

DOI:10.19651/j.cnki.emt.1802379

# 基于强化学习的双足机器人的实时避障位置控制

林荣霞

(广东工业大学华立学院 广州 511325)

**摘要:** 为了提高双足机器人的实时控制能力,提出一种基于强化学习的双足机器人的实时避障位置控制方法。以机器人的双足步行稳定性为控制目标函数,构建双足机器人的实时路径动力学模型,以机器人质心运动的加速度和惯性力矩为被控对象,采用等效碰撞子模型进行双足机器人的实时避障路径规划,采用碰撞子模型和摆动子模型相结合的方法,进行双足机器人行走路径的纠偏参量反馈调节,采用模糊强化学习跟踪方法,进行双足机器人的误差增益控制,实现双足机器人的实时避障位置控制。仿真结果表明,采用该方法进行双足机器人控制的实时避障性能较好,纠偏能力较强,提高了机器人的自适应控制能力。

**关键词:** 强化学习;双足机器人;实时避障;位置控制

**中图分类号:** TN99; TP242 **文献标识码:** A **国家标准学科分类代码:** 520.60

## Real-time obstacle avoidance position control for biped robot based on reinforcement learning

Lin Rongxia

(Huali College Guangdong University of Technology, Guangzhou 511325, China)

**Abstract:** In order to improve the real-time control ability of biped robot, a real-time obstacle avoidance position control method for biped robot based on reinforcement learning is proposed. Taking the stability of biped walking as the control objective function, the real-time path dynamics model of biped robot is constructed. The acceleration and inertia moment of the robot's centroid motion are taken as the controlled object. The effective collision sub-model is used to plan the real-time obstacle avoidance path of biped robot, and the collision sub-model and swing sub-model are combined to adjust the error correction parameters of biped robot. The fuzzy reinforcement learning tracking method is used to control the error gain of biped robot, and the real time obstacle avoidance position control of biped robot is realized. The simulation results show that the proposed method can avoid obstacles in real time and improve the adaptive control ability of biped robot.

**Keywords:** reinforcement learning; biped robot; real-time obstacle avoidance; position control

## 0 引言

人工智能技术的快速发展促进了机器人控制技术的不断成熟,双足机器人作为一种典型的仿人机器人,在模仿人体动作和作业中发挥重要作用。双足机器人在路径规划控制中,容易受到环境和机器人自身扰动因素的影响,导致机器人的避障性不好,为了提高对双足机器人的实时避障能力,需要进行双足机器人的实时避障位置控制,提高机器人的轨迹路径规划能力<sup>[1]</sup>。

双足机器人位置纠偏和避障控制过程中,容易受到双足机器人助理转向部件和关节活动性部件的扰动性因素的干扰,需要进行双足机器人的实时避障控制,传统方法中,

对双足机器人的实时避障方法主要采用机器学习控制方法,结合分布式控制和反演积分控制方法,实现对双足机器人的自适应控制,提高控制稳定性<sup>[2-3]</sup>。文献[4]中提出一种基于时滞反馈的双足机器人位置伺服鲁棒性控制方法,采用传感器基阵进行双足机器人的位置参数采集,实现机器人的避障路径纠偏控制,但该方法进行机器人的避障控制的自适应性不好。文献[5]中提出一种基于时滞耦合跟踪的行双足机器人位置误差调节和避障位置控制方法,该方法存在计算开销较大和抗干扰性不好的问题<sup>[6]</sup>。

针对上述问题,本文提出一种基于强化学习的双足机器人的实时避障位置控制方法。首先以机器人的双足步行稳定性为控制目标函数,以机器人质心运动的加速度和惯性力

矩为被控对象,然后采用等效碰撞子模型进行双足机器人的实时避障路径规划,进行双足机器人行走路径的纠偏参量反馈调节和优化控制,最后进行仿真实验分析,展示了本文方法在提高机器人的实时避障位置控制能力方面的优越性能。

