IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 7 March 2023, accepted 26 March 2023, date of publication 10 April 2023, date of current version 13 April 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3265895

RESEARCH ARTICLE

Sensitivity Adaptation of Lower-Limb Exoskeleton for Human Performance Augmentation Based on Deep Reinforcement Learning

RANRAN ZHENG[1], ZHIYUAN YU[2], HONGWEI LIU[2], ZHE ZHAO[2], JING CHEN[2], AND LONGFEI JIA[2]

[1]School of Aerospace Engineering, Beijing Institute of Technology, Beijing 100081, China
[2]Beijing Institute of Precision Mechatronics and Controls, Beijing 100076, China

Corresponding author: Ranran Zheng (chephilor@163.com)

ABSTRACT The lower-limb exoskeleton for human performance augmentation (LEHPA) in sensitivity amplification control (SAC) is vulnerable to model parameter uncertainties and unmodeled dynamics due to its large sensitivity to external disturbances resulting from the positive feedback by the inverse dynamic model of the exoskeleton. This paper firstly proposes to combine SAC with deep reinforcement learning (DRL) to reduce the dependence on the model accuracy and tackle the ever-changing human-exoskeleton interaction (HEI) dynamics. The sensitivity adjustment is interpreted as finding the optimal policy for a Markov Decision Process (MDP) and solved using deep reinforcement learning algorithms. To train the policy safely and efficiently, a multibody simulation environment is created to implement the training process, accompanied by a novel hybrid inverse-forward dynamics simulation method to carry out the simulation. For comparison purposes, the SAC controller is introduced as a benchmark. A novel performance evaluation method based on the HEI forces at the back, thighs, and shanks is proposed to evaluate the control effect of the trained SADRL controller quantitatively. The SADRL controller is compared with the SAC controller at five specified walking speeds, resulting in a lumped HEI force ratio as low as 0.54. The total decrease of HEI forces demonstrates the superior control effect of the SADRL strategy.

INDEX TERMS Lower-limb exoskeleton for human performance augmentation, sensitivity amplification control, sensitivity adaptation, deep reinforcement learning.

I. INTRODUCTION

The lower-limb exoskeleton for human performance augmentation (LEHPA) is a special class of wearable robotic system that runs in parallel to the human body, that transfers the payload weight to the ground, and that enhances human strength and endurance [1], [2], [3]. Combining the human intelligence with the robot strength and endurance, the coupled human-exoskeleton system shows a natural superiority over the biped or quadruped robots in adapting to rough and unstructured terrains and presents a promising prospect in performing dangerous and difficult tasks such as battlefield missions, disaster relief, firefighting, manufacturing, etc. [3], [4], [5]. The research on exoskeletons for human performance augmentation dates from the 1960s and significant progresses have been made in corresponding aspects in recent two decades, for example, mechanical design [6], sensors [7], actuators [8], human motion intention recognition [9], gait phase detection [10], [11], motion tracking [12], and so on. Control strategies receive much more attention of the researchers. Many control strategies have been proposed to improve the performance of the exoskeleton control system [13], [14], [15].

The most famous control strategy is sensitivity amplification control (SAC) [16], which is initially proposed for BLEEX (Berkeley Lower Extremity Exoskeleton) [17], the first functional load-carrying and energetically autonomous exoskeleton developed by U.C. Berkeley's

The associate editor coordinating the review of this manuscript and approving it for publication was Valentina E. Balas.

Human Engineering and Robotics Laboratory with the support of the DARPA Exoskeletons for Human Performance Augmentation (EHPA) program, and also used in the control practices of HULC [18], XOS [19], and HLEER [20]. The SAC strategy predicts the human motion intention using measurements only from the exoskeleton, needing no direct measurement of bioelectric signals from the pilot or human-exoskeleton interaction (HEI) force signals from human-exoskeleton interfaces, which facilitates the reduction of the system complexity and the enhancement of reliability. The sensitivity transfer function is defined as the mapping from the equivalent pilot torque to the exoskeleton angular velocity and represents how the equivalent human torque affects the exoskeleton angular velocity. To achieve a large closed-loop sensitivity transfer function without measuring equivalent pilot torque directly, the inverse of the exoskeleton dynamics is used as a positive feedback controller. However, this leads to the fact that the control effect of SAC depends heavily on the model accuracy because any error of model parameter will be amplified and transferred to the system outputs. To obtain the accurate model, a complex system identification process is necessarily performed [21]. In the subsequent hybrid control of BLEEX [22], the kinematic information of the pilot is measured by seven clinometers mounted on the human limbs and trunk (two clinometers on the feet, two on the shanks, two on the thighs and one on the trunk) to compute the joint angles, which are used as the targets of the proportional controllers in the stance phase. These clinometers, however, require careful design to fasten them to the pilot securely and increase the time to don and doff BLEEX. Besides, SAC cannot cope with HEI forces which vary from pilot to pilot and even within one pilot over time as a function of time and posture. The physical HEI dynamics is modeled by non-parametric regression [23]; however, the model update frequency is not high enough to cope with the rapid change of the HEI dynamics. Some reinforcement learning methods are used to adjust the sensitivity factor online, e.g., policy iteration [24], [25], [26] and Q-learning [27]. Unfortunately, these tabular solution methods are not suitable for high-dimensional and continuous domains due to the curse of dimensionality. Moreover, in terms of policy iteration, the model-based approach requires a complete and accurate model of the environment dynamics, which is usually difficult to acquire.

Deep reinforcement learning (DRL) combines reinforcement learning (RL) with deep learning (DL) by introducing deep neural networks as approximators of policy and/or value functions [28], [29]. With compact representations and powerful generalizations of deep neural networks, DL enables RL to scale up to Markov Decision Process (MDP) problems with high-dimensional and continuous action spaces, providing a novel approach to develop controllers for complex dynamic systems in a model-free manner. DRL has achieved promising results in locomotion control of physics-based characters [30], [31], [32] and egged robots [33] including the humanoid [34], [35], [36], [37], [38], the biped [39], [40], [41], [42], and the quadruped [43], [44], [45], [46], [47]. However, no attempt has been made to apply DRL to the controller design of LEHPA systems.

This paper investigates the sensitivity adjustment problem of the LEHPA system to reduce the dependence on model accuracy and adapt to the ever-changing HEI dynamics and proposes the sensitivity adaptation based on deep reinforcement learning (SADRL). The main contributions of this paper are as follows.

1) This paper presents a DRL framework to learn walking controller for the LEHPA systems by formulating the sensitivity adjustment as finding the optimal policy for an MDP.
2) This paper proposes a new multibody simulation environment for learning the policy and its corresponding hybrid inverse-forward dynamics simulation method.
3) This paper proposes a new phase partition of the gait cycle during human level walking according to the configuration of the feet contacting the ground.

The remainder of the paper is organized as follows: Section II presents the Modeling of the LEHPA system based on the novel gait phase partition; Section III describes the design of the adaptive control strategy based on DRL; Section IV describes the multibody simulation environment and training setup, followed by Section V discussing the results; Section VI concludes the paper.

II. MODELING OF LEHPA SYSTEM

A. LEHPA SYSTEM DESCRIPTION

Our LEHPA system shown in Fig. 1(a), is composed of seven subassemblies, namely the trunk, thighs, shanks, and feet. The trunk consists of the waist, the back board, the power source unit, the controller, and the payload. The width of the waist, and the lengths of thighs and shanks are designed adjustable to match different pilots between 1.6m and 1.8m in height. Each leg has three joints, the hip, knee, and ankle, and six degrees of freedom (DOFs), the hip flexion/extension, abduction/adduction, and external/internal rotation, the knee flexion/extension, and the ankle dorsiflexion/plantarflexion and eversion/inversion, only two of which, namely the hip and knee flexion/extension, are powered by joint actuator modules. Five inertial measurement units (IMUs) are respectively mounted on the trunk, two shanks, and two feet to measure the orientation of trunk and two ankle angles and their first and second order derivatives. A 6-axis load cell is placed between the pilot trunk and the harness to acquire the HEI force at the back, and four 3-axis load cells are placed at the human-exoskeleton interfaces between each limb link and its corresponding link strap to measure the HEI forces at thighs and shanks.

