航空科学技术
Aeronautical Science & Technology
Jan. 25 2020 Vol. 31 No.01 68-75

基于深度强化学习的无人机着陆轨迹跟踪控制

宋欣屿[1,*]，王英勋[1]，蔡志浩[1]，赵江[1]，陈小龙[2]，宋栋梁[2]
1.北京航空航天大学 自动化科学与电气工程学院，北京 100191
2.航空工业自控所 飞行器控制一体化技术国防科技重点实验室，陕西 西安 710065

摘 要：本文针对固定翼无人机自主着陆控制问题，提出了基于深度强化学习(DRL)的无人机着陆轨迹跟踪控制方法。首先，搭建了小型固定翼无人机Ultra Stick 25E的仿真模型，设计了满足过程和终端约束的着陆参考轨迹。其次，提出了基于深度确定性策略梯度(DDPG)的无人机一体化控制框架，设计了考虑跟踪误差和轨迹平稳性的奖励函数。最后，通过离线训练，得到了轨迹跟踪一体化控制器。仿真试验结果表明，本文提出的方法比传统PID控制方法精度更高。

关键词：固定翼无人机；自主着陆；轨迹跟踪控制；深度强化学习；深度确定性策略梯度

中图分类号：TP273　　**文献标识码**：A　　**DOI**：10.19452/j.issn1007-5453.2020.01.009

固定翼无人机没有人员伤亡的风险，还有着机动性能强、飞行速度快、航程远、成本低、重量(质量)轻等多种优点，在很多领域得到了广泛的应用。在民用上，固定翼无人机可被用于资源探测、森林防火、城市规划、大气监测、边境及海岸线巡逻等领域。在军用上，可执行空中侦察、战情评估、电子干扰、对地攻击、拦截巡航导弹，甚至空中格斗等多种任务[1]。随着科技的发展，无人机的自主飞行技术日益成熟，但自主着陆仍然是一大难点。据统计，起飞和着陆是无人机最容易发生事故的阶段，而其中着陆最为严重[2]。在很大程度上，轨迹跟踪控制如果不够精确，无人机的飞行安全、任务的完成效果都会受到极大的影响。在飞行安全要求极高的着陆过程中，轨迹跟踪控制显得格外重要[3,4]。

针对无人机的轨迹跟踪控制问题，众多学者提出了多种不同的控制方法，如最为经典的PID控制方法、鲁棒性较好的滑模控制方法、在线辨识改变控制器参数的自适应控制方法等，这些方法虽然都通过了试验的验证，但都有着各自的不足。PID算法最早被用于实际应用中，也最为经典，但它需要人工整定参数，需要经过大量的尝试不断调整，十分繁琐。滑模控制方法虽然响应速度很快，鲁棒性也较强，但它存在抖动的问题，需要配合其他方法一起使用。2016年人工智能机器人AlphaGo战胜了围棋世界冠军李世石，这场人机大战使人工智能走进了人们的视野。随着人工智能的不断发展，作为人工智能重要组成部分的强化学习算法的研究也日益深入，目前强化学习在诸多领域都取得了成功的应用，如机器人控制领域[5,6]、多智能体编队控制问题[7,8]等。

最早将强化学习应用到无人机控制领域中的是斯坦福大学的吴恩达教授[9]，他选择了Yamaha R-50无人直升机作为控制对象。这一直升机全长约3.6m，负载可高达20kg，机上载有一台飞控计算机和多种传感器[10]。吴恩达根据卡内基梅隆大学的Bernard Mettler团队的方法建立了Yamaha R-50无人直升机的12阶模型[11,12]，利用强化学习中随机策略搜索算法的思想对直升机进行训练，使其可以维持稳定的悬停状态，试验结果显示了强化学习悬停控制器具有很好的控制效果。此后他又利用强化学习使直升机能完成其他如原地转弯、倒飞、360°翻转等难度更高的动作，均取得了良好的效果。他的学生Pieter Abbeel利用强化学习中学徒学习的算法，从专业飞手的任务演示中获取

收稿日期：2019-10-30；退修日期：2019-11-05；录用日期：2019-12-05
基金项目：航空科学基金(20175851032)
*通信作者. Tel.：13121239357　　E-mail：songxinyu@buaa.edu.cn
引用格式：Song Xinyu, Wang Yingxun, Cai Zhihao, et al. Landing trajectory tracking control of unmanned aerial vehicle by deep reinforcement learning[J].Aeronautical Science & Technology,2020,31(01):68-75.宋欣屿,王英勋,蔡志浩,等.基于深度强化学习的无人机着陆轨迹跟踪控制[J].航空科学技术,2020,31(01):68-75.