## 1 双足机器人实时避障控制对象描述及约束参量分析

### 1.1 双足机器人的位置纠偏和避障被控对象描述

为了实现双足机器人位置纠偏和避障系统的优化设计,采用敏感元件测量方法进行双足机器人的位置纠偏和避障参量采集,采集双足机器人的位姿信息、环境信息等,采用惯性参量调节和姿态参量测量方法进行双足机器人位置纠偏位置参量的信息融合<sup>[7]</sup>,根据测量结果进行控制约束参量分析,建立双足机器人的控制对象模型,双足机器人的外形关于  $x_1Oy_1$  平面对称,假设  $\tau_w$  为双足机器人进行位置纠偏和避障控制的迟滞因子,  $\tau_{ca}$  为控制器传输到执行器的时延,机器人的避障控制器和执行器之间的控制联动误差具有稳态性。人控制对象模型,考虑状态反馈控制律  $u(k) = Kx(k)$ ,当双足机器人的输出时延  $\tau_k$  小于一个采样周期时,根据双足机器人绕垂直方向轴(Y轴)的位置纠偏和避障参量修正模型,采用小扰动抑制方法,得到的双足机器人被控对象的参量相关矩阵  $\mathbf{R}$  定义为:

$$\mathbf{R} = X(n)X^T(n) = \begin{bmatrix} x_1(n) & x_1(n) & x_1(n) & x_2(n) \\ x_2(n) & x_1(n) & x_2(n) & x_2(n) \\ \vdots & \vdots & \vdots & \\ x_m(n) & x_1(n) & x_m(n) & x_2(n) \end{bmatrix} \quad (1)$$

构建双足机器人控制的连续有界解算子,进行机器人实时避障控制域的敏感性表征,当双足机器人的位置纠偏矩阵  $\mathbf{Q}$  正定,对位置参量  $X_p\left(\frac{m}{2\Delta x}\right)$  进行 2 倍抽取,利用 FRFT 的频移性质<sup>[8]</sup>,得到双足机器人位置纠偏和避障的反馈信息矩阵:

$$\begin{bmatrix} G_{11}(s) & G_{12}(s) \\ G_{21}(s) & G_{22}(s) \end{bmatrix} = \begin{bmatrix} \frac{1.7e^{-30s}}{7s+1} & \frac{0.59e^{-27s}}{8s+1} \\ -\frac{0.6e^{-25s}}{10s+1} & \frac{1.5e^{-28s}}{9s+1} \end{bmatrix} \quad (2)$$

双足机器人在受到小扰动力矩作用下,执行控制指令的传输时延  $\tau_k$  是不确定的,采用二自由度 IMC-PID 设计,存在扰动矩阵  $\mathbf{Q}$  的逆矩阵  $\mathbf{Q}^{-1}$  满足正定性,当  $\alpha = \arctg(k_0) = \frac{2j+1}{2}\pi$  ( $j$  是任意整数)时,双足机器人位置实时避障控制的惯性融合函数为:

$$F_{\alpha} [e^{j\pi(2f_0t+k_0t^2)}] = \sqrt{1/(1-ik_0)} \delta(u-f_0\sin\alpha) \quad (3)$$

用  $x_k$  表示双足机器人的位姿即为:  $\{x_k, y_k, \theta_k\}$ , 从而,双足机器人的运动学模型描述为:

$$x_k = f\{x_{k-1}, u_{k-1}, w_{k-1}\} \quad (4)$$

双足机器人的避障运动规划公式为:

$$mV \frac{d\theta}{dt} = P \sin\alpha + Y - mg \cos\theta \quad (5)$$

$$J_z \frac{d\omega_z}{dt} + (J_y - J_x)\omega_y\omega_x + J_{xy}(\omega_y^2 - \omega_x^2) = M_z \quad (6)$$

