2021 NDIA GROUND VEHICLE SYSTEM ENGINEERING AND TECHNOLOGY SYMPOSIUM
AUTONOMY, ARTIFICIAL INTELLIGENCE & ROBOTICS (AAIR) TECHNICAL SESSION
AUGUST 10-12, 2021 - NOVI, MICHIGAN

DEEP REINFORCEMENT LEARNING FOR SIMULTANEOUS PATH PLANNING AND STABILIZATION OF OFFROAD VEHICLES

Ameya Salvi[1], Jake Buzhardt[2], Phanindra Tallapragada[2], Venkat Krovi[1], Mark Brudnak[3], Jonathon M. Smereka[3]

[1]Automotive Engineering, Clemson University, Clemson, SC
[2]Mechanical Engineering, Clemson University, Clemson, SC
[3]US Army CCDC, Ground Vehicles Systems Center (GVSC), Warren, MI

ABSTRACT

Motion planning algorithms for vehicles in an offroad environment have to contend with the significant vertical motion induced by the uneven terrain. Besides the obvious problems related to driver comfort, for autonomous vehicles, such "bumpy" vertical motion can induce significant mechanical noise in the real time data acquired from onboard sensors such as cameras to the point that perception becomes especially challenging. This paper advances a framework to address the problem of vertical motion in offroad autonomous motion control for vehicular systems. This framework is first developed to demonstrate the stabilization of the sprung mass in a modified quarter-car tracking a desired velocity while traversing a terrain with changing height. Even for an idealized model such as the quarter-car the dynamics turn out to be nonlinear and a model-based controller is not obvious. We therefore formulate this control problem as a Markov decision process and solve it using deep reinforcement learning. The control inputs that are learned are the torque on the wheel and the stiffness of the active suspension. It is demonstrated here that a time-varying velocity can be tracked with reduced chassis oscillations using these control inputs. We anticipate that reducing such oscillations will lead to sensor stabilization, which will improve perception and reduce the required frequency of recalibration. The deep reinforcement learning approach advanced in this paper remains useful for offroad motion planning when complex terramechanics and uncertain model parameters are introduced or the vehicle model increases in complexity.

Citation: A. Salvi, J. Buzhardt, P. Tallapragada, V. Krovi, M. Brudnak, J. M. Smereka, "Deep reinforcement learning for simultaneous path planning and stabilization of offroad vehicles", In *Proceedings of the Ground Vehicle Systems Engineering and Technology Symposium* (GVSETS), NDIA, Novi, MI, Aug. 10-12, 2021.

1. INTRODUCTION

Traditional path planning algorithms like A*, Dijkstra's algorithm, and RRT provide a real time solution to complex path planning problems by leveraging high performance computing [1]. While doing so, such algorithms rarely consider the detailed dynamics of the vehicle or the impact those dynamics might have in executing the plan from a global, non-reactive path planner. At present, global motion planning algorithms work siloed, with low level controllers being used for vehicle stabilization or disturbance rejection.

Reinforcement learning is a powerful technique which can incorporate the low-level dynamics of a physical system without its explicit knowledge [2, 3, 4]. Having a trained reinforcement learning agent working in tandem with the high-level motion planner eliminates the need for a traditional onboard controller. Such a framework increases the system's robustness to unknown environment dynamics in comparison to traditional model-based methods, which rely on the accuracy of the assumed model [2, 5]. A deep reinforcement learning agent can be used in conjunction with a high-level path planner to reduce the complexity of the deep reinforcement learning problem and training [6]. Here, we assume that a path is given from a global planner, thus reducing the motion planning problem to control of longitudinal vehicle velocity with simultaneous stabilization of the chassis. Reinforcement learning for control of mobile robots designed specifically for continuous action and observation space is an area much to be explored [7]. In this paper we propose a reinforcement learning agent which considers the vertical dynamics of a vehicle while traversing an off-road terrain setting. Most traditional algorithms do not consider vertical vehicle dynamics with high-level path planning in a closed loop, likely due to the increased computational cost [8]. This is significant, as the vertical dynamic response of a vehicle to the off-road terrain can make an optimal path become suboptimal or even dangerous at high speeds.

