-1}\right\}+tr\left\{\tilde{\omega}_{oi}^{T}\|\tilde{y}_i\|(\omega_{oi}-\tilde{\omega}_{oi})\right\}\end{aligned}\tag{13}$$

由于tr(AB$^T$)＝tr(BA$^T$)＝BA$^T$,(13)改写成：

$$\dot{L}_{oi}(t)=-\frac{1}{2}q_i\tilde{x}_i^T\tilde{x}_i+\tilde{x}_i^TP_i\left(\tilde{\omega}_{oi}^T\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right)$$
$$+\eta_i^{-1}\beta_i\tilde{y}_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\tilde{\omega}_{oi}^T\varphi_{oi}(\hat{x}_i)+\|\tilde{y}_i\|\left(\omega_{oi}-\tilde{\omega}_{oi}\right)\tilde{\omega}_{oi}^T \quad (14)$$

因为$\omega_{oi}$、$\varphi_{oi}(\hat{x}_i)$、$\xi_{oi}(x_i)$有界,式(15)整理为：

$$\dot{L}_{oi}(t)\le-\frac{1}{2}q_i\|\tilde{x}_i\|^2+\|\tilde{x}_i\|\cdot\|P_i\|\left(\|\tilde{\omega}_{oi}\|\varphi_{omi}+\xi_{omi}\right)$$
$$+\eta_i^{-1}\beta_i\|\tilde{x}_i\|\cdot\left\|C_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\right\|\cdot\|\tilde{\omega}_{oi}\|\varphi_{omi} \quad (15)$$
$$+\|C_i\|\cdot\|\tilde{x}_i\|\left(\|\omega_{oi}\|\cdot\|\tilde{\omega}_{oi}\|-\|\tilde{\omega}_{oi}\|^2\right)$$

所以：$\dot{L}_{0i}(t)\le-\frac{1}{2}q_i\|\hat{x}_i\|^2$

$$+\|\hat{x}_i\|^2+\|\hat{x}_i\|\cdot\|C_i\|\left(\frac{\|P_i\|}{\|C_i\|}\xi_{omi}+2\gamma\|\hat{\omega}_{oi}\|-\|\hat{\omega}_{oi}\|^2\right) \quad (16)$$

为使$\dot{L}_{oi}(t)<0$，只需令$\frac{\|P_i\|}{\|C_i\|}\xi_{omi}+2\gamma_i\|\tilde{\omega}_{oi}\|-\|\tilde{\omega}_{oi}\|^2<0$；即只要满足：

$$\|\hat{x}(t)\|>\frac{2\left(\|P_i\|\xi_{omi}+\gamma^2\|C_i\|\right)}{q} \quad (17)$$

机器人的输出$y(t)=C\tilde{x}(t)$,代价函数写成下面的形式：

$$J\left(\tilde{x}(t),\ u(t)\right)=\int_t^{\infty}\left[\left(\tilde{x}^T(\tau)Q_c(\tau)+U(u(\tau))\right)\right]d\tau \quad (18)$$

其中,$Q_c$＝$C^T$QC半正定的；

利用牛顿-莱布尼茨公式对式(18)中时间t求导得到贝尔曼方程：

$$\dot{J}\left(\tilde{x}(t),\ u(t)\right)=-\tilde{x}^T(t)Q_c\tilde{x}(t)-U(u(t)) \quad (19)$$

联立(3)(19)得：

$$\tilde{x}^T(t)Q_c\tilde{x}(t)+2\int_0^u\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^TRd\vartheta+\dot{J}\left(\tilde{x}(t),u(t)\right)=0 \quad (20)$$

定义Hamiltonian方程为：

$$H\left(\tilde{x},\ u,\nabla J(x)\right)=\tilde{x}^T(t)Q_c\tilde{x}(t)+2\int_0^u\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^TRd\vartheta$$
$$+(\nabla J)^T\left(F\left(y,\tilde{x}\right)\right)+g(x)u(t)=0 \quad (21)$$

令最优代价函数为$J^*\left(\tilde{x}(t)\right)$：

$$J^*\left(\tilde{x}(t)\right)=\min_{u\in\Omega}\int_t^{\infty}\left(\tilde{x}^{T}(\tau)Q_c\,\tilde{x}(\tau)+U(u(\tau))\right)d\tau \qquad (22)$$

则根据(21)中Hamiltonian方程,得到如下HJB方程

$$H\left(\tilde{x},\ u,\nabla J^*\right)=\tilde{x}^{T}(\tau)Q_c\,\tilde{x}(\tau)+2\int_0^{u^*}\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^{T}Rd\vartheta$$

$$+\nabla J^{*T}\left(F\left(y,\tilde{x}\right)+g(x)u^*(t)\right)=0 \qquad (23)$$

当稳定性条件$\frac{\partial H\left(\tilde{x},\ u,\nabla J\right)}{\partial u(t)}=0$时,得到如下最优控制输入:

$$u^*(t)=\arg\min_{u\in\Omega}\left[H\left(\tilde{x},u,\nabla J^*\right)\right]=-\lambda\tanh\left(\frac{1}{2\lambda}R^{-1}g^{T}(x)\nabla J^*\right) \qquad (24)$$

由于HJB方程很难求解,在该算法中采用IRL的策略迭代来求解上述HJB方程;首先将(18)中的值函数写成下面贝尔曼方程的形式:

$$J\left(\tilde{x}(t)\right)=\int_t^{t+T}\left(\tilde{x}^{T}(\tau)Q_c\,\tilde{x}(\tau)\right)+U(u(\tau))d\tau+J\left(\tilde{x}(t+T)\right) \qquad (25)$$

得到下面基于策略迭代的在线IRL算法:

算法:基于策略迭代的在线IRL算法求解HJB方程

步骤1:利用下式解出$J^{(i)}(x(t))$

$$J^{(i)}(x(t))=\int_t^{t+T}\left(x^{T}(\tau)Q_c x(\tau)\right)+U(u(\tau))d\tau+J^{(i)}(x(t+T)) \qquad (12)$$

步骤2:通过下式更新控制策略:

$$u^{(i+1)}x(t)=-\lambda\tanh\left(\frac{1}{2\lambda}R^{-1}g^{T}(x)\nabla J^{(i)}(x)\right) \qquad (13)$$

步骤3:令$u_i^j=u_i^{j+1}$,返回步骤1,直到$J^{(i)}x(t)$收敛到最小值。

2.根据权利要求1所述的一种基于强化学习的最优输出反馈控制器的导盲机器人,其特征在于:所述4个伺服电机采用24V供电,通过DSP编码,将上层发布的轮速信息处理后执行;

采用24V 10AH的锂电池作为机器人的底层供电电源;其中,伺服电机驱动器为24V供电,DSP为5V供电;稳压模块调节电压,使其输出一个5V电压。

3.根据权利要求1所述的一种基于强化学习的最优输出反馈控制器的导盲机器人,其特征在于:所述感知层由视觉识别和语音识别两部分组成;

其中,视觉感知部分为:

1)基于realsense D435i深度摄像机的导盲机器人视觉识别系统的实现

根据机器人与识别目标的位置,第一摄像头向下倾斜30安装,第二摄像头向上倾斜20°安装;机器人后方安装第三摄像头实现主人面部识别与跟踪;深度相机通过USB与上位机连

接,激光雷达通过以太网与上位机通信;

2)基于ROS系统和realsense D435i深度摄像机实现导盲机器人的目标图像信息采集

通过ROS系统中的Master发布命令,运行realsense D435i深度摄像机启动节点,读入图像或视频流,通过OPENCV和ROS的接口完成图像格式转换,将采集到的图像储存,使用Python构建深度学习数据集,安装REQUESTS包,创建Python脚本下载图像,配置环境,然后修剪深度学习图像数据集;

3)基于YOLOV3深度学习和realsense D435i深度摄像机的导盲机器人目标识别算法实现

准备数据:使用yolo_mark对图片进行标注

修改配置文件:修改训练数据、验证数据、物体名称文件路径,修改神经网络的详细构建参数

训练及输出:训练网络,输出参数进行保存;

测试:验证模型效果

导盲机器人系统搭建在ROS机器人操作系统下,视觉图像数据采集储存在ROS系统中,需要在Ubuntu系统下构建YOLOV3深度学习网络;

导盲机器人通过第一摄像头和第二摄像头与YOLOV3深度学习网络识别出前方物体信息,将物体的具体识别信息以及位置坐标回传给上位机,通过第三摄像头来识别主人信息,再通过上层决策信息来决定机器人运动;

盲道识别:

采用基于颜色区域的图像分割的方法,筛选出盲道区域,并对盲道区域进行边缘提取实现盲道的识别;首先将图像由RGB转为HSI色彩空间,RGB色彩空间到HSI色彩空间的转换关系如下:

$$H=\begin{cases}\cos^{-1}\left\{\dfrac{(R-G)+(R-B)}{\sqrt{2(R-B)+(R-B)(G-B)}}\right\}, & (B\le G)\text{且}(R\neq B\text{或}R\neq G)\\ 2\pi-H\end{cases}$$

$$S=1-\frac{3\min(R,G,B)}{R+G+B}$$

$$I=\frac{1}{3}(R+G+B)$$

通过转换后得到在HSI色彩空间上的图像;

语音识别部分为:

1)基于ROS的语音交互系统搭建

语音云服务平台是位于云端的服务器,包括语音识别、语义理解和语音合成;除去语音云服务平台系统分为三层:其中最底层为Linux内核,为系统运行环境;其次是中间层,该层主要是第三方库以及ROS系统;基于ROS的人机语音交互系统从采集语音一直到机器人做出响应,划分为如下几个功能节点:语音识别节点、语义分析节点、实时性信息获取节点、语音合成节点和音频播放节点;

在ROS中实现的语音交互主要功能包括:语音信息采集、语音识别节点、语音合成节点、语义分析节点、实时性信息获取节点、机器人控制功能节点;

语音信息采集:通过机器人外置麦克风采集语音信息,将采集的语音信息存储为音频文件;

语音识别节点:语音识别节点负责将采集的语音信息识别为文字信息;

语音合成节点:语音合成节点负责将请求信息合成为音频;

语义分析节点:语义分析节点具有对从语音识别节点接收到的请求信息进行理解,以判决机器人应该执行何种操作的功能;

实时性信息获取节点:通过实时性信息获取节点能得到实时变化的信息内容;

机器人控制功能节点:机器人控制功能节点包括控制机器人行走、避障、到达指定位置节点;

2)语音人机交互具体需实现的功能

盲人以语音的形式唤醒导盲机器人;

盲人以语音形式控制导盲机器人选择模式;

导盲机器人遇到障碍时,播报“前方有障碍物,请注意通行”;

导盲机器人在识别到盲道时,播报“前方盲道,请沿盲道行走”;

导盲机器人识别出红绿灯时,播报“前方红绿灯,请等待”;

