基于强化学习补偿的地面无人战车行进间跟瞄自适应控制

Wei, Lianzhen; Gong, Jianwei; Chen, Huiyan; Li, Zirui; Gong, Cheng

DOI
10.12382/bgxb.2021.0786

Publication date
2022

Document Version
Final published version

Published in
Binggong Xuebao/Acta Armamentarii

基于强化学习补偿的地面无人战车行进间跟瞄自适应控制

魏连震[1,2], 龚建伟[1], 陈慧岩[1], 李子睿[1,3], 龚乘[1]
(1. 北京理工大学 机械与车辆学院, 北京 100081; 2. 北京理工大学 长三角研究院, 浙江 嘉兴 314019;
3. 代尔夫特理工大学 交通与规划系, 荷兰 代尔夫特 2628 CN)

摘要: 针对底盘运动和路面起伏对地面无人战车行进间跟瞄带来的非线性干扰问题,提出一种基于强化学习补偿的地面无人战车行进间跟瞄自适应控制方法。该跟瞄控制方法由主控制器与补偿控制器两部分构成,主控制器利用 PID 控制算法结合当前跟瞄误差得到主控制量,补偿控制器利用 Dueling Q 网络强化学习算法对战车当前状态和局部规划路径附近的路面起伏信息进行处理得到补偿控制量。建立地面无人战车一体化运动学模型,对基于强化学习的补偿控制算法进行阐述;基于 V-REP 动力学软件在三维场景中进行仿真验证。实验结果表明:基于强化学习补偿的跟瞄控制方法对底盘运动和路面起伏具备较好的自适应能力,有效地提升了无人战车行进间跟瞄的准确性与稳定性。

关键词: 地面无人战车; 行进间跟瞄; 强化学习; 自适应控制; 补偿控制
中图分类号: TJ810.2 **文献标志码**: A **文章编号**: 1000-1093(2022)08-1947-09
DOI: 10.12382/bgxb.2021.0786

Tracking and Aiming Adaptive Control for Unmanned Combat Ground Vehicle on the Move Based on Reinforcement Learning Compensation

WEI Lianzhen[1,2], GONG Jianwei[1], CHEN Huiyan[1], LI Zirui[1,3], GONG Cheng[1]
(1. School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China;
2. Yangtze Delta Region Academy, Beijing Institute of Technology, Jiaxing 314019, Zhejiang, China;
3. Department of Transport and Planning, Delft University of Technology, Delft 2628 CN, The Netherlands)

Abstract: To deal with the nonlinear interference caused by chassis movement and road surface undulations with the tracking and aiming of unmanned combat ground vehicles, a tracking and aiming adaptive control method for unmanned combat ground vehicles on the move based on reinforcement learning compensation is proposed. This method consists of a main controller and a compensation controller. The main controller uses the PID control algorithm combined with the current tracking error to obtain the main control quantity, and the compensation controller uses the Dueling DQN reinforcement learning network to process the current state of the combat vehicle as well as the road surface undulation

收稿日期: 2021-11-18
基金项目: 武器装备预先研究项目(301060701)
作者简介: 魏连震(1998—), 男, 硕士研究生。E-mail:3120200396@bit.edu.cn
通信作者: 龚建伟(1969—), 男, 教授, 博士生导师。E-mail:gongjianwei@bit.edu.cn

information near the local planning path to obtain the compensation control quantity. Firstly, the integrated kinematics model of the unmanned combat ground vehicle is established. Then, the compensation control algorithm based on reinforcement learning is described. Finally, simulation and verification are performed in three-dimensional scenes based on the V-REP dynamic software. The experimental results show that the tracking and aiming control method based on reinforcement learning compensation has good adaptive ability for chassis movement and road surface undulations, which effectively improves the tracking/aiming accuracy and stability of unmanned combat vehicles.

Keywords: unmanned combat ground vehicle; tracking and aiming on the move; reinforcement learning; adaptive control; compensation control

0 引言

现代局部战争的实践反复证明,高新技术已经成为现代战争的制胜因素。随着自主智能、网络协同、云处理等高新技术的发展,作战模式正在发生重要转变,以地面无人战车为代表的无人作战系统能够执行多种特殊任务,是应对未来不确定形势的重要突破口,具有广泛的应用前景[1]。

在执行打击任务时,地面无人战车通常可采取静态射击与行进间射击两种作战方式。相比静态射击的作战方式,行进间射击能够缩短任务完成时间以提升作战效率,降低被反装甲武器命中的概率从而提升战场生存能力,是地面无人战车未来发展的重要方向[2]。行进间射击的关键技术之一是跟瞄镜对目标准确、稳定地跟瞄。现代坦克主流采用稳像式火控系统:火炮与瞄准镜分别稳定,瞄准镜对目标实时跟瞄并调动火炮,火控计算机根据跟瞄角速度、目标距离、炮弹弹种、风速等值计算射击诸元以实现射击[3]。然而,无论跟瞄系统处于稳像状态还是自动跟踪状态,底盘运动和路面起伏都会对瞄准带来平移误差,这给跟瞄控制系统带来了挑战[4]。