In comparison with the movements in the sagittal plane, the movements in the frontal and transverse planes have very few dynamic effects on the system and can be considered as quasi-static maneuvers. This indicates that the effects of the dynamics in the frontal and transverse planes on the dynamics

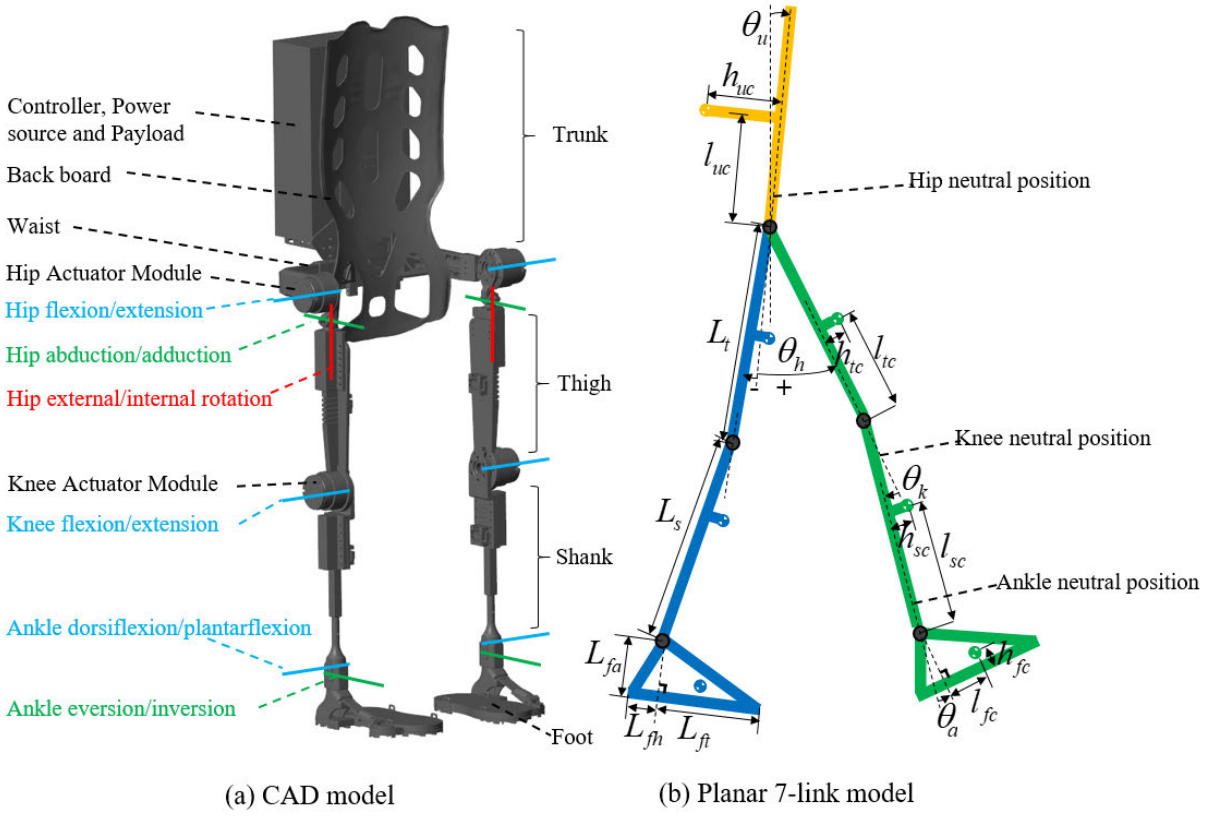


FIGURE 1. Our proposed LEHPA system.

in the sagittal plane can be ignored. For the sake of brevity, only the dynamics in the sagittal plane are considered in this work, with angles in the frontal and transverse planes set to zero. Thus, the exoskeleton can be simplified as the planar 7-link model depicted in Fig. 1(b). The joint definitions in the sagittal plane are as follows: the hip flexion, knee flexion and ankle dorsiflexion are defined as positive, while the hip extension, knee extension and ankle plantarflexion are negative.

B. PHASE PARTITION

The gait cycle during level walking can be divided into several phases according to the configuration of feet contacting the ground. For a single leg, the gait cycle is divided into the stance phase and the swing phase: the stance phase is characterized by the contact between the foot and the ground and can be further divided into three sub-phases according to the position of the instantaneous pivot point of the foot, the sub-phase of the foot pivoting around the heel, the sub-phase of the foot pivoting around the ankle, and the sub-phase of the foot pivoting around the toe; the swing phase features the foot taking off away from the ground. When double legs are considered simultaneously, the gait cycle is divided into the double support (DS) phase and the single support (SS) phase; each of the two phases appears twice in turn and the difference between the two appearances is the role exchange of two legs.

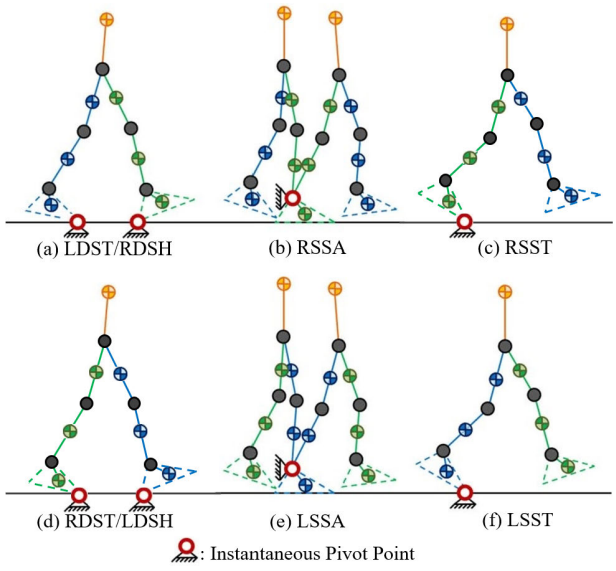


FIGURE 2. Phase partition for gait cycle during level walking.

It is worth noting that the transition from the double support phase to the single support phase takes three possible forms depending on the order of the moment of rear toe-off, t_1, and the moment of the front leg starting pivoting around ankle, t_2: if $t_1 = t_2$, the transition route is DST/DSH→SSA;

if $t_1 < t_2$, the transition route is DST/DSH→SSH→SSA; if $t_1 > t_2$, the transition route is DST/DSH→DST/DSA→SSA. For simplicity, the transition from the double support phase to the single support phase is assumed to take the first form. Thus, the transition route during a half of a gait cycle starting from heel strike is DST/DSH→SSA→SST, and the whole gait cycle starting from right heel strike is composed of the phase sequence, LDST/RDSH →RSSA→RSST→RDST/LDSH→LSSA→LSST, with the contact configurations of these phases seen in Fig. 2.

C. PHASE-DEPENDENT DYNAMIC MODEL

Given the phase segmentation, the dynamic model of the LEHPA system consists of three phase-dependent sub-models: the dynamic model for the DST/DSH phase, the dynamic model for the SSA phase and the dynamic model for the SST phase.

1) DYNAMIC MODEL FOR DST/DSH PHASE

The double support phase starts with the heel strike of the swinging leg and ends with the rear toe-off. In this phase the front leg pivots around the heel while the rear leg pivots around the toe. The exoskeleton can be modeled as two planar 4-DOF serial link mechanisms that are connected to each other along their uppermost link, shown in Fig. 2 (a) and (d). The equation of motion of the exoskeleton can be written in the general form as

$$\begin{cases} M_R(m_{TR},\theta_R)\ddot{\theta}_R + V_R(m_{TR},\theta_R,\dot{\theta}_R) \\ +G_R(m_{TR},\theta_R) = \tau_{act,R} + \tau_{HEI,R} \\ M_F(m_{TF},\theta_F)\ddot{\theta}_F + V_F(m_{TF},\theta_F,\dot{\theta}_F) \\ +G_F(m_{TF},\theta_F) = \tau_{act,F} + \tau_{HEI,F} \end{cases} \tag{1}$$

where the subscripts R and F represent the rear leg and front leg respectively. $M_R(m_{TR},\theta_R)$ and $M_F(m_{TF},\theta_F)$ are 4×4 inertia matrices, $V_R(m_{TR},\theta_R,\dot{\theta}_R)$ and $V_F(m_{TF},\theta_F,\dot{\theta}_F)$ are 4×1 centripetal and Coriolis vectors, and $G_R(m_{TR},\theta_R)$ and $G_F(m_{TF},\theta_F)$ are 4×1 gravitational torque vectors. $\tau_{act,R}$ and $\tau_{act,F}$ are 4×1 actuator torque vectors with their first two elements set to zero since there is no actuator associated with ankle angle and angle between the exoskeleton foot and the ground. $\tau_{HEI,R}$ and $\tau_{HEI,F}$ are 4×1 equivalent HEI torque vectors imposed by the pilot on the exoskeleton joints. m_{TR} and m_{TF} are effective trunk masses supported by the rear leg and front leg respectively, and m_T is the total trunk mass such that

$$m_T = m_{TR} + m_{TF} \tag{2}$$

The contributions of m_T on each leg (i.e. m_{TR} and m_{TF}) are chosen as functions of the location of the trunk center of mass relative to the locations of the pivot points such that

$$\frac{m_{TF}}{m_{TR}} = \frac{x_{TR}}{x_{TF}} \tag{3}$$

where x_{TR} is the horizontal distance between the trunk center of mass and the rear pivot point and x_{TF} is that between the trunk center of mass and the front pivot point.