训练集进行学习，设计出了直升机的控制器，这一算法极大地提高了直升机控制的自主性，抗干扰能力也较强。

强化学习方法虽然获得了一些有效的应用，但是大多数特征状态需要人工设定，在面对高维数据所表示的复杂环境时，难以找到合适的特征表达方法，容易陷入维数灾难问题。而深度学习网络无须人类干预，可以自动进行特征提取[13]。因而将深度学习与强化学习相结合，由强化学习定义任务的模型目标及优化的方向，深度学习给出表征问题以及解决问题的方式，就可以更好地解决高维连续空间的控制问题。

针对固定翼无人机着陆轨迹跟踪控制问题，本文基于大量的训练设计了深度强化学习轨迹跟踪控制器，仿真试验结果表明，这一控制方法实现了对固定翼无人机着陆轨迹的一体化跟踪控制且控制精度优于PID控制方法。

1 固定翼无人机建模

本文选取了Ultra Stick 25E无人机（见图1）作为参考对象。这是一架小型商用无线电遥控固定翼无人机，该无人机具有传统的水平和垂直尾翼，采用对称翼型机翼，并具有副翼和襟翼操纵面。其所有操纵面均由Hitec伺服系统驱动，推进系统由一台功率为600W的E-Flite电动机和APC 12×6的螺旋桨组成[14]。

图1 Ultra Stick 25E无人机

Fig.1 UAV Ultra Stick 25E

无人机是一个十分复杂的多输入多输出的非线性系统，若考虑飞行过程中所有的因素会给建模带来极大的难度。由于本文的小型固定翼无人机在大气层内飞行，飞行速度和高度有限，因此可以做出合理地假设以简化模型。作用在无人机上的重力、推力、空气动力和力矩是因为不同的原因而产生的，合理地选择坐标系分析受力有助于简化计算。本文选择了地面坐标系来分析无人机受到的重力作用，选择机体坐标系分析无人机受到的推力和力矩的影响，选择气流坐标系来描述无人机受到的空气动力和力矩。

在分析无人机所受力与力矩时，主要分为了三个部分。

(1) 重力

无人机受到的重力是一个惯性矢量，由于假设质量和重力加速度不变，因此大小恒定，由于重力作用于无人机的质心，因此不会产生力矩。

(2) 推力

本文所研究的无人机的推力由电机驱动螺旋桨转动获得，由于电机数据无法从制造商处获得，因此利用商业软件MotorCalc[15]获取数据对推进系统进行建模。选择具有零空速的静态飞行条件，模拟油门输入从0逐步增加到1，步长为0.1时，无人机受到的推力。利用Matlab对这些数据进行三次多项式插值处理，获得无人机受到的推力T与油门输入δ_T的关系。

(3) 空气动力

本文研究的无人机的主要操纵面有升降舵、副翼和方向舵，操纵面主要通过影响空气动力来控制无人机的飞行状态。设总空气动力沿气流坐标系各轴的分量分别为X_A、Y_A、Z_A，总空气动力矩沿机体坐标系各轴的分量分别为$\bar{L}_A$、M_A、N_A。通常用D、L、Y分别表示阻力、升力和侧力，于是有$D=X_A$，$L=-Z_A$，$Y=Y_A$。设ρ为空气密度(取$\rho=1.225\text{kg/m}^3$)，V为空速，S_W为机翼参考面积，b为机翼展长，c为机翼平均几何弦长，q为动压($q=\frac{1}{2}\rho V^2$)，p、q、r分别为滚转、俯仰、偏航角速度，δ_e、δ_r、δ_a分别为升降舵、方向舵、副翼偏转角，则有：

(1) 气流坐标系的下空气动力

升力：

$$L = qS_wC_L$$
$$C_L = C_{L_0} + C_{L_\alpha}\alpha + C_{L_{\delta_e}}\delta_e + C_{L_q}\frac{qc}{2V} \tag{1}$$

阻力：

$$D = qS_wC_D$$
$$C_D = C_{D_0} + C_{D_{\delta_e}}\delta_e + C_{D_{\delta_r}}\delta_r + \frac{(C_L - C_{L\min})}{\pi \cdot e \cdot AR} \tag{2}$$

侧力：

$$Y = qS_WC_Y$$
$$C_Y = C_{Y_\beta}\beta + C_{Y_{\delta_r}}\delta_r + C_{Y_p}\frac{pb}{2V} + C_{Y_r}\frac{rb}{2V} \tag{3}$$

式中：C_L、C_{L_0}、C_{L_α}、$C_{L_{\delta_e}}$、C_{L_q}为升力系数，C_D、C_{D_0}、$C_{D_{\delta_e}}$、$C_{D_{\delta_r}}$为阻力系数，AR为展弦比，e为梯形比，C_Y、C_{Y_β}、$C_{Y_{\delta_r}}$、C_{Y_p}、C_{Y_r}为侧力系数。

(2) 机体坐标系下的空气动力矩

滚转力矩：

$$\bar{L}_A = qS_W C_l b$$
$$C_l = C_{l_\beta}\beta + C_{l_{\delta_a}}\delta_a + C_{l_{\delta_r}}\delta_r + C_{l_p}\frac{pb}{2V} + C_{l_r}\frac{rb}{2V} \quad (4)$$

