Learning Physical Human–Robot Interaction With Coupled Cooperative Primitives for a Lower Exoskeleton

Rui Huang, *Member, IEEE*, Hong Cheng, *Senior Member, IEEE*, Jing Qiu, and Jianwei Zhang

***Abstract*— Human-powered lower exoskeletons have received considerable interests from both academia and industry over the past decades, and encountered increasing applications in human locomotion assistance and strength augmentation. One of the most important aspects in those applications is to achieve robust control of lower exoskeletons, which, in the first place, requires the proactive modeling of human movement trajectories through physical human–robot interaction (pHRI). As a powerful representative tool for motion trajectories, dynamic movement primitives (DMP) have been used to model human movement trajectories. However, canonical DMP only offers a general representation of human movement trajectory and may neglects the interactive term, therefore it cannot be directly applied to lower exoskeletons which need to track human joint trajectories online, because different pilots have different trajectories and even same pilot might change his/her motion during walking. This paper presents a novel coupled cooperative primitive (CCP) strategy, which aims at modeling the motion trajectories online. Besides maintaining canonical motion primitives, we model the interaction term between the pilot and exoskeletons through impedance models, and propose a reinforcement learning method based on policy improvement and path integrals (PI^2) to learn the parameters online. Experimental results on both a single degree-of-freedom platform and a HUman-powered Augmentation Lower EXoskeleton (HUALEX) system demonstrate the advantages of our proposed CCP scheme.**

***Note to Practitioners*—This paper was motivated by the problem of lower exoskeleton when it interacts with different pilots. In both military and industrial applications of lower exoskeleton for strength augmentation, a most challenge problem is how to deal with the pHRI caused by different pilots. This paper suggests a new learning-based strategy, which modeled the pilot's motion with movement primitives and update through the pHRI between the pilot and the lower exoskeleton with online reinforcement learning method. In order to employ the proposed CCP into the real-time application, we also combine the CCP with a hierarchical control framework, and applied on a lower exoskeleton system which we built for strength augmentation application (which named as HUALEX). In the experiments of this paper, we validate the proposed CCP on different pilots with HUALEX system, the proposed CCP also achieve a good performance on the online learning and adaptation of the pilot's gait. In the future, we will extend this algorithm for adapting complex environment in both industrial and military applications, such as in different terrains, stairs, and slopes scenarios, and so on.**

***Index Terms*— Coupled cooperative primitive (CCP), dynamic movement primitive (DMP), lower exoskeletons, physical human–robot interaction (pHRI), reinforcement learning.**

I. Introduction

Lower extremity exoskeletons are wearable robotic systems which integrate human intelligence and robot power. They have evolved from the stuff of science fiction to quasi-commercialized products over the past few decades. With recent developments of wearable technologies, lower extremity exoskeletons have drawn considerable interests from researchers and raised increasing applications in human augmentation related scenarios [1]–[7], especially in industrial applications. As lower exoskeletons are tightly coupled with human beings in human augmentation related applications, they need to possess the ability to follow the pilot's motion with little interaction force. One prerequisite is that the lower exoskeleton needs to understand where the pilot intends to move. Therefore, it is crucial to model and predict human movement trajectories through physical human–robot interaction (pHRI) [8], [9].

Ijspeert *et al.* [10]–[12] proposed a powerful tool named dynamic movement primitives (DMP) for representing rhythmic and discrete trajectories. In imitation learning-related applications, DMP achieves good performances through learning by demonstration, especially in multidegree-of-freedom (DOFs) robot control scenarios [13]. Moreover, Gams *et al.* [14]–[16] employed additional interaction term in DMP to make the robot able to interact with the environment. With the interaction tasks with human, Amor *et al.* [17] proposed interaction primitives to interact with human with given conditions. Furthermore, compliant movement primitives are proposed to model and learn task-related movements through human demonstration [18]–[21]. However, these methods are only suitable for particular tasks.

Manuscript received August 26, 2018; revised November 22, 2018; accepted December 5, 2018. Date of publication January 10, 2019; date of current version October 4, 2019. This paper was recommended for publication by Associate Editor H. Liu and Editor H. Liu upon evaluation of the reviewers' comments. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1302300, in part by the National Natural Science Foundation of China under Grant 6150020696 and Grant 61503060, in part by the Sichuan Science and Technology Major Projects of New Generation Artificial Intelligence under Grant 2018GZDZX0037, and in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2015J148. *(Corresponding author: Hong Cheng.)*

R. Huang, H. Cheng, and J. Qiu are with the School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: hcheng@uestc.edu.cn).

J. Zhang is with the Department of Computer Science, University of Hamburg, 22527 Hamburg, Germany.

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2018.2886376

In applications of strength augmentation with lower exoskeletons, the pilot always move with different patterns, which leads that the pilot's motion should be predicted online [8]. In human augmentation related applications, model-based control strategies are quite efficiency since these controllers always need less sensory feedback. The sensitivity amplification control is the most famous model-based control method which proposed by Kazerooni *et al.* [22] in Berkeley lower extremity exoskeleton system. The SAC method only needs sensory feedback of joint encoders. However, the complicated identification process is needed for the SAC method [23]. In our previous work, we have proposed a hierarchical interactive learning (HIL) framework [24], [25] based SAC method, in which we use DMP to model the pilot's motion trajectories in motion learning hierarchy. The proposed HIL framework can update sensitivity factors of SAC controllers online through the reinforcement learning process. However, the pilot's motion trajectories can only be updated after one full gait cycle in the previous work. This means that when the pilot changes his/her gait pattern, he/she has to endure high interaction forces from lower exoskeletons during the same gait cycle. The interaction force can only be reduced after at least one full gait cycle. This may be undesirable as the interaction force might be unbearable for the pilot when he/she change his motion patterns frequently. All of these motivate us to innovate an algorithm which adapts the pilot's motion trajectory online.

In this paper, we present a novel coupled cooperative primitive (CCP) method to model and learn the motion trajectories online through pHRI between the pilot and the exoskeleton. Compared with the DMP used in HIL framework, the proposed CCP is able to adapt the predicted pilot's motion trajectory online. This feature enables the lower exoskeleton's application to scenarios where the pilot changes his/her motion patterns frequently. Different from conventional DMP, the proposed CCP incorporates an interaction model between the pilot and the exoskeleton as the modulation term. The modulation term, represented based on the impedance model, can be used to describe the interaction change between pilot and the exoskeleton. When the pHRI between the pilot and the exoskeleton changes over time, we employ a reinforcement learning method based on policy improvement and path integrals (PI^2) to adapt/tune the parameters of modulation term in CCP online. For different pilots of the lower exoskeleton, less than 30 gait cycles would be taken to achieve optimal CCP for online adaptation. This paper's contribution can be summarized as follows.