导盲机器人识别绿灯剩余时间时,播报“绿灯时间不足,请等待下次通行”。

4.根据权利要求1所述的一种基于强化学习的最优输出反馈控制器的导盲机器人,其特征在于:所述策略层中,导盲机器人在路径规划中的相关动作决策,包括接受视觉传来的障碍物信息后改变电机转向绕开障碍物、在红灯时控制电机停止、在红灯转绿灯时启动电机、盲人通过语音唤醒时启动导盲机器人以及相应控制算法的实现;

为确保导盲机器人因故障无法自主控制时,还设置手势杆操作器;

手势杆操作器输入功能:在自由散步模式下,通过手势杆操作器控制机器人同时为方便盲人使用手势杆操作器,手势杆操作器的按键应设计得更适合盲人使用;在手势杆操作器中箭头方向表示机器人运动的方向,中间圆形键表示为暂停键;当运行在自由散步模式下时,通过手势杆操作器方向实现对机器人的运动控制;

将底层信息及里程计和陀螺仪信息传到机器人,通过RS232通讯线使机器人PC传递位置信息给主控DSP320F2812;选择DSP320F2815作为主控芯片,DSP320F281含有多种外接接口,输出PWM波和脉冲信号的功能,通过RS232通讯线接受PC端传递来的信息。

# 一种基于强化学习的最优输出反馈控制器的导盲机器人

## 技术领域

[0001] 本发明属于机器人技术领域,涉及一种基于强化学习的最优输出反馈控制器的导盲机器人。

## 背景技术

[0002] 目前,导盲机器类型并不是很多,其主要类型有(1)导盲手杖:视觉障碍者最普遍的就是手握一根白色手杖,但是手杖结构简单,并不智能,其正在被一种叫做镭射手杖的导盲机器所取代;(2)穿戴式导盲辅助工具:可分为引导式和全景式;引导式主要是避障,而全景式在避障的功能要求之上加入了超声波,试图对视觉障碍者的周边环境进行构图。(3)移动式导盲机器人:其主要原理是以移动机器人为基础,加入红外传感及超声波模块来探测周围障碍物。上述三种导盲机器类型中,最智能化的就是移动式导盲机器人,但是现今常见的导盲机器人中,大多采用的是红外传感器和超声波来探测障碍物,超声波的原理是通过超声波碰到杂质或分界面会产生显著反射形成反射成回波,通过接受回波判断前方是否有物体,以及物体的距离的,但是在盲人行驶道路的复杂环境中并不适用。

[0003] 在本专利中设计的导盲机器人采用YOLOV3深度学习算法和深度摄像机数据的目标辨识方法,对深度摄像机数据集进行标注,然后对采用的YOLOV3深度学习网络进行训练,将训练完成的参数输出,使用测试集对模型进行目标检测效果测试。这种目标识别方法更加精确也更加灵活,可以在盲人行驶道路上识别移动障碍物和静止障碍物,基于这种识别方式的导盲机器人更加智能化。

[0004] 在现今众多导盲机器人应用中,很少考虑设计一种稳定,有效的控制算法使导盲机器人在行驶和人机交互时更加稳定。因为导盲机器人在引导盲人行走时,会遇见许多突发事件,如突然袭来的自行车或人;道路不平;上坡或下坡;与人进行语音交互时突然受到外来信号干扰等,这些外来干扰都会影响导盲机器人的品质及控制的稳定性。所以设计一种有效的控制算法对导盲机器人进行控制就显得十分重要。所以在本发明中采用模型完全未知的积分强化学习算法构造控制器对导盲机器人进行控制。强化学习算法(RL)是建立在成功的控制策略应该被记住的想法上,而后通过一个强化信号使得它们可以在第二次使用。强化学习算法求解最优控制问题的主要优点是不需要知道系统动力学知识及相关辨识的基础上,只在系统能够获得足够的数据,则可根据预定义的性能指标函数逼近最优控制策略。强化学习算法(RL)通常基于策略迭代(PI)技术,在策略评估和策略改进之间进行迭代。而积分强化学习(IRL)是在线性和非线性强化学习算法的基础上,将积分步骤中的时间间隔(t,t+T)视为强化信号,这个算法放宽了对输入耦合动力学知识的局限,即对系统是完全未知的。IRL算法对传统强化学习算法的策略评估和策略改进分别都进行了优化。在本发明中,我采用的是一种在线同步策略迭代技术,其critic和actor是同时更新的,通过在actor的优化中加入一个额外的约束条款,可以保证闭环系统的动态稳定性。

[0005] 但是考虑到导盲机器人在引导行驶过程中会发生许多未知的变故(一般在实际装置中,都普遍存在振幅约束,即约束输入或执行器饱和),所以在控制器设计过程中必须考

虑约束控制输入，导盲机器人在实际情况中系统状态的不可测，所以在本发明中设计了一种基于强化学习的最优输出反馈控制器。

## 发明内容

[0006] 有鉴于此，本发明的目的在于提供一种基于强化学习的最优输出反馈控制器的导盲机器人。

[0007] 为达到上述目的，本发明提供如下技术方案：

[0008] 一种基于强化学习的最优输出反馈控制器的导盲机器人，包括底层硬件层、感知层和策略层；

[0009] 采用分层控制，基于ROS机器人操作系统，采用4个伺服电机配套4个万向轮的轮式机器人；

[0010] 底层用于完成机器人本体的硬件平台搭建；

[0011] 其中DSP作为底层的控制器，用于采集陀螺仪和里程计信息，并且控制伺服电机的运动；

[0012] 感知层和策略层的PC用于感知层和策略层的信息采集与计算。

[0013] 可选的，所述4个伺服电机采用24V供电，通过DSP编码，将上层发布的轮速信息处理后执行；

[0014] 采用24V 10AH的锂电池作为机器人的底层供电电源；其中，伺服电机驱动器为24V供电，DSP为5V供电；稳压模块调节电压，使其输出一个5V电压。

[0015] 可选的，所述感知层由视觉识别和语音识别两部分组成；

[0016] 其中，视觉感知部分为：

[0017] 1)基于realsense D435i深度摄像机的导盲机器人视觉识别系统的实现

[0018] 根据机器人与识别目标的位置，第一摄像头向下倾斜30安装，第二摄像头向上倾斜20°安装；机器人后方安装第三摄像头实现主人面部识别与跟踪；深度相机通过USB与上位机连接，激光雷达通过以太网与上位机通信；

[0019] 2)基于ROS系统和realsense D435i深度摄像机实现导盲机器人的目标图像信息采集

[0020] 通过ROS系统中的Master发布命令，运行realsense D435i深度摄像机启动节点，读入图像或视频流，通过OPENCV和ROS的接口完成图像格式转换，将采集到的图像储存，使用Python构建深度学习数据集，安装REQUESTS包，创建Python脚本下载图像，配置环境，然后修剪深度学习图像数据集；

[0021] 3)基于YOLOV3深度学习和realsense D435i深度摄像机的导盲机器人目标识别算法实现

[0022] 准备数据：使用yolo_mark对图片进行标注

[0023] 修改配置文件：修改训练数据、验证数据、物体名称文件路径，修改神经网络的详细构建参数

[0024] 训练及输出：训练网络，输出参数进行保存；

[0025] 测试：验证模型效果

[0026] 导盲机器人系统搭建在ROS机器人操作系统下，视觉图像数据采集储存在ROS系统

中,需要在Ubuntu系统下构建YOLOV3深度学习网络;

[0027] 导盲机器人通过第一摄像头和第二摄像头与YOLOV3深度学习网络识别出前方物体信息,将物体的具体识别信息以及位置坐标回传给上位机,通过第三摄像头来识别主人信息,再通过上层决策信息来决定机器人运动;

[0028] 盲道识别:

[0029] 采用基于颜色区域的图像分割的方法,筛选出盲道区域,并对盲道区域进行边缘提取实现盲道的识别;首先将图像由RGB转为HSI色彩空间,RGB色彩空间到HSI色彩空间的转换关系如下:

[0030] $$H=\begin{cases}\cos^{-1}\left\{\dfrac{(R-G)+(R-B)}{\sqrt{2(R-B)+(R-B)(G-B)}}\right\}, (B\le G)\text{且}(R\ne B\text{或}R\ne G)\\ 2\pi-H\end{cases}$$

[0031] $$S=1-\frac{3\min(R,G,B)}{R+G+B}$$

[0032] $$I=\frac{1}{3}(R+G+B)$$

[0033] 通过转换后得到在HSI色彩空间上的图像;

[0034] 语音识别部分为:

[0035] 1)基于ROS的语音交互系统搭建

[0036] 语音云服务平台是位于云端的服务器,包括语音识别、语义理解和语音合成;除去语音云服务平台系统分为三层:其中最底层为Linux内核,为系统运行环境;其次是中间层,该层主要是第三方库以及ROS系统;基于ROS的人机语音交互系统从采集语音一直到机器人做出响应,划分为如下几个功能节点:语音识别节点、语义分析节点、实时性信息获取节点、语音合成节点和音频播放节点;

[0037] 在ROS中实现的语音交互主要功能包括:语音信息采集、语音识别节点、语音合成节点、语义分析节点、实时性信息获取节点、机器人控制功能节点;

[0038] 语音信息采集:通过机器人外置麦克风采集语音信息,将采集的语音信息存储为音频文件;

[0039] 语音识别节点:语音识别节点负责将采集的语音信息识别为文字信息;

[0040] 语音合成节点:语音合成节点负责将请求信息合成为音频;

[0041] 语义分析节点:语义分析节点具有对从语音识别节点接收到的请求信息进行理解,以判决机器人应该执行何种操作的功能;

[0042] 实时性信息获取节点:通过实时性信息获取节点能得到实时变化的信息内容;

[0043] 机器人控制功能节点:机器人控制功能节点包括控制机器人行走、避障、到达指定位置等节点;

[0044] 2)语音人机交互具体需实现的功能

[0045] 盲人以语音的形式唤醒导盲机器人;

[0046] 盲人以语音形式控制导盲机器人选择模式;

[0047] 导盲机器人遇到障碍时,播报“前方有障碍物,请注意通行”;

[0048] 导盲机器人在识别到盲道时,播报“前方盲道,请沿盲道行走”;

[0049] 导盲机器人识别出红绿灯时，播报“前方红绿灯，请等待”；

[0050] 导盲机器人识别绿灯剩余时间时，播报“绿灯时间不足，请等待下次通行”。

[0051] 可选的，所述策略层中，导盲机器人在路径规划中的相关动作决策，包括接受视觉传来的障碍物信息后改变电机转向绕开障碍物、在红灯时控制电机停止、在红灯转绿灯时启动电机、盲人通过语音唤醒时启动导盲机器人以及相应控制算法的实现；

[0052] 为确保导盲机器人因故障无法自主控制时，还设置手势杆操作器；

[0053] 手势杆操作器输入功能：在自由散步模式下，通过手势杆操作器控制机器人同时为方便盲人使用手势杆操作器，手势杆操作器的按键应设计得更适合盲人使用；在手势杆操作器中箭头方向表示机器人运动的方向，中间圆形键表示为暂停键；当运行在自由散步模式下时，通过手势杆操作器方向实现对机器人的运动控制；