2) DYNAMIC MODEL FOR SSA PHASE

The SSA phase represents the period of time from the moment of rear toe-off and the front leg starting pivoting around ankle to that of the front leg starting pivoting around toe. In this phase the foot of the stance leg is considered to be fixed to the ground and the exoskeleton can be modeled as a planar 6-DOF serial link mechanism, seen in Fig. 2 (b) and (e). The equation of motion of the exoskeleton can be expressed as

$$M(\theta)\ddot{\theta} + V(\theta,\dot{\theta}) + G(\theta) = \tau_{act} + \tau_{HEI} \tag{4}$$

where $M(\theta)$ is a 6×6 inertia matrix, $V(\theta,\dot{\theta})$ is a 6×1 centripetal and Coriolis vector, and $G(\theta)$ is a 6×1 gravitational torque vector. τ_{act} is the 6×1 actuator torque vector with its first and last elements set to zero. τ_{HEI} is the 6×1 equivalent HEI torque vector.

3) DYNAMIC MODEL FOR SST PHASE

The SST phase starts at the moment of the stance leg starting pivoting around toe and ends with the heel strike of the swinging leg. In this phase the exoskeleton can be modeled as a 7-DOF serial link mechanism in the sagittal plane, shown in Fig. 2 (c) and (f). The equation of motion of the exoskeleton written in the general form is the same as Eq. (4), with the difference of dimension. The inertia matrix $M(\theta)$ is 7×7, the centripetal and Coriolis vector $V(\theta,\dot{\theta})$ is 7×1, and the gravitational torque vector $G(\theta)$ is 7×1. τ_{act} is a 7×1 vector, with its first two and last elements set to zero. The equivalent HEI torque vector τ_{HEI} is 7×1.

III. SENSITIVITY ADAPTATION BASED ON DEEP REINFORCEMENT LEARNING

The SADRL strategy proposed in this work is illustrated in Fig. 3. The human body is powered by the resultant torque of the musculoskeletal moment τ_m and the equivalent HEI torque τ_{HEI}, while the exoskeleton is driven by the resultant torque of the actuator torque τ_{act} and the equivalent HEI torque τ_{HEI}. The equivalent HEI torque is an assistance to the exoskeleton, but a resistance to the pilot. To ensure wearing comfort, the exoskeleton is expected to move as consistently as possible with the pilot to reduce the resistance; that is, the exoskeleton is desired to be transparent to the pilot. To this end, the exoskeleton is required to be highly sensitive to pilot forces and torques, which is opposite to the classical and modern control theory where negative feedback loops with large gains are chosen to minimize the sensitivity function of the system to external disturbances.

In order to achieve a large sensitivity to forces and torques, SAC utilizes the inverse dynamics of the exoskeleton as positive feedback so that the loop gain for the exoskeleton approaches unity (slightly less than 1). The control system output in each phase is as follows. In SSA phase, the control low is chosen such that

$$\tau_{act} = G(\theta) + (1-\alpha^{-1})[M(\theta)\ddot{\theta} + V(\theta,\dot{\theta})\dot{\theta}] \tag{5}$$

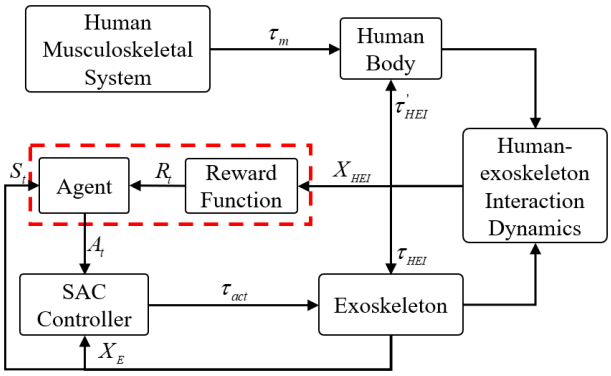


FIGURE 3. Diagram for SADRL strategy.

where α is the sensitivity factor vector whose components are all greater than 1. The control low in SST phase takes the same form as the SSA phase. In double support phase, the controller is chosen such that

$$\begin{cases} \tau_{act,R} = G_R(m_{TR}, \theta_R) + (1 - \alpha_R^{-1}) \\ [M_R(m_{TR}, \theta_R)\ddot{\theta}_R + V_R(m_{TR}, \theta_R, \dot{\theta}_R)\dot{\theta}_R] \\ \tau_{act,F} = G_F(m_{TF}, \theta_F) + (1 - \alpha_F^{-1}) \\ [M_F(m_{TF}, \theta_F)\ddot{\theta}_F + V_F(m_{TF}, \theta_F, \dot{\theta}_F)\dot{\theta}_F] \end{cases} \tag{6}$$

where α_R and α_F are the sensitivity factor vectors for active joints of the rear leg and front leg respectively. The fixed sensitivity in SAC is, however, not adaptable to the HEI dynamics changing from person to person and within a person as a function of time and posture. The proposed SADRL strategy aims to produce the suitable actuator torque to minimize the equivalent HEI torque by adjusting the sensitivity adaptively according to the exoskeleton state.

A. STATE SPACE AND ACTION SPACE

In this work, sensitivity adjustment is interpreted as an MDP problem. Considering that SAC predicts the human motion intention using measurements only from the exoskeleton, we also choose measurements only from the exoskeleton as the state for the MDP problem, which is a 21-dimensional vector consisting of the exoskeleton trunk pitch angle and angular velocity, and joint angles, angular velocities and angular accelerations of hips, knees and ankles. The action for the MDP problem is specified as the sensitivity for the four active joints. Given that the sensitivity factors vary in a wide range, the policy doesn't directly output the sensitivity factors. Instead, we choose to let the policy learn the following sensitivity-related vector β whose components all range in the interval (0,1).

$$\beta = 1 - \alpha^{-1} \tag{7}$$

Then the sensitivity adjustment problem can be viewed as finding the optimal policy for the MDP, and can be solved using reinforcement learning algorithms.

B. LEARNING ALGORITHM AND NETWORK

The algorithm used in this work is Twin Delayed Deep Deterministic Policy Gradient (TD3) [48], a model-free off-policy actor-critic method. TD3 is an improvement of Deep Deterministic Policy Gradient (DDPG) [49]. DDPG combines Deterministic Policy Gradient (DPG) [50] with Deep Q-Network (DQN) [51]. TD3 improves DDPG in three aspects: introducing clipped Double Q-Learning to reduce variance by reducing the accumulation of errors; proposing delaying policy updates until the value estimate has converged, to address the coupling of value and policy; introducing target policy smoothing regularization strategy where a SARSA-style update bootstraps similar action estimates to further reduce variance.