1) A novel CCP is proposed, which learns the pilot's motion trajectories online through pHRI; the interaction force between the pilot and the exoskeleton can be drastically reduced through online adaptation of CCP.
2) We employ a reinforcement learning method based on PI^2 to adapt the modulation term online, which does not require preassumed human-movement models, and can be applicable for different pilots.
3) The proposed CCP scheme has been verified on both a single DOF platform and the HUman-powered Augmentation Lower EXoskeleton (HUALEX) system. We first validate the proposed CCP on a single DOF exoskeleton platform, and then test it on a HUALEX system.

Experimental results show the online learning feature of the proposed CCP and its advantages over traditional offline DMP representation.

The structure of this paper is organized as follows. The materials and methods are introduced in Section II. In Section II, we first introduce the proposed CCP in Section II-A, and then lay down the reinforcement learning algorithm for online learning of modulation term in Section II-C. Experimental results on both the single DOF platform and the HUALEX system are presented and analyzed in Section III. This paper ends with conclusions and future work in Section IV.

II. Materials and Methods

A. Coupled Cooperative Primitives

This section presents the methodology details of the proposed CCP. We will first introduce the basic theory of DMP and its imitation learning process based on locally weighted regression (LWR) in Section II-A1. Then, we give a brief introduction of how to apply proposed CCP into HIL framework. In Section II-A2, we introduce CCP, which adds an impedance model on top of basic DMP to describe the pHRI between the pilot and the exoskeleton. In human-coupled lower exoskeleton applications, the added modulation term is essential for the online adaptive representation of the pilot's motion trajectories.

1) Dynamic Movement Primitives: DMP has been widely employed in robotic applications with a focus on imitation learning [11], [13]. Many types of DMP have been introduced in the literature [26]. In this paper, we employ a discrete DMP formulation provided in [10], since the terminate angle of joints may not be the same every gait cycle during walking. For a motion angle trajectory denoted by y, a DMP is defined as the following a set of nonlinear differential equations:

$$\tau \dot{z} = \alpha_z(\beta_z(g - y) - z) \tag{1}$$

$$\tau \dot{y} = z + f(x) \tag{2}$$

where $f(x)$ is a combination of nonlinear functions with Gaussian kernels

$$f(x) = \frac{\sum_{i=1}^{N} \omega_i \Psi_i(x)}{\sum_{i=1}^{N} \Psi_i(x)} x \tag{3}$$

$$\Psi_i(x) = \exp(-h_i(x - c_i)^2). \tag{4}$$

Here, N is the number of Gaussian kernels Ψ_i which have centers c_i and widths h_i. The parameters α_z and β_z in (1) are chosen as $\alpha_z = 4\beta_z$, and τ determines duration of the DMP, g is the goal parameter of the DMP. The phase variable x in the nonlinear function (3) is utilized to avoid direct dependence of f on time. In this paper, we use the first-order dynamics defined as

$$\tau \dot{x} = -\alpha_x \cdot x \tag{5}$$

where α_x is a positive constant, and the initial condition of x is $x(0) = 1$.

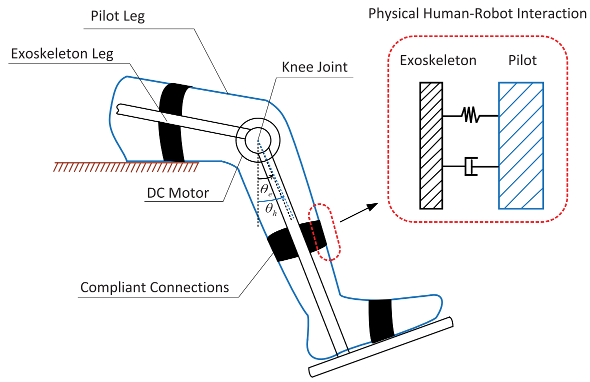


Fig. 1. Schematic of pHRI in single DOF exoskeleton system.

The shape can be seen as the most important feature of a motion trajectory, which is determined by a set of Gaussian kernel weights ω_i in (3). The imitation learning process of DMP is to use the LWR method [27] to optimize the weights ω_i through a given motion trajectory.

2) Modulation With pHRI: DMP has already been used in robotic applications for interacting with the environment [14], [28]. However, exoskeleton is a tightly coupled robot, we need to use the pHRI between pilot and the exoskeleton to represent the modulation term. In this paper, we propose a new primitive which blends both original DMP and the modulation term, and name it as CCP, because the new primitive brings the coupled relationship between pilot and the exoskeleton together in a cooperative way. We employ an impedance model to describe pHRI between the pilot and the exoskeleton. The schematic of pHRI in single DOF exoskeleton system is shown in Fig. 1.

As shown in Fig. 1, the impedance model can be defined as a spring–damper model which based on angle and angular error between the pilot and the exoskeleton

$$I = k_p(\theta_h - \theta_e) + k_d(\dot{\theta}_h - \dot{\theta}_e) \tag{6}$$

where k_p and k_d are the stiffness and damper parameters, respectively. θ_h and $\dot{\theta}_h$ are the pilot's angle and angular velocity, θ_e and $\dot{\theta}_e$ are the angle and angular velocity of exoskeleton joint.

In nonlinear differential equations of original DMP (1) and (2), $\dot{y}$ denotes the velocity level of output y and $\dot{z}$ denotes the acceleration level [14]. In applications of human augmentation exoskeleton systems, the change of human-exoskeleton motion usually caused by acceleration of the pilot. Hence, in this paper, the proposed CCP coupled the interaction model in both the velocity and acceleration levels

$$\tau \dot{z} = \alpha_z(\beta_z(g - y) - z) + c_d \dot{I} \tag{7}$$

$$\tau \dot{y} = z + f(x) + c_k I \tag{8}$$

where c_k and c_d are the scaling factors.