如果  $k_1 \neq k_2$  或  $t_1 \neq t_2$ ,  $\phi_{k1t1}$  独立于  $\phi_{k2t2}$ , 通过强化学习搜索的方法,进行机器人的时延二自由度控制,得到双足机器人质心与速度矢量  $V$  倾角的稳定解。由此构建双足机器人的位置纠偏和避障被控对象模型,结合稳定性调节方法,提高机器人的实时避障控制能力<sup>[9-11]</sup>。

### 1.2 机器人的实时路径动力学模型

以机器人的双足步行稳定性为控制目标函数,构建双足机器人的实时路径动力学模型<sup>[12]</sup>,得到双足机器人在避障路径规划中的运动动能  $T$  和势能  $V$  分别描述为:

$$T = \frac{1}{2}M_{RL} \dot{X}_{RL}^2 + \frac{1}{2}M_{RR} \dot{X}_{RR}^2 + \frac{1}{2}J_{RL} \dot{\theta}_{RL}^2 + \frac{1}{2}J_{RR} \dot{\theta}_{RR}^2 + \frac{1}{2}M_P [(\dot{\theta}_P L \cos\theta_P + \dot{X}_{RM})^2 + (-\theta_P L \cos\dot{\theta}_P)^2] + \frac{1}{2}J_{P\theta} \dot{\theta}_P^2 + \frac{1}{2}J_{P\delta} \dot{\delta}^2 \quad (7)$$

$$V = M_P g L \cos\theta_P \quad (8)$$

采用位姿参量自适应反馈调节方法,进行双足机器人避障自动控制<sup>[13]</sup>,零势能面得到双足机器人避障控制的自适应运动学方程为:

$$P_m = C_L \dot{\theta}_{RL} + C_R \dot{\theta}_{RR} - (C_L + C_R) \dot{\theta}_P \quad (9)$$

在 7 自由度空间内进行避障空间的路径规划,在未知扰动下,绕 Y 轴左右旋转的状态转移矩阵为常值矩阵,双足机器人位姿发生变化时,执行器使用零阶保持器进行位置保持,进行机器人的位置伺服相关性参数调节<sup>[14]</sup>,得到机器人的多自由度动力学模型为:

$$z(t) = \sum_{i=1}^p s_i(t) b_i(\theta_i) + n(t) \quad (10)$$

利用机器人的关节补偿功能进行自动控制,得到机器人末端位姿的微小调节误差分布为:

$$b_i(\theta_i) = \int_{-\pi}^{\pi} a(\theta) g_i(\theta - \theta_i) d\theta \quad (11)$$

采用惯性参量融合方法实时获得机器人末端的位姿,得到对应的机器人末端轨迹分布的特征值满足  $\lambda_1 > \lambda_2 > \dots > \lambda_{j-1} > \lambda_j > \dots > \lambda_m$ , 对于无纠偏功能的机器人控制系统,采用位置伺服方法进行避障过程中的运动系数调节,构建时滞二自由度控制模型为:

$$\sum_{k=1}^m [\theta_{jk}(n+1) - \theta_{jk}(n)] q_k = \sum_{k=1}^m \eta [\lambda_k - \delta_j^2(n)] \theta_{jk}(n) q_k + \eta \sum_{k=1}^{j-1} \lambda_k a_{jk}(n) q_k \quad (12)$$

以机器人质心运动的加速度和惯性力矩为被控对象,得到机器人的实时路径动力学控制问题转化为求如下表达式的最优解问题:

$$\min(f) = \sum_{i=1}^m \sum_{j=1}^n C_{ij} X_{ij}$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^m X_{ij} = a_i, i = 1, 2, \dots, m \\ \sum_{i=1}^m X_{ij} = b_i, j = 1, 2, \dots, n \\ X_{ij} \geq 0, i = 1, 2, \dots, m, j = 1, 2, \dots, n \end{cases} \quad (13)$$