俯仰力矩：

$$M_A = qS_W C_m c$$
$$C_m = C_{m_0} + C_{m_\alpha}\alpha + C_{m_{\delta_e}}\delta_e + C_{m_q}\frac{qc}{2V} \quad (5)$$

偏航力矩：

$$N_A = qS_W C_n b$$
$$C_n = C_{n_\beta}\beta + C_{n_{\delta_a}}\delta_a + C_{n_{\delta_r}}\delta_r + C_{n_p}\frac{pb}{2V} + C_{n_r}\frac{rb}{2V} \quad (6)$$

式中：C_l、C_{l_β}、$C_{l_{\delta_a}}$、$C_{l_{\delta_r}}$、C_{l_p}、C_{l_r}为滚转力矩系数，C_m、C_{m_0}、C_{m_α}、$C_{m_{\delta_e}}$、C_{m_q}为俯仰力矩系数，C_n、C_{n_β}、$C_{n_{\delta_a}}$、$C_{n_{\delta_r}}$、C_{n_p}、C_{n_r}为偏航力矩系数[16]。

将这三部分整合到一起，结合无人机的运动学与动力学方程，即可根据每一时刻无人机的状态矢量$\boldsymbol{x}=(u,v,w,\phi,\theta,\psi,p,q,r,x_g,y_g,h)^{\mathrm{T}}$和控制输入矢量$\boldsymbol{u}=(\delta_{\mathrm{T}},\delta_{\mathrm{e}},\delta_{\mathrm{r}},\delta_{\mathrm{a}})^{\mathrm{T}}$得知任何时刻无人机的运动状态，仿真模型示意图如图2所示。

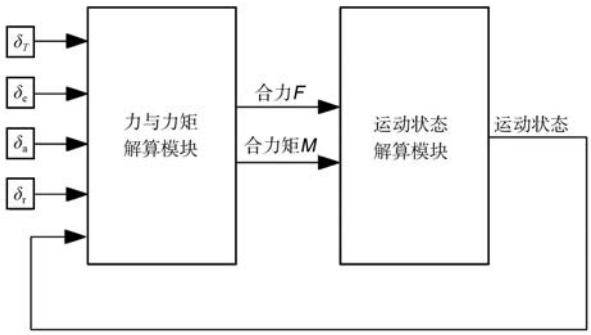


图2 无人机仿真模型

Fig. 2 Simulation model of UAV

2 着陆轨迹跟踪控制方法

2.1 无人机着陆轨迹设计

由于固定翼无人机一般有着固定的航迹切换点，在切换时对速度和姿态也有着一定的要求，所以进近段的着陆轨迹无论是形式还是参数都较为固定，适合离线规划[17]。本文在设计无人机的进近段着陆轨迹时，主要考虑了如图3所示的4个阶段，即定高、下滑、拉平及滑跑。

根据无人机的有关参数，本文设计的着陆轨迹定高飞行的高度H_1为15m，进场速度V_{enter}为15m/s，下滑段下滑角γ为5°，拉平段选择指数拉平，其中拉平时间常数τ为

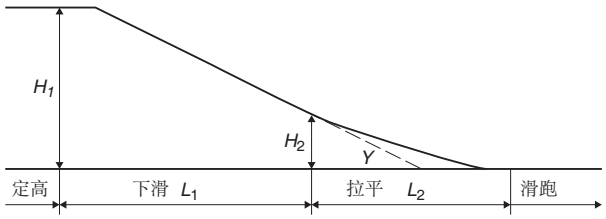


图3 着陆过程示意图

Fig.3 Diagram of landing process

2.6970s。无人机接地后进入滑跑阶段，这一阶段只需要调整偏航角使无人机能对准跑道中心即可，不需要进行轨迹规划，本文不对这一阶段进行控制和研究。根据上述内容，为Ultra Stick 25E无人机设计的着陆轨迹如图4所示。

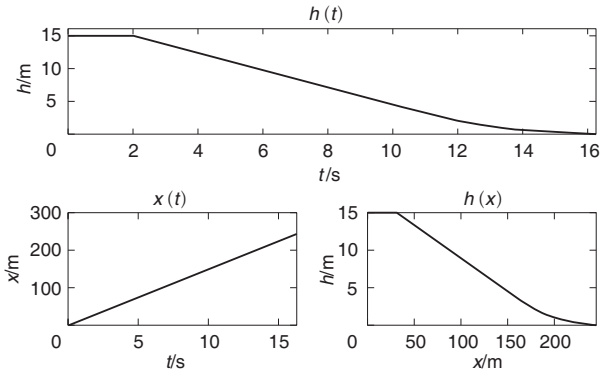


图4 着陆轨迹

Fig.4 Landing trajectory

2.2 PID轨迹跟踪控制器

由于本文主要研究纵向着陆的轨迹跟踪控制，所以假设固定翼无人机横侧向所受的力与力矩始终为0，滚转角、偏航角、滚转角速度、偏航角速度、横侧向的速度和位移也始终保持0，在设计PID轨迹跟踪控制器时，也只考虑了纵向。