In order to achieve the accurate representation of motion trajectories through pHRI between the pilot and the exoskeleton, the scaling factors c_k and c_d must be tuned properly as well as parameters k_p and k_d in the impedance model. Sections II-B and II-C will introduce how to apply proposed CCP into HIL framework and how to utilize reinforcement learning methods for online adaptation of those parameters.

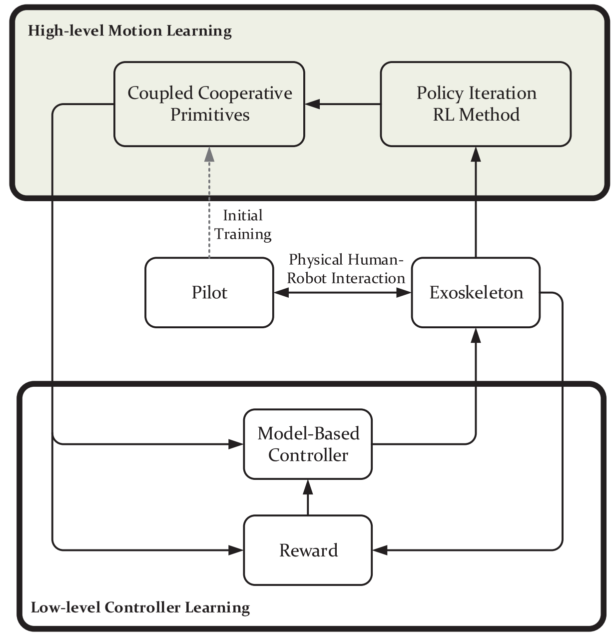


Fig. 2. Combining proposed CCP into HIL framework.

B. Combining CCP With HIL Framework

In the implementation of proposed CCP in lower exoskeleton control, we combine the proposed CCP into HIL framework which presented in our previous work [25]. In the original HIL framework, two learning hierarchies are included for learning motion trajectories of the pilot and the exoskeleton controller simultaneously.

Fig. 2 illustrates the framework that combines CCP and HIL. In the low-level controller learning hierarchy, model-based controller with reinforcement learning progress is implemented, which is the same as in original framework [25]. In the high-level motion learning hierarchy, we utilized the proposed CCP to replace the traditional DMP method, and employ policy iteration reinforcement learning method to learn parameters of the CCP. The learning process of policy iteration RL method will be introduced in detail in Section II-C.

C. Reinforcement Learning for Parameters Online Adaptation

In the previous work of modulation DMP for a particular interaction task, iteration learning control method has been employed to learn the parameters of modulation parts offline, which is based on the collected sensor information [14]. However, in the application of human-powered lower exoskeletons, the pHRI is changing from pilot to pilot even one pilot over time. Hence, the parameters in CCP need to be learned online to ensure that the modulation of motion trajectories is suitable for the pHRI between the pilot and the exoskeleton.

PI^2 algorithm is first proposed to learn the parameters of DMP via trial and error [29], and then it is applied in the learning-based impedance control strategy to obtain the impedance parameters of controllers for particular manipulating tasks [30]. In this paper, we expand the PI^2 algorithm for the online learning of the CCP-related parameters, namely, c_k, c_d, k_p, and k_d.

Parameterized policies are employed in CCP based on the PI^2 algorithm

$$\mathbf{a}_t = \mathbf{G}_t^T(\Theta + \epsilon_t) \tag{9}$$

where Θ is a parameter vector $[c_k, c_d, k_p, k_d]^T$ and ϵ_t is interpreted as exploration noise. $\mathbf{G}_t$ represent the basis functions with Gaussian kernels w

$$[\mathbf{G}_t]_j = \frac{w_j}{\sum_i^n w_i}. \tag{10}$$

The jth average weight of total n is calculated following (10), where n is the number of parameters that needs to be learned. The Gaussian kernels w are set to be the same formula as in (4).

In the implementation of lower exoskeleton, the immediate reward function is calculated from the sensory feedback. For example, in single DOF exoskeleton, the immediate cost function is defined as follows:

$$r_t = \beta_1\left(\theta_t - \theta_t^*\right)^2 + \beta_2\left(\dot{\theta}_t - \dot{\theta}_t^*\right)^2 \tag{11}$$

where θ and θ^* are the sensory feedbacks from the encoder and the inclinometer, which represent the angle of the exoskeleton θ_e and the pilot θ_h, respectively. The differentials represent the angular velocities.

After defining the actions and cost function, we implement the PI^2 algorithm to update the parameters of CCP of single DOF exoskeleton online. The pseudocode of improved PI^2 algorithm for CCP of a single DOF exoskeleton is given in Table I. As described in Table I, the parameters of CCP will be updated every K gait cycles. The update rule is given in Table I with the following equations:

$$\mathbf{M}_{t_i,k} = \frac{H^{-1}\mathbf{G}_{t_i,k}\mathbf{G}_{t_i,k}^T}{\mathbf{G}_{t_i,k}^T H^{-1}\mathbf{G}_{t_i,k}} \tag{12}$$

$$\mathbf{S}_{i,k} = \sum_{j=i}^{n-1} r_{t_j,k} + \frac{1}{2}\sum_{j=i+1}^{n-1}(\Theta + \mathbf{M}_{t_j,k}\epsilon_{t_j,k})^T \times H(\Theta + \mathbf{M}_{t_j,k}\epsilon_{t_j,k}) \tag{13}$$

$$\mathbf{P}_{i,k} = \frac{e^{-\frac{1}{\lambda}S_{i,k}}}{\sum_{k=1}^K\left[e^{-\frac{1}{\lambda}S_{i,k}}\right]} \tag{14}$$

$$\Delta\Theta_{t_i} = \sum_{k=1}^K [\mathbf{P}_{i,k}\mathbf{M}_{t_i,k}\epsilon_{t_i,k}] \tag{15}$$

$$[\Delta\Theta]_j = \frac{\sum_{i=1}^\rho(\rho - i)w_{j,t_i}[\Delta\Theta_{t_i}]_j}{\sum_{i=1}^\rho w_{j,t_i}(\rho - i)}. \tag{16}$$

The matrix H in (12) is a positive semidefinite weight matrix, and the parameter λ in (14) is set automatically within $(0, 1]$. After obtaining the updated parameters Θ in CPPs, the exoskeleton should take one noiseless gait cycle (here, the term "noiseless" means that we update parameters Θ without exploration noise ϵ_t) to compute the trajectory cost R, which determines whether the optimization process is terminated. The trajectory cost R is computed as follows:

TABLE I
PSEUDOCODE OF IMPROVED PI^2 ALGORITHM FOR UPDATING PARAMETERS OF CCP OF A SINGLE DOF EXOSKELETON

1.	Initialize the parameter vector Θ
2.	Initialize the basis function $\mathbf{G}_{t_i}$ according (10)
3.	**Repeat**
4.	Run K gait cycles (roll-outs) of the exoskeleton using stochastic parameters $\Theta + \varepsilon_t$ at every time step
5.	**For** all gait cycles $k \in [1, K]$:
6.	Compute the projection matrix $\mathbf{M}$ through (12)
7.	Compute the stochastic cost $\mathbf{S}$ through (13)
8.	Compute the probability $\mathbf{P}$ through (14)
9.	**For** all time steps $i \in [1, \rho]$:
10.	Compute $\Delta\Theta_{t_i}$ for each time step through (15)
11.	Normalize $\Delta\Theta$ according (16)
12.	Update $\Theta \leftarrow \Theta + \Delta\Theta$
13.	Run one noiseless gait cycle to compute the trajectory cost R through (17)
14.	**Until** Trajectory cost R is converged

$$R = \sum_{i=1}^\rho r_{t_i}. \tag{17}$$

With the online adaptation of CCP parameters, the pHRI between the exoskeleton and the pilot can be reduced since the motion trajectories of the pilot have been predicted through CCP.

III. EXPERIMENTS

In this section, we validate the proposed CCP scheme through experiments on both single DOF exoskeleton platform and the HUALEX system. Experimental setup, results, and analysis are provided in Sections III-A and III-B.

A. Prior Verification in Single DOF Exoskeleton

1) Single DOF Exoskeleton Platform: Fig. 1 illustrates the schematic of the single DOF exoskeleton platform. In single DOF exoskeleton, the pilot's leg is attached with the exoskeleton in the thigh and shank, which makes only the swing movement in the knee joint possible (i.e., single DOF).

The exoskeleton is powered by a direct current (dc) motor, which provides active torques on the exoskeleton system at the knee joint. An encoder is embedded in the knee joint to measure the angle information of the exoskeleton. The pilot is bounded on the exoskeleton via compliant connections at lower links. Because of the compliant connections, the pilot can make contact with exoskeleton flexibly and impose forces to the exoskeleton. In order to measure the pilot's motion, two inclinometers are installed on the compliant connections of shank and thigh (black areas in Fig. 1).

2) Simulation Experimental Setup: The single DOF exoskeleton platform is built in the Simulink environment with a imported control plant from ADAMS software. A motion generator in Simulink is utilized to generate the pilot's motion of simulation experiments, which should be imposed to the exoskeleton plant via a physical compliant model in ADAMS. In order to compare the proposed CCP with the original DMP, the HIL framework [24] is employed in the experiment. HIL is a learning-based control framework which contains two learning hierarchies, namely, motion learning hierarchy and controller learning hierarchy. In the original HIL framework, DMP is utilized to model and learn the motion trajectories in the motion learning hierarchy. In the experiments of this paper, we employ the proposed CCP into the motion learning hierarchy in HIL framework (called CCP-HIL). Note that in this paper, we only focus on the results and discussions of the motion learning hierarchy.

In the numerical simulation experiments on single DOF exoskeleton, the pilot's motion is set as periodic sine waves but with different frequencies and amplitudes. Before controlling the exoskeleton via CCP, the weights of the Gaussians [ω_i in (3)] and the frequency τ need to be trained through sampled pilot motion trajectories. Here, we initialize the parameters of modulation parts in CCP (parameter vector Θ) as zero vector, and the number of Gaussians in CCP is set to be 25. The LWR method [27] is employed to learn the Gaussian weights incrementally. During the learning process of CCP parameters, the frequency τ is also updated through the time period of every gait cycle.

After obtaining the weights parameters in CCP, the exoskeleton will take several gait cycles to learn the optimal parameters Θ according to Table I, which makes the trajectory cost R converge. In the experiments, the exoskeleton runs for four ($K = 4$) gait cycles to update the parameters and one gait cycle to calculate the trajectory cost R. Thus, the parameters Θ will be updated every five gait cycles. The time steps ρ are normalized to be 50 based on the time interval of each gait cycle. Table II shows the values and meanings of parameters in CCP and the learning process.

3) Experimental Results: In the numerical simulation of single DOF exoskeleton, two experiments are carried out to validate the proposed CCP in the HIL framework. The first experiment is to evaluate the efficiency of the reinforcement learning algorithm in CCP. In this experiment, the exoskeleton runs to follow the pilot's motion for several gait cycles in order to make the trajectory cost R convergence. The learning curve of the proposed CCP in single DOF exoskeleton is shown in Fig. 3.

As shown in Fig. 3, the trajectory R converges after the exoskeleton has been running for about 100 gait cycles (the parameters Θ are updated 20 times. Fig. 4 shows the updating process of parameters in CCP before the trajectory cost R converged (in total 100 gait cycles).

After the online learning of the parameters of the modulation part in CCP, we compare the performance of CCP-HIL and the original HIL strategy with DMP. The pilot's motion is set to a total of 11 gait cycles with different frequencies and amplitudes in this experiment. Fig. 5 shows the performances of the CCP-HIL and original HIL strategy with DMP in single DOF exoskeleton, where the interaction force between the pilot and the exoskeleton is calculated through (6) with the learned impedance parameters. In the original HIL strategy with DMP in Fig. 5(a), the interaction force in the transition gait cycles increases significantly when the pilot changes the gait cycle and the force will decrease only after the full gait cycle. This is because DMP needs a whole gait cycle to imitate the motion trajectories. While Fig. 5(b) shows that the

TABLE II

VALUES AND MEANINGS OF PARAMETERS IN CCP AND THE LEARNING PROCESS

Parameters	Values	Meanings
N	25	number of Gaussians in CCP
τ	1	initial value of frequency of CCP
α_z	4	dynamic parameter of CCP
β_z	1	dynamic parameter of CCP
α_x	2	oscillator parameter of CCP
σ_i	0.0386	variance of Gaussians
λ	0.9	discount factor of learning process
T	50	normalized time steps
β_1	1000	parameter of reward function
β_2	500	parameter of reward function
K	4	number of noise gait cycles

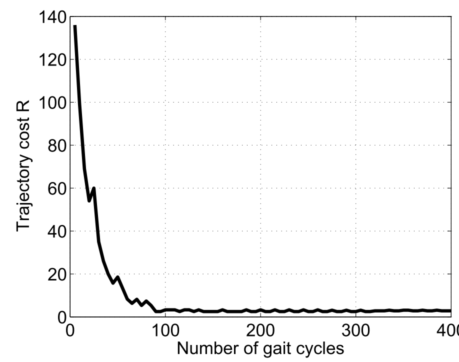


Fig. 3. Learning curve of the proposed CCP in single DOF exoskeleton.