The general reinforcement learning framework requires the formulation of the problem as a Markov decision process [9] in order to model the available vehicle movement options. Here the state transition dynamics are derived from a reduced order model of a military vehicle on uneven terrain. While the bicycle model has become a standard model in many vehicle applications [10, 11], here we expand the focus to the vertical dynamics of the vehicle, by considering instead a reduced longitudinal dynamics model, but including suspension forces, the normal reaction at the wheel, and vertical oscillations of the vehicle chassis. These dynamics are modeled as a single quarter-car suspension with linear stiffness and damping between the wheel and the sprung mass representing the chassis. This allows for the consideration of resistance and reaction forces at the ground and in the suspension, which can become significant when applications require aggressive or agile maneuvers of the vehicle over rough terrain.

With this in mind, we develop a reinforcement learning formulation with an objective function that rewards stabilization of the sprung mass and tracking of a desired velocity. In addition, constraints are also imposed on the vehicle velocities and accelerations, the input torque, and the normal reaction at the wheel to prevent loss of contact. The control actions are taken to be the torque applied at the wheel and the stiffness of the suspension. We use numerical simulation to demonstrate that this proposed model, when coupled with the previously described reinforcement learning framework can yield an effective control strategy on uneven terrains. The observations are assumed to be the full state of the quarter car and preview of the terrain elevation over a look-ahead horizon. Such a preview is crucial, as it allows for disturbance rejection while accounting for actuator delay associated with changes to suspension stiffness.

The rest of the paper is organized as follows: Section 2 discusses the derivation of the vehicle

dynamics model and terrain interactions. Section 3 discusses the formulation of the velocity tracking and stabilization problem within the deep reinforcement learning framework. Section 4 presents numerical results for the trained reinforcement learning agent navigating the vehicle on an uneven terrain while tracking a time-varying velocity. Section 5 concludes the paper with discussion of the results and avenues for future work.

2. MODELING

2.1. Vehicle Model

A full, high-fidelity vehicle model that considers the complete dynamics of an offroad vehicle would be too computationally expensive to use in controller design. Instead a model is implemented that captures the most significant aspects of the vehicle dynamics. The model considered here is a quarter car suspension for the vertical dynamics coupled with a model for the longitudinal motion of the vehicle due to a torque produced at the wheel. A schematic of this model and the coordinates employed is shown in Fig. 1.

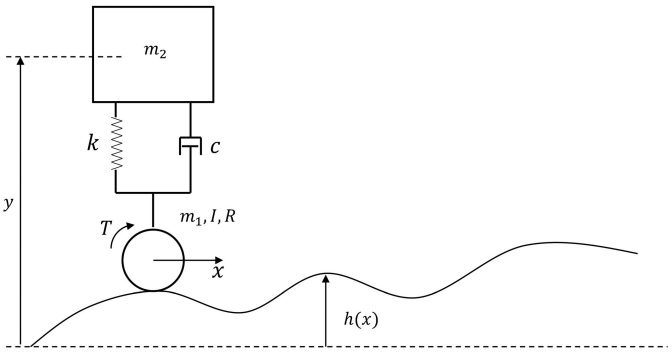


Figure 1: Schematic of quarter-car vehicle model

The equations of motion in terms of the generalized coordinates depicted in Fig. 1 are given below. These are derived by considering a sum of forces acting on each of the masses.

$$\left(m_1 + m_2 + \frac{I}{R^2}\right)\ddot{x} - \frac{T}{R} + N\frac{\partial h}{\partial x} + F_f = 0 \quad (1)$$

$$m_2\,\ddot{y} + k\,(y - h(x) - L_0) + c\left(\dot{y} - \frac{\partial h}{\partial x}\dot{x}\right) + m_2 g = 0 \quad (2)$$

Eq. (1) describes the longitudinal motion of the vehicle in terms of the coordinate x, defined as of the longitudinal displacement of the wheel. Here, the parameters m_1, I, and R are the mass, mass moment of inertia, and radius of the wheel, respectively; m_2 is the one quarter of the mass of the vehicle chassis; and T is the torque applied at the wheel. For this effort, the terrain elevation is assumed to be given by a smooth function $h(x)$, whose first and second derivatives with respect to x are also known.