轨迹控制(外环)是建立在姿态控制(内环)基础上的，在控制高度时，首先要对俯仰角进行控制，然后在此基础上设计纵向下降速度的控制器，在最外环设计高度的跟踪控制器，高度控制原理如图5所示。在控制纵程时也是同理，先设计了速度的控制器，在外环设计纵程跟踪控制器(见图6)。

2.3 深度强化学习轨迹跟踪控制方法

在强化学习中，通常将可以通过学习来自动获取有价值的信息的机器称作智能体，应具备必要的计算能力。强化学习的基本原理如图7所示，智能体在完成某一项任务时，首先要通过产生一个动作a_t来与环境进行交互，在动作a_t和环

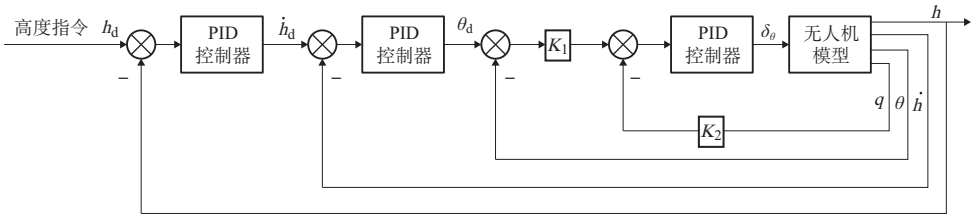


图5 高度控制框图

Fig.5 Block diagram of height control

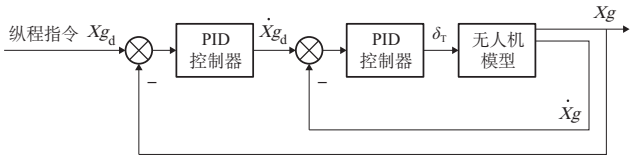


图6 纵程控制框图

Fig.6 Block diagram of longitudinal length control

境的共同作用下，智能体会产生新的状态s_{t+1}，环境会给出一个同步的回报r_{t+1}，智能体根据新的状态s_{t+1}产生新的动作a_{t+1}，继续与环境交互。按照这种方式不断循环下去，在智能体和环境不断交互的过程中（见图8）会产生大量的数据，强化学习算法，利用这些数据修改自身产生动作的策略，再与环境交互，进而产生大量新的数据，并利用新的数据进一步学习以改善自身的动作策略。经过多次的迭代和学习后，智能体最后就可以学到能完成期望的任务所对应的最优的动作策略。

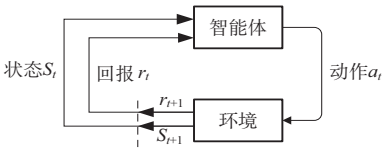


图7 强化学习基本框架图

Fig. 7 Basic framework of reinforcement learning

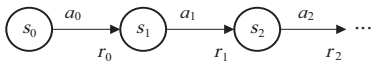


图8 智能体与环境的交互过程示意图

Fig.8 Schematic diagram of the interaction process between the agent and the environment

根据动作输出连续还是离散，可以将强化学习算法分为值函数方法和策略梯度方法。采用值函数近似的方法，需要将输出的动作进行离散化，但对固定翼无人机输出的舵偏和油门指令进行离散会产生很大的动作空间，很难保证训练结果一定收敛。一方面会导致输出的舵偏和油门指令不够准确，另一方面过于离散的控制指令也不符合无人机的机械特性。同时，由于强化学习具有较强的决策能力，但对感知问题束手无策，而深度学习具有较强的感知能力，但是缺乏一定的决策能力。将深度学习的感知能力和强化学习的决策能力相结合，令二者优势互补，可以直接从高维原始数据学习控制策略。因此针对固定翼无人机的着陆轨迹跟踪控制问题时，本文设计了基于深度确定性策略梯度算法（Deep Deterministic Policy Gradient，DDPG[18]）的无人机着陆轨迹跟踪控制器，既确保了无人机输出的控制指令的连续性，也便于对高维连续数据的处理。

DDPG是一种基于Actor-Critic框架的算法，可以用于解决连续动作空间上的深度强化学习问题，基本框架如图9所示。单独采用Critic网络低方差，但基于贪婪策略无法处理连续的动作域，单独使用Actor网络通过参数化可以处理连续动作域，但方差很高。Actor-Critic结合两者优点，使用参数化的Actor来根据当前状态产生动作，并能处理连续动作域，使用Critic的低方差的值函数来评估Actor产生的动作，产生一个更好的梯度估计值，改善局部优化的问题。