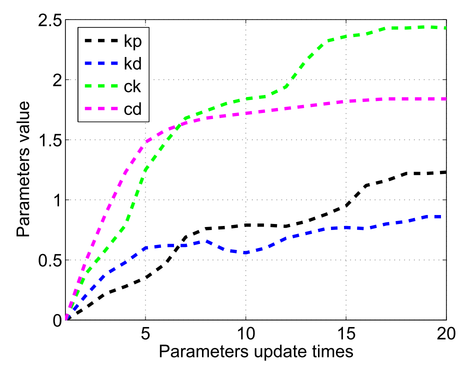


Fig. 4. Updating process of parameters in CCP during the previous 100 gait cycles (20 times of updates).

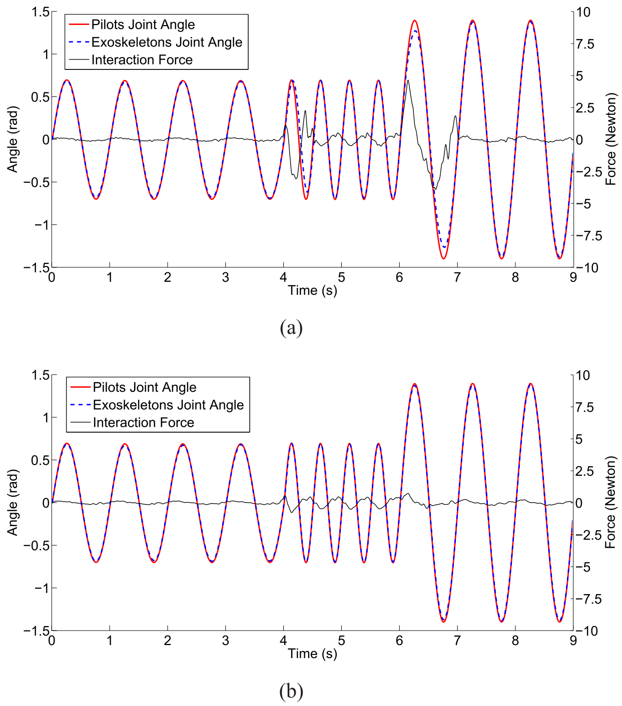


Fig. 5. Comparison of performances of (a) original HIL strategy and (b) combining CCP in HIL framework.

TABLE III
INTERACTION FORCE OF THE TRANSITION GAIT CYCLES IN SINGLE DOF SIMULATION

Interaction force (N) (DMP-HIL|CCP-HIL)	Maximum		Mean	
5^{th} gait cycle	3.04	0.84	2.54	0.36
9^{th} gait cycle	4.63	0.76	3.18	0.25

proposed CCP can reduce the pHRI within the same transition gait cycle because the motion trajectories are corrected online through the modulation parts in CCP.

Table III shows the comparison results of the interaction force in the transition gait cycles (fifth and ninth gait cycles in our experiment). In Table III, we can see the interaction forces of the proposed CCP are significantly smaller than the original DMP (both in terms of the maximal force and mean force).

During normal walking of the pilot when operate the exoskeleton system, the gait of the pilot will not always be regular, which means the motion trajectories are different with the near gait cycles. Therefore, we utilize irregular movements with tightly difference in each gait cycle to simulate the pilot's motion. Fig. 6 shows the performance of CCP-HIL with irregular movements. The experimental results show that the proposed CCP has the ability to deal with frequent changing movements with slightly difference (the mean of interaction forces during total six gait cycles is $0.24N$).

B. Experiments on the HUALEX System

1) HUALEX System: HUALEX system is designed as an ergonomic, robust, and lightweight equipment for human

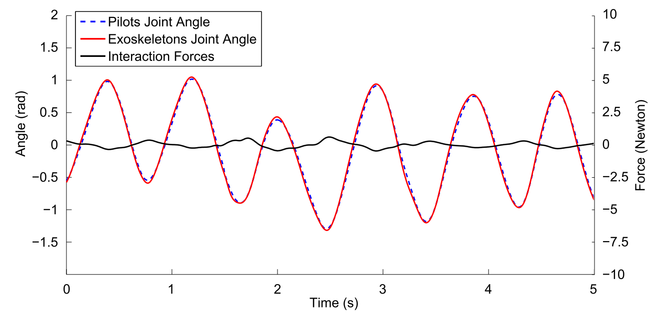


Fig. 6. Performances of CCP-HIL with irregular movements in single DOF simulation. The mean of the interaction torques during total six gait cycles is $0.24N$.

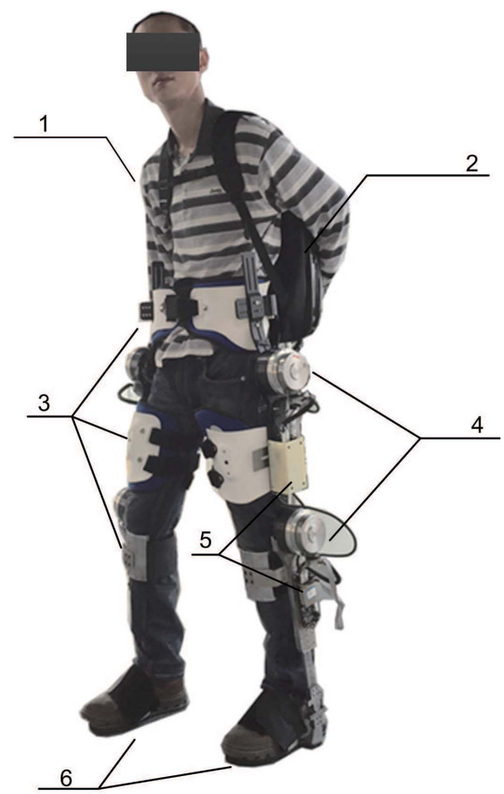


Fig. 7. HUALEX with the pilot. 1: pilot. 2: load backpack with the power unit and main controller (rigid connection with the HUALEX spline). 3: semirigidly connecting HUALEX to the pilot (waist, thighs, shanks, and feet). 4: active joints with dc servo motors (hip joints and knee joints). 5: node controllers for active joints. 6: smart shoes with plantar sensors.