Eq. (2) describes the vertical motion of the vehicle chassis in terms of the coordinate, y, which measures the absolute displacement of the upper mass from a fixed datum. The coefficients k and c are stiffness and damping coefficients associated with the suspension, and the $m_2\,g$ term represents the force due to gravity.

There are three forcing terms in Eq. (1): one due to the applied torque, T/R, one due to a dissipative friction F_f at the wheel/soil interface, and one due to a projection of the normal force due to the gradient of the road elevation $N\partial h/\partial x$. The normal force, N is found by considering the vertical dynamics of the two masses and by assuming that the wheel remains in contact with the ground. Mathematically, this can be written as a constraint on the vertical position of the wheel, y_1:

$$y_1 = h(x) + R \quad (3)$$

With this constraint, the normal force is given by

$$\begin{aligned} N = m_1 &\left(\frac{\partial^2 h}{\partial x^2}\dot{x}^2 + \frac{\partial h}{\partial x}\ddot{x} + g\right) \\ &- c\left(y - \frac{\partial h}{\partial x}\dot{x}\right) - k\,(y - h(x) - L_0) \end{aligned} \quad (4)$$

The dissipative forcing F_f is taken to account for rolling resistance, air drag, and friction forces at the wheel, all which resist the vehicle's longitudinal motion. This force is parameterized in terms of the

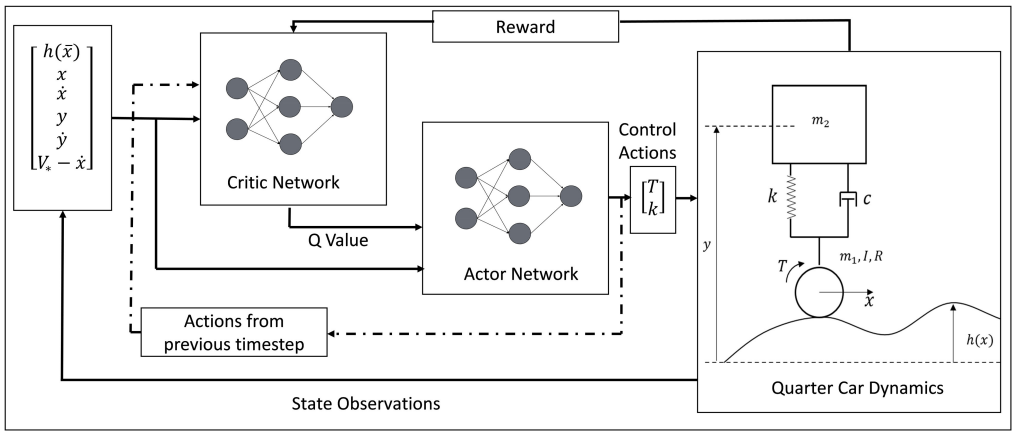


Figure 2: Interfacing the actor critic networks with system dynamics

the forward velocity of the vehicle, $\dot{x}$ as

$$F_f = -C_{v1}\dot{x} - C_{v2}\,\mathrm{sign}\,(\dot{x})\dot{x}^2 \tag{5}$$

where C_{v_1} and C_{v_2} are positive constants.

The values used here for simulation and training of the reinforcement learning agent are chosen to be representative of the parameter values for a full-scale offroad vehicle. The parameter values used here are summarized in Table 1.