[0054] 将底层信息及里程计和陀螺仪信息传到机器人，通过RS232通讯线使机器人PC传递位置信息给主控DSP320F2812；选择DSP320F2815作为主控芯片，DSP320F281含有多种外接接口，输出PWM波和脉冲信号的功能，通过RS232通讯线接受PC端传递来的信息。

[0055] 可选的，所述导盲机器人的动态模型为：

[0056] $$\begin{cases}\dot{x}(t)=f(x(t))+g(x(t))u(t)\\ y(t)=Cx(t)\end{cases} \quad (1)$$

[0057] 其中$x(t)\in\Re^{n}$是不可测系统状态向量，$u(t)\in\Re^{n*m}$是系统的控制输入，y(t)是系统唯一输出；

[0058] 现假定f(0)＝0，f(x)是未知的且满足$||f(x)||\leq b_f||x||$，$b_f$是一个常量；g(x)是已知且有界的，$0<||g(x)||\leq b_g$，$b_g$是一个常量；

[0059] 定义导盲机器人系统的代价函数：

[0060] $$J(y(t),\ u(t))=\int_{t}^{\infty}[Q(y(\tau))+U(u(\tau))]d\tau \quad (2)$$

[0061] 其中，$\tau=[\tau_1,\ \tau_2,\cdots,\ \tau_m]^T\in\Re^m$，$Q(y(\tau))=y^T(\tau)Qy(\tau)$是正定且连续可微的；U(u(τ))是被积函数；考虑系统的输入受限，定义以下一个非二次性能函数：

[0062] $$U(u)=2\int_{0}^{u}(\lambda\beta^{-1}(\vartheta/\lambda))^T Rd\vartheta \quad (3)$$

[0063] 其中，$\vartheta\in\Re^m$，β(•)＝tanh(•)，λ是饱和有界的；$R=diag(r_1,r_2,\cdots,r_m)>0$是对角型；

[0064] 通过设置基于输出反馈的神经网络观测器，导盲机器人运行时，将实时状态传给设计的控制器进行处理后使系统稳定；

[0065] 系统状态x(t)不可测，基于输出反馈的状态观测器的动态模型如下：

[0066] $$\begin{cases}\dot{\tilde{x}}(t)=f(\tilde{x}(t))+g(\tilde{x}(t))u(t),\\ \tilde{y}(t)=C\tilde{x}(t).\end{cases} \quad (4)$$

[0067] 由于系统动态模型$f_i(x_i)$的内部函数未知，用神经网络来估计$f_i(x_i)$：

[0068] $$f_i(x_i)=A_{oi}x_i+\omega_{oi}^T\varphi_{oi}(x_i)+\varepsilon_{oi}(x_i) \quad (5)$$

[0069] 其中$A_{oi}\in\mathbb{R}^{n\times n}$是赫尔维茨矩阵；$\omega_{oi}\in\mathbb{R}^{l_0\times n}$为神经网络权重，且$||\omega_{oi}||\leq||\omega_{omi}||$；$\varphi_{oi}(x_i)\in\mathbb{R}^{l_0}$为神经网络激活函数，且$\varphi_{oi}(x_i)\leq\varphi_{omi}$；

[0070] 不妨将系统的动态模型写成：

[0071] $$\begin{cases}\dot{x}(t)=A_{0i}(x(t))+n(x(t))+g(x(t))u(t),\\ y(t)=Cx(t).\end{cases} \quad (6)$$

[0072] 其中，$n(x(t))=f(x)-A_{0i}x(t)$，$A_{oi}\in\Re^{n*n}$是赫尔维茨矩阵；

[0073] 则观测器的动态模型为：

[0074] $$\begin{cases}\dot{\tilde{x}}(t)=F\left(y(t),\tilde{x}(t)\right)+g\left(\tilde{x}(t)\right)u(t),\\ \tilde{y}(t)=C\tilde{x}(t).\end{cases} \quad (7)$$

[0075] 其中，$F\left(y(t),\tilde{x}(t)\right)=A_{0i}\tilde{x}(t)+\tilde{\omega}_0{}^{T}\varphi_0\left(\tilde{x}(t)\right)+K\left(y(t)-\tilde{y}(t)\right)$，$\tilde{x}(t)$和$\tilde{y}(t)$是所设计观测器的状态；K是观测器的增益，$(A_{0i}-KC)$是赫尔维茨矩阵；系统满足：

[0076] $$(A_{0i}-KC)^TP+P(A_{0i}-KC)=-qI \quad (8)$$

[0077] 其中，q是正常量，I是一个单位矩阵，P是一个对称正定矩阵；

[0078] 定义观测器误差为$\hat{x}_i=x_i-\tilde{x}_i$，则：

[0079] $$\dot{\hat{x}}(t)=(A_{0i}-KC)\hat{x}+\hat{\omega}_0{}^{T}\varphi_0\left(\tilde{x}(t)\right)+\xi(x(t)) \quad (9)$$

[0080] 其中，$\hat{\omega}_0=\omega_0-\tilde{\omega}_0$是构造的神经网络观测器的估计误差；

[0081] 选择下面李雅普诺夫函数：

[0082] $$L_{oi}(t)=\frac{1}{2}\tilde{x}_i^TP_i\tilde{x}_i+\frac{1}{2}tr\left\{\tilde{\omega}_{oi}^T\eta_i^{-1}\tilde{\omega}_{oi}\right\} \quad (10)$$

[0083] 将上面李雅普诺夫函数进行求导：

[0084] $$\dot{L}_{oi}(t)=\frac{1}{2}\dot{\tilde{x}}_i^TP_i\tilde{x}_i+\frac{1}{2}\tilde{x}_i^TP_i\dot{\tilde{x}}_i+tr\left\{\tilde{\omega}_{oi}^T\eta_i^{-1}\dot{\tilde{\omega}}_{oi}\right\} \quad (11)$$

[0085] 根据观测器测得的误差$\hat{\omega}_0=\omega_0-\tilde{\omega}_0$，知：

[0086] $$\dot{\hat{\omega}}=\beta_i\varphi_0\left(\tilde{x}_i\right)\hat{y}_iC_i(A_0-KC)^{-1}+\eta_i\|\hat{y}\|\left(\omega_0-\hat{\omega}_0\right) \quad (12)$$

[0087] 将(8)(9)(12)带入(11)得：

$$\begin{aligned}\dot{L}_{oi}(t)=&\frac{1}{2}\left[\left(A_{oi}-K_iC_i\right)\tilde{x}_i+\tilde{\omega}_{oi}^T\varphi_{oi}\left(\hat{x}_i\right)+\xi_{oi}\left(x_i\right)\right]^T P_i\tilde{x}_i\\&+\frac{1}{2}\tilde{x}_i^T P_i\left[\left(A_{oi}-K_iC_i\right)\tilde{x}_i+\tilde{\omega}_{oi}^T\varphi_{oi}\left(\hat{x}_i\right)+\xi_{oi}\left(x_i\right)\right]\\&+tr\left\{\tilde{\omega}_{oi}^T\eta_i^{-1}\left[\beta_i\varphi_{oi}\left(\hat{x}_i\right)\tilde{y}_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}+\eta_i\left\|\tilde{y}_i\right\|\left(\omega_{oi}-\tilde{\omega}_{oi}\right)\right]\right\}\\=&-\frac{1}{2}q_i\tilde{x}_i^T\tilde{x}_i+\tilde{x}_i^TP_i\left(\tilde{\omega}_{oi}^T\varphi_{oi}\left(\hat{x}_i\right)+\xi_{oi}\left(x_i\right)\right)\\&+tr\left\{\eta_i^{-1}\beta_i\tilde{\omega}_{oi}^T\varphi_{oi}\left(\hat{x}_i\right)\tilde{y}_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\right\}+tr\left\{\tilde{\omega}_{oi}^T\left\|\tilde{y}_i\right\|\left(\omega_{oi}-\tilde{\omega}_{oi}\right)\right\}\end{aligned}\tag{13}$$

[0088]

[0089] 由于tr(AB$^T$)=tr(BA$^T$)=BA$^T$,(13)改写成:

$$\begin{aligned}\dot{L}_{oi}(t)=&-\frac{1}{2}q_i\tilde{x}_i^T\tilde{x}_i+\tilde{x}_i^TP_i\left(\tilde{\omega}_{oi}^T\varphi_{oi}\left(\hat{x}_i\right)+\xi_{oi}\left(x_i\right)\right)\\&+\eta_i^{-1}\beta_i\tilde{y}_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\tilde{\omega}_{oi}^T\varphi_{oi}\left(\hat{x}_i\right)+\left\|\tilde{y}_i\right\|\left(\omega_{oi}-\tilde{\omega}_{oi}\right)\tilde{\omega}_{oi}^T\end{aligned}\tag{14}$$

[0090]

[0091] 因为$\omega_{oi}$、$\varphi_{oi}\left(\hat{x}_i\right)$、$\xi_{oi}(x_i)$有界,式(15)整理为:

$$\begin{aligned}\dot{L}_{oi}(t)\le&-\frac{1}{2}q_i\left\|\tilde{x}_i\right\|^2+\left\|\tilde{x}_i\right\|\cdot\left\|P_i\right\|\left(\left\|\tilde{\omega}_{oi}\right\|\varphi_{omi}+\xi_{omi}\right)\\&+\eta_i^{-1}\beta_i\left\|\tilde{x}_i\right\|\cdot\left\|C_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\right\|\cdot\left\|\tilde{\omega}_{oi}\right\|\varphi_{omi}\\&+\left\|C_i\right\|\cdot\left\|\tilde{x}_i\right\|\left(\left\|\omega_{oi}\right\|\cdot\left\|\tilde{\omega}_{oi}\right\|-\left\|\tilde{\omega}_{oi}\right\|^2\right)\end{aligned}\tag{15}$$

[0092]

[0093] 所以:$\dot{L}_{0i}(t)\le-\frac{1}{2}q_i\|\hat{x}_i\|^2$

[0094] $$+\|\hat{x}_i\|^2+\|\hat{x}_i\|\cdot\|C_i\|\left(\frac{\|P_i\|}{\|C_i\|}\xi_{omi}+2\gamma\|\hat{\omega}_{oi}\|-\|\hat{\omega}_{oi}\|^2\right)\tag{16}$$

[0095] 为使$\dot{L}_{oi}(t)<0$,只需令$\frac{\|P_i\|}{\|C_i\|}\xi_{omi}+2\gamma_i\|\tilde{\omega}_{oi}\|-\|\tilde{\omega}_{oi}\|^2<0$;

[0096] 即只要满足:

[0097] $$\|\hat{x}(t)\|>\frac{2\left(\|P_i\|\xi_{omi}+\gamma^2\|C_i\|\right)}{q}\tag{17}$$

[0098] 机器人的输出$y(t)=C\tilde{x}(t)$,代价函数写成下面的形式:

[0099] $$J\left(\tilde{x}(t),\ u(t)\right)=\int_t^{\infty}\left[\left(\tilde{x}^T(\tau)Q_c(\tau)+U(u(\tau))\right)\right]d\tau\tag{18}$$