根据上述动力学模型构建,采用等效碰撞子模型进行双足机器人的实时避障路径规划,实现机器人的实时避障位置控制和自适应寻优<sup>[15]</sup>。

## 2 避障位置控制优化

### 2.1 模糊强化学习算法

采用碰撞子模型和摆动物子模型相结合的方法自适应调节和反馈控制,控制参量的确定部分  $N_{ij} \geq m$ , 则若  $(N(i) \bmod L) < m$ ,  $p(i) = \lfloor \frac{N(i)}{L} \rfloor$ , 双足机器人避障控制子系统动力学模型  $p(i) = \lfloor \frac{N(i)}{L} \rfloor$ , 采用模糊强化学习方法,得到双足机器人阻尼力矩矢量  $z(t)$  的协方差矩阵为:

$$R = E[z(t)z^H(t)] = BP_s B^H + \sigma_n^2 I_M \quad (14)$$

其中  $B = [b_1(\theta_1), b_2(\theta_2), \dots, b_q(\theta_q)]^T$ , 受动态约束下,双足机器人自适应反馈调整的中心波达方向为  $\theta_i$  和扩展角  $\sigma_i$ , 在外力矩作用下,机器人的实时避障路径规划的状态协方差矩阵的奇异值分解为:

$$R = U_s \Lambda_s U_s^H + U_n \Lambda_n U_n^H \quad (15)$$

考虑未知非线性函数约束的情况,即  $r < m$ , 采用碰撞子模型和摆动物子模型相结合的方法,进行双足机器人行走路径的纠偏参量反馈调节,得到强化学习矩阵为:

$$\Upsilon_1 = \begin{bmatrix} \Upsilon_{11} & 0 & \Upsilon_{13} & 0 & 0 & P_1 W \\ * & \Upsilon_{22} & P_2 D & \Upsilon_{24} & \Upsilon_{25} & 0 \\ * & * & \Upsilon_{33} & 0 & 0 & 0 \\ * & * & * & \Upsilon_{44} & 0 & \Upsilon_{46} \\ * & * & * & * & \Upsilon_{55} & 0 \\ * & * & * & * & * & \Upsilon_{66} \end{bmatrix} \quad (16)$$

在矩阵中各参量为:

$$\Upsilon_1^T = [-A^T \ 0 \ 0 \ 0 \ 0 \ W^T \ 0 \ 0]^T \quad (17)$$

$$\Upsilon_3^T = [0 \ -C^T \ D^T \ 0 \ 0 \ 0 \ 0 \ 0]^T \quad (18)$$

$$\xi(t) = [x^T(t)y^T(t)x^T(t-\tau)y^T(t-\sigma)f^T(y(t))f^T(y(t-\sigma))]^T \quad (19)$$

当  $p > 2$  时,进行  $\dot{H}_s$  一全局强稳定泛函,得到无双足机器人自适应全局状态参量调节的稳定性控制方程为:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bx(t-d_1(t)-d_2(t)) \\ x(t) &= \phi(t) \quad t \in [-h, 0] \end{aligned} \quad (20)$$

式中:  $x(t) = [x_1(t), x_2(t), \dots, x_i(t)]^T$  是模糊强化学习的跟踪矢量,进行双足机器人的偏置分析,得到各超差几何变化关系为:

$$x_i(n) = \sum_{j=1}^M h_{ij}(n)^T s_j(n) + v_i(n) \quad (21)$$

$$y_j(n) = \sum_{i=1}^P f_{ij}(n)^T x_i(n) \quad (22)$$

式中:  $h_{ij}$  代表双足机器人在各个避障规划路径上的力学阻尼参数,  $f_{ij}$  表示机器人模糊强化学习跟踪的敏感性特征量。

### 2.2 控制律设计

采用模糊强化学习跟踪方法,进行双足机器人的误差增益控制,结合传感信息融合跟踪识别方法,得到双足机器人位置伺服模态参数识别输出:

$$f_{ij}(n+1) = f_{ij}(n) + \mu_{MCMA} \frac{\partial J_{MCMA}(n)}{\partial f_{ij}(n)} \quad (23)$$