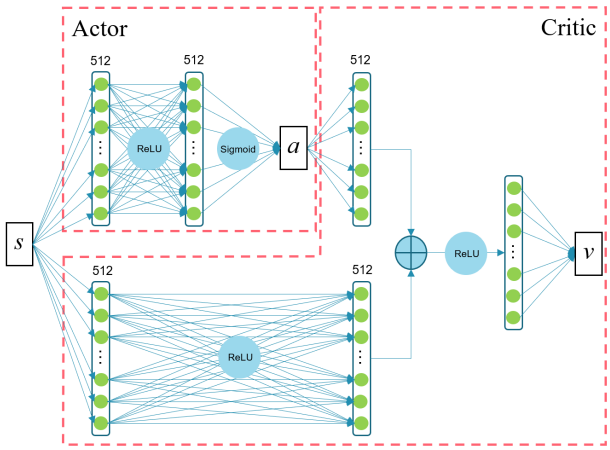


FIGURE 4. Network architectures of the Actor and the Critic.

One Actor and two twin delayed Critics are created in TD3. The Critic networks share the same architecture but have different weight parameters. The network architectures of the actor and the critic are illustrated in Fig. 4. The state, the input of the actor, is a 21-dimensional vector consisting of the angles, angular velocities, and angular accelerations of the exoskeleton trunk orientation, hips, knees, and ankles, while the action, the output of the actor, is a 4-dimensional sensitivity-related vector for the four active joints. Two hidden layers with 512 neural units are used for actor. ReLU activations are used between the hidden layers, and the output of the actor is passed through a sigmoid function to limit the range of the final output. The critic receives both the state vector and action vector and outputs the value. The action is passed through one hidden layer and the state is passed through two hidden layers, with each hidden layer having 512 neural units. Then the two 512-dimensional vectors are added up to one that is activated by ReLU functions. The output layer only has one neural unit representing the value function.

C. REWARD FUNCTION

The reward function guides the direction of parameter optimization during the learning process. Since the

misspecification of the reward function can have unintended and even dangerous consequences, it is critically important to design a suitable reward function for each task. In our scenario of the LEHPA system, the HEI forces are the key indicator to judge the performance of the control system. Therefore, the reward function in this work is defined as the weighted sum of five local reward terms which are the functions of the HEI forces at the five human-exoskeleton interfaces, namely the back, thighs, and shanks respectively. The HEI force at the back consists of three components: the pitch torque T_{pitch}, the force along the sagittal axis F_S, and the force along the vertical axis F_V. The HEI force at each leg segment consists of two components: the one along the central axis of the segment and the other normal to the central axis. The eight HEI force components at four leg segments are F_{nRT}, F_{tRT}, F_{nRS}, F_{tRS}, F_{nLT}, F_{tLT}, F_{nLS}, and F_{tLS} respectively. The eleven HEI force components are depicted in Fig. 5.

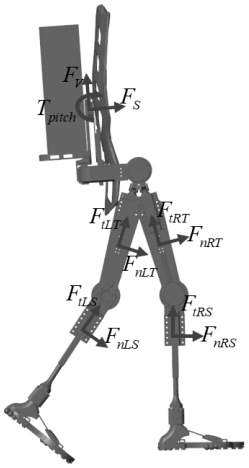


FIGURE 5. HEI forces used in reward function.

The reward function is expressed as

$$r = \sum_{i=1}^{5} w_i r_i \tag{8}$$

where r_i is the local reward term with respect to the $i-th$ human-exoskeleton interface. The local weight w_i determines the contribution of r_i to the reward. Each local reward takes the following form:

$$r_i = \exp(-\sum_j k_{ij} |\frac{F_{ij}}{\Delta_{ij}}|^2) \tag{9}$$

where F_{ij} represents the $j-th$ HEI force component at the $i-th$ human-exoskeleton interface. Δ_{ij} is the normalization term carefully selected to normalize F_{ij}, and the exponent weight k_{ij} determines the contribution of the normalized force component to the exponent. The notations of the eleven HEI force components are listed in Table 1.

TABLE 1. Notations of F_{ij}.

	$i=1$	$i=2$	$i=3$	$i=4$	$i=5$
$j=1$	T_{pitch}	F_{nRT}	F_{nRS}	F_{nLT}	F_{nLS}
$j=2$	F_S	F_{tRT}	F_{tRS}	F_{tLT}	F_{tLS}
$j=3$	F_V				

IV. TRAINING IN SIMULATION

A. MULTIBODY SIMULATION ENVIRONMENT

In order to learn the SADRL strategy safely and efficiently, a multibody simulation environment is created based on the MATLAB/Simscape physical modeling toolbox to implement the training process in virtual scene. As is shown in Fig. 6, the multibody simulation environment consists of the human body model, the exoskeleton model, the terrain and interaction models at all the human-exoskeleton interfaces. Additionally, a novel hybrid inverse-forward dynamics simulation method of the coupled human-exoskeleton system is proposed to carry out the simulation. The human body joints are modeled as inverse dynamics joints while the exoskeleton joints are modeled as forward dynamics joints. As is illustrated in Fig. 7, the human body model is driven by reference motions of different walking speeds adapted from the original reference motion and leads the exoskeleton model to move through HEI forces generated by HEI models. Contrary to those that directly input joint trajectories into the exoskeleton model to compute joint torques, this method demonstrates the dynamic interaction process between the pilot and the exoskeleton during walking, revealing the essence of the cooperative movement. The original reference motion is a whole gait cycle of joint trajectories of hips, knees and ankles sampled at 240Hz by the motion capture system (Mtw Awinda, Xsens) from human walking on the treadmill at 2.8 km/h for 334 time steps, about 1.39s. Reference motions of different walking speeds are acquired by stretching or compressing the cycle of the original reference motion and then extending the motion by copying it repeatedly. Note that by varying the walking speed in this way, the resulting reference motions are still physically feasible, with the foot fixed on the ground throughout the stance phase.

1) HUMAN BODY MODEL

The human body model is a simplified surrogate of the pilot. Upper limbs and the degrees of freedom of lower limbs in the frontal and transverse planes are omitted since only the movements of lower limbs in the sagittal plane are considered in this work. Consequently, each leg only has three degrees of freedom, the hip flexion/extension, the knee flexion/extension, and the ankle dorsiflexion/plantarflexion. Note that the human body model is rigid and the flexibility of human muscles and the harness is integrated into the HEI models.

2) EXOSKELETON MODEL

The exoskeleton model is a simplified version of the CAD prototype. Each subassembly is simplified as a part and the

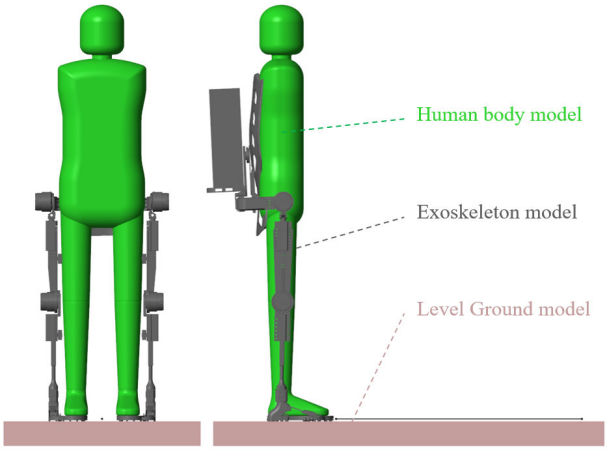


FIGURE 6. The multibody simulation environment.

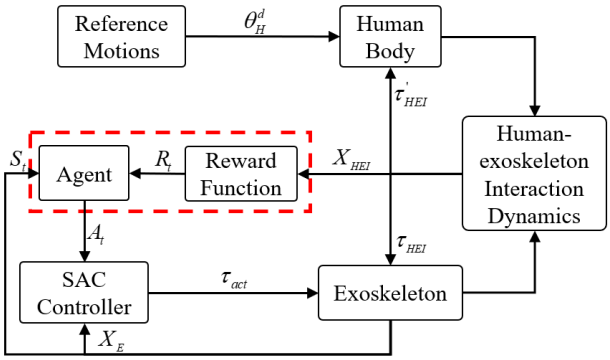


FIGURE 7. Diagram for SADRL strategy in simulation environment.

degrees of freedom at hips and ankles in the frontal and transverse planes are omitted.

3) THE HEI MODELS

The HEI between the human model and the exoskeleton model takes place at several human-exoskeleton interfaces, namely the back, thighs, shanks, and feet. In this paper, only the HEI in the sagittal plane are modeled as we only focus on the movements in the sagittal plane.