augmentation applications. As shown in Fig. 7, HUALEX has total four active joints in hips and knees to provide driving torques, which are activated by dc servo motors (90 W, 150 gear ratio). The ankle joints of HUALEX are designed as an energy storage mechanism which stores energy in the stance phase and release it in the swing phase during walking. Besides the joints and rigid links, many semirigid connections at waist, thighs, shanks, and feet are provided to connect HUALEX to the pilot, which aiming to make the connections compliantly.

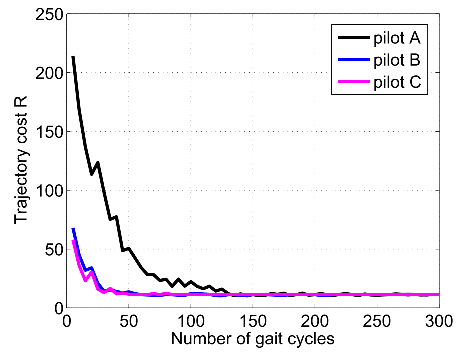
A distributed control system is embedded in HUALEX system, which consists of a main controller and four node controllers. The main controller is set at the backpack, which runs the perception and control algorithms. Near each active joint, a node controller is installed for two purposes: 1) collecting sensor data for the main controller and 2) executing commands from the main controller.

In order to ensure the real-time performance of the control algorithm, node controllers communicate with the main controllers via controller area network. HUALEX has a total of three kinds of sensors for its current state monitoring. Encoders are integrated in joint actuators, which measure the current state of each joint. Moreover, inclinometers are installed on the compliant connections which aiming to measure the motion of the pilot. An accelerometer is set at the backpack to record the walking velocity of the pilot. Plantar sensors are placed in the sole for foot pressure measurement.

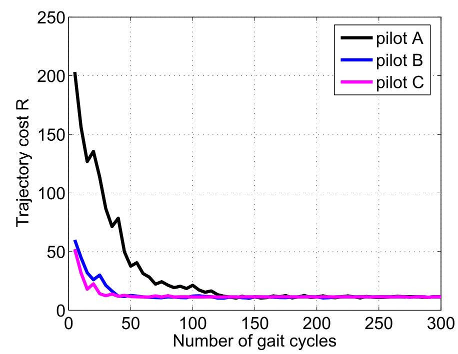
2) Experimental Setup: In the experiments of human-powered exoskeleton system, three pilots (A∼170 cm/75 kg/25 age, B∼176 cm/85 kg/27 age, and C∼180 cm/100 kg/27 age) with different heights are chosen to operate the HUALEX system in sequence, which means that the learned parameters of CCP with pilot A will be regarded as initial values of CCP with pilot B. The HUALEX system is operated in different environment situations (flat, slope, and stair) with different walking speeds. The proposed CCP is running in each node controller independently, and the HIL framework is also embedded in the node controllers for achieving better control performance. The plantar sensors are utilized to obtain the start–end time point of each gait cycle, and the accelerometer measures the walking speed of the pilot which relate to the frequency parameters in CCP.

Before operating the HUALEX system, the Gaussian weights in each CCP are trained through normal walking motion trajectories from clinical gait analysis data set. During the online reinforcement learning process of HUALEX system, the interaction models in CCP will be influenced by the variation of compliant connections, such as locomotion caused by the pHRI. Therefore, the termination condition (also the start condition of reinforcement learning process after convergence) of each CCP in PI^2 algorithm (14 line in Table I) is changed to $R < \eta$ (where η is a positive threshold). Other experimental settings are set as the same as in experiments of single DOF exoskeleton, and the threshold η is chosen as 10 for each CCP. The parameter settings are set as the same in Table II except the number of the Gaussian kernel in CCP. In the experiments of HUALEX system, the number of the Gaussian kernel in CCP was chosen as 10 in order to ensure the real-time performance on HUALEX system.

3) Results and Discussions: Fig. 8 shows the learning curves of CCP in HUALEX system with different pilots (the joints of right lower limb). As discussed in the experimental setup, pilot A operates the HUALEX system first so that the online learning process of CCP need to spend more gait cycles (almost 130 gait cycles as shown in Fig. 8). After obtaining the optimal parameters Θ from pilot A, the HUALEX system only need less than 50 gait cycles to convergence when the operator changes to pilots B and C. Table IV shows the optimal parameters in CCP of the right knee joint.



(a)



(b)

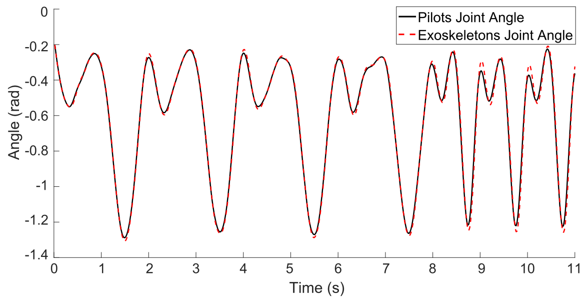
Fig. 8. Learning curves of CCP in the HUALEX system with different pilots at (a) right knee joint and (b) right hip joint.