为提升战车行进间跟瞄的准确性与稳定性,不同研究人员提出了各自的技术方案。如钟洲等[5]建立了车载防空导弹的行进和发射一体化多柔性体动力学模型,并分析了路面和车速对防空导弹行进间发射精度的影响,但仅重点关注动力学模型的创建与分析,并未给出合适的控制方法。慕巍等[6]利用光电跟踪仪、火炮、载体惯导系统、视频跟踪器和激光测距机输出的相关参数,完成瞄准线坐标系下方位速度环和俯仰速度环跟踪前馈补偿参数的计算,以提升对高速目标跟瞄控制的准确性。熊珍凯等[7]针对机动快速目标的跟踪问题,采用基于当前统计模型的改进卡尔曼滤波算法预测出目标运动状态参数,并采用自适应滑模的解算控制方法,实现伺服系统的位置控制,提升跟瞄精度。这些方法没有涉及本车运动状态的分析,在动对静、动对动场景受限。郝强等[4]采集目标距离、火炮相对车体角度和车体速度等信息,循环解算瞄准线的补偿角速度,减小了跟瞄误差。但是,该方法仅考虑底盘速度影响,忽略了路面起伏影响,在地形复杂的越野场景中跟瞄补偿的效果不佳。张卫民等[8]以自行火炮与敌遭遇时紧急直瞄场景为研究对象,提出一种自行火炮自动直瞄控制方法,以提高火炮直瞄时快速反应能力和射击精度。然而,该方法侧重于瞄准的快速性,没有充分考虑各种非线性干扰对瞄准稳定性的影响。朱斌等[9]考虑系统内部扰动和外部扰动对稳瞄系统速度跟踪精度的影响,提出了采用自抗扰的控制方案。不过,该方法侧重于稳定性,仍然没有有效消除底盘运动与路面起伏因素带来的瞄准线平移误差。

针对跟瞄控制存在的上述问题,本文从整车角度进行研究,提出一种基于强化学习补偿的地面无人战车行进间跟瞄自适应控制方法。将感知模块感知得到的地形信息与规划模块规划得到的未来轨迹传输至上装跟瞄控制模块,上装跟瞄控制模块利用Dueling深度Q网络(DQN)强化学习算法对这些信息处理后得到补偿控制量,以削弱底盘运动与路面起伏对跟瞄的影响,提升战车跟瞄的准确性与稳定性。首先建立地面无人战车一体化运动学模型,之后对补偿控制方法进行细节性描述,最后利用仿真实验证明方法的有效性。

1 系统模型

针对地面无人战车行进间跟瞄自适应控制问

题,提出问题场景模型、地面无人战车一体化运动学模型以及强化学习模型。

1.1　问题场景描述

地面无人战车行进间跟瞄平面示意如图 1 所示。无人战车接收上级指挥端下发的打击任务,从起点位置规划战车的运动轨迹,而后自主跟踪运动轨迹并且实时搜索打击目标,跟瞄系统对可疑目标识别并在自动跟踪状态对其瞄准。跟瞄控制的目标是迅速、准确、稳定地减小跟瞄镜与打击目标随动角度误差 θ[10]。

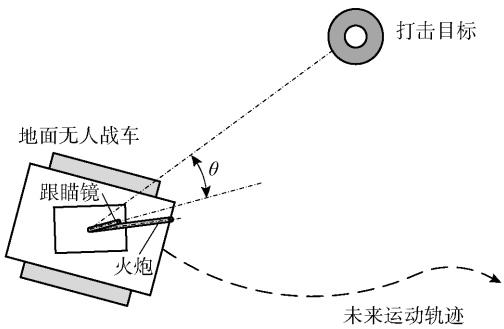


图 1　问题场景描述

Fig. 1　Problem scenario description

1.2　地面无人战车一体化运动学模型

地面无人战车采用履带式移动底盘,可通过调节左、右两侧主动轮的转速或转矩控制整车航向和速度。战车配备无人炮塔,其中升降式搜索镜用于识别周围可疑目标,跟瞄镜对搜索到的敌方目标实时跟瞄,火炮随动,而后火控计算机计算射击诸元,控制火炮在阈值内完成射击。考虑战车底盘的平移、俯仰、横摆、侧倾等会对上装跟瞄与打击模块产生影响,基于履带式无人车运动学模型[11],推导出右手坐标系的地面无人战车底盘与上装一体化运动学模型,如图 2 所示。