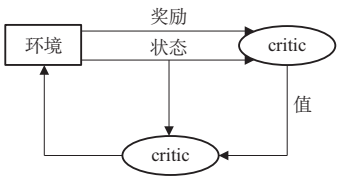


图9 DDPG原理框图

Fig.9 Block diagram of DDPG algorithm

DDPG算法中共有4种网络：(1)当前Actor网络$\mu(s;\theta^{\mu})$；(2)当前Critic网络$Q(s,a;\theta^{Q})$；(3)目标Actor网络$\mu(s;\theta^{\mu'})$；(4)目标Critic网络$Q(s,a;\theta^{Q'})$。其中，Actor网络以状态为输入，动作为输出；Critic网络以状态和动作为输入，Q值为输出。在训练完一组最小批量的数据之后，更

新当前网络的参数，然后再通过软更新算法更新目标网络的参数。目标网络参数变化小，算法更为稳定，训练易于收敛。

对无人机着陆轨迹跟踪控制器进行训练的过程如图10所示，主要分为以下几个步骤：(1) 初始化Actor和Critic当前网络的参数：θ^{μ}和θ^{Q}；(2) 将当前网络的参数拷给对应的目标网络：$\theta^{\mu'} \leftarrow \theta^{\mu}$，$\theta^{Q'} \leftarrow \theta^{Q}$；(3) 初始化经验缓存。

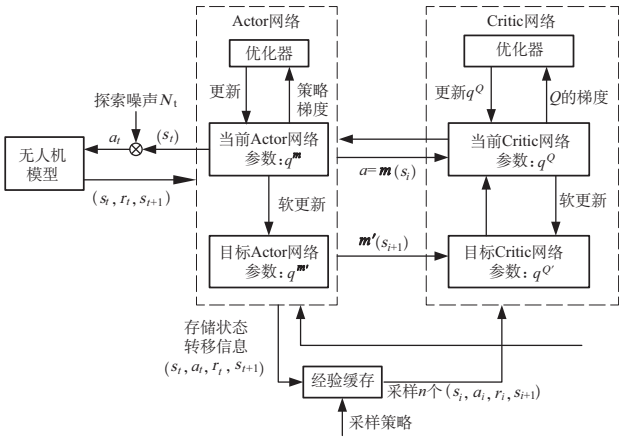


图10 训练过程原理图

Fig.10 Schematic diagram of the training process

对于每个回合：

(1) 初始化Uhlenbeck-Ornstein(UO)随机过程；

(2) 获得无人机初始状态s_1；

(3) 重复以下过程直至到达最大步长：

(a) Actor网络根据当前策略选择一个动作$\mu(s_t)$，引入UO随机过程产生的噪声N_t，下达指令$a_t = \mu(s_t|\theta^{\mu}) + N_t$给无人机模型；

(b) 无人机执行这一指令，返回奖励r_t和新的状态s_{t+1}；

(c) 将状态转移信息(s_t, a_t, r_t, s_{t+1})存入经验缓存，作为训练当前网络的数据集；

(d) 从经验缓存中，随机采样N个数据，作为当前Actor网络和当前Critic网络的训练数据，用(s_i, a_i, r_i, s_{i+1})表示单个状态转移数据；

(e) 通过最小化Critic网络的损失函数

$$L = \frac{1}{N}\sum_i (y_i - Q(s_i, a_i|\theta^{Q}))^2$$

更新目标Critic网络(采用Adam优化器更新θ^{Q})。

(f) 根据Actor网络的策略梯度

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{N}\left[\sum_i \nabla_{\theta^{\mu}} Q(s, a|\theta^{Q})|_{s=s_i, a=\mu(s_i,\theta^{\mu})}\right] =$$

$$\frac{1}{N}\left[\sum_i \nabla_a Q(s, a|\theta^{Q})|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^{\mu}} \mu(s, \theta^{\mu})|_{s=s_i}\right]$$

更新当前Actor网络(采用Adam优化器更新θ^{μ})。

(g) 更新目标网络

$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$$

$$\theta^{\mu'} \leftarrow \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$$

式中：$0 < \tau < 1$。

在训练无人机的着陆轨迹跟踪控制器时，本文采用的状态为$s = [u, w, \theta, q, Xg, h]$。由于无人机在着陆阶段主要控制的是高度的变化，横向速度基本恒定，因此油门仍由PID控制器控制，而升降舵舵偏作为训练的动作，即$a = [\delta_e]$。利用训练好的智能体控制无人机进行着陆轨迹跟踪的示意图如图11所示。

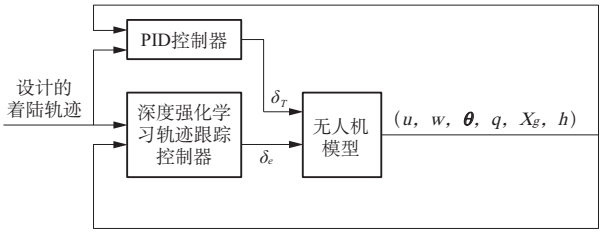


图11 深度强化学习控制器原理框图

Fig.11 Block diagram of controller based on DRL

3 仿真试验与分析

3.1 PID轨迹跟踪控制

将设计的着陆轨迹中高度和纵程随时间的变化数据，作为无人机的高度指令输入，初始条件为$\phi = 0, \theta = 0, \psi = 0$，$u = 15\text{m/s}, v = 0, w = 0, p = 0, q = 0, r = 0, x_g = 0, y_g = 0, h = 15\text{m}$，轨迹跟踪仿真结果如图12所示。