TABLE IV
OPTIMAL PARAMETERS IN CCP OF THE HUALEX SYSTEM IN RIGHT KNEE JOINT

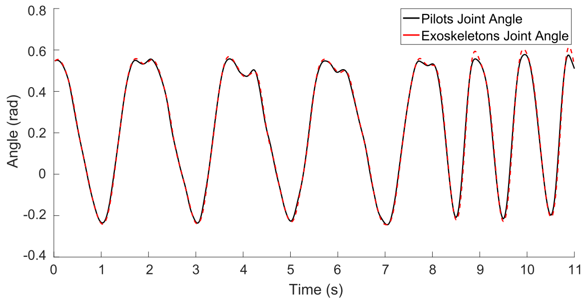
Parameter values (Knee|Hip)	k_p	k_d	c_k	c_d
pilot A	1.16 | 0.98	1.89 | 1.77	3.04 | 3.88	2.24 | 3.45
pilot B	1.34 | 1.16	1.34 | 2.75	2.65 | 2.81	2.03 | 4.55
pilot C	1.45 | 1.87	1.62 | 2.05	2.62 | 3.24	2.58 | 3.88

After the online learning of the parameters in CCP, the HUALEX system has a good performance in dealing with different walking patterns. Fig. 9 illustrates the control performance of the proposed CCP combining with the HIL strategy on pilot C after obtaining the optimal parameters in CCP. As shown in Fig. 9, the proposed algorithm can make the exoskeleton system follow the pilot's motion with little tracking error in the transition cycles, which indicates that the pHRI between the pilot and the HUALEX system can be reduced significantly in the transition gait cycles.

Through Table IV, we can also infer that the proposed CCP has the ability to adapt different pilots according to update the parameters of CCP online. Table V shows the performances of CCP on pilots B and C, which including both the results before learning and after learning. Pilot A is not concluded in this experiment, since before online learning of CCP with pilot A, the initial parameters of CCP are both zero (which means the modulation terms of CCP is not initialized). The



(a)



(b)

Fig. 9. Control performance of CCP combining with the HIL strategy on HUALEX system with pilot C in a total of 11-s walking (period of whole experiment) at the (a) right knee joint and (b) right hip joint.

TABLE V
PERFORMANCES OF CCP BEFORE AND AFTER LEARNING ON PILOT B AND C

nMSE of Motion Trajectories (rad)	Before Learning	After Learning
pilot B	0.115	0.034
pilot C	0.095	0.026

experimental results in Table V show that the CCP has the ability to adapt different pilot and achieve better performances after online learning, which can track the pilot's motion faster (with less nMSE in motion trajectories' tracking).

The experimental results on both single DOF exoskeleton and HULEX show that, through online reinforcement learning of parameters, the CCP is able to adapt with the changing pHRI between different pilots and the exoskeleton.

IV. CONCLUSION

This paper has proposed a CCP for the robust control of a HUALEX robotic system. The CCP can model and learn the pilot's motion trajectories online through the pHRI between the pilot and the exoskeleton. Reinforcement learning method based on PI^2 algorithm is employed to update the parameters of modulation terms in CCP, which adapts with the changing pHRI online. The performance of the CCP has been evaluated on a single DOF exoskeleton platform as well as the HUALEX robotic system. Experimental results show that the proposed CCP is able to model and learn the pilot's motion trajectories online through pHRI between the pilot and the exoskeleton.

In the future, we will build a hybrid agent framework to model the human-coupled exoskeleton, which has both inner interactions (interaction between the pilot and the exoskeleton) and external interactions (interaction with the environment). In this framework, the motion models and controllers need to consider both interactions with the pilot and the environment. Moreover, understanding the environment with extra sensors is also a critical issue for robust control in complex environments.

REFERENCES

[1] H. Kazerooni, A. Chu, and R. Steger, "That which does not stabilize, will only make us stronger," *Int. J. Robot. Res.*, vol. 26, no. 1, pp. 75–89, 2007.

[2] C. J. Walsh, D. Paluska, K. Pasch, W. Grand, A. Valiente, and H. Herr, "Development of a lightweight, underactuated exoskeleton for load-carrying augmentation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2006, pp. 3485–3491.

[3] 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./Oct. 2015, pp. 6409–6415.

[4] K. Gui, Y. Ren, and D. Zhang, "Online brain-computer interface controlling robotic exoskeleton for gait rehabilitation," in *Proc. IEEE Int. Conf. Rehabil. Robot. (ICORR)*, Aug. 2015, pp. 931–936.

[5] Y. Ren and D. Zhang, "FEXO Knee: A rehabilitation device for knee joint combining functional electrical stimulation with a compliant exoskeleton," in *Proc. IEEE Int. Conf. Biomed. Robitcs Biomechatron.*, Aug. 2014, pp. 683–688.

[6] X. Wu, D.-X. Liu, M. Liu, C. Chen, and H. Guo, "Individualized gait pattern generation for sharing lower limb exoskeleton robot," *IEEE Trans. Autom. Sci. Eng.*, vol. 15, no. 4, pp. 1459–1470, Oct. 2018.

[7] H. Liu, F. Sun, and X. Zhang, "Robotic material perception using active multi-modal fusion," *IEEE Trans. Ind. Electron.*, to be published, doi: 10.1109/TIE.2018.2878157.

[8] 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. 1–11, 2014.

[9] J. Huang, W. Huo, W. Xu, S. Mohammed, and Y. Amirat, "Control of upper-limb power-assist exoskeleton using a human-robot interface based on motion intention recognition," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1257–1270, Oct. 2015.

[10] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Movement imitation with nonlinear dynamical systems in humanoid robots," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2002, pp. 1398–1430.

[11] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Learning rhythmic movements by demonstration using nonlinear oscillators," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2002, pp. 958–963.

[12] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: Learning attractor models for motor behaviors," *Neural Comput.*, vol. 25, no. 2, pp. 328–373, 2013.

[13] J. Nakanishi, J. Morimoto, G. Endo, G. Cheng, S. Schaal, and M. Kawato, "Learning from demonstration and adaptation of biped locomotion," *Robot. Auto. Syst.*, vol. 47, pp. 79–91, Jun. 2004.

[14] A. Gams, B. Nemec, A. J. Ijspeert, and A. Ude, "Coupling movement primitives: Interaction with the environment and bimanual tasks," *IEEE Trans. Robot.*, vol. 30, no. 4, pp. 816–830, Aug. 2014.

[15] A. Gams *et al.*, "Adaptation and coaching of periodic motion primitives through physical and visual interaction," *Robot. Auto. Syst.*, vol. 75, pp. 340–351, Jan. 2016.

[16] A. Gams, T. Petrič, B. Nemec, and A. Ude, "Learning and adaptation of periodic motion primitives based on force feedback and human coaching interaction," in *Proc. IEEE Int. Conf. Humanoid Robots*, Nov. 2014, pp. 166–171.