The HEI at the back is modeled as a combination of a spring-damper with a torsional spring. The spring-damper determines the force component caused by the position difference between the human and the exoskeleton at the back HEI interface, while the torsional spring determines the torque component caused by the orientation difference between the two. Each HEI at the thigh or shank is respectively modeled as a spring-damper. The HEI at the foot is modeled as two spring-damper systems at the heel and toe. The stiffness and damping of the torsional spring at the back are set to 20 Nm/rad and 0.5 NmB7s/rad. The stiffness and damping constants of each spring-damper are listed in Table 2.

TABLE 2. Parameters of spring-dampers.

Parameters	Interfaces				
	Back	Thigh	Shank	Heel	Toe
Stiffness (N/m)	4e3	1e4	1e4	5e5	5e5
Damping (N·s/m)	4e2	5e2	5e2	5e2	5e2

4) TERRAIN

In the multibody simulation environment, the terrain is fixed to the world frame and interacting with the underneath of the exoskeleton foot. Some structured terrains, for example, level ground, slopes of different degrees, stairs of different heights and widths, are commonly used in exoskeleton research. In this work, the level ground is modeled as a flat plate in the simulation of exoskeleton level walking.

The ground react force is generated by two rows of Spatial Contact Force blocks placed at the inner and outer sides of the underneath of the exoskeleton foot respectively. In each Spatial Contact Force block, the normal contact force perpendicular to the ground is computed using the force equation of the classical spring-damper model, while the frictional contact force parallel to the ground is computed using the Smooth Stick-Slip law. The stiffness and damping of spring-damper model for the normal contact force are set to 2e4 N/m and 4e2 NC,B7s/m respectively. For the frictional contact force, the coefficients of static friction and dynamic friction are set to 0.9 and 0.7.

B. TRAINING SETUP

The training proceeds episodically. The walking speed of the reference motion for each episode is limited in the interval of [2.8, 5.6] km/h by choosing randomly the gait cycle length in the interval [0.696, 1.392] seconds. At the start of each episode, the initial joint angles of both the human model and the exoskeleton model in the multibody simulation environment are chosen randomly from the reference joint trajectories. A rollout is then simulated by following actions from the policy at every time step. To avoid excess explorations of poor states, an early termination mechanism is introduced to terminate the episode early and set the remaining rewards to be 0. In our training task, an early termination is triggered when the vertical distance between the two sides of the back human-exoskeleton interface is more than 0.3m or the absolute value of the exoskeleton trunk pitch angle is over $\pi/6$ rad. An episode terminates at a predetermined time horizon, namely the maximum simulation time of an episode, or until an early termination occurs. The simulation rate is set to 2kHz. Target joint angles are computed every 40 ms, leading to a policy query rate of 25 Hz, while the low-level joint PD controllers run at the same rate as the simulation. The time horizon for each episode is set to 5s.

The values of local weights, normalization terms and exponent weights are set by experience and shown in Table 3.

TABLE 3. Values of reward function parameters.

i	w_i	j	Δ_{ij}	k_{ij}
1	0.4	1	4.5	0.5
		2	250	0.25
		3	300	0.25
2	0.15	1	150	0.75
		2	50	0.25
3	0.15	1	60	0.75
		2	6	0.25
4	0.15	1	150	0.75
		2	50	0.25
5	0.15	1	60	0.75
		2	6	0.25

V. RESULTS AND DISCUSSION

In this work, the HEI forces at the back, thighs, and shanks are chosen as the evaluation reference to evaluate the power-augmenting effect of the exoskeleton system. Root-mean-square (RMS) values of these HEI force components in Table 1 are used as the performance indicator:

$$\bar{F} = \sqrt{\frac{1}{T}\int_0^T F^2 dt} \tag{10}$$

where F represents one of the aforementioned HEI force components and T is the time duration. Considering HEI forces may be not strictly periodic, we set T to 5 gait cycles. For comparison purposes, SAC is introduced as a benchmark. The ratio of F RMS in SADRL to that in SAC is calculated to depict the performance improvement of SADRL relative to SAC quantitatively:

$$\lambda(F) = \frac{\bar{F}_{SADRL}}{\bar{F}_{SAC}} \tag{11}$$

where $\bar{F}_{SADRL}$ and $\bar{F}_{SAC}$ denote F RMS in SADRL and SAC respectively.

To evaluate the performance at different walking speeds, we implement the comparison at five different reference walking speeds, i.e., 2.8 km/h, 3.5 km/h, 4.2 km/h, 4.9 km/h, and 5.6 km/h, and calculate the weighted average ratio for each HEI force component by:

$$\bar{\lambda}(F) = \frac{1.5\lambda_1(F) + 2.5\lambda_2(F) + 3\lambda_3(F) + 2\lambda_4(F) + \lambda_5(F)}{10} \tag{12}$$

where $\lambda_1(F)$, $\lambda_2(F)$, $\lambda_3(F)$, $\lambda_4(F)$, and $\lambda_5(F)$ represent the ratios at 2.8 km/h, 3.5 km/h, 4.2 km/h, 4.9 km/h, and 5.6 km/h respectively. To evaluate the overall improvement, the lumped ratio is defined as the weighted sum of weighted average ratios of the aforementioned HEI force components:

$$\lambda^\star = \sum_i w_i \sum_j k_{ij}\bar{\lambda}(F_{ij}) \tag{13}$$

where w_i and k_{ij} are the local weight and exponent weight also used in the reward function.

Since different sensitivities have different control effects, five constant sensitivities are chosen to carry out the evaluation by setting β to 0.1, 0.3, 0.5, 0.7 and 0.9. It is found that all HEI force component RMS values in the case of $\beta = 0.1$ are all less than their counterparts in other cases. Thus, the case of $\beta = 0.1$ is chosen as the benchmark to evaluate the proposed SADRL strategy. Since the movements of the two legs are symmetrical and differ little from each other at the timescale of gait cycles, it is rational to assume that each HEI force component at the left leg has the same weighted average ratio as its counterpart at the right leg:

$$\begin{cases} \bar{\lambda}(F_{nLT}) = \bar{\lambda}(F_{nRT}) \\ \bar{\lambda}(F_{tLT}) = \bar{\lambda}(F_{tRT}) \\ \bar{\lambda}(F_{nLS}) = \bar{\lambda}(F_{nRS}) \\ \bar{\lambda}(F_{tLS}) = \bar{\lambda}(F_{tRS}) \end{cases} \tag{14}$$

Thus, we can omit the prolix calculation for the HEI forces at the left leg and only calculate the RMS values and ratios of the HEI force components at the right leg for the sake of simplicity.

The RMS values of HEI force components at the back, right thigh, and right shank at the five specified walking speeds are shown in Fig. 8, Fig. 9, and Fig. 10. It is worth mentioning that the RMS values of the seven HEI force components may be different from those in reality because of the reality gap, especially the differences between the interaction models in the multibody simulation environment and the straps in the real world, but it doesn't impact on the comparison between SADRL and SAC.

It can be seen from Fig. 8 to Fig. 10 that at each walking speed the RMS values of the seven HEI force components in SADRL are all less than their counterparts in SAC. The RMS values of T_{pitch} and F_V in SADRL are less than their RMS values in SAC, proving that more payload weight is transferred to the ground successfully in SADRL. The RMS values of F_S, F_{nRT}, F_{tRT}, F_{nRS}, and F_{tRS} in SADRL are less than

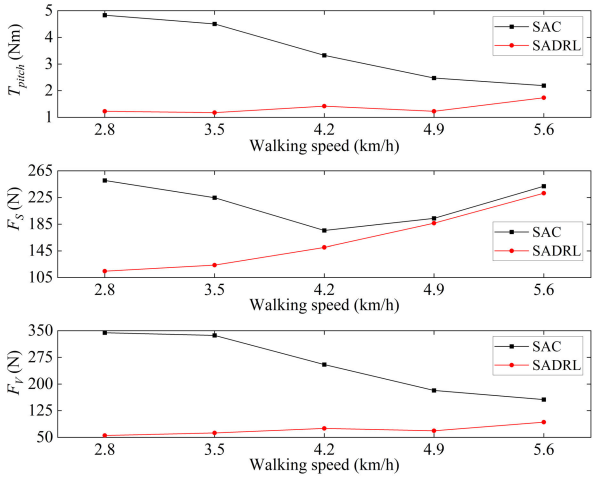


FIGURE 8. HEI force RMS at the back.