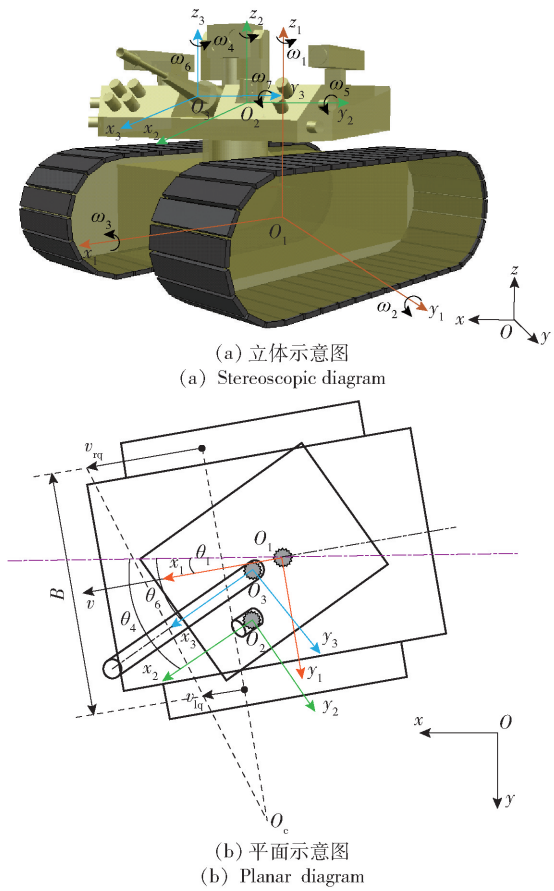


图 2　地面无人战车一体化运动学模型

Fig. 2　Integrated kinematics model of unmanned combat ground vehicle

图 2 中,$Oxyz$ 为世界坐标系,$O_1x_1y_1z_1$ 为底盘坐标系,$O_2x_2y_2z_2$ 为跟瞄坐标系,$O_3x_3y_3z_3$ 为火炮坐标系。如 2(a)中同时给出了可旋转方向,记 ω_1 代表底盘在世界坐标系中的横摆角速度,ω_2 代表底盘在世界坐标系中的俯仰角速度,ω_3 代表底盘在世界坐标系中的侧倾角速度,ω_4 代表跟瞄镜在底盘坐标系中的方位角速度,ω_5 代表跟瞄镜在底盘坐标系中的高低角速度,ω_6 代表火炮在底盘坐标系中的方位角速度,ω_7 代表火炮在底盘坐标系中的高低角速度。图 2(b)中 v_{lq}、v_{rq} 分别为左、右两侧履带或驱动轮的牵连速度,θ_1 为底盘在世界坐标系中的横摆角,θ_4 为跟瞄镜在世界坐标系中的方位角,θ_6 为火炮在世界坐标系中的方位角,B 为战车底盘履带中心距,O_c 为底盘瞬时转向中心,v 为底盘运动速度。

由于差速转向战车在转向时,两侧履带或驱动轮不可避免地会发生滑移滑转[12],定义左右两侧的滑移滑转系数分别为

$$f_l=\frac{v_{lq}-v_{ls}}{v_{lq}},f_r=\frac{v_{rq}-v_{rs}}{v_{rq}} \tag{1}$$

式中:v_{ls}、v_{rs} 分别为左、右两侧履带或驱动轮相对于车体的卷绕纵向线速度。考虑到滑转滑移,底盘的运动速度、横摆角速度分别为

$$v=\frac{v_{lq}+v_{rq}}{2} \tag{2}$$

$$\omega_1=\frac{v_{rq}-v_{lq}}{B}=\frac{v_{rs}/(1-f_r)-v_{ls}/(1-f_l)}{B} \tag{3}$$

由上述定义与推导,可得地面无人战车的数学模型为

$$\begin{bmatrix}\dot{x}\\ \dot{y}\\ \dot{\theta}_1\\ \dot{\theta}_2\\ \dot{\theta}_3\\ \dot{\theta}_4\\ \dot{\theta}_5\\ \dot{\theta}_6\\ \dot{\theta}_7\end{bmatrix}=\begin{bmatrix}0 & 0 & 0 & 0&0&0&0 & \cos\theta_1\\ 0 & 0 & 0 & 0&0&0&0 & \sin\theta_1\\ 1 & 0 & 0 & 0&0&0&0 & 0\\ 0 & 1 & 0 & 0&0&0&0 & 0\\ 0 & 0 & 1 & 0&0&0&0 & 0\\ 1 & 0 & 0 & 1&0&0&0 & 0\\ 0 & \cos(\theta_4-\theta_1) & \sin(\theta_4-\theta_1) & 0&1&0&0 & 0\\ 1 & 0 & 0 & 0&0&1&0 & 0\\ 0 & \cos(\theta_6-\theta_1) & \sin(\theta_6-\theta_1) & 0&0&0&1 & 0\end{bmatrix}\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\\ \omega_4\\ \omega_5\\ \omega_6\\ \omega_7\\ v\end{bmatrix} \tag{4}$$

式中：θ_2、θ_3、θ_5、θ_7 分别为底盘在世界坐标系中的俯仰角、侧倾角、跟瞄镜在世界坐标系中的高低角以及火炮在世界坐标系中的高低角。

1.3 强化学习模型

强化学习是机器学习的一个重要分支，它模拟的是生物学中的行为主义，即自然界中的生物体在一定的正向或负向刺激下，通过不断学习形成一套应对刺激的策略，从而实现自身利益最大化[13]。强化学习任务通常利用马尔可夫决策过程（MDP）进行描述，它满足马尔可夫性质：系统下一时刻状态只与当前时刻状态有关，与过往时刻状态无关[14]。MDP 的基本组成是五元组$(\boldsymbol{S},A,P,\gamma,R)$，其中 $\boldsymbol{S}$ 为智能体在交互环境中的状态集，A 为智能体在交互环境中对应的动作集，P 为智能体的状态转移概率，γ 为奖励的折现因子，R 为智能体在交互环境中采取特定动作的回报奖励[13]。强化学习过程是智能体从初始状态开始，不断从动作集中选取动作进行状态的转移，之后利用奖赏函数对选取的动作进行评价从而更新参数直到累计奖励最大化的过程，核心思想是试错与学习，具体如图 3 所示。

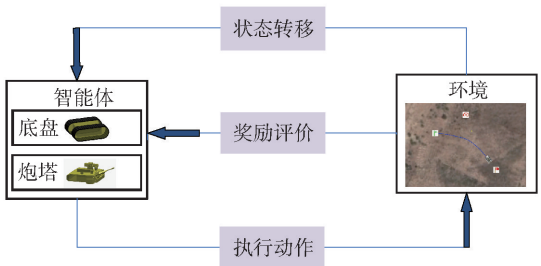


图 3 强化学习过程

Fig. 3 Process of reinforcement learning

强化学习主体框架包括智能体、环境、动作、奖励 4 个内容[13]。本文主要涉及地面无人战车跟瞄控制方法：由强化学习控制的智能体为地面无人战车的炮塔；环境指代的是战车周围态势；动作指代的是炮塔方位角控制量、炮塔高低角控制量；奖励指代的是人为设定的奖赏函数。通过奖赏函数的奖赏值引导智能体进行学习，下面阐述了强化学习模型的基本要素：

1）累积奖励。智能体每次执行动作后系统都会对该步操作进行评价，该评价值是单步奖励，累积奖励是智能体在一个回合之后所有动作单步奖励的折扣加权和，如（5）式所示：

$$G_t=R_{t+1}+\gamma R_{t+2}+\cdots=\sum_{k=0}^{\infty}\gamma^k R_{t+k+1} \tag{5}$$