Table 1: Parameter descriptions and nominal values

Parameter	Description	Value
m_1	wheel mass	75 kg
m_2	$1/4$ chassis mass	300 kg
R	wheel radius	0.3 m
I	wheel moment of inertia	3.375 kg·m^2
k	suspension stiffness	variable
c	suspension damping	10^3 N·s/m
L_0	undeformed spring length	0.5 m
C_{v_1}	friction coefficient	5 N·s/m
C_{v_2}	friction coefficient	0.6 N· s^2/m^2

3. REINFORCEMENT LEARNING FORMULATION

The success of any reinforcement learning problem is determined by the choice of the agent and strategic shaping of the reward function. Depending on the complexity of the control challenge, the construction of the reinforcement learning environment can also have an impact on both the training time and the controller performance. For instance, the environment may take form of a physical robot interacting in the real world, or a high fidelity simulation engine or a reduced order model of the vehicle's dynamics, with each platform having its own set of challenges. Fig. 2. gives a high level overview of the observations and actions involved and how the agent is interfaced with the environment.

The development of the three subsections - the agent, the reward function and the environment have been discussed below.

3.1. Agent

The choice of the agent is largely dependent on the application of the controller [12]. For developing reinforcement learning based controllers for robotics applications or any other controls applications with high non-linearities in the system's physics, it is useful to use neural networks as universal function approximators for the value function approximations. For the specific problem of velocity tracking and sprung mass oscillation damping, it is essential to have a continuous observation space and action space for fine controller resolution. Both the criteria of using neural network as function approximator and

having a continuous action and observation space are satisfied by the Deep Deterministic Policy Gradient algorithm (DDPG) [13, 14]. The basis of the DDPG agent is formed by two neural networks termed as the actor network and critic network. The observations received from the environment are the quarter-car's states for velocities and displacements along horizontal and vertical axes for the sprung mass (x, $\dot{x}$, y, $\dot{y}$) and terrain preview ($h(x)$). These coupled with the actions – wheel torque and suspension stiffness (T, k) serve as an input to the critic network. The critic network generates and stores an expected long term return value known as the Q-value for the input state-action pair. The actions for the next time step are then generated by the actor and are based on observations and the Q-values provided by the critic from the past experience.

3.2. Reward Function

The weights of both the actor and critic networks are updated based on the current reward and past experiences randomly sampled from experience buffer [14]. As a result, a poorly shaped reward function could incentivize negative actions or delay the overall training process [15]. For the current problem, the horizontal velocity to be tracked and the vertical velocity of the sprung mass parameterize the reward function. The trade-off between these two penalty terms to achieve maximum reward is a classic optimization routine for which the agent is trained to provide a solution. Eq.(6) describes the reward function where desired velocity was denoted as V_* and W_1, W_2 were the assigned weights for the function parameters.

$$Reward = -W_1(V_* - \dot{x})^2 - W_2\,\dot{y}^2 \qquad (6)$$

The choice of the weights does not come naturally but is a systematic trade-off depending on which parameter is desired to be tracked more accurately. In this work, the values of W_1 and W_2 are chosen heuristically to balance the error in longitudinal and vertical velocities, relative to their respective nominal values. For example, if both objectives, velocity tracking and oscillation damping, are desired to be tracked with equal importance, the difference in their error magnitudes needs to be taken in account. The velocity tracking error in real time usually has an error magnitude of 10 where as the oscillations are of the magnitude of 10^{-1} or 10^{-2}. Specifying equal weight in such a scenario will give conversely give more importance to velocity tracking and less to oscillations damping.