[0100] 其中,$Q_c=C^TQC$半正定的;

[0101] 利用牛顿-莱布尼茨公式对式(18)中时间t求导得到贝尔曼方程:

[0102] $$\dot{J}\left(\tilde{x}(t),\ u(t)\right)=-\tilde{x}^T(t)Q_c\tilde{x}(t)-U(u(t))\tag{19}$$

[0103] 联立(3)(19)得:

[0104] $$\tilde{x}^{T}(t)Q_c\tilde{x}(t)+2\int_0^u\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^T Rd\vartheta+J\left(\tilde{x}(t),u(t)\right)=0 \qquad (20)$$

[0105] 定义Hamiltonian方程为：

[0106] $$H\left(\tilde{x},\ u,\nabla J(x)\right)=\tilde{x}^T(t)Q_c\tilde{x}(t)+2\int_0^u\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^T Rd\vartheta$$
$$+(\nabla J)^T\left(F\left(y,\tilde{x}\right)\right)+g(x)u(t)=0 \qquad (21)$$

[0107] 令最优代价函数为$J^*\left(\tilde{x}(t)\right)$：

[0108] $$J^*\left(\tilde{x}(t)\right)=\min_{u\in\Omega}\int_t^\infty\left(\tilde{x}^T(\tau)Q_c\tilde{x}(\tau)+U(u(\tau))\right)d\tau \qquad (22)$$

[0109] 则根据(21)中Hamiltonian方程，得到如下HJB方程

[0110] $$H\left(\tilde{x},\ u,\nabla J^*\right)=\tilde{x}^T(\tau)Q_c\tilde{x}(\tau)+2\int_0^{u^*}\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^T Rd\vartheta$$
$$+\nabla J^{*T}\left(F\left(y,\tilde{x}\right)+g(x)u^*(t)\right)=0 \qquad (23)$$

[0111] 当稳定性条件$\frac{\partial H\left(\tilde{x},\ u,\nabla J\right)}{\partial u(t)}=0$时，得到如下最优控制输入：

[0112] $$u^*(t)=\arg\min_{u\in\Omega}\left[H\left(\tilde{x},u,\nabla J^*\right)\right]=-\lambda\tanh\left(\frac{1}{2\lambda}R^{-1}g^T(x)\nabla J^*\right) \qquad (24)$$

[0113] 由于HJB方程很难求解，在该算法中采用IRL的策略迭代来求解上述HJB方程；

[0114] 首先将(18)中的值函数写成下面贝尔曼方程的形式：

[0115] $$J\left(\tilde{x}(t)\right)=\int_t^{t+T}\left(\tilde{x}^T(\tau)Q_c\tilde{x}(\tau)\right)+U(u(\tau))d\tau+J\left(\tilde{x}(t+T)\right) \qquad (25)$$

[0116] 得到下面基于策略迭代的在线IRL算法：

[0117] 算法：基于策略迭代的在线IRL算法求解HJB方程

[0118] 步骤1：利用下式解出$J^{(i)}(x(t))$

[0119] $$J^{(i)}(x(t))=\int_t^{t+T}\left(x^T(\tau)Q_cx(\tau)\right)+U(u(\tau))d\tau+J^{(i)}(x(t+T)) \qquad (12)$$

[0120] 步骤2：通过下式更新控制策略：

[0121] $$u^{(i+1)}x(t)=-\lambda\tanh\left(\frac{1}{2\lambda}R^{-1}g^T(x)\nabla J^{(i)}(x)\right) \qquad (13)$$

[0122] 步骤3：令$u_i^j=u_i^{j+1}$，返回步骤1，直到$J^{(i)}x(t)$收敛到最小值。

[0123] 本发明的有益效果在于：

[0124] 1、本设计采用多传感器融合技术，以DSP320F2815作为主控芯片，可以实现导盲机器人的多功能协同处理；

[0125] 2、本设计采用HOKUYO激光雷达和realsense D435i深度摄像机共同对障碍物信息进行处理，提高了识别的精确性，使得导盲机器人无论是在识别障碍物还是在识别盲道、红绿灯方面的精度都有了很大的提高；

[0126] 3、本设计对YOLOv3网络结构的改进，其检测的精度更高，通过前方摄像头与YOLOV3深度学习网络识别出前方物体信息通过后方摄像头来识别主人信息，再通过上层决策信息来决定机器人运动，通过采用ROS系统可以很方便地处理上传的具体识别信息；

[0127] 4、本设计采用基于ROS系统的语音处理模块，通过使用ROS提供的话题、服务方式实现系统中相关模块之间的通信，同时定义通信时的信息格式。通过调用ROS中已经开源的语音交互功能包可以很好的实现盲人与导盲机器人之间的语音交互。解决了目前大多数导盲机器人在人机交互上的缺陷；

[0128] 5、本设计提出了一种基于强化学习的最优输出反馈控制器。在导盲机器人状态未知的情况下，采用基于策略迭代的在线IRL算法求解HJB方程，得到输出最优的反馈控制器，解决了导盲机器人在运行过程易受外界干扰的问题，使导盲机器人能够稳定的工作。

[0129] 6、本发明中设计了基于输出反馈的神经网络状态观测器来观测跟随者的状态。可以使系统在不稳定的情况下也能实时观测系统的状态，性能十分稳定。

[0130] 本发明的其他优点、目标和特征在某种程度上将在随后的说明书中进行阐述，并且在某种程度上，基于对下文的考察研究对本领域技术人员而言将是显而易见的，或者可以从本发明的实践中得到教导。本发明的目标和其他优点可以通过下面的说明书来实现和获得。

## 附图说明

[0131] 为了使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明作优选的详细描述，其中：

[0132] 图1为导盲机器人硬件平台；

[0133] 图2为激光雷达、深度相机与上位机通信；

[0134] 图3为导盲机器人视觉感知模块；

[0135] 图4为语音识别模块与master之间的通信架构；

[0136] 图5为语音播报功能实现；

[0137] 图6为手势杆；

[0138] 图7为平台原理图；

[0139] 图8为本发明计算机运行流程图；

[0140] 图9为语音信息采集流程图；

[0141] 图10为本发明流程图。

## 具体实施方式

[0142] 以下通过特定的具体实例说明本发明的实施方式，本领域技术人员可由本说明书

所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用，本说明书中的各项细节也可以基于不同观点与应用，在没有背离本发明的精神下进行各种修饰或改变。需要说明的是，以下实施例中所提供的图示仅以示意方式说明本发明的基本构想，在不冲突的情况下，以下实施例及实施例中的特征可以相互组合。

[0143] 其中，附图仅用于示例性说明，表示的仅是示意图，而非实物图，不能理解为对本发明的限制；为了更好地说明本发明的实施例，附图某些部件会有省略、放大或缩小，并不代表实际产品的尺寸；对本领域技术人员来说，附图中某些公知结构及其说明可能省略是可以理解的。

[0144] 本发明实施例的附图中相同或相似的标号对应相同或相似的部件；在本发明的描述中，需要理解的是，若有术语“上”、“下”、“左”、“右”、“前”、“后”等指示的方位或位置关系为基于附图所示的方位或位置关系，仅是为了便于描述本发明和简化描述，而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作，因此附图中描述位置关系的用语仅用于示例性说明，不能理解为对本发明的限制，对于本领域的普通技术人员而言，可以根据具体情况理解上述术语的具体含义。

[0145] 请参阅图1～图10，为一种基于强化学习的最优输出反馈控制器的导盲机器人，通过采用realsense D435i深度摄像机作为视觉传感器，能够准确且高效的获取导盲机器人在前进引导过程中的实时环境信息。为了增强导盲机器人的人机交互，在本发明中还设计了一种语音系统，在导盲机器人中加入语音模块不仅能使导盲机器人更加智能，而且还可以解决机器人无法灵活将路况信息传送给盲人的缺陷。同时，为解决导盲机器人在移动过程中所面临诸多不稳定因素的问题，设计了一种基于ADP方法的无模型同步积分强化学习控制器，通过构建基于强化学习的导盲机器人系统的代价函数，建立所构建代价函数的HJB(Hamilton Jacobi Bellman)方程，通过基于同步强化学习的方法求解HJB方程，最后通过迭代的方法得到最优解，实现导盲机器人系统的最优控制。并设计了一整套适用于盲人引导环境的导盲机器人软硬件系统。

[0146] 本发明设计的导盲机器人采用分层设计主要分为底层(硬件层)、感知层、策略层。

[0147] 导盲机器人是基于ROS机器人操作系统、采用4个万向轮的轮式机器人。采用分层控制，其中DSP作为底层的控制器，主要采集陀螺仪和里程计信息，并且控制伺服电机的运动。上层的PC主要用于感知层和策略层的信息采集与计算。

[0148] 第一部分导盲机器人的底层设计

[0149] 底层主要是完成机器人本体的硬件平台搭建，如图1所示。

[0150] ①基于万向轮的伺服底盘系统实现

[0151] 为实现机器人的灵活运动，本方案采用4个伺服电机配套4个万向轮作小车的移动执行机构，万向轮的布局方式采用对角线式。4个伺服电机采用24V供电，通过DSP编码，将上层发布的轮速信息处理后执行。

[0152] ②HOKUYO激光雷达

[0153] 可用于高速运动机器人避障和位置识别；高精度、高分辨率、宽视场设计给自主导航机器人提供了良好的环境识别能力；紧凑型设计节约了安装空间，低重量、低功耗。在本发明中，采用HOKUYO激光雷达可以十分灵敏的检测到前方障碍物，将障碍物的大小及距离

信息上传至上位机处理，通过与视觉信息融合处理后可以在精度非常高的情况下实现物体识别及避障处理。

[0154] ③配供电系统实现

[0155] 采用24V 10AH的锂电池作为机器人的底层供电电源。其中，伺服电机驱动器为24V供电，DSP为5V供电。因此，需要稳压模块调节电压，使其输出一个5V电压。

[0156] 第二部分导盲机器人感知层方案设计

[0157] 导盲系统的感知层主要由视觉识别和语音识别两部分组成。

[0158] 一、视觉感知部分

[0159] (1)视觉部分需实现的功能

[0160] ①基于ROS系统和realsense D435i深度摄像机实现导盲机器人的目标图像信息采集

[0161] ROS(机器人操作系统)是当今十分流行的一种机器人软件编写架构，本设计中，在ROS系统搭建导盲机器人的视觉感知模块，可以十分方便的处理信息传递不及时以及信息处理帧率慢的缺陷。搭建基于ROS的目标识别系统框架，通过ROS系统建立分析系统和realsense D435i深度摄像机节点的连接，读入图像或视频流以及深度信息，完成采集数据的格式转换的等数据预处理工作。进行实验设计，利用建立的采集系统采集数据，构造训练数据集。

[0162] ②基于YOLOV3深度学习和realsense D435i深度摄像机的导盲机器人目标识别算法研究与实现

[0163] 探索基于YOLOV3深度学习和深度摄像机数据的目标辨识方法，对深度摄像机数据集进行标注，然后对采用的YOLOV3深度学习网络进行训练，将训练完成的参数输出，使用测试集对模型进行目标检测效果测试。