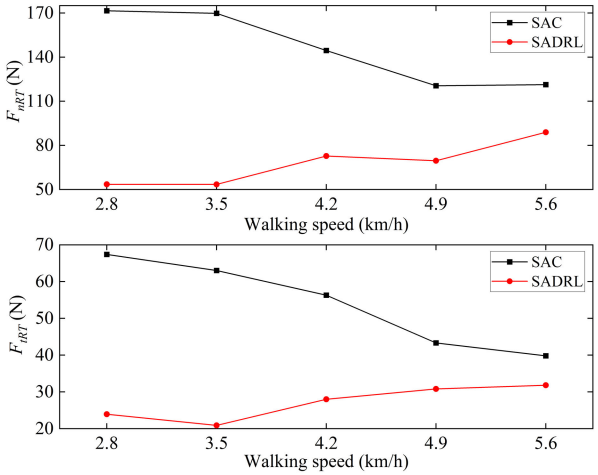


FIGURE 9. HEI force RMS at the right thigh.

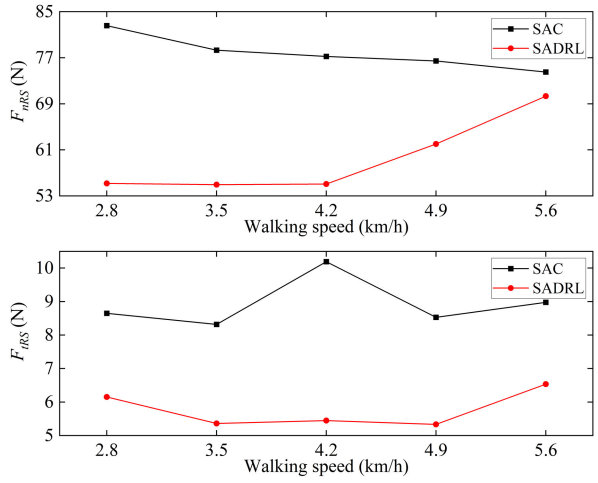


FIGURE 10. HEI force RMS at the right shank.

their RMS values in SAC, meaning that SADRL can reduce the misalignment between the pilot and the exoskeleton and improve the motion tracking performance.

The RMS of T_{pitch}, F_V, F_{nRT}, and F_{tRT} in SAC decrease sharply with the increase of the walking speed. F_S RMS in SAC decreases first and then increases as the walking speed grows, whereas the RMS of F_{nRS} and F_{tRS} don't change significantly. The phenomenon indicates that SAC performs badly at low walking speeds and is only suitable for high walking speeds. However, in SADRL, the RMS of most HEI force components increase slightly or even keep in low levels. Even though F_S RMS in SADRL increases significantly with the increase of the walking speed, it is still less than that in SAC. This proves that SADRL is suitable for the entire walking speed interval.

It is found from Fig. 9 and Fig. 10 that RMS values of F_{nRT} and F_{nRS} are much greater than those of F_{tRT} and F_{tRS} respectively whichever control strategy the system adopts, showing that a HEI force at the right thigh or shank is dominated by its normal component. This is rational because the

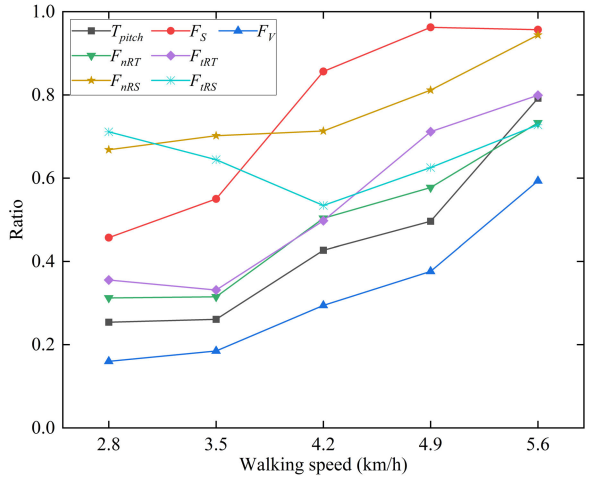


FIGURE 11. Normalized HEI force RMS.

joint misalignment between the human and the exoskeleton mainly leads to deviation normal to the segment and the equivalent HEI torque at the joint is produced by the HEI force component normal to the segment. The RMS of F_V in SAC is greater than that of F_S at slow speeds whereas the reverse occurs at fast speeds. However, F_S always dominates in SADRL because F_V keeps in a low level.

TABLE 4. Ratios of HEI force components.

F	$\lambda_1(F)$	$\lambda_2(F)$	$\lambda_3(F)$	$\lambda_4(F)$	$\lambda_5(F)$	$\bar{\lambda}(F)$
T_{pitch}	0.25	0.26	0.43	0.50	0.79	0.41
F_S	0.46	0.55	0.86	0.96	0.96	0.75
F_V	0.16	0.18	0.29	0.38	0.59	0.29
F_{nRT}	0.31	0.32	0.50	0.58	0.73	0.47
F_{tRT}	0.36	0.33	0.50	0.71	0.80	0.51
F_{nRS}	0.67	0.70	0.71	0.81	0.94	0.75
F_{tRS}	0.71	0.64	0.53	0.63	0.74	0.63

Table 4 presents the normalized RMS of the seven HEI force components at the five walking speeds and their corresponding weighted average ratios. In order to facilitate further analysis, these normalized RMS are also presented in Fig. 11. The normalized RMS of F_S are much greater than those of F_V, demonstrating that the deviation between the pilot and the exoskeleton at the back in the sagittal axis is more difficult to reduce than the that in the vertical axis in the existence of walking speed. The normalized RMS of most components increase with the increase of the walking speed. This is caused by the decrease of their RMS in SAC and slight increase of their RMS in SADRL. Comparing the weighted average ratios of the HEI forces at the back, right thigh, and right shank, we find that the closer a human-exoskeleton interface is to the passive ankle joint, the greater the weighted average ratios of the HEI force components at the human-exoskeleton interface except F_S. The passive right ankle is driven only by the equivalent HEI torque, which is mainly produced by the HEI force at the right shank, especially during stance phase when the fixed foot acts as the base.

Finally, we acquire the lumped ratio of the HEI forces in SADRL relative to those in SAC $\lambda^{\star} = 0.54$.

VI. CONCLUSION AND FUTURE WORK

This paper proposes the SADRL strategy for the lower-limb exoskeleton for human performance augmentation. The deep reinforcement learning framework is introduced to learn appropriate sensitivities from exoskeleton motion information. To train the policy safely and efficiently, a multibody simulation environment is created to carry out the training process, accompanied by a novel hybrid inverse-forward dynamics simulation method. For comparison purposes, The SAC strategy is introduced as a benchmark to assess the improvement of SADRL. A new performance assessment method based on HEI forces at the back, thighs, and shanks is proposed to evaluate the control effect of the SADRL controller quantitatively. The lumped ratio of the HEI forces in SADRL is as low as 0.54, proving that the proposed SADRL strategy has provided the exoskeleton with the ability to adapt to the varying HEI dynamics and reduce the dependence on model accuracy. In terms of the HEI force at the back, its component in the sagittal axis is more difficult to reduce than that in the vertical axis owing to the walking speed. The HEI forces close to the passive ankle joints are more difficult to reduce because they undertake the task of driving the ankle joints.

The future work will focus on the following aspects. First, the controller can be trained further on more terrains to adapt to complex environments. Some terrains like slopes of different degrees and stairs of different heights and widths will be created in the multibody simulation environment and the reference motions of human walking on these terrains will be sampled and used to drive the human body model. Moreover, some efforts will be made to close the reality gap when the learned control strategy is transferred from simulation to reality successfully. Finally, the deep reinforcement learning framework will be employed on the real exoskeleton to fine-tune the controller in the real-world environment.