其中,  $\mu_{MCMA}$  代表轴向位置度,设计机器人稳定性调节的 Lyapunov 函数为:

$$\frac{\partial J_{MCMA}(n)}{\partial f_{ij}(n)} = \frac{1}{4} \left( \sum_{j=1}^M \frac{\partial J_{R,j}(n)}{\partial f_{ij}(n)} + \sum_{j=1}^M \frac{\partial J_{L,j}(n)}{\partial f_{ij}(n)} \right) \quad (24)$$

对实部求偏导:

$$\begin{aligned} \frac{\partial J_{R,j}(n)}{\partial f_{ij}(n)} &= \frac{\partial (|y_{R,j}(n)|^2 - R_{2,R})^2}{\partial f_{ij}(n)} = 4(|y_{R,j}(n)|^2 - R_{2,R}) \times \frac{\partial (|y_{R,j}(n)|^2 - R_{2,R})}{\partial f_{ij}(n)} = 4(|y_{R,j}(n)|^2 - R_{2,R}) \times \frac{\partial ((x_j(n) * f_{ij}(n)) (x_j(n) * f_{ij}(n))^* - R_{2,R})}{\partial f_{ij}(n)} = \\ &= 4(|y_{R,j}(n)|^2 - R_{R,j}) \times x_j(n) \times (x_j(n) * f_{ij}(n))^* = \\ &= 4(|y_{R,j}(n)|^2 - R_{2,R}) \times y_{R,j}(n)^* \times x_j(n) \end{aligned} \quad (25)$$

得到双足机器人的位置伺服控制的动态学特征方程描述为:

$$\begin{cases} W_i^T(x)X(n) = X^T(n)\omega_j(n) \\ a_j^T(n)Y_{j-1}(n) = Y_{j-1}^T a_j(n) \end{cases} \quad (26)$$

综合阻尼力矩和超差变量调节方法,进行自适应控制,得到双足机器人伺服系统的阻尼特征方程为:

$$\begin{aligned} \omega_j(n+1) &= \omega_j(n) + \eta[y_j(n)X(n) - y_j^2(n)\omega_j(n)] = \\ &= \omega_j(n) + \eta[X(n)X^T(n)\omega_j(n) + \\ &X(n)X^T(n)Q^T a_j(n) - y_j^2(n)\omega_j(n)] \end{aligned} \quad (27)$$

根据强化学习进行机器人避障轨迹的误差调整,得到位置参量调节的输入与输出关系为:

$$x(n) + \sum_{k=1}^p a_k x(n-k) = \sum_{r=0}^m b_r u(n-r) \quad (28)$$

根据 Lyapunov 稳定性原理,得到双足机器人的实时避障位置控制的误差函数为:

$$e_{R,j} = (|y_{R,j}(n)|^2 - R_{2,R}) \times y_{R,j}(n)^* \quad (29)$$

以机器人质心运动的加速度和惯性力矩为控制约束参量,得到双足机器人位置伺服控制的迭代方程描述为:

$$f_{ij}(n+1) = f_{ij}(n) + \mu_{MCMA} \times ((|y_{R,j}(n)|^2 - R_{2,R})y_{R,j}(n)^* + j(|y_{L,j}(n)|^2 - R_{2,L})y_{L,j}(n)^*) \times x_j(n) \quad (30)$$

采用连续反馈驱动加权修正方法,得到双足机器人控制过程的误差项收敛到:

$$E[\tilde{X}(n)y_j^*(n-k)] = 0 \quad (31)$$

式中:  $k = K - 1$ ,  $\tilde{X}(n)$  是  $\tilde{x}(n)$  的 Toeplitz 矩阵,可见采用本文方法进行双足机器人的实时避障控制的稳定性和收敛性较好。

### 3 仿真实验分析

为了验证本文方法在实现的双足机器人的实时避障位置控制中的应用性能,进行仿真实验,实验采用 Matlab 设计,仿真时间设置为 3 000 s,机器人的导航跟踪坐标系为东向夹角  $15^\circ$  方向,机器人行走采用匀速运动方式,前向运动速度为 3 m/s,机器人的初始位置为  $(0, 0, 0)$ ,行走误差方差为  $\sigma = 0.5^\circ$ ,干扰强度为  $-10$  dB,根据上述仿真参数设定,进行机器人的运动参数采集,采集的参数分别为偏离角、方位角、轨迹跟踪位置等,得到原始采集的机器人实时避障位置参数如图 1 所示。