[0164] (2)视觉层具体设计方案

[0165] 1)基于realsense D435i深度摄像机的导盲机器人视觉识别系统的实现

[0166] 由于机器人前方需要识别红路灯、盲道、斑马线等物体，根据机器人与识别目标的位置，尽量减小其他因素的干扰，一个摄像头需向下倾斜30安装，另外一个摄像头需要向上倾斜20°安装。机器人后方需要安装一个摄像头实现主人面部识别与跟踪。深度相机通过USB与上位机连接，激光雷达通过以太网与上位机通信，如图2所示。

[0167] 2)基于ROS系统和realsense D435i深度摄像机实现导盲机器人的目标图像信息采集

[0168] ROS具有交叉编译、开源、分布式管理等优点，逐步成为机器人研发领域的通用平台，ROS的出现加强了机器人代码的复用率和模块化，降低了智能机器人开发中不必要的重复劳动。通过ROS系统中的Master发布命令，运行realsense D435i深度摄像机启动节点，读入图像或视频流，通过OPENCV和ROS的接口完成图像格式转换，将采集到的图像储存，使用Python构建深度学习数据集，先安装REQUESTS包，创建Python脚本下载图像，配置环境，然后修剪深度学习图像数据集。ROS系统实现导盲机器人视觉感知模块如图3所示。

[0169] 3)基于YOLOV3深度学习和realsense D435i深度摄像机的导盲机器人目标识别算法实现

[0170] A.准备数据

[0171] 使用yolo_mark对图片进行标注

[0172] B.修改配置文件

[0173] 修改训练数据、验证数据、物体名称文件路径，修改神经网络的详细构建参数

[0174] C.训练及输出

[0175] 训练网络，输出参数进行保存。

[0176] D.测试

[0177] 验证模型效果

[0178] 导盲机器人系统搭建在ROS机器人操作系统下，视觉图像数据采集储存在ROS系统中，需要在Ubuntu系统下构建YOLOV3深度学习网络，首先需要安装对应版本的CUDA和CUDNN，配置编译环境。

[0179] 导盲机器人通过前方摄像头与YOLOV3深度学习网络识别出前方物体信息，将物体的具体识别信息以及位置坐标回传给上位机，通过后方摄像头来识别主人信息，再通过上层决策信息来决定机器人运动。

[0180] 盲道识别：

[0181] 盲道的颜色通常很鲜艳，因此可以通过盲道的颜色特征来进行检测。本文采用基于颜色区域的图像分割的方法，能够筛选出盲道区域，并对盲道区域进行边缘提取实现盲道的识别。首先将图像由RGB转为HSI色彩空间，相对RGB的彩色空间而言，HSI色彩空间同人对色彩的感知一致，符合人的视觉感知，不易受到周围环境的影响。RGB色彩空间到HSI色彩空间的转换关系如下：

[0182] $$H=\begin{cases}\cos^{-1}\left\{\dfrac{(R-G)+(R-B)}{\sqrt{2(R-B)+(R-B)(G-B)}}\right\}, (B\le G)\text{且}(R\ne B\text{或}R\ne G)\\ 2\pi-H\end{cases}$$

[0183] $$S=1-\frac{3\min(R,G,B)}{R+G+B}$$

[0184] $$I=\frac{1}{3}(R+G+B)$$

[0185] 通过上述转换后可以得到在HSI色彩空间上的图像。

[0186] 二、语音识别及人机交互部分

[0187] (1)基于ROS的语音交互系统搭建

[0188] 机器人操作系统ROS使用简单，在确定了人机语音交互系统应具有的功能模块之后，使用ROS提供的话题、服务方式实现系统中相关模块之间的通信，同时定义通信时的信息格式。通过调用ROS中已经开源的语音交互功能包可以很好的实现盲人与导盲机器人之间的语音交互。

[0189] 语音云服务平台是位于云端的服务器，它为系统提供一系列支持，包括语音识别、语义理解、语音合成等。除去语音云服务平台系统主要分为三层：其中最底层为Linux内核，为系统运行环境；其次是中间层，该层主要是第三方库以及ROS系统。基于ROS的人机语音交互系统从采集语音一直到机器人做出响应，主要划分为如下几个功能节点：语音识别节点，语义分析节点，实时性信息获取节点，语音合成节点，音频播放节点。在ROS中其与master之间的通信架构如图4所示。

[0190] 在ROS中实现的语音交互主要功能包括:语音信息采集、语音识别节点、语音合成节点、语义分析节点、实时性信息获取节点、机器人控制功能节点。

[0191] • 语音信息采集:通过机器人外置麦克风采集语音信息,将采集的语音信息存储为音频文件。

[0192] • 语音识别节点:语音识别节点负责将采集的语音信息识别为文字信息。

[0193] • 语音合成节点:语音合成节点负责将请求信息合成为音频。

[0194] • 语义分析节点:语义分析节点具有对从语音识别节点接收到的请求信息进行理解,以判决机器人应该执行何种操作的功能。

[0195] • 实时性信息获取节点:通过实时性信息获取节点能得到实时变化的信息内容。

[0196] • 机器人控制功能节点:机器人控制功能节点包括控制机器人行走、避障、到达指定位置等节点。

[0197] (2)语音人机交互具体需实现的功能

[0198] 1、盲人以语音的形式唤醒导盲机器人:如“小明,请一键启动”;

[0199] 2、盲人以语音形式控制导盲机器人选择模式(自由散步、好友散步):如“小明,请带我到张三家”;

[0200] 3、导盲机器人遇到障碍时,播报“前方有障碍物,请注意通行”;

[0201] 4、导盲机器人在识别到盲道时,播报“前方盲道,请沿盲道行走”;

[0202] 5、导盲机器人识别出红绿灯时,播报“前方红绿灯,请等待”;

[0203] 6、导盲机器人识别绿灯剩余时间时,播报“绿灯时间不足,请等待下次通行”;

[0204] 图5为语音播报功能实现。

[0205] 第三部分 导盲机器人决策层方案设计

[0206] 1、策略层主要实现导盲机器人在路径规划中的相关动作决策(接受视觉传来的障碍物信息后改变电机转向绕开障碍物、在红灯时控制电机停止、在红灯转绿灯时启动电机、盲人通过语音唤醒时启动导盲机器人)以及相应控制算法的实现。

[0207] 为确保导盲机器人因故障无法自主控制时,还设计了一种手势杆操作器,如图6所示:

[0208] 手势杆操作器输入功能:在自由散步模式下,主人可以通过手势杆操作器控制机器人同时为方便盲人使用手势杆操作器,手势杆操作器的按键应设计得更适合盲人使用。在手势杆操作器中箭头方向表示机器人运动的方向,中间圆形键表示为暂停键。当运行在自由散步模式下时,主人只需要通过手势杆操作器方向实现对机器人的运动控制。加入手势杆设计后可以很好的解决自主控制故障的问题,可以让视觉障碍者更加灵活的控制。

[0209] 2、PC端和机器人的通讯:本发明需要将底层信息及里程计和陀螺仪信息传到机器人,通过RS232通讯线使机器人PC可以传递位置信息给主控DSP320F2812。为实现自主学习,所以对主控芯片有一定的要求,经过分析选择了DSP320F2815作为主控芯片,DSP320F281含有多种外接接口,可以很好地完成输出PWM波和脉冲信号的功能,同时还可以通过RS232通讯线接受PC端传递来的信息,而且由于它的时钟频率达到150MHZ,其处理的速度较快。

[0210] 第四部分 机器人控制算法

[0211] 因为本发明中的设计一种新型导盲机器人在引导行驶过程中会发生许多未知的变故(一般在实际装置中,都普遍存在振幅约束,即约束输入或执行器饱和),所以在控制器

设计过程中必须考虑约束控制输入，导盲机器人在实际情况中系统状态的不可测，所以在本发明中设计了一种基于强化学习的最优输出反馈控制器。

[0212] 机器人的动态模型为：

[0213] $$\begin{cases}\dot{x}(t)=f(x(t))+g(x(t))u(t),\\ y(t)=Cx(t).\end{cases} \quad (1)$$

[0214] 其中$x(t)\in\Re^{n}$是不可测系统状态向量，$u(t)\in\Re^{n*m}$是系统的控制输入，y(t)是系统唯一输出。

[0215] 现假定f(0)＝0，f(x)是未知的且满足||f(x)||≤$b_f$||x||，$b_f$是一个常量；g(x)是已知且有界的，0＜||g(x)||≤$b_g$，$b_g$是一个常量。

[0216] 定义导盲机器人系统的代价函数：

[0217] $$J(y(t),\ u(t))=\int_{t}^{\infty}[Q(y(\tau))+U(u(\tau))]d\tau \quad (2)$$

[0218] 其中，$\tau=[\tau_1,\ \tau_2,\cdots,\ \tau_m]^T\in\Re^m$，$Q(y(\tau))=y^T(\tau)Qy(\tau)$是正定且连续可微的。U(u(τ))是被积函数。所以考虑到系统的输入受限，可以定义以下一个非二次性能函数：

[0219] $$U(u)=2\int_{0}^{u}(\lambda\beta^{-1}(\vartheta/\lambda))^{T}Rd\vartheta \quad (3)$$

[0220] 其中，$\vartheta\in\Re^m$，β(•)＝tanh(•)，λ是饱和有界的；R＝diag($r_1$，$r_2$，…，$r_m$)＞0是对角型。

[0221] 考虑到系统是基于完全无模型的设计，所以系统的状态是不可测的，所以在本发明中了一个基于输出反馈的神经网络观测器。通过设计观测器实时观测导盲机器人运行时的状态，将实时状态传给设计的控制器进行处理后使系统稳定。

[0222] 因为系统状态x(t)不可测，在这里构造基于输出反馈的状态观测器，其状态观测器的动态模型如下：

[0223] $$\begin{cases}\dot{\tilde{x}}(t)=f(\tilde{x}(t))+g(\tilde{x}(t))u(t),\\ \tilde{y}(t)=C\tilde{x}(t).\end{cases} \quad (4)$$

[0224] 由于系统动态模型$f_i(x_i)$的内部函数未知，在此我们用神经网络来估计$f_i(x_i)$：

[0225] $$f_i(x_i)=A_{oi}x_i+\omega_{oi}^{T}\varphi_{oi}(x_i)+\varepsilon_{oi}(x_i) \quad (5)$$

[0226] 其中$A_{oi}\in\mathbb{R}^{n\times n}$是赫尔维茨矩阵；$\omega_{oi}\in\mathbb{R}^{l_0\times n}$为神经网络权重，且||$\omega_{oi}$||≤||$\omega_{omi}$||；$\varphi_{oi}(x_i)\in\mathbb{R}^{l_0}$为神经网络激活函数，且$\varphi_{oi}(x_i)\le\varphi_{omi}$；