在利用PID控制器跟踪着陆轨迹的过程中，高度的最大误差为0.4361m。可见所设计的PID轨迹跟踪控制器在控制过程中各状态量较为平稳，且能够以较小的误差对预先设计好的着陆轨迹进行跟踪，效果良好。

3.2 深度强化学习轨迹跟踪控制

在训练时，首先初始化网络参数和经验缓存，令无人机的初始状态为$s_0 = (15,0,0,0,0,15)$，也就是在离地15m的空中以15m/s的速度平飞。然后对Actor网络选择动作添加一个方差为5的随机噪声，方差随训练次数增多逐渐减少。将产生的动作输入无人机模型，返回新的状态和回报。这里的回报函数设置为：

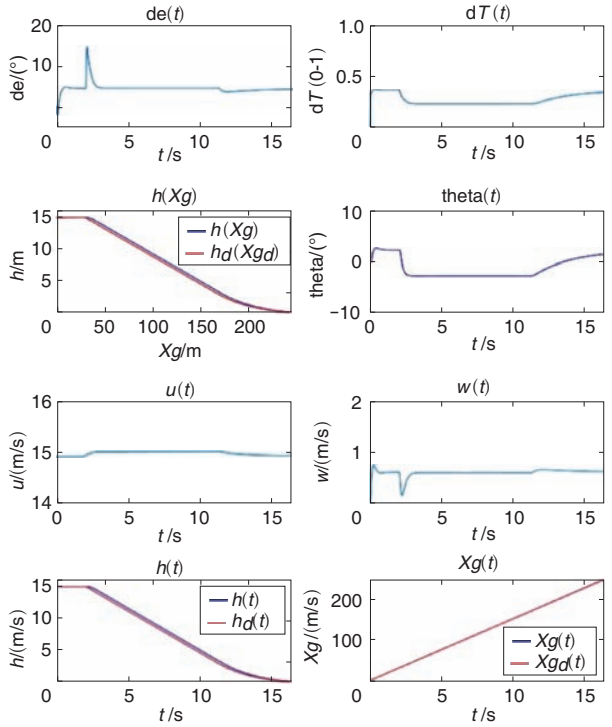


图12 PID轨迹跟踪仿真结果

Fig.12 Results of trajectory tracking simulation using PID

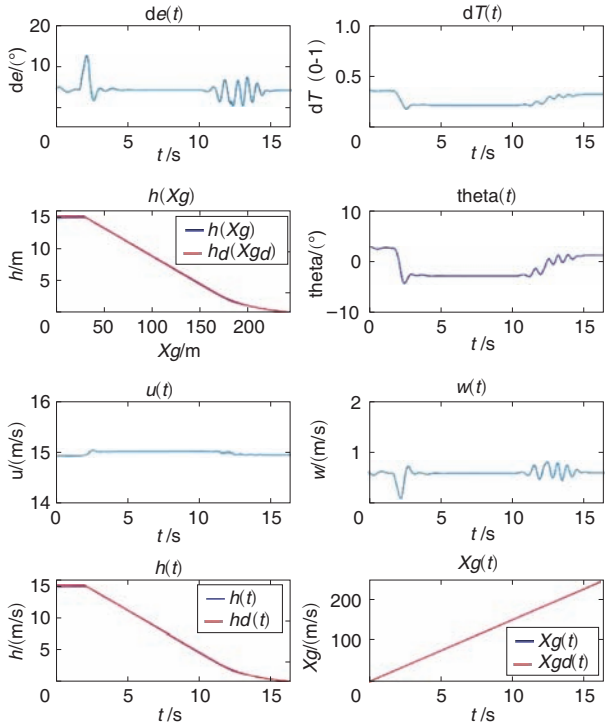


图13 DDPG轨迹跟踪仿真结果

Fig.13 Results of trajectory tracking simulation based on DDPG

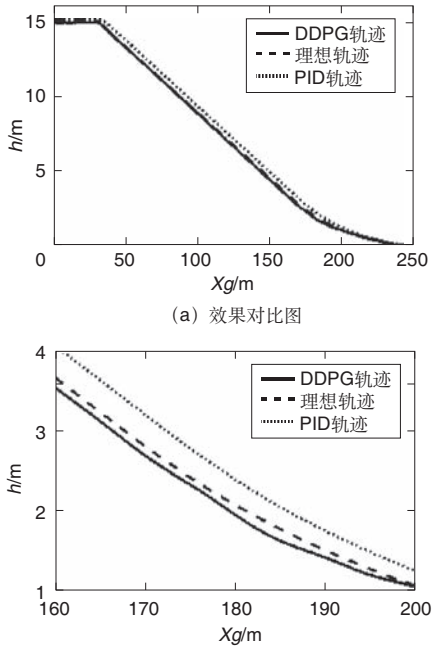
$$\text{reward} = -(w_1 \cdot |h - h_d| + w_2 \cdot (0.05 \cdot |\delta_{e\_new} - \delta_{e\_last}| + 0.05|\dot{q}| + 0.05\sqrt{\dot{u}^2 + \dot{v}^2}))$$