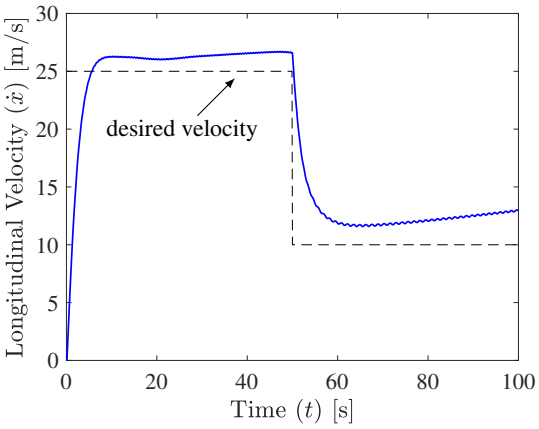
3.3. Environment

The reinforcement learning environment interfaces with the agent and provides the observations as a feedback for the actions commanded by the agent. In this project, the environment is defined by the combined quarter-car dynamics and the terrain model. The torque and stiffness updates provided by the DDPG agent are updated in real time in the quarter-car's dynamic equations. The equations of motion are then simulated for a time-step and the updated vehicle states are sent back to the agent from the environment. The observations received by the DDPG agent are a combination of the vehicle states, terrain preview and the tracking error. The terrain preview has been constructed as values of the terrain elevation function $h(x)$ sampled over some finite look-ahead distance.

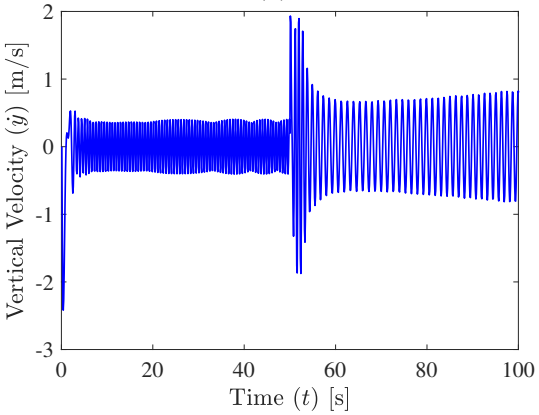
The simulation framework has been setup using the MATLAB-Simulink Reinforcement Learning toolbox [16].

4. RESULTS

In this section, we present the results of numerical simulation in which the reinforcement learning algorithm described in the previous section is applied to the velocity tracking and stabilization problem for the quarter-car vehicle model moving over an uneven terrain. The terrain considered in



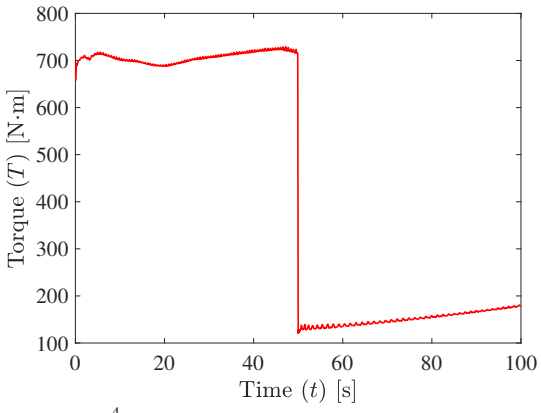
(a)



(b)

Figure 3: Vehicle velocities in simulation of the trained reinforcement learning agent for velocity tracking over an uneven terrain. (a) Longitudinal velocity of the vehicle (blue) with the desired velocity (dotted black). (b) Vertical velocity of the vehicle chassis.

the training of the reinforcment learning agent is described by constant friction parameters, while the height of the terrain is described by summation of cosine functions in space of the form $h(x) = \sum_{i=1}^{N=4} H_i \cos(\omega_i x)$. For clarity of demonstration, in the simulations presented here, we consider a terrain elevation described by a single cosine wave, where the amplitude and frequency are taken to be $H_1 = 0.1$ m and $\omega_1 = 0.4$ rad/m. The reward function for the problem formulation is described by Eq. 6, where the



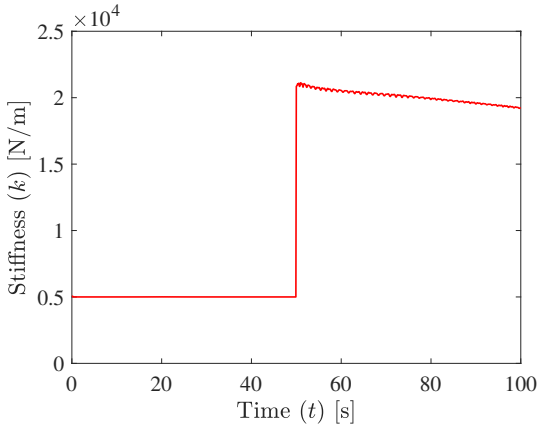


Figure 4: Control sequence specified by trained actor network for velocity tracking over an uneven terrain. (a) Wheel torque (b) Suspension stiffness.

longitudinal velocity to be tracked varies with time as

$$V_*(t) = \begin{cases} 25 \text{ m/s} & t \leq 50 \text{ s} \\ 10 \text{ m/s} & t > 50 \text{ s} \end{cases} \tag{7}$$

as shown by the dotted line in Fig. 3 (a).