[0227] 不妨将系统的动态模型写成：

[0228] $$\begin{cases}\dot{x}(t)=A_{0i}(x(t))+n(x(t))+g(x(t))u(t),\\ y(t)=Cx(t).\end{cases} \quad (6)$$

[0229] 其中，$n(x(t))=f(x)-A_{0i}x(t)$，$A_{oi}\in\Re^{n*n}$是赫尔维茨矩阵；

[0230] 则观测器的动态模型为：

[0231] $$\begin{cases}\dot{\tilde{x}}(t)=F\left(y(t),\tilde{x}(t)\right)+g\left(\tilde{x}(t)\right)u(t),\\ \tilde{y}(t)=C\tilde{x}(t).\end{cases}\qquad(7)$$

[0232] 其中，$F\left(y(t),\tilde{x}(t)\right)=A_{0i}\tilde{x}(t)+\tilde{\omega}_0^{T}\varphi_0\left(\tilde{x}(t)\right)+K\left(y(t)-\tilde{y}(t)\right)$，$\tilde{x}(t)$和$\tilde{y}(t)$是所设计观测器的状态。K是观测器的增益，$(A_{0i}-KC)$是赫尔维茨矩阵。所以系统满足：

[0233] $(A_{0i}-KC)^{T}P+P(A_{0i}-KC)=-qI$ (8)

[0234] 其中，q是正常量，I是一个单位矩阵，P是一个对称正定矩阵。

[0235] 定义观测器误差为$\hat{x}_i=x_i-\tilde{x}_i$，则：

[0236] $$\dot{\hat{x}}(t)=(A_{0i}-KC)\hat{x}+\hat{\omega}_0^{T}\varphi_0\left(\tilde{x}(t)\right)+\xi(x(t))\qquad(9)$$

[0237] 其中，$\hat{\omega}_0=\omega_0-\tilde{\omega}_0$是构造的神经网络观测器的估计误差。

[0238] 选择下面李雅普诺夫函数：

[0239] $$L_{oi}(t)=\frac{1}{2}\tilde{x}_i^{T}P_i\tilde{x}_i+\frac{1}{2}tr\left\{\tilde{\omega}_{oi}^{T}\eta_i^{-1}\tilde{\omega}_{oi}\right\}\qquad(10)$$

[0240] 将上面李雅普诺夫函数进行求导：

[0241] $$\dot{L}_{oi}(t)=\frac{1}{2}\dot{\tilde{x}}_i^{T}P_i\tilde{x}_i+\frac{1}{2}\tilde{x}_i^{T}P_i\dot{\tilde{x}}_i+tr\left\{\tilde{\omega}_{oi}^{T}\eta_i^{-1}\dot{\tilde{\omega}}_{oi}\right\}\qquad(11)$$

[0242] 根据观测器测得的误差$\hat{\omega}_0=\omega_0-\tilde{\omega}_0$，可知：

[0243] $$\dot{\hat{\omega}}=\beta_i\varphi_0\left(\tilde{x}_i\right)\hat{y}_i C_i(A_0-KC)^{-1}+\eta_i\|\hat{y}\|\left(\omega_0-\hat{\omega}_0\right)\qquad(12)$$

[0244] 将(8)(9)(12)带入(11)得：

[0245] $$\begin{aligned}\dot{L}_{oi}(t)=&\frac{1}{2}\left[(A_{oi}-K_iC_i)\tilde{x}_i+\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right]^{T}P_i\tilde{x}_i\\&+\frac{1}{2}\tilde{x}_i^{T}P_i\left[(A_{oi}-K_iC_i)\tilde{x}_i+\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right]\\&+tr\left\{\tilde{\omega}_{oi}^{T}\eta_i^{-1}\left[\beta_i\varphi_{oi}(\hat{x}_i)\tilde{y}_i^{T}C_i(A_{oi}-K_iC_i)^{-1}+\eta_i\|\tilde{y}_i\|(\omega_{oi}-\tilde{\omega}_{oi})\right]\right\}\qquad(13)\\=&-\frac{1}{2}q_i\tilde{x}_i^{T}\tilde{x}_i+\tilde{x}_i^{T}P_i\left(\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right)\\&+tr\left\{\eta_i^{-1}\beta_i\tilde{\omega}_{oi}^{T}\varphi_{oi}(\hat{x}_i)\tilde{y}_i^{T}C_i(A_{oi}-K_iC_i)^{-1}\right\}+tr\left\{\tilde{\omega}_{oi}^{T}\|\tilde{y}_i\|(\omega_{oi}-\tilde{\omega}_{oi})\right\}\end{aligned}$$

[0246] 由于$tr(AB^{T})=tr(BA^{T})=BA^{T}$，所以(13)可改写成：

[0247]
$$\dot{L}_{oi}(t)=-\frac{1}{2}q_i\tilde{x}_i^T\tilde{x}_i+\tilde{x}_i^TP_i\left(\tilde{\omega}_{oi}^T\varphi_{oi}(\hat{x}_i)+\xi_{oi}(x_i)\right)$$
$$+\eta_i^{-1}\beta_i\tilde{y}_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\tilde{\omega}_{oi}^T\varphi_{oi}(\hat{x}_i)+\|\tilde{y}_i\|(\omega_{oi}-\tilde{\omega}_{oi})\tilde{\omega}_{oi}^T \quad (14)$$

[0248] 因为$\omega_{oi}$、$\varphi_{oi}(\hat{x}_i)$、$\xi_{oi}(x_i)$有界，所以式(4.15)可整理为：

[0249]
$$\dot{L}_{oi}(t)\le-\frac{1}{2}q_i\|\tilde{x}_i\|^2+\|\tilde{x}_i\|\cdot\|P_i\|\left(\|\tilde{\omega}_{oi}\|\varphi_{omi}+\xi_{omi}\right)$$
$$+\eta_i^{-1}\beta_i\|\tilde{x}_i\|\cdot\left\|C_i^TC_i\left(A_{oi}-K_iC_i\right)^{-1}\right\|\cdot\|\tilde{\omega}_{oi}\|\varphi_{omi} \quad (15)$$
$$+\|C_i\|\cdot\|\tilde{x}_i\|\left(\|\omega_{oi}\|\cdot\|\tilde{\omega}_{oi}\|-\|\tilde{\omega}_{oi}\|^2\right)$$

[0250] 所以：$\dot{L}_{0i}(t)\le-\frac{1}{2}q_i\|\hat{x}_i\|^2$

[0251]
$$+\|\hat{x}_i\|^2+\|\hat{x}_i\|\cdot\|C_i\|\left(\frac{\|P_i\|}{\|C_i\|}\xi_{omi}+2\gamma\|\hat{\omega}_{oi}\|-\|\hat{\omega}_{oi}\|^2\right) \quad (16)$$

[0252] 为使$\dot{L}_{oi}(t)<0$，只需令$\frac{\|P_i\|}{\|C_i\|}\xi_{omi}+2\gamma_i\|\tilde{\omega}_{oi}\|-\|\tilde{\omega}_{oi}\|^2<0$；

[0253] 即只要满足：

[0254]
$$\|\hat{x}(t)\|>\frac{2\left(\|P_i\|\xi_{omi}+\gamma^2\|C_i\|\right)}{q} \quad (17)$$

[0255] 由于机器人的输出$y(t)=C\tilde{x}(t)$，所以代价函数也可以写成下面的形式：

[0256]
$$J\left(\tilde{x}(t),\ u(t)\right)=\int_t^{\infty}\left[\left(\tilde{x}^T(\tau)Q_c(\tau)+U(u(\tau))\right)\right]d\tau \quad (18)$$

[0257] 其中，$Q_c=C^TQC$半正定的。

[0258] 利用牛顿-莱布尼茨公式对式(18)中时间t求导得到贝尔曼方程：

[0259]
$$\dot{J}\left(\tilde{x}(t),\ u(t)\right)=-\tilde{x}^T(t)Q_c\tilde{x}(t)-U(u(t)) \quad (19)$$

[0260] 联立(3)(19)可得：

[0261]
$$\tilde{x}^T(t)Q_c\tilde{x}(t)+2\int_0^u\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^TRd\vartheta+\dot{J}\left(\tilde{x}(t),u(t)\right)=0 \quad (20)$$

[0262] 定义Hamiltonian方程为：

[0263]
$$H\left(\tilde{x},\ u,\nabla J(x)\right)=\tilde{x}^T(t)Q_c\tilde{x}(t)+2\int_0^u\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^TRd\vartheta$$
$$+(\nabla J)^T\left(F\left(y,\tilde{x}\right)\right)+g(x)u(t)=0 \quad (21)$$

[0264] 令最优代价函数为$J^*\left(\tilde{x}(t)\right)$：

[0265] $$J^*\left(\tilde{x}(t)\right)=\min_{u\in\Omega}\int_t^{\infty}\left(\tilde{x}^T(\tau)Q_c\tilde{x}(\tau)+U(u(\tau))\right)d\tau \quad (22)$$

[0266] 则根据(21)中Hamiltonian方程，可得到如下HJB(Hamilton Jacobi Bellman)方程

[0267] $$H\left(\tilde{x},\ u,\nabla J^*\right)=\tilde{x}^T(\tau)Q_c\tilde{x}(\tau)+2\int_0^{u^*}\left(\lambda\tanh^{-1}(\vartheta/\lambda)\right)^T Rd\vartheta$$

[0268] $$+\nabla J^{*T}\left(F\left(y,\tilde{x}\right)+g(x)u^*(t)\right)=0 \quad (23)$$

[0269] 当稳定性条件$\frac{\partial H\left(\tilde{x},\ u,\nabla J\right)}{\partial u(t)}=0$时，可以得到如下最优控制输入：

[0270] $$u^*(t)=\arg\min_{u\in\Omega}\left[H\left(\tilde{x},u,\nabla J^*\right)\right]=-\lambda\tanh\left(\frac{1}{2\lambda}R^{-1}g^T(x)\nabla J^*\right) \quad (24)$$

[0271] 由于HJB方程很难求解，所以在该算法中采用IRL的策略迭代来求解上述HJB方程。

[0272] 首先将(18)中的值函数写成下面贝尔曼方程的形式：

[0273] $$J\left(\tilde{x}(t)\right)=\int_t^{t+T}\left(\tilde{x}^T(\tau)Q_c\tilde{x}(\tau)\right)+U(u(\tau))d\tau+J\left(\tilde{x}(t+T)\right) \quad (25)$$

[0274] 得到下面基于策略迭代的在线IRL算法：

[0275] 算法：基于策略迭代的在线IRL算法求解HJB方程

[0276] 步骤1：(策略评估)利用下式解出$J^{(i)}(x(t))$

[0277] $$J^{(i)}(x(t))=\int_t^{t+T}\left(x^T(\tau)Q_cx(\tau)\right)+U(u(\tau))d\tau+J^{(i)}(x(t+T)) \quad (12)$$

[0278] 步骤2：(策略改进)通过下式更新控制策略：

[0279] $$u^{(i+1)}x(t)=-\lambda\tanh\left(\frac{1}{2\lambda}R^{-1}g^T(x)\nabla J^{(i)}(x)\right) \quad (13)$$