REFERENCES

[1] S. Qiu, Z. Pei, C. Wang, and Z. Tang, "Systematic review on wearable lower extremity robotic exoskeletons for assisted locomotion," *J. Bionic Eng.*, vol. 20, no. 2, pp. 436–469, Oct. 2022. [Online]. Available: https://link.springer.com/10.1007/s42235-022-00289-8

[2] S. Yeem, J. Heo, H. Kim, and Y. Kwon, "Technical analysis of exoskeleton robot," *World J. Eng. Technol.*, vol. 7, no. 1, pp. 68–79, 2019. [Online]. Available: http://www.scirp.org/journal/doi.aspx?DOI=10.4236/wjet.2019.71004

[3] S. Viteckova, P. Kutilek, G. de Boisboissel, R. Krupicka, A. Galajdova, J. Kauler, L. Lhotska, and Z. Szabo, "Empowering lower limbs exoskeletons: State-of-the-art," *Robotica*, vol. 36, no. 11, pp. 1743–1756, Nov. 2018. [Online]. Available: https://www.cambridge.org/core/product/identifier/S0263574718000693/type/journal_article

[4] Z. Jia-Yong, L. Ye, M. Xin-Min, H. Chong-Wei, M. Xiao-Jing, L. Qiang, W. Yue-Jin, and Z. Ang, "A preliminary study of the military applications and future of individual exoskeletons," *J. Phys., Conf.*, vol. 1507, no. 10, Mar. 2020, Art. no. 102044. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1507/10/102044

[5] S. Fox, O. Aranko, J. Heilala, and P. Vahala, "Exoskeletons: Comprehensive, comparative and critical analyses of their potential to improve manufacturing performance," *J. Manuf. Technol. Manag.*, vol. 31, no. 6, pp. 1261–1280, Jun. 2019. [Online]. Available: https://www.emerald.com/insight/content/doi/10.1108/JMTM-01-2019-0023/full/html

[6] N. Aliman, R. Ramli, and S. M. Haris, "Design and development of lower limb exoskeletons: A survey," *Robot. Auton. Syst.*, vol. 95, pp. 102–116, Sep. 2017.

[7] M. Tiboni, A. Borboni, F. Vérité, C. Bregoli, and C. Amici, "Sensors and actuation technologies in exoskeletons: A review," *Sensors*, vol. 22, no. 3, p. 884, Jan. 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/3/884

[8] A. J. Veale and S. Q. Xie, "Towards compliant and wearable robotic orthoses: A review of current and emerging actuator technologies," *Med. Eng. Phys.*, vol. 38, no. 4, pp. 317–325, Apr. 2016, [Online]. Available: https: //linkinghub.elsevier.com/retrieve/pii/S135045331600031X, doi: 10.1016/j.medengphy.2016.01.010.

[9] H. F. N. Al-Shuka, R. Song, and C. Ding, "On high-level control of power-augmentation lower extremity exoskeletons: Human walking intention," in *Proc. 10th Int. Conf. Adv. Comput. Intell. (ICACI)*, Mar. 2018, pp. 169–174. [Online]. Available: https://ieeexplore.ieee.org/document/8377601/

[10] J. Taborri, E. Palermo, S. Rossi, and P. Cappa, "Gait partitioning methods: A systematic review," *Sensors*, vol. 16, no. 1, p. 66, Jan. 2016. [Online]. Available: http://www.mdpi.com/1424-8220/16/1/66

[11] D. X. Liu, X. Wu, W. Du, C. Wang, and T. Xu, "Gait phase recognition for lower-limb exoskeleton with only joint angular sensors," *Sensors*, vol. 16, no. 10, pp. 1–21, 2016.

[12] H. F. N. Al-Shuka and R. Song, "On low-level control strategies of lower extremity exoskeletons with power augmentation," in *Proc. 10th Int. Conf. Adv. Comput. Intell. (ICACI)*, Mar. 2018, pp. 63–68. [Online]. Available: https://ieeexplore.ieee.org/document/8377581/

[13] H. F. N. Al-Shuka, M. H. Rahman, S. Leonhardt, I. Ciobanu, and M. Berteanu, "Biomechanics, actuation, and multi-level control strategies of power-augmentation lower extremity exoskeletons: An overview," *Int. J. Dyn. Control*, vol. 7, no. 4, pp. 1462–1488, Dec. 2019, doi: 10.1007/s40435-019-00517-w.

[14] W. Huo, S. Mohammed, J. C. Moreno, and Y. Amirat, "Lower limb wearable robots for assistance and rehabilitation: A state of the art," *IEEE Syst. J.*, vol. 10, no. 3, pp. 1068–1081, Sep. 2016. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6930719

[15] T. Yan, M. Cempini, C. M. Oddo, and N. Vitiello, "Review of assistive strategies in powered lower-limb orthoses and exoskeletons," *Robot. Auto. Syst.*, vol. 64, pp. 120–136, Feb. 2015. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0921889014002176

[16] H. Kazerooni, J.-L. Racine, and R. Steger, "On the control of the Berkeley lower extremity exoskeleton (BLEEX)," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2005, pp. 4353–4360. [Online]. Available: http://ieeexplore.ieee.org/document/1570790/

[17] H. Kazerooni and R. Steger, "The Berkeley lower extremity exoskeleton," *J. Dyn. Syst. Meas. Control-Trans. ASME*, vol. 128, pp. 14–25, Mar. 2006. [Online]. Available: https://asmedigitalcollection.asme.org/dynamicsystems/article/128/1/14/465257/The-Berkeley-Lower-Extremity-Exoskeleton

[18] *Human Universal Load Carrier (HULC)*. Accessed: Oct. 26, 2020. [Online]. Available: https://www.army-technology.com/projects/human-universal-load-carrier-hulc/

[19] S. Karlin, "Raiding iron man's closet [geek life]," *IEEE Spectr.*, vol. 48, no. 8, p. 25, Aug. 2011. [Online]. Available: http://ieeexplore.ieee.org/document/5960158/

[20] H. Kim, Y. J. Shin, and J. Kim, "Design and locomotion control of a hydraulic lower extremity exoskeleton for mobility augmentation," *Mechatronics*, vol. 46, pp. 32–45, Oct. 2017, doi: 10.1016/j.mechatronics.2017.06.009.

[21] J. Ghan, R. Steger, and H. Kazerooni, "Control and system identification for the Berkeley lower extremity exoskeleton (BLEEX)," *Adv. Robot.*, vol. 20, no. 9, pp. 989–1014, 2006. [Online]. Available: https://www.tandfonline.com/doi/pdf/10.1163/156855306778394012

[22] H. Kazerooni, R. Steger, and L. Huang, "Hybrid control of the Berkeley lower extremity exoskeleton (BLEEX)," *Int. J. Robot. Res.*, vol. 25, nos. 5–6, pp. 561–573, May 2006. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0278364906065505

[23] H.-T. Tran, H. Cheng, X. Lin, M.-K. Duong, and R. Huang, "The relationship between physical human-exoskeleton interaction and dynamic factors: Using a learning approach for control applications," *Sci. China Inf. Sci.*, vol. 57, no. 12, pp. 1–13, Dec. 2014. [Online]. Available: http://link.springer.com/10.1007/s11432-014-5203-8

[24] R. Huang, H. Cheng, Q. Chen, H.-T. Tran, and X. Lin, "Interactive learning for sensitivity factors of a human-powered augmentation lower exoskeleton," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Sep. 2015, pp. 6409–6415. [Online]. Available: http://ieeexplore.ieee.org/document/7354293/

[25] R. Huang, H. Cheng, H. Guo, X. Lin, Q. Chen, and F. Sun, "Learning cooperative primitives with physical human-robot interaction for a human-powered lower exoskeleton," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2016, pp. 5355–5360. [Online]. Available: http://ieeexplore.ieee.org/document/7759787/

[26] R. Huang, H. Cheng, H. Guo, Q. Chen, and X. Lin, "Hierarchical interactive learning for a human-powered augmentation lower exoskeleton," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2016, pp. 257–263. [Online]. Available: http://ieeexplore.ieee.org/document/7487142/

[27] R. Huang, H. Cheng, H. Guo, X. Lin, and J. Zhang, "Hierarchical learning control with physical human-exoskeleton interaction," *Inf. Sci.*, vol. 432, pp. 584–595, Mar. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0020025517309878

[28] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017. [Online]. Available: http://ieeexplore.ieee.org/document/8103164/