将数据存入经验缓存中，从经验缓存随机采样一组数据进行网络的训练，训练时状态和动作都进行了归一化处理，并在纵程达到阈值或达到最大步长MAX_EP_STEPD时停止这一回合，计算当前回合的总的代价。重复上述过程，直到达到设置的最大训练回合数MAX_EPISODES。以下为训练20000个回合中回报函数最大的网络控制无人机着陆轨迹的仿真结果，图13为在DDPG轨迹跟踪控制器控制下的无人机着陆轨迹跟踪结果。

可以观察到利用DDPG算法训练出的智能体可以控制无人机对预先设计好的着陆轨迹进行跟踪，在控制过程中各状态量较为平稳，高度控制的最大误差为0.2491m。

3.3 仿真结果分析

与传统PID控制器高度误差最大达0.4361m相比，DDPG控制方法最大高度误差仅0.2491m，控制精度优于PID控制方法。但由于DDPG的Actor网络输出与前一刻的动作无关，所以输出的动作连续性较差，导致中间的状态量波动较大。与PID控制器的效果对比如图14所示。



（a）效果对比图

（b）局部放大效果对比图

图14 DDPG与PID控制结果对比

Fig.14 Comparison diagram of DDPG and PID control

仿真试验结果表明，本文设计的基于深度强化学习方法的无人机着陆轨迹跟踪控制器不仅免去了手动调整参数的繁琐过程，而且在控制精度上要优于传统PID控制方法，具有研究价值。

4 结论

针对固定翼无人机着陆轨迹跟踪控制问题，本文对Ultra Stick 25E小型固定翼无人机进行了适当的简化与运动假设，对该型无人机在不同坐标系下进行受力与力矩分析，结合无人机的运动学与动力学方程，搭建了该无人机的仿真模型。同时，根据所建模型的特性，为其离线设计了可以保证其安全着陆的着陆轨迹。本文采用深度强化学习的思想，设计了合理的奖励函数和控制方式，通过大量的训练得到了深度强化学习轨迹跟踪控制器，实现了对固定翼无人机着陆轨迹的一体化跟踪控制。为了检测其控制效果，本文同时利用PID控制方法实现了对固定翼无人机着陆轨迹的跟踪控制。仿真试验结果表明，深度强化学习着陆轨迹跟踪控制方法具有比传统PID轨迹跟踪控制方法更高的精度。 AST

参考文献

[1] 郑积仕,蒋新华,陈兴武.基于CFD方法的固定翼无人机着陆控制建模[J].信息与控制,2012,41(1):51-56.
Zheng Jishi, Jiang Xinhua, Chen Xingwu. The landing control modeling for fixed-wing UAV based on CFD method[J]. Information and Control, 2012,41(1):51-56. (in Chinese)

[2] Arrabito G R, Ho G, Lambert A, et al. Human factors issues for controlling uninhabited aerial vehicles[R]. Defense Research and Development Canada, Toronto, 2010.

[3] González I, Salazar S, Torres J, et al. Real-time attitude stabilization of a mini-UAV quad-rotor using motor speed feedback[J]. Journal of Intelligent & Robotic Systems, 2013, 70 (1):93-106.

[4] Zou A M, Kumar K D, Hou Z G, et al. Finite-time attitude tracking control for spacecraft using terminal sliding mode and chebyshev neural network[J]. IEEE Transactions on Systems Man & Cybernetics Part B, 2011, 41(4):950-963.

[5] 张汝波,周宁,顾国昌,等.基于强化学习的智能机器人避碰方法研究[J].机器人,1999(3):45-50.
Zhang Rubo, Zhou Ning, Gu Guochang, et al. Reinforcement-Learning-Based obstacle avoidance learning for intelligent robot[J]. Robot, 1999(3):45-50. (in Chinese)

[6] 刘春阳,谭应清,柳长安,等.多智能体强化学习在足球机器人中的研究与应用[J].电子学报,2010,38(8):1958-1962.
Liu Chunyang, Tan Yingqing, Liu Changan, et al. Application of multi-agent reinforcement learning in robot soccer[J]. Chinese Journal of Electronics, 2010, 38(8): 1958-1962. (in Chinese)