[0280] 步骤3：令$u_i^j=u_i^{j+1}$，返回步骤1，直到$J^{(i)}x(t)$收敛到最小值。

[0281] 最后说明的是，以上实施例仅用以说明本发明的技术方案而非限制，尽管参照较佳实施例对本发明进行了详细说明，本领域的普通技术人员应当理解，可以对本发明的技术方案进行修改或者等同替换，而不脱离本技术方案的宗旨和范围，其均应涵盖在本发明的权利要求范围当中。

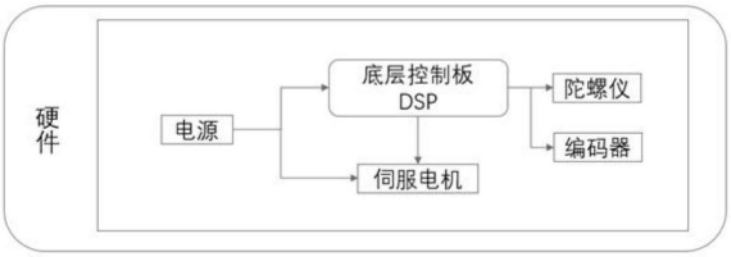
硬件
电源
底层控制板
DSP
陀螺仪
编码器
伺服电机

图1

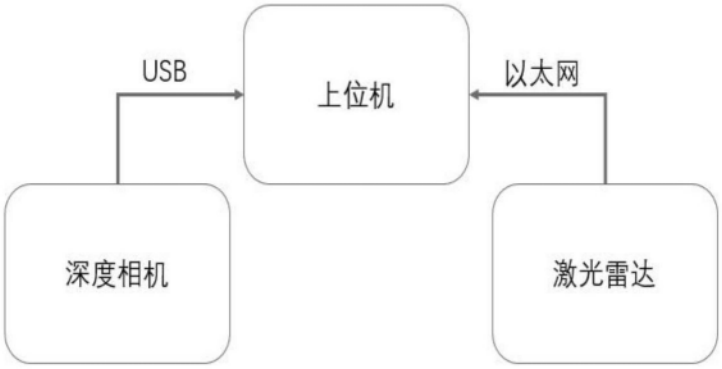
USB
上位机
以太网
深度相机
激光雷达

图2

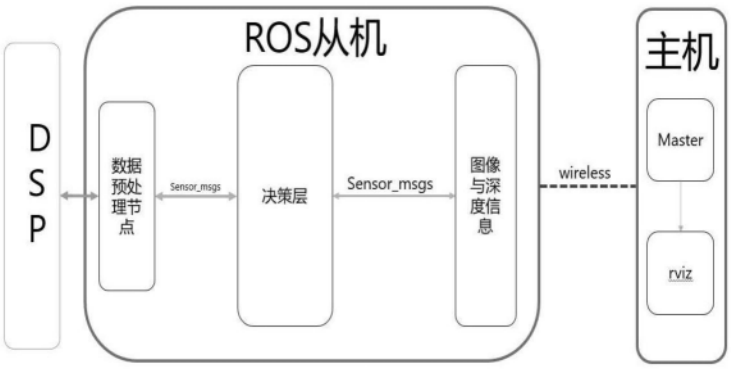
ROS从机
主机
D
S
P
数据预处理节点
Sensor_msgs
决策层
Sensor_msgs
图像与深度信息
wireless
Master
rviz

图3

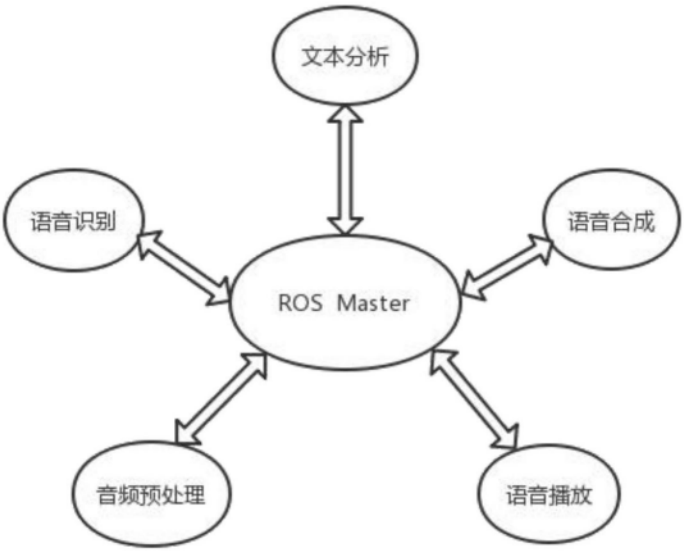
文本分析
语音识别
语音合成
ROS Master
音频预处理
语音播放

图4

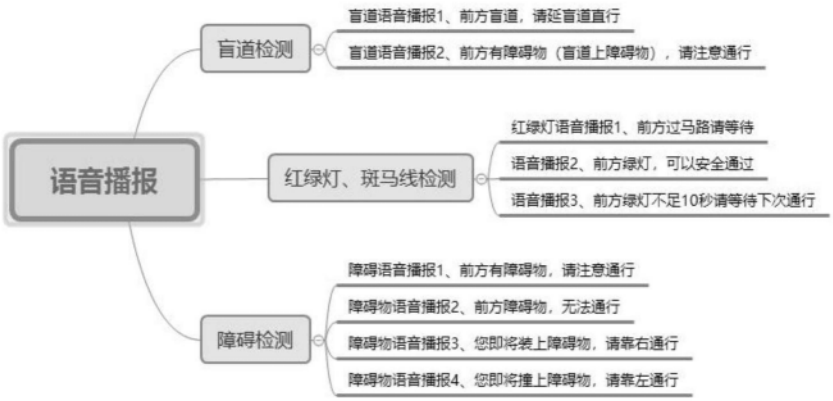
语音播报
盲道检测
盲道语音播报1、前方盲道，请延盲道直行
盲道语音播报2、前方有障碍物（盲道上障碍物），请注意通行
红绿灯、斑马线检测
红绿灯语音播报1、前方过马路请等待
语音播报2、前方绿灯，可以安全通过
语音播报3、前方绿灯不足10秒请等待下次通行
障碍检测
障碍语音播报1、前方有障碍物，请注意通行
障碍物语音播报2、前方障碍物，无法通行
障碍物语音播报3、您即将装上障碍物，请靠右通行
障碍物语音播报4、您即将撞上障碍物，请靠左通行

图5

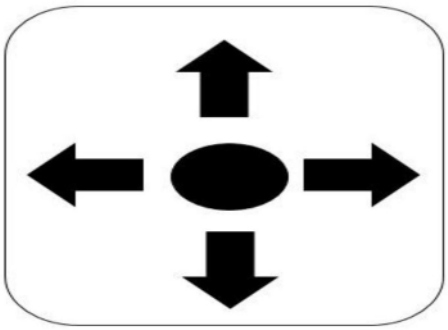

图6

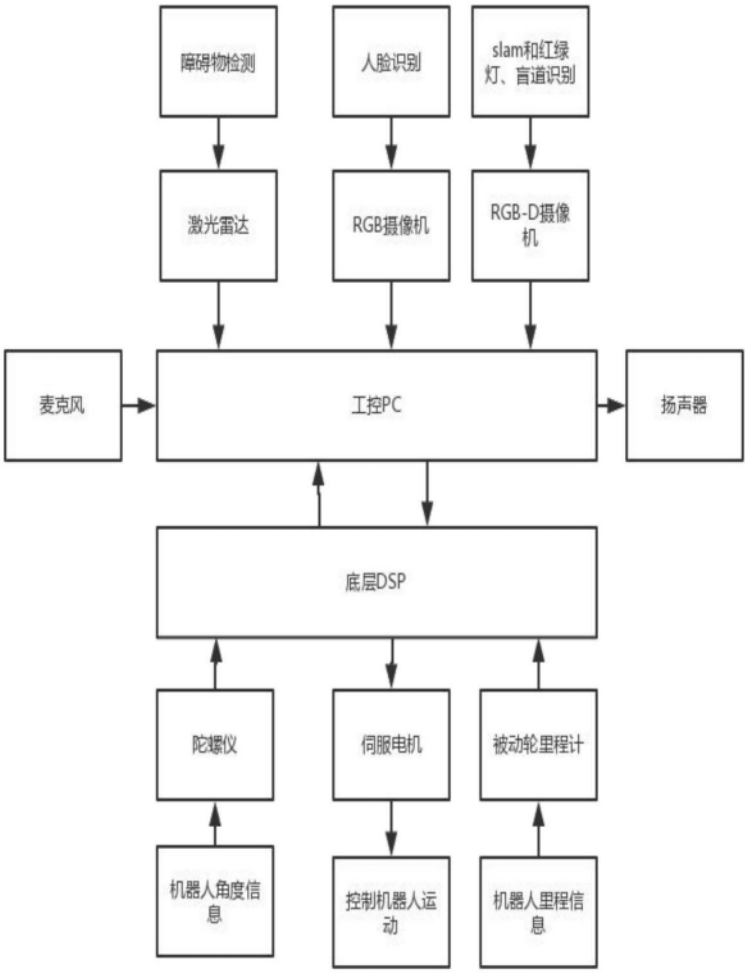
障碍物检测
人脸识别
slam和红绿灯、盲道识别
激光雷达
RGB摄像机
RGB-D摄像机
麦克风
工控PC
扬声器
底层DSP
陀螺仪
伺服电机
被动轮里程计
机器人角度信息
控制机器人运动
机器人里程信息