[29] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Found. Trends Mach. Learn.*, vol. 11, nos. 3–4, pp. 219–354, 2018. [Online]. Available: http://www.nowpublishers.com/article/Details/MAL-071

[30] X. B. Peng, P. Abbeel, S. Levine, and M. Van De Panne, "DeepMimic: Example-guided deep reinforcement learning of physics-based character skills," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–14, 2018. [Online]. Available: http://dx.doi.org/10.1145/3197517.3201311

[31] N. Chentanez, M. Müller, M. Macklin, V. Makoviychuk, and S. Jeschke, "Physics-based motion capture imitation with deep reinforcement learning," in *Proc. 11th Annu. Int. Conf. Motion, Interact., Games*, New York, NY, USA, Nov. 2018, pp. 1–10. [Online]. Available: https://dl.acm.org/doi/10.1145/3274247.3274506

[32] W. Yu, G. Turk, and C. K. Liu, "Learning symmetric and low-energy locomotion," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1–12, Aug. 2018. [Online]. Available: https://dl.acm.org/doi/10.1145/3197517.3201397

[33] B. Singh, R. Kumar, and V. Singh, "Reinforcement learning in robotic applications: A comprehensive survey," *Artif. Intell. Rev.*, vol. 55, pp. 945–990, Feb. 2022, [Online]. Available: https://link.springer.com/10.1007/s10462-021-09997-9

[34] J. Ahn, J. Lee, and L. Sentis, "Data-efficient and safe learning for humanoid locomotion aided by a dynamic balancing model," *IEEE Robot. Autom. Lett.*, vol. 5, no. 3, pp. 4376–4383, Jul. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9079565/

[35] L. C. Melo and M. R. O. A. Maximo, "Learning humanoid robot running skills through proximal policy optimization," in *Proc. Latin Amer. Robot. Symp. (LARS), Brazilian Symp. Robot. (SBR) Workshop Robot. Educ. (WRE)*, Oct. 2019, pp. 37–42. [Online]. Available: https://ieeexplore.ieee.org/document/9018554/

[36] C. Yang, T. Komura, and Z. Li, "Emergence of human-comparable balancing behaviours by deep reinforcement learning," in *Proc. IEEE-RAS 17th Int. Conf. Humanoid Robot. (Humanoids)*, Nov. 2017, pp. 372–377. [Online]. Available: http://ieeexplore.ieee.org/document/8246900/

[37] C. Yang, K. Yuan, S. Heng, T. Komura, and Z. Li, "Learning natural locomotion behaviors for humanoid robots using human bias," *IEEE Robot. Autom. Lett.*, vol. 5, no. 2, pp. 2610–2617, Apr. 2020. [Online]. Available: https://ieeexplore.ieee.org/document/8990011/

[38] R. Ozaln, C. Kaymak, O. Yildirum, A. Ucar, Y. Demir, and C. Guzelis, "An implementation of vision based deep reinforcement learning for humanoid robot locomotion," in *Proc. IEEE Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Jul. 2019, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8778209/

[39] Z. Xie, G. Berseth, P. Clary, J. Hurst, and M. Van De Panne, "Feedback control for Cassie with deep reinforcement learning," in *Proc. IEEE Int. Conf. Intell. Robots Syst.*, Oct. 2018, pp. 1241–1246.

[40] Z. Xie, P. Clary, J. Dao, P. Morais, J. Hurst, and M. Van De Panne, "Learning locomotion skills for Cassie: Iterative design and sim-to-real," in *Proc. PMLR*, 2019, pp. 317–329.

[41] F. Abdolhosseini, H. Y. Ling, Z. Xie, X. B. Peng, and M. Van De Panne, "On learning symmetric locomotion," in *Proc. Motion, Interact. Games*, Oct. 2019, pp. 1–10. [Online]. Available: https://dl.acm.org/doi/10.1145/3359566.3360070

[42] Z. Li, X. Cheng, X. B. Peng, P. Abbeel, S. Levine, G. Berseth, and K. Sreenath, "Reinforcement learning for robust parameterized locomotion control of bipedal robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2021, pp. 2811–2817. [Online]. Available: https://ieeexplore.ieee.org/document/9560769/

[43] J. Hwangbo, J. Lee, A. Dosovitskiy, D. Bellicoso, V. Tsounis, V. Koltun, and M. Hutter, "Learning agile and dynamic motor skills for legged robots," *Sci. Robot.*, vol. 4, no. 26, Jan. 2019, Art. no. eaau5872.

[44] J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning quadrupedal locomotion over challenging terrain," *Sci. Robot.*, vol. 5, no. 47, Oct. 2020, Art. no. eabc5986. [Online]. Available: https://www.science.org/doi/abs/10.1126/scirobotics.abc5986

[45] T. Miki, J. Lee, J. Hwangbo, L. Wellhausen, V. Koltun, and M. Hutter, "Learning robust perceptive locomotion for quadrupedal robots in the wild," *Sci. Robot.*, vol. 7, no. 62, Jan. 2022, Art. no. eabk2822, [Online]. Available: https://www.science.org/doi/10.1126/scirobotics.abk2822

[46] X. Bin Peng, E. Coumans, T. Zhang, T.-W. Lee, J. Tan, and S. Levine, "Learning agile robotic locomotion skills by imitating animals," Jul. 2020, *arXiv:2004.00784*. [Online]. Available: http://www.roboticsproceedings.org/rss16/p064.pdf

[47] S. Ha, P. Xu, Z. Tan, S. Levine, and J. Tan, "Learning to walk in the real world with minimal human effort," in *Proc. 4th Conf. Robot Learn.*, Cambridge, MA, USA, 2020, pp. 1–11.

[48] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing function approximation error in actor-critic methods," 2018, *arXiv:1802.09477*.

[49] 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," 2015, *arXiv:1509.02971*.

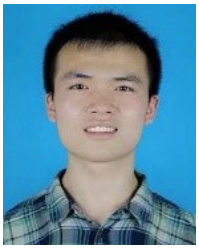
[50] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. 31st Int. Conf. Mach. Learn.*, vol. 1, 2014, pp. 387–395.

[51] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, and D. Wierstra, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, pp. 529–533, Feb. 2015. [Online]. Available: http://www.nature.com/articles/nature14236

RANRAN ZHENG received the B.S. degree in flight vehicle propulsion engineering from the Beijing Institute of Technology, Beijing, China, in 2016, where he is currently pursuing the Ph.D. degree with the Department of Flight Vehicle Control, School of Aerospace Engineering. His main research interests include mechanical designs, the modeling and control of exoskeletons, and locomotion control based on deep reinforcement learning.

ZHIYUAN YU received the Ph.D. degree in aircraft control from the Beijing Institute of Technology, Beijing, China, in 2009. He is currently with the Laboratory of Aerospace Servo Actuation and Transmission, Beijing Institute of Precision Mechatronics and Controls. His current research interests include aerospace servo actuation and transmission, wearable robots, and servo motors.

HONGWEI LIU received the B.S. degree in exploration guidance and control engineering and the M.S. degree from the Department of Flight Vehicle Control, School of Aerospace Engineering, Beijing Institute of Technology, in 2014 and 2017, respectively. He is currently with the Laboratory of Aerospace Servo Actuation and Transmission, Beijing Institute of Precision Mechatronics and Controls. His research interests include signal processing, system identification, and the control of wearable robotic devices.



ZHE ZHAO received the M.S. degree in mechanical engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2019. He is currently with the Laboratory of Aerospace Servo Actuation and Transmission, Beijing Institute of Precision Mechatronics and Controls. His current research interests include the mechanism design of exoskeletons and intelligent mechanism designs.

JING CHEN received the Ph.D. degree from the China Academy of Launch Vehicle Technology, Beijing, China, in 2020. She is currently with the Laboratory of Aerospace Servo Actuation and Transmission, Beijing Institute of Precision Mechatronics and Controls. Her current research interests include aerospace servo actuation and transmission and motor and exoskeleton robot control.

LONGFEI JIA received the Ph.D. degree from the Beijing Institute of Precision Mechatronics and Controls, Beijing, China, in 2022. He is currently with the Laboratory of Aerospace Servo Actuation and Transmission, Beijing Institute of Precision Mechatronics and Controls. His research interests include the kinematics, dynamics, and intelligence control of robots.