The resulting trajectories of the longitudinal and vertical velocities of the vehicle in the simulation of the trained reinforcement learning agent over this terrain are shown in Fig. 3. It can be seen that the desired longitudinal velocity is tracked well over time, with small steady-state error. The vertical velocity also experiences only one brief period of high velocity oscillations, associated with the sudden change in desired forward velocity, which occurs at $t = 50$ s.

The control trajectories of the wheel torque and the suspension stiffness chosen by the trained actor network to achieve these velocities are shown in Fig. 4. In the training of the reinforcement learning agent, as well as in its simulation, the torque values are limited to a range of $[0, 10^3$ N$\cdot$m$]$ and the suspension stiffness values are limited to a range of $[5\times10^3, 2.5\times 10^4$ N/m$]$. We see that given these action ranges, the agent selects a large value of torque at around 700 N$\cdot$m for the first half of the simulation while trying to track the 25 m/s velocity, before settling to a lower torque of below 200 N$\cdot$m for tracking the 10 m/s velocity.

While the forward velocity is mostly affected by the choice of wheel torque, the velocity and magnitude of oscillations of the vehicle chassis can be affected by choice of suspension stiffness. In particular, the stiffness should be chosen so that the frequency of the forcing imparted on the suspension by the terrain is far from the resonant frequency of the suspension system in order to reduce the magnitude of oscillations. That is, by changing the suspension stiffness, the resonant frequency of the system can be altered to avoid oscillations and stabilize the vehicle body.

For a terrain with a single frequency of oscillations, the spatial frequency of the terrain directly translates to a forcing frequency on the suspension when the vehicle is travelling with constant forward velocity. Thus, the optimal choice of suspension stiffness for traversing such a terrain can be understood in terms of the suspension's frequency response for the given terrain. Such a response curve is computed from constant torque simulation of the equations of motion on this terrain for several constant values of the suspension stiffness, k and the resulting curves are shown in Fig. 5.

From Fig. 5, it can be seen that for tracking a velocity of $V_* = 25$ m/s, the optimal choice of stiffness is the minimum value of the allowable range. However, the velocity value of $V_* = 10$ m/s

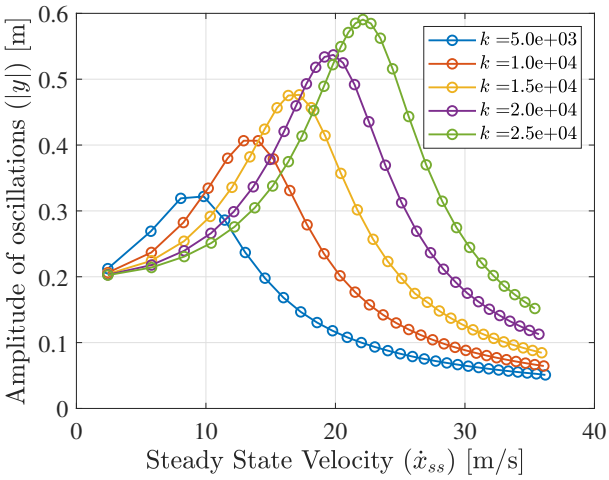


Figure 5: Frequency response of the suspension under constant wheel-torque simulation for varying values of suspension stiffness, k. Amplitude of chassis oscillations versus steady state longitudinal velocity over terrain of constant spatial frequency.