[7] 王醒策,张汝波,顾国昌.多机器人动态编队的强化学习算法研究[J].计算机研究与发展,2003(10):1444-1450.
Wang Xingce, Zhang Rubo, Gu Guochang. Research on dynamic team formation of multi-robots reinforcement learning [J]. Journal of Computer Research and Development, 2003(10):1444-1450. (in Chinese)

[8] 王醒策,张汝波,顾国昌.基于强化学习的多机器人编队方法研究[J].计算机工程,2002(6):15-16,98.
Wang Xingce, Zhang Rubo, Gu Guochang. Research on multi-agent team formation based on reinforcement learning [J]. Computer Engineering, 2002(6):15-16, 98. (in Chinese)

[9] Ng A Y , Kim H J , Jordan M I , et al. Autonomous helicopter flight via reinforcement learning[Z]. German: Springer, Berlin, Heidelberg,2006.

[10] Hyunchul S. Hierarchical flight control system synthesis for rotorcraft-based unmanned aerial vehicles[R]. AIAA-2000-4057, 2000.

[11] Mettler B, Kanade T. System identification modeling of a model-scale helicopter[R]. American Helicopter Society, 2000.

[12] Mettler B, Kanade T. System identification of small-size unmanned helicopter dynamics[R]. American Helicopter Society Forum, 1999.

[13] 多南讯,吕强,林辉灿,等.迈进高维连续空间:深度强化学习在机器人领域中的应用[J].机器人,2019,41(2):276-288.
Duo Nanxun, Lv Qiang, Lin Huican, et al. Step into high-dimensional and continuous action space: a survey on applications of deep reinforcement learning to robotics[J]. Robot, 2019, 41 (2):276-288. (in Chinese)

[14] Paw Y C. Synthesis and validation of flight control for UAV [D]. University of Minnesota, Minneapolis and St. Paul,2009.

[15] MotoCalc. Homepage. [EB/OL]. Available: http://www.motocalc.com.

[16] 吴森堂,费玉华.飞行控制系统[M].北京:北京航空航天大学出版社,2005.
Wu Sentang, Fei Yuhua. Flight control system[M]. Beijing: Beihang University Press, 2005. (in Chinese)

[17] Rezaee A. Determining PID controller coefficients for the

moving motor of a welder robot using fuzzy logic[J]. Automatic Control and Computer Sciences, 2017, 51(2): 124-132.

[18] Silver D, Lever G, Heess N, et al. Deterministic policy gradient algorithms[C]// International Conference on International Conference on Machine Learning, 2014.

（责任编辑 皮卫东）

作者简介

宋欣屿(1997—)女，硕士研究生。主要研究方向：无人机建模与控制、深度强化学习。

Tel：13121239357

E-mail：songxinyu@buaa.edu.cn

王英勋(1964—)男，教授，博士生导师。主要研究方向：无人机的总体技术、自主飞行控制技术。

E-mail：wangyx@buaa.edu.cn

蔡志浩(1979—)男，高级工程师，硕士生导师。主要研究方向：创新布局飞行器系统综合设计、短距/垂直起降无人机建模与控制、无人机自主与协同控制。

E-mail：czh@buaa.edu.cn

赵江(1986—)男，讲师，硕士生导师。主要研究方向：飞行器轨迹优化与制导、飞行器协同控制与决策、智能控制理论与应用。

E-mail：jzhao@buaa.edu.cn

陈小龙(1988—)男，高级工程师。主要研究方向：先进飞行控制理论与应用。

E-mail：keylab@facri.com

宋栋梁(1988—)男，高级工程师。主要研究方向：先进飞行控制理论与应用。

E-mail：keylab@facri.com

Landing Trajectory Tracking Control of Unmanned Aerial Vehicle by Deep Reinforcement Learning

Song Xinyu[1,*], Wang Yingxun[1], Cai Zhihao[1], Zhao Jiang[1], Chen Xiaolong[2], Song Dongliang[2]

1. School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

2. National Key Laboratory of Science and Technology on Aircraft Control, Facri, Xi'an 710065, China

Abstract: Focusing on the problem of autonomous landing control of fixed-wing UAVs, this paper proposes a tracking control method for UAV landing trajectory based on Deep Reinforcement Learning (DRL). First, we built a simulation model of the small fixed-wing UAV Ultra Stick 25E and designed a landing reference trajectory that satisfies the process and terminal constraints. Second, we proposed a UAV-integrated control framework based on Deep Deterministic Policy Gradient (DDPG) and designed a reward function considering tracking error and trajectory stability. Finally, through the offline training, we obtained the trajectory tracking integrated controller. The simulation results show that the proposed method is more accurate than the traditional PID control method.

Key Words: fixed-wing UAV; autonomous landing; trajectory tracking control; DRL; DDPG

Received: 2019-10-30； **Revised:** 2019-11-05； **Accepted:** 2019-12-05

Foundation item: Aeronautical Science Foundation of China(20175851032)

***Corresponding author.Tel.:** 13121239357 **E-mail:** songxinyu@buaa.edu.cn