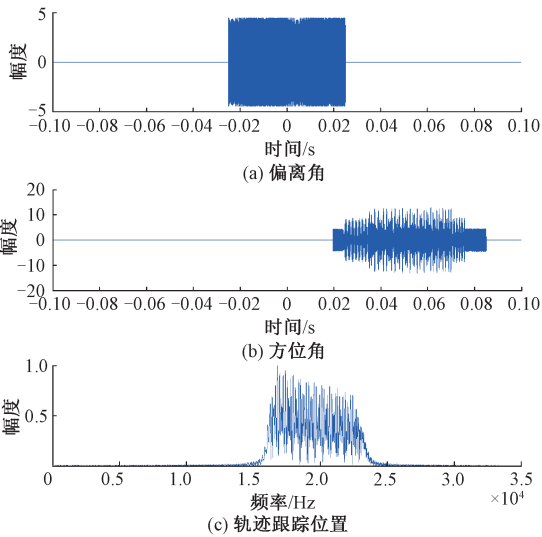
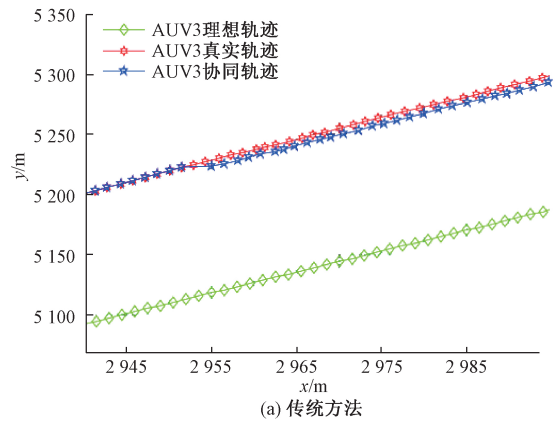


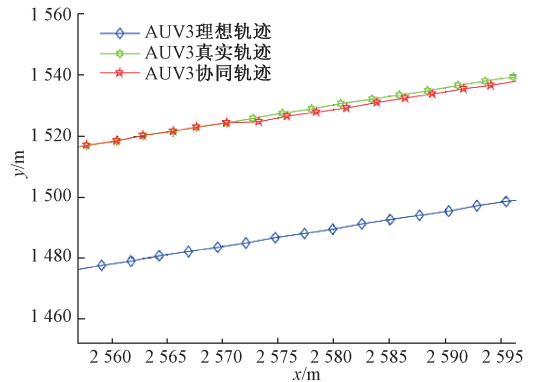
图 1 机器人实时避障位置参数

以图 1 所示的机器人避障位置参数为输入,采用碰撞子模型和摆动物子模型相结合的方法,进行双足机器人行走路径的纠偏参量反馈调节,采用模糊强化学习跟踪方法,进行双足机器人的误差增益控制,得到采用本文方法和传统方法进行机器人的实时避障位置控制的轨迹跟踪结果如图 2 所示。

分析图 2 得知,采用本文方法进行机器人避障控制的轨迹跟踪性能较好,控制误差较低,测试在不同的避障位置点数条件下进行机器人避障控制的误差,得到对比结果见表 1。分析表 1 得知,随着障碍物的增大,位置跟踪的误差随之增大,本文方法进行避障控制,机器人轨迹跟踪误差较低,比传统方法具有明显的改善,轨迹纠偏性能较好。



(a) 传统方法



(b) 本文方法

图 2 机器人的实时避障位置控制的轨迹跟踪结果

表 1 避障控制误差 (mm)

避障位置点数	本文方法	PID	滑膜控制
10	0.012	0.087	0.076
20	0.023	0.102	0.082
30	0.046	0.244	0.134
40	0.048	0.356	0.201