式中：G_t 代表 t 时刻后开始的累积奖励；R_{t+1} 代表 $t+1$ 时刻的单步奖励。需要注意的是：累积奖赏实际上是一个随机变量，对它求期望可以得到价值函数。

2）策略。策略代表智能体在每种状态下执行某种动作的概率，是状态空间到动作空间的映射，如（6）式所示：

$$\pi(a|\boldsymbol{s})=P[A_t=a|\boldsymbol{S}_t=\boldsymbol{s}] \tag{6}$$

式中：$\pi(a|\boldsymbol{s})$ 为状态 $\boldsymbol{s}$ 时执行动作 a 的概率；A_t 为 t 时刻可选动作集；$\boldsymbol{S}_t$ 为 t 时刻状态集。

3）状态价值函数。为评价智能体所在状态的优劣，需获得智能体从当前状态转移到结束状态的累积奖励，在当前状态下按照一个固定策略求得的累积奖励期望是状态价值函数，如（7）式所示：

$$v_\pi(\boldsymbol{s})=\mathrm{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}|\boldsymbol{S}_t=\boldsymbol{s}\right] \tag{7}$$

4）动作价值函数。在当前状态下执行某个动作后按照某固定策略求得的累积奖励期望即是动作价值函数，如（8）式所示：

$$q_\pi(\boldsymbol{s},a)=\mathrm{E}_\pi\left[\sum_{k=0}^{\infty}\gamma^k R_{t+k+1}|\boldsymbol{S}_t=\boldsymbol{s}|A_t=a\right] \tag{8}$$

5）贝尔曼方程。贝尔曼方程是将多层决策转化为多个决策的动态规划过程，根据迭代公式求解状态价值函数与动作价值函数，状态价值函数与动作价值函数对应的贝尔曼方程分别为

$$v_\pi(\boldsymbol{s})=\sum_{a\in A}\pi(a|\boldsymbol{s})\left(R_s^a+\gamma\sum_{s'\in S}P_{ss'}^a v_\pi(\boldsymbol{s}')\right) \tag{9}$$

$$q_\pi(\boldsymbol{s},a)=R_s^a+\gamma\sum_{\boldsymbol{s}'\in\boldsymbol{S}}P_{ss'}^a\sum_{a'\in A}\pi(a'|\boldsymbol{s}')q_\pi(\boldsymbol{s}',a') \tag{10}$$

式中：$P_{ss'}^a$为智能体在当前状态 $\boldsymbol{s}$ 时执行动作 a 后转移到下一状态 $\boldsymbol{s}'$的概率；a'为在状态 $\boldsymbol{s}'$时采取的动作；R_s^a 代表智能体在当前状态 $\boldsymbol{s}$ 时执行动作 a 后获得的单步奖励。

2　控制方法

跟瞄控制问题的核心在于跟瞄系统能够快速、准确、稳定地对目标实时瞄准，其难点在于目标点运动、己方战车运动、路面起伏等因素带来的非线性干扰。针对此，本文提出一种基于强化学习补偿的地面无人战车跟瞄控制方法，以减小跟瞄误差，提升跟瞄性能。

控制方法架构如图 4 所示。PID 控制器根据当前跟瞄偏差得到主控制量；Dueling DQN 控制器将底盘局部规划路径点与目标的相对位置、局部规划路径点附近的起伏梯度、车辆运动速度、当前跟瞄误差等信息作为输入，利用神经网络处理得到补偿控制量；主控制量与补偿控制量加权之和为最终控制量，共包括方位控制量与高低控制量两个输出。主控制量保证跟瞄的大致方向性，补偿控制量用于对主控制量进行修正，从而提升地面无人战车行进间跟瞄对底盘速度变化以及路面起伏的自适应能力。需要说明的是：该控制方法得到的控制量是跟瞄系统下一时刻相对转动的角度增量，并非底层的转矩控制量。本文中强化学习算法的学习机制与网络结构能够针对复杂动态信息分析和处理，并且具备持续学习效果，随着训练次数的增多，跟瞄效果的准确性与稳定性可逐步提升[15]。图 4 中，e_1、e_2 分别为方位角度偏差值与高低角度偏差值，k_{p1}、k_{i1}、k_{d1}、k_{p2}、k_{i2}、k_{d2}分别为方位角和高低角对应的比例、积分、微分权重系数，u_{rot}是方位角增量，u_{pit}是高低角增量。

战车对目标的实时跟瞄偏差角度值可以由目标在跟瞄坐标系中位置求解得到，角度计算如(11)式[16]所示：

$$\begin{cases}e_1=\arctan\left(\dfrac{y_0-y}{x_0-x}\right)-\theta_4\\ e_2=\arctan\left(\dfrac{z_0-z}{\sqrt{(x_0-x)^2+(y_0-y)^2}}\right)-\theta_5\end{cases} \tag{11}$$