图7

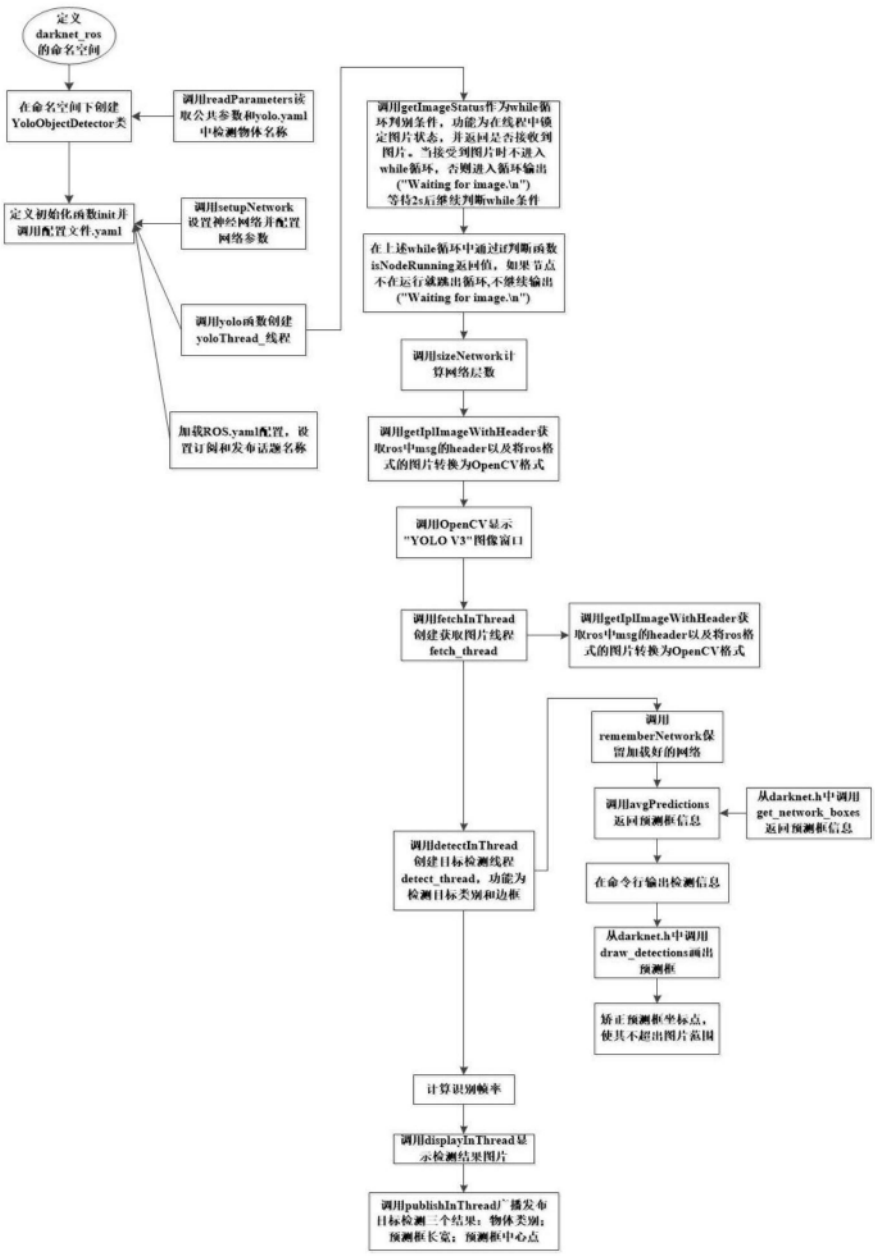
定义darknet_ros的命名空间
在命名空间下创建YoloObjectDetector类
调用readParameters读取公共参数和yolo.yaml中检测物体名称
定义初始化函数init并调用配置文件.yaml
调用setupNetwork设置神经网络并配置网络参数
调用yolo函数创建yoloThread_线程
加载ROS.yaml配置，设置订阅和发布话题名称
调用getImageStatus作为while循环判别条件，功能为在线程中锁定图片状态，并返回是否接收到图片。当接受到图片时不进入while循环，否则进入循环输出("Waiting for image.\n")等待2s后继续判断while条件
在上述while循环中通过if判断函数isNodeRunning返回值，如果节点不在运行就跳出循环,不继续输出("Waiting for image.\n")
调用sizeNetwork计算网络层数
调用getIplImageWithHeader获取ros中msg的header以及将ros格式的图片转换为OpenCV格式
调用OpenCV显示"YOLO V3"图像窗口
调用fetchInThread创建获取图片线程fetch_thread
调用getIplImageWithHeader获取ros中msg的header以及将ros格式的图片转换为OpenCV格式
调用detectInThread创建目标检测线程detect_thread，功能为检测目标类别和边框
调用rememberNetwork保留加载好的网络
调用avgPredictions返回预测框信息
从darknet.h中调用get_network_boxes返回预测框信息
在命令行输出检测信息
从darknet.h中调用draw_detections画出预测框
矫正预测框坐标点，使其不超出图片范围
计算识别帧率
调用displayInThread显示检测结果图片
调用publishInThread广播发布目标检测三个结果：物体类别；预测框长宽；预测框中心点

图8

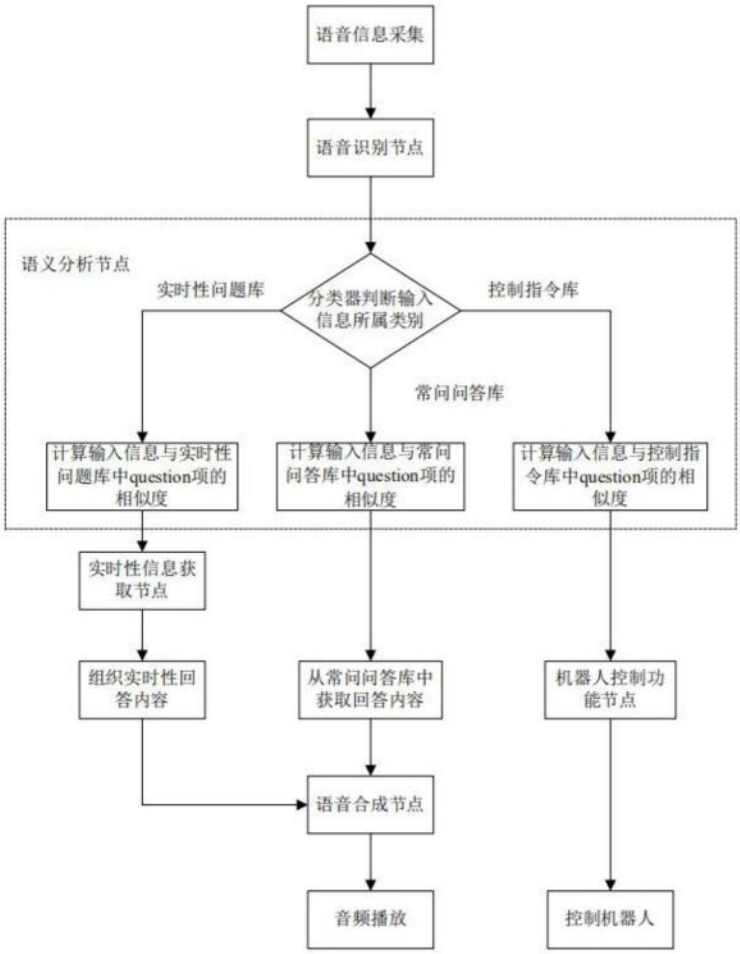
语音信息采集
语音识别节点
语义分析节点
实时性问题库
分类器判断输入信息所属类别
控制指令库
常问问答库
计算输入信息与实时性问题库中question项的相似度
计算输入信息与常问问答库中question项的相似度
计算输入信息与控制指令库中question项的相似度
实时性信息获取节点
组织实时性回答内容
从常问问答库中获取回答内容
机器人控制功能节点
语音合成节点
音频播放
控制机器人

图9

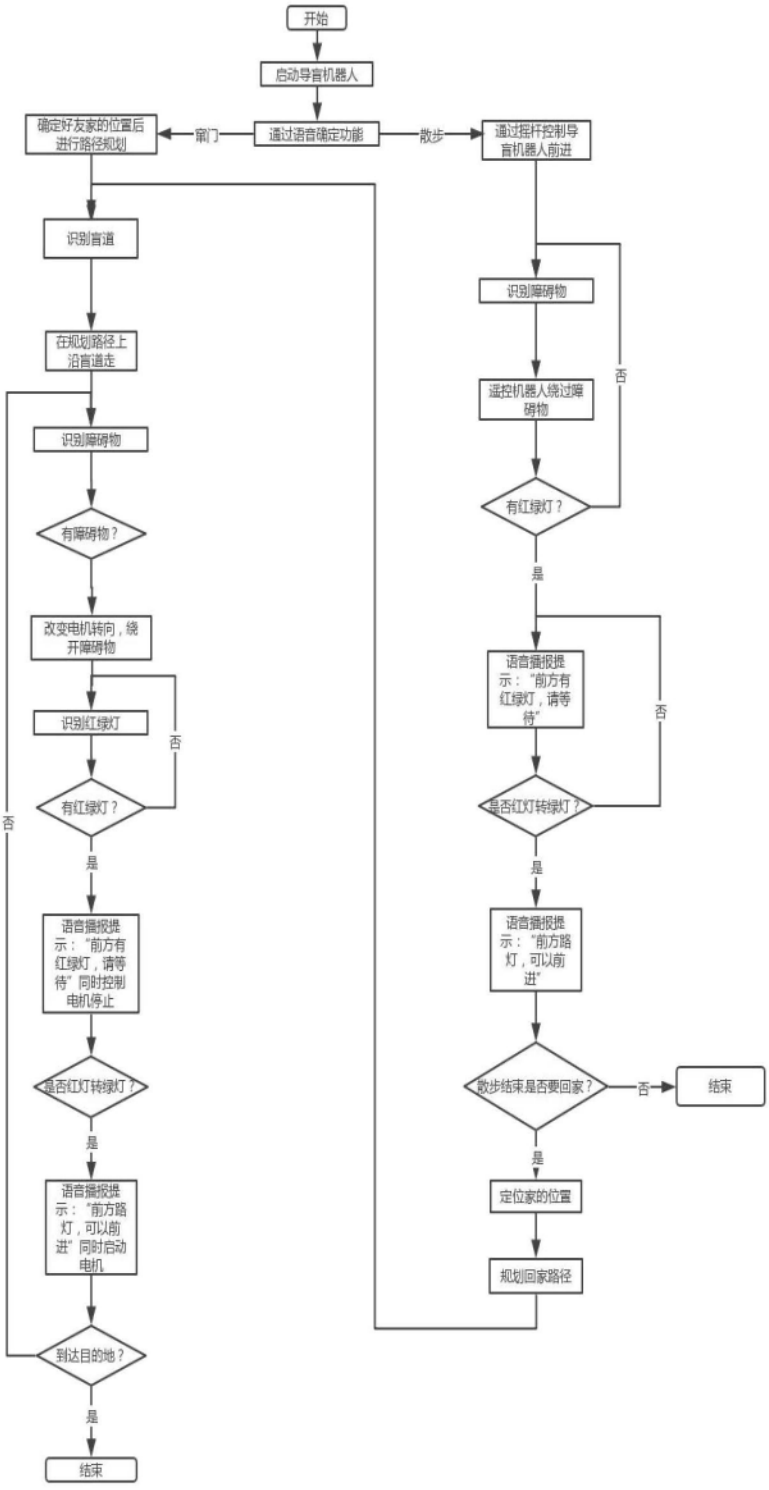
开始
启动导盲机器人
通过语音确定功能
窗门
散步
确定好友家的位置后进行路径规划
通过握杆控制导盲机器人前进
识别盲道
在规划路径上沿盲道走
识别障碍物
有障碍物？
改变电机转向，绕开障碍物
识别红绿灯
有红绿灯？
否
是
否
语音播报提示："前方有红绿灯，请等待"同时控制电机停止
是否红灯转绿灯？
是
语音播报提示："前方路灯，可以前进"同时启动电机
到达目的地？
是
否
结束
识别障碍物
遥控机器人绕过障碍物
有红绿灯？
否
是
语音播报提示："前方有红绿灯，请等待"
是否红灯转绿灯？
否
是
语音播报提示："前方路灯，可以前进"
散步结束是否要回家？
否
结束
是
定位家的位置
规划回家路径

图10