occurs near the resonant peaks of the curves with lower stiffnesses, and thus it is optimal to choose a higher stiffness value. This agrees with the action sequence chosen by the trained actor network, as it minimizes the stiffness in the early part of the simulation while tracking a high velocity, and nearly maximizes the allowable stiffness while tracking a lower velocity. This can be seen in Fig. 4 (b). This result indicates that using an active suspension in conjunction with a velocity tracking controller allows for significantly reduced oscillations of the upper mass. Fig. 5 shows that the stiffness change selected by the reinforcement learning agent enables better stabilization than any single stiffness value held constant through the simulation. As detailed previously, this improved stabilization can lead to improved rider comfort, enhanced perception from onboard cameras and other sensors, and decreased mechanical noise through the system.

5. CONCLUSION AND FUTURE WORK

The formulation and results presented here lay the groundwork for future work in more complex

scenarios, such as on dynamic, deformable terrains with unknown or partially known properties or planning for the coordination of multiple vehicles. This work has applications to more than a few problems for the military and may be especially useful for non-conventional unmanned ground vehicles. Such systems can also serve a part of a larger cyber physical system which works in sync with aerial vehicles that assist in finding a suitable path for a mobile robot. Data from an aerial vehicle could provide extended look-ahead information, which the reinforcement learning algorithm can consider to improve the ground vehicle's traversal approach. Further, we plan to expand our verification and understanding of this approach to more rigorous datasets, and demonstrate the effectiveness of the proposed methods experimentally on a physical rover or scaled vehicle in future work. We also expect that this work can be usefully extended to include input from camera sensors in a vision-based deep reinforcement learning framework, as the deep reinforcement learning algorithms employed here have been previously shown to be effective when using raw pixel data as input. Such a framework would allow for more advanced motion planning and maneuvering.

6. REFERENCES

[1] M. A. Djojo and K. Karyono, "Computational load analysis of Dijkstra, A*, and Floyd-Warshall algorithms in mesh network," in *2013 International Conference on Robotics, Biomimetics, Intelligent Computational Systems*, pp. 104–108, 2013.

[2] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[3] J. Hwangbo, I. Sa, R. Siegwart, and M. Hutter, "Control of a quadrotor with reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 2, no. 4, pp. 2096–2103, 2017.

[4] N. O. Lambert, D. S. Drew, J. Yaconelli, S. Levine, R. Calandra, and K. S. Pister, "Low-level control of a quadrotor with deep model-based reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4224–4230, 2019.

[5] R. S. Sutton, A. G. Barto, and R. J. Williams, "Reinforcement learning is direct adaptive optimal control," *IEEE Control Systems Magazine*, vol. 12, no. 2, pp. 19–22, 1992.

[6] T. Manderson, S. Wapnick, D. Meger, and G. Dudek, "Learning to Drive Off Road on Smooth Terrain in Unstructured Environments Using an On-Board Camera and Sparse Aerial Images," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1263–1269, 2020.

[7] V. Tsounis, M. Alge, J. Lee, F. Farshidian, and M. Hutter, "Deepgait: Planning and control of quadrupedal gaits using deep reinforcement learning," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 3699–3706, 2020.

[8] J. Liu, P. Jayakumar, J. L. Stein, and T. Ersal, "A study on model fidelity for model predictive control-based obstacle avoidance in high-speed autonomous ground vehicles," *Vehicle System Dynamics*, vol. 54, no. 11, pp. 1629–1650, 2016.

[9] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.

[10] R. Rajamani, *Vehicle dynamics and control*. Springer Science & Business Media, 2011.

[11] W. F. Milliken and D. L. Milliken, *Race car vehicle dynamics*. Society of Automotive Engineers, 1995.

[12] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 26–38, 2017.

[13] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," *31st International Conference on Machine Learning, ICML 2014*, vol. 1, pp. 605–619, 2014.

[14] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2019.

[15] M. Grześ, "Reward shaping in episodic reinforcement learning," in *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 565–573, 2017.

[16] The MathWorks, Inc., *Reinforcement Learning Toolbox*. Natick, Massachusetts, United States, 2019.