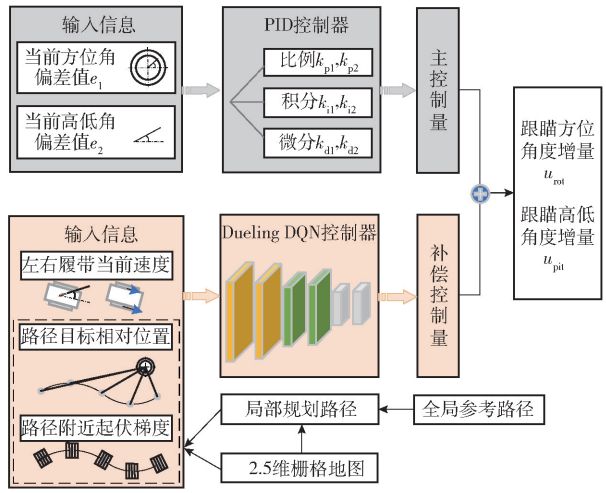


图 4　基于强化学习的补偿控制方法架构图

Fig. 4　Framework of compensation control method based on reinforcement Learning

式中：x_0、y_0、z_0 代表跟瞄目标在世界坐标系中坐标；x、y、z 代表车辆跟瞄镜在世界坐标系中坐标。

最终的控制量(当前控制时刻相对于上一控制时刻，其跟瞄方位角度增量与跟瞄高低角度增量)的数学表达如(12)式所示：

$$\begin{cases}u_{rot}=k_1u_1+k_{r1}u_{r1}\\ u_{pit}=k_2u_2+k_{r2}u_{r2}\\ u_1=k_{p1}e_1(t)+k_{i1}\displaystyle\int_0^T e_1(t)\,\mathrm{d}t+k_{d1}\frac{\mathrm{d}e_1(t)}{\mathrm{d}t}\\ u_2=k_{p2}e_2(t)+k_{i2}\displaystyle\int_0^T e_2(t)\,\mathrm{d}t+k_{d2}\frac{\mathrm{d}e_2(t)}{\mathrm{d}t}\\ u_{r1}=f_{r1}(\boldsymbol{s})\\ u_{r2}=f_{r2}(\boldsymbol{s})\end{cases} \tag{12}$$

式中：k_1、k_2 分别为方位角和高低角主控制量权重系数；u_1、u_2 分别为方位角和高低角主控制量；k_{r1}、k_{r2}分别为方位角和高低角补偿控制量权重系数；u_{r1}、u_{r2}分别为方位角和高低角补偿控制量；T 代表积分时间；$f_{r1}(\boldsymbol{s})$、$f_{r2}(\boldsymbol{s})$分别为强化学习神经网络拟合的方位角和高低角非线性函数。

本文采用的强化学习算法参考了 Dueling DQN 算法思路[17-18]，它属于值迭代算法的一种，是基于传统 DQN 算法的一种改进算法，如图 5 所示。图 5 中，$L(j)$代表第 j 条数据对应的误差值，M 代表一次性处理的数据条数。

图 5 中，估计网络与目标网络在网络结构上一致，区别在于估计网络实时更新参数，目标网络非实

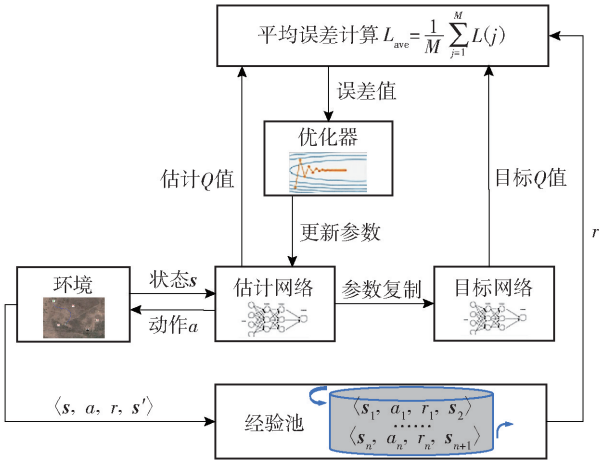


图 5 强化学习算法思路图

Fig. 5 Algorithm diagram of reinforcement learning

时更新,算法 Q 值计算如(13)式所示:

$$Q(\boldsymbol{s},a|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta})=V_Q(\boldsymbol{s}|\boldsymbol{w},\boldsymbol{\alpha})+\left[A_Q(\boldsymbol{s},a|\boldsymbol{w},\boldsymbol{\beta})-\frac{1}{C}\sum_{a'\in A}A_Q(\boldsymbol{s},a'|\boldsymbol{w},\boldsymbol{\beta})\right] \tag{13}$$

式中:$V_Q(\boldsymbol{s}|\boldsymbol{w},\boldsymbol{\alpha})$ 为状态值函数,用于衡量状态价值,仅与状态 $\boldsymbol{s}$ 有关,$\boldsymbol{w}$ 为公有网络参数,$\boldsymbol{\alpha}$ 为状态值函数特有网络参数;$A_Q(\boldsymbol{s},a|\boldsymbol{w},\boldsymbol{\beta})$ 是动作优势函数,用于衡量不同动作相对于所处状态的价值,同时与状态 $\boldsymbol{s}$ 以及动作 a 有关,$\boldsymbol{\beta}$ 是动作优势函数特有网络参数;C 为离散动作空间元素个数。

本文中使用的神经网络结构如图 6 所示,其中方位角度补偿控制网络与高低角度补偿控制网络类似,区别在于神经网络的输入信息、输出信息以及神经元个数。方位角度补偿控制网络的输入为底盘局部规划路径点与目标的相对位置、左右履带速度、方位跟瞄误差;高低角度补偿控制网络的输入为局部规划路径点附近的起伏梯度、左右履带速度、高低跟瞄误差。其中,路径附近起伏梯度指的是“一定数目的未来路径点以及对应的左右偏移路径点集合”前后相邻点之间高度差值构成的矩阵。输入信息先经过若干层全连接层,之后分为状态值网络以及动作值网络,最后得到每种动作对应的 Q 值。此外,本文对部分全连接层进行了 *Dropout* 处理,即在训练阶段随机将部分神经元丢弃从而削弱训练中的发生过拟合现象[19]。