[17] H. B. Amor, G. Neumann, S. Kamthe, O. Kroemer, and J. Peters, "Interaction primitives for human-robot cooperation tasks," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2014, pp. 2831–2837.

[18] T. Petrič, L. Colasanto, A. Gams, A. Ude, and A. J. Ijspeert, "Bio-inspired learning and database expansion of compliant movement primitives," in *Proc. IEEE Int. Conf. Humanoid Robots*, Nov. 2015, pp. 346–351.

[19] T. Petrič, A. Ude, T. Borangiu, and A. J. Ijspeert, "Autonomous learning of internal dynamic models for reaching tasks," *Adv. Robot Des. Intell. Control*, vol. 371, pp. 439–447, May 2015.

[20] M. Deniša, A. Gams, A. Ude, and T. Petrič, "Learning compliant movement primitives through demonstration and statistical generalization," *IEEE/ASME Trans. Mechatronics*, vol. 21, no. 5, pp. 2581–2594, Oct. 2015.

[21] Z. Cao, H. Guo, J. Zhang, F. A. Oliehoek, and U. Fastenrath, "Maximizing the probability of arriving on time: A practical Q-learning method," in *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2017, pp. 4481–4487.

[22] H. Kazerooni, J.-L. Racine, L. Huang, and R. Steger, "On the control of the Berkeley lower extremity exoskeleton (BLEEX)," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Apr. 2005, pp. 4353–4360.

[23] 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.

[24] 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.

[25] 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.

[26] S. Schaal, J. Peters, J. Nakanishi, and A. J. Ijspeert, "Learning rhythmic movements by demonstration using nonlinear oscillators," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2003, pp. 1–21.

[27] S. Schaal and C. Atkeson, "Constructive incremental learning from only local information," *Neural Comput.*, vol. 10, no. 8, pp. 2047–2084, 1998.

[28] H. Hoffmann, P. Pastor, D. H. Park, and S. Schaal, "Biologically-inspired dynamical systems for movement generation: Automatic real-time goal adaptation and obstacle avoidance," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2009, pp. 2587–2592.

[29] E. Theodorou, J. Buchli, and S. Schaal, "A generalized path integral control approach to reinforcement learning," *J. Mach. Learn. Res.*, vol. 11, pp. 3137–3181, Nov. 2010.

[30] J. Buchli, F. Stulp, E. A. Theodorou, and S. Schaal, "Learning variable impedance control," *Int. J. Robot. Res.*, vol. 30, pp. 820–833, Apr. 2011.

Rui Huang (M'18) received the Ph.D. degree in control science and engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in July 2018.

He was a joint training doctoral student with TAMS, University of Hamburg, Hamburg, Germany, from 2016 to 2017. He is a post-doctor at the Center for Robotics, School of Automation Engineering, UESTC. His current research interests include reinforcement learning, exoskeleton, human–robot interaction, and robot control.

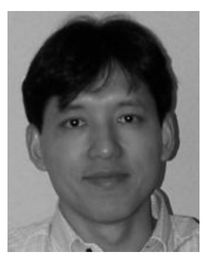
Hong Cheng (M'06–SM'14) received the Ph.D. degree in pattern recognition and intelligent systems from Xi'an Jiaotong University (XJTU), Xi'an, China, in 2003.

He is currently a Full Professor with the School of Automation and the Vice Director of the Center for Robotics, University of Electronic Science and Technology of China, Chengdu, China, (UESTC), where he is currently the Founding Director of the Machine Intelligence Institute. Before this, he held a post-doctoral position at the Computer Science School, Carnegie Mellon University, Pittsburgh, PA, USA, from 2006 to 2009. He has been an Associate Professor with XJTU since 2005. Since 2000, he has been with XJTU, where he has been a Team Leader of the Intelligent Vehicle Group, Institute of Artificial Intelligence and Robotics, before going to the USA. The team that he was leading in XJTU developed an intelligent driving platform–Springrobot, which has an important social effect in China. His current research interests include computer vision and machine learning, robotics, human–computer interaction, and multimedia signal processing.

Dr. Cheng is a Senior Member of the ACM. He is a reviewer of many important journals and conferences (IEEE TITS, MAV, CVPR, ICCV, ITSC, IVS, and ACCV). He serves as the Registration Chair for 2005 IEEE ICVES, the Local Arrangement Chair for VLPR 2012, and the Finance Chair for ICME 2014. He is also an Associate Editor of the *IEEE Computational Intelligence Magazine*.

Jing Qiu received the Ph.D. degree in human factors and ergonomics from the Darmstadt University of Technology, Darmstadt, Germany, in 2010.

She is currently an Associate Professor with the Center for Robotics, School of Mechanical Engineering, University of Electronic Science and Technology of China, Chengdu, China. Her current research interests include human factors in exoskeleton, human–robot interaction, and rehabilitation robot.



Jianwei Zhang received the B.E. (Hons.) and M.E. degrees from the Department of Computer Science, Tsinghua University, Beijing, China, in 1986 and 1989, respectively, the Ph.D. degree from the Department of Computer Science, Institute of Real-Time Computer Systems and Robotics, University of Karlsruhe, Karlsruhe, Germany, in 1994, and the Habilitation degree from the Faculty of Technology, University of Bielefeld, Bielefeld, Germany, in 2000.

He is currently a Professor and the Head of the TAMS, Department of Informatics, University of Hamburg, Hamburg, Germany. He is also a life-long Academician of the Academy of Sciences in Hamburg, Germany. He has published about 300 journal and conference papers, technical reports, four book chapters, and three books in his research areas. He holds 37 patents on service robot components and systems. He is the coordinator of the DFG/NSFC Transregional Collaborative Research Centre SFB/TRR169 for crossmodal learning and several EU robotics projects. His research interests are service robotics, sensor fusion, service robotics and multimodal machine learning, and cognitive computing of Industry 4.0.

Dr. Zhang has received several awards, including the IEEE ROMAN Best Paper Award in 2002, the IEEE AIM Best Paper Award 2008, the IEEE ROBIO Best Conference Paper Award 2013, and the ROBIO Best Tarn Paper on Robotics 2016. He is the General Chair of the IEEE MFI 2012 and the IEEE/RSJ IROS 2015, and the IEEE Robotics and Automation Society AdCom (2013–2015).