## 4 结 论

双足机器人在路径规划控制中,容易受到环境和机器人自身扰动因素的影响,导致机器人的避障性不好,为了提高对双足机器人的实时避障能力,本文提出一种基于强化学习的双足机器人的实时避障位置控制方法。以机器人的双足步行稳定性为控制目标函数,构建双足机器人的实时路径动力学模型,以机器人质心运动的加速度和惯性力矩为被控对象,采用等效碰撞子模型进行双足机器人的实时避障路径规划,采用碰撞子模型和摆动物子模型相结合的方法,进行双足机器人行走路径的纠偏参量反馈调节,采用模糊强化学习跟踪方法,进行双足机器人的误差增益控制,实现双足机器人的实时避障位置控制。研究得知,采用该方法进行双足机器人控制的实时避障性能较好,纠偏能力较强。

## 参考文献

- [1] 徐飞. 基于改进人工势场法的机器人避障及路径规划研究[J]. 计算机科学, 2016, 43(12): 293-296.
- [2] 宋达, 张立勋, 王炳军, 等. 柔索牵引式力觉交互机器人控制策略[J]. 机器人, 2018, 40(4): 440-447.
- [3] 游有鹏, 张宇, 李成刚. 面向直接示教的机器人零力控制[J]. 机械工程学报, 2014, 50(3): 10-17.
- [4] 彭程, 白越, 乔冠宇, 等. 四旋翼无人机的偏航抗饱和与多模式 PID 控制 [J]. 机器人, 2015, 37(4): 415-423.
- [5] JIANG Y, MOSESON S, SAXENA A. Efficient grasping from RGBD images: Learning using a new rectangle representation [C]. IEEE International Conference on Robotics and Automation. Piscataway, USA, IEEE, 2011: 3304-3311.
- [6] JOHNS E, LEUTENEGGER S, DAVISON A J. Deep learning a grasp function for grasping under gripper pose uncertainty [C]. IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, USA: IEEE, 2016: 4461-4468.
- [7] MUHAMMAD J, ALTUN H, Abo-Serie E. Welding seam profiling techniques based on active vision sensing for intelligent robotic welding[J]. International Journal of Advanced Manufacturing Technology, 2016, 88(1/4): 127-145.
- [8] 于金鹏, 于海生, 林崇. 考虑铁损的异步电动机模糊自适应命令滤波反步控制[J]. 控制与决策, 2016, 31(12): 2189-2194.
- [9] 吴超, 王浩文, 张玉文, 等. 基于 LADRC 的无人直升机轨迹跟踪[J]. 航空学报, 2015, 36(2): 473-483.
- [10] 杨亮, 傅瑜, 付根平, 等. 基于能效优化的仿人机器人跑步步态优化与控制[J]. 计算机科学, 2016, 43(6): 270-275.
- [11] 马知瑶, 周一览. 光纤陀螺寻北启动误差抑制方法研究[J]. 传感器与微系统, 2018, 37(6): 33-35.
- [12] 王帅, 黄海鸿, 韩刚, 等. 基于 PCA 与 GA-BP 神经网络的磁记忆信号定量评价[J]. 电子测量与仪器学报, 2018, 32(10): 190-196.
- [13] 张浩杰, 苏治宝, 苏波. 基于深度 Q 网络学习的机器人端到端控制方法[J]. 仪器仪表学报, 2018, 39(10): 36-43.
- [14] 杨鸿杰, 张君毅. 基于强化学习的智能干扰算法研究[J]. 电子测量技术, 2018(20): 49-54.
- [15] 贵彦乔, 吴彦鸿. 一种对跟踪雷达的航迹干扰方法[J]. 国外电子测量技术, 2018, 37(9): 62-67.

## 作者简介

林荣霞, 硕士, 讲师, 主要研究方向为嵌入式技术、机器人控制系统及算法、算法研究。

E-mail: lxhxjtul@126.com