程序训练过程:先随机探索一定步数以获得多组数据并将其存储在经验池中,每一次从经验池中抽出若干条数据并不断更新网络参数值,直至模型

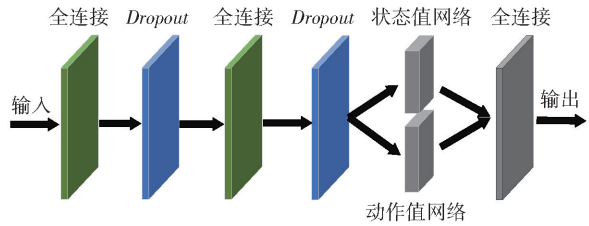


图 6 Dueing DQN 神经网络结构图

Fig. 6 Structure of Dueing DQN neural network

满足要求或训练次数达到阈值。Dueling DQN 算法是通过最小化时序差分误差实现网络更新,其数学表达如(14)式所示:

$$L=(R+\gamma\max_{a'}Q'(\boldsymbol{s}',a'|\boldsymbol{w}',\boldsymbol{\alpha}',\boldsymbol{\beta}')-Q(\boldsymbol{s},a|\boldsymbol{w},\boldsymbol{\alpha},\boldsymbol{\beta}))^2 \tag{14}$$

式中:Q'代表下一状态的目标 Q 值。因实际进行参数更新是同时对若干条数据进行处理,平均后的误差值如(15)式所示:

$$L_{ave}=\frac{1}{M}\sum_{j=1}^{M}L(j) \tag{15}$$

利用 TD 误差对网络参数的更新原理是借助梯度下降算法,本文在实验时采用了 Adam 优化器实现参数梯度下降,相比传统的随机梯度下降算法能够更快地实现参数收敛。

3 仿真实验与结果分析

3.1 V-REP 三维仿真实验设置

底盘运动是影响地面无人战车行进间跟瞄误差的一个重要非线性干扰,当速度大小或者速度方向发生变化时会对跟瞄的稳定性产生影响,即使战车保持匀速直线运动,也会对战车跟瞄带来瞄准线的平移[4]。路面起伏是影响地面无人战车行进间跟瞄误差的另一个重要非线性干扰因素。基于单独 PID 控制的跟瞄算法不能对战车未来阶段的起伏信息进行预判,这种被动跟随控制策略在起伏路面时跟瞄效果不佳;并且,由于路面起伏的复杂性,传统的前馈补偿方法难以针对性开展设计。本章基于 V-REP 动力学仿真软件进行强化学习网络参数训练与测试[20],通过观察训练过程中奖赏值的上升和对比单独 PID 控制方法与补偿控制方法跟瞄误差角数值来验证本文提出的补偿控制方法有效性,仿真实验流程如图 7 所示,仿真软硬件环境如表 1 所示。

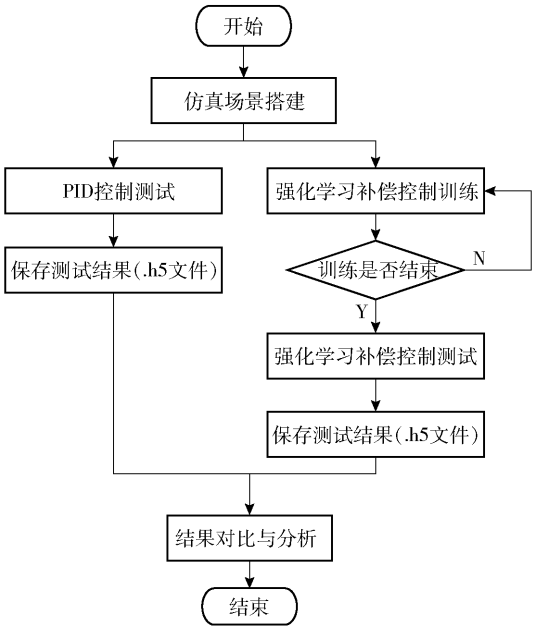


图 7　仿真实验流程图

Fig. 7　Flow chart of simulation

表 1　仿真软硬件环境

Table 1　Software and hardware of the simulation

参数	数值
操作系统	Ubuntu16. 04 + ROS Kinetic
图形处理器	NVIDIA GeForce GTX1650
内存/GB	8. 0
机器学习框架	TensorFlow
程序语言	Python 和 C + +
动力学仿真软件	V-REP EDU 3. 6. 2

为在 V-REP 动力学软件中搭建路面起伏环境，采用 Perlin 噪声算法构建近似于自然环境的起伏路面[21]，并将地形文件、车辆模型、打击目标导入 V-REP 仿真软件，再利用 ROS 接口实现与程序端的通信，最终完成起伏路面仿真环境搭建，如图 8 所示。仿真中设定车辆运动速度为 15 km/h，方位角速度阈值为 40°/s，高低角速度阈值为 40°/s。设计两个强化学习神经网络分别对方位角与高低角进行补偿控制，强化学习的基本信息如表 2 所示。

3. 2　实验结果分析

由表 2 可以看出，奖赏函数是关于目标跟瞄角误差值的二次函数，当误差角越小时对应的奖赏值越大，因此可通过观察训练过程中奖赏值变化分析跟瞄效果。图 9 绘制出了无人战车从起始位置自主运动到目标位置的前 500 次训练过程中高低角网络平均奖赏值的变化情况，为便于观察进行了均值滤波。由图 9 看出：随着训练次数地增多，平均奖赏值呈现整体上升的趋势，这代表 Dueling DQN 控制器对于跟瞄误差补偿效果随着训练增多而提升。

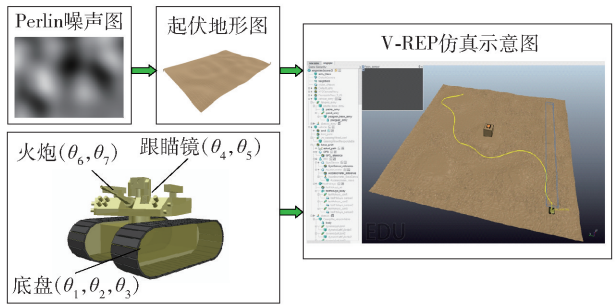


图 8　三维仿真环境搭建过程

Fig. 8　Construction process of 3D simulation environment

表 2　强化学习基本设置

Table 2　Basic settings of reinforcement learning

参数	数值
方位网络输入信息	方位跟瞄误差、履带速度、局部规划路径点与目标相对位置
高低网络输入信息	高低跟瞄误差、履带速度、局部规划路径点附近的起伏梯度
方位网络输出信息/(°)	离散补偿角度集合中的一个：-0. 3，-0. 2，-0. 1，-0. 05，0，0. 05，0. 1，0. 2，0. 3
高低网络输出信息/(°)	离散补偿角度集合中的一个：-0. 3、-0. 2、-0. 1、-0. 05、0、0. 05、0. 1、0. 2、0. 3
方位奖赏函数	$10-e_1^2$
高低奖赏函数	$10-e_2^2$

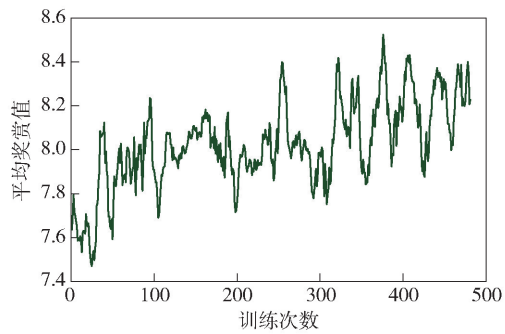


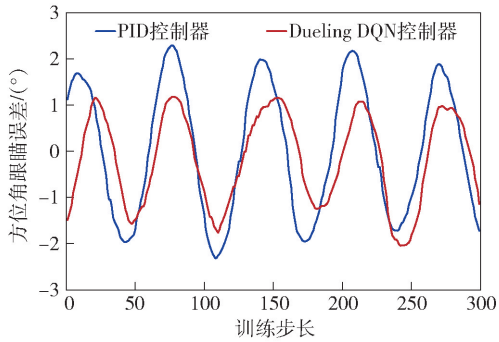
图 9　平均奖赏值变化图

Fig. 9　Variation diagram of average reward values

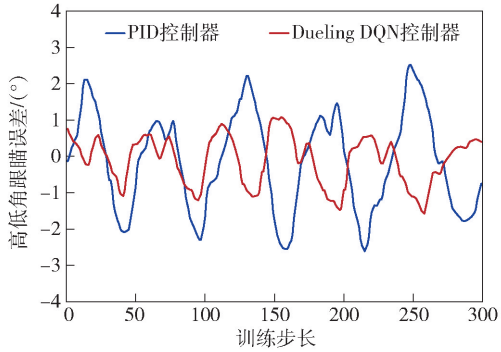
地面无人战车在从起点位置到终点位置的运行中，不同跟瞄控制方法对应的跟瞄角度误差均值能够反映控制效果的好坏。

将战车从跟瞄稳定位置到终点位置运动过程中

上装跟瞄角度误差的变化情况进行记录,并对比基于 PID 控制与强化学习补偿控制两种方法的跟瞄角度误差变化情况,对比结果如图 10 所示,其中图 10(a)为方位角度误差变化,图 10(b)为高低角度误差变化。由图 10 可知:基于强化学习补偿的控制方法平均跟瞄误差明显更小,控制效果更优。



(a) 方位角误差变化
(a) Variation of azimuth error



(b) 高低角误差变化
(b) Variation of elevation error

图 10 跟瞄角误差变化图

Fig. 10 Variation diagram of tracking/aiming error

4 结论

本文提出一种基于强化学习补偿的地面无人战车行进间跟瞄自适应控制方法,有效地提升了地面无人战车的动态作战性能。首先建立地面无人战车一体化运动学模型以及强化学习模型,然后具体介绍了基于强化学习补偿的跟瞄控制方法架构,最后基于 V-REP 动力学仿真软件进行了控制方法效果对比,得出结论:强化学习补偿能够较好地削弱速底盘运动以及路面起伏对上装跟瞄的非线性干扰。不过,目前的工作仍是初步的:1)在跟瞄系统建模方面采用了简单运动学模型,后续会针对该模型进行完善并深入分析底盘运动与路面起伏对跟瞄性能的影响特性;2)后续将补充开展与上装载荷任务相协同的底盘运动规划研究。

参考文献(References)

[1] 陈慧岩, 张玉. 军用地面无人机动平台技术发展综述[J]. 兵工学报, 2014, 35(10): 1696 – 1706.
CHEN H Y, ZHANG Y. An overview of research on military unmanned ground vehicles [J]. Acta Armamentarii, 2014, 35(10): 1696 – 1706. (in Chinese)

[2] 陈建荣, 郭齐胜, 刘军. 地面运动目标攻击命中概率模型及仿真[J]. 火力与指挥控制, 2007, 32(7): 43 – 46.
CHEN J R, GUO Q S, LIU J. Hit probability model of attacking the mobile ground target and simulation [J]. Fire Control and Command Control, 2007, 32(7): 43 – 46. (in Chinese)

[3] 罗来科, 王耀北, 王顺岭. 稳像火控系统误差分析[J]. 火力与指挥控制, 2002, 27(5): 30 – 32, 35.
LUO L K, WANG Y B, WANG S L. Error analysis of the image-stabilized fire control system [J]. Fire Control and Command Control, 2002, 27(5): 30 – 32, 35. (in Chinese)

[4] 郝强, 南立军, 刘斌, 等. 坦克火控系统瞄准线平移的补偿方法[J]. 火炮发射与控制学报, 2018, 39(3): 71 – 75.
HAO Q, NAN L J, LIU B, et al. Compensation method of aiming line translation of tank fire control system [J]. Journal of Gun Launch & Control, 2018, 39(3): 71 – 75. (in Chinese)

[5] 钟洲, 姜毅, 刘群. 车载防空导弹行进间发射过程动力学数值分析[J]. 兵工学报, 2014, 35(1): 83 – 87.
ZHONG Z, JIANG Y, LIU Q. Dynamics numerical analysis of vehicle-mounted antiaircraft missile launching on the move [J]. Acta Armamentarii, 2014, 35(1): 83 – 87. (in Chinese)

[6] 慕巍, 张宝宜, 王新明, 等. 适用于光电跟踪仪的高速目标跟踪控制算法[J]. 激光与红外, 2020, 50(4): 468 – 474.
MU W, ZHANG B Y, WANG X M, et al. High speed target tracking control algorithm fbr electro-optical tracker[J]. Laser & Infrared, 2020, 50(4): 468 – 474. (in Chinese)

[7] 熊珍凯, 陈汀峰. 精确跟瞄控制技术研究[J]. 强激光与粒子束, 2012, 24(6): 1339 – 1343.
XIONG Z K, CHEN T F. High precision tracking and pointing control technique [J]. High Power Laser and Particle Beams, 2012, 24(6): 1339 – 1343. (in Chinese)

[8] 张卫民, 梁建奇, 马红卫, 等. 自行火炮自动直瞄控制方法研究[J]. 兵工学报, 2015, 36(1): 182 – 186.
ZHANG W M, LIANG J Q, MA H W, et al. An automatic direct aiming control method of self-propelled artillery [J]. Acta Armamentarii, 2015, 36(1): 182 – 186. (in Chinese)

[9] 朱斌, 谢杰, 孙皓泽, 等. 某新型坦克稳瞄系统自抗扰控制器的设计[J]. 计算机工程与应用, 2013, 49(增刊3): 71 – 75.
ZHU B, XIE J, SUN H Z, et al. Design of active disturbance rejection controller for some new type tank steady sighting system [J]. Computer Engineering and Applications, 2013, 49(S3): 71 – 75. (in Chinese)

[10] 张文丽, 郭俊文, 曲俊海, 等. 基于自适应差分进化算法的武器稳定系统参数辨识[J]. 火力与指挥控制, 2020,

45(5): 119 – 124.
ZHANG W L, GUO J W, QU J H, et al. Parameter identification of weapon stability system based on adaptive differential evolution algorithm[J]. Fire Control and Command Control, 2020, 45(5): 119 – 124. (in Chinese)

[11] 鲁浩. 基于瞬时转向中心实时估计的滑动转向车辆运动轨迹预测方法研究[D]. 北京:北京理工大学, 2016.
LU H. Trajectory prediction based on estimation of instantaneous centers of rotation in real time for skid-steer vehicles[D]. Beijing: Beijing Institute of Technology, 2016. (in Chinese)

[12] 盖江涛, 刘春生, 马长军, 等. 考虑履带滑转滑移的电驱动履带车辆转向控制[J]. 兵工学报, 2021, 42(10): 2092 – 2101.
GAI J T, LIU C S, MA C J, et al. Steering control of electric drive tracked vehicle considering tracks' skid and slip[J]. Acta Armamentarii, 2021, 42(10): 2092 – 2101. (in Chinese)

[13] 李壮. 基于深度强化学习的六自由度机械臂避障规划[D]. 北京理工大学, 2020.
LI Z. Obstacle avoidance planning of six degrees of freedom manipulator based on deep reinforcement learning[D]. Beijing: Beijing Institute of Technology, 2020. (in Chinese)

[14] 冷鹏飞, 徐朝阳. 一种深度强化学习的雷达辐射源个体识别方法[J]. 兵工学报, 2018, 39(12): 2420 – 2426.
LENG P F, XU C Y. Specific emitter identification based on deep reinforcement learning[J]. Acta Armamentarii, 2018, 39(12): 2420 – 2426. (in Chinese)

[15] LIU B Y, WANG L J, LIU M. Lifelong federated reinforcement learning: a learning architecture for navigation in cloud robotic systems[J]. IEEE Robotics Automation Letters, 2019, 4(4): 4555 – 4562.

[16] GONG C, LI Z, ZHOU X, et al. Orientation-aware planning for parallel task execution of omni-directional mobile robot[C]// Proceedings of 2021 International Conference on Intelligent Robots and Systems. Prague, Czech: IEEE/RSJ, 2021: 6891 – 6898.

[17] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529 – 533.

[18] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]// Proceedings of the 33rd International Conference on Machine Learning. New York, NY, US: PMLR, 2016: 1995 – 2003.

[19] DADAY M J A, MILLADO K F M R. Enhanced reinforcement learning with targeted dropout[C]// Proceedings of 2019 International Conference on Digitization (ICD). Sharjah, Emirate: IEEE, 2019: 207 – 211.

[20] ROHMER E, SINGH S P N, FREESE M. V-REP: A versatile and scalable robot simulation framework[C]// Proceedings of 2013 International Conference on Intelligent Robots and Systems (IROS). Tokyo, Japan: IEEE/RSJ, 2013: 1321 – 1326.

[21] PERLIN K. Improving noise[J]. ACM Transactions on Graphics, 2002, 21(3): 681 – 682.