Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Optimized Control for Human-Multi-Robot Collaboration via Multi-Agent Adaptive Dynamic Programming

Xing Liu * **Shuzhi Sam Ge** **

* *Shaanxi Key Laboratory of Intelligent Robots, Xi'an Jiaotong University, Xi'an, Shaanxi 710049 China, (e-mail: liuxing1990.ok@163.com).*
** *Department of Electrical and Computer Engineering, National University of Singapore, Singapore, 117576, and the Institute for Future, Qingdao University, Qingdao, China, 266071 (e-mail: samge@nus.edu.sg)*

Abstract: In this paper we consider the problem of controlling the dynamic behavior of the robot agents while collaborating with the human worker. The presented dynamic behavior control method leads to achieving optimized interaction performance of the human-multi-robot collaboration system. We investigate in depth the dynamics equation of the robot agents collaborating with the human worker. Considering the unknown parameters in the system dynamics, the adaptive dynamic programming method is utilized to deal with the optimized interaction control problems during human-multi-robot collaboration process. To achieve the coordination of the multi robot agents, multi-agent adaptive dynamic programming method is employed in this paper. The neural networks with one hidden layer are utilized to approximate both the unknown system dynamics as well as the optimized cost function. The simulation studies verify the effectiveness of the presented algorithm.

Keywords: Human-multi-robot collaboration, optimized control, dynamic behavior control, multi-agent adaptive dynamic programming, neural network approximation.

1. INTRODUCTION

The collaborative robots integrate into human society and conduct close and complex interactions with human beings, e.g., physical interactions such as medical care, rehabilitaion, multi-robot collaboration, and human-robot collaboration in the manufacturing industry Li et al. (2011). Human-robot collaboration control and multi-robot collaboration control have been extensively investigated in many fields Li et al. (2015b,a); Reed and Peshkin (2008); Moertl et al. (2012); Lawitzky et al. (2010); Noohi et al. (2016); Wu et al. (2016). Human-robot collaboration integrates the advantages of the human being and the robot, to better carry out the tasks. For multi-robot collaboration, the capability of a single robot is expanded. Human-multi-robot collaboration further combines the advantages of human-robot collaboration and multi-robot collaboration, and will be utilized in more and more application scenarios, such as human-multi-robot interaction Sklar et al. (2013), human-multi-robot search and rescue Cacace et al. (2016), in which the human worker leads, supervises, or coordinates the cooperation among multi robot agents. However, there are few researches in the field of physical human-multi-robot collaboration, which is the most important motivation of this paper.

⋆ This work was supported in part by the National Natural Science Foundation of China under Grant 91748208, and in part by the China Scholarship Council under Grant 201706280378.

Impedance control has been widely used in physical human-robot collaboration control and multi-robot collaboration control Albu-Schäffer et al. (2007); Ge et al. (2014); Li et al. (2015b,a), which regulates the dynamic behavior at the interaction points rather than the interaction force Hogan (1985). Impedance control is a special case of dynamic behavior control. The impedance parameters describe a simple and compact relationship between the robot motion and the interaction force at the interaction point. Whereas in practice, there are many kinds of dynamic behavior models, which cannot be limited to impedance model. A typical example is the human-multi-robot interaction control problem. In human-multi-robot interaction, the controller for one robot should consider not only the states of this robot, but also the states of other robots and the human worker. Obviously, this cannot be described by the traditional impedance model and a more general dynamic behavior model is needed.

For human-multi-robot collaboration control, optimization plays an important role because the control objective of interaction control includes both the force regulation and trajectory tracking and usually it is the tradeoff of these two objectives Ge et al. (2014). This mimics the human beings' adaptation of force and impedance as a concurrent minimization of instability, motion error and metabolic cost in muscle space Franklin et al. (2008). In the previous literature, the Linear Quadratic Regulator (LQR) is selectedin to determe the impedance parameters but the environment dynamics needs to be known Matinfar and Hashtrudi-Zaad (2005). For practical applications, this always does not hold. Adaptive dynamic programming (ADP) method has been widely studied in the previous literature Vrabie et al. (2009), Gao et al. (2016) to realize optimal control of the systems with unknown dynamics. Several publications utilized ADP method for optimized interaction control of robot manipulator Ge et al. (2014), Wang et al. (2015), Li et al. (2015b). For human-multi-robot collaboration, the human worker is deemed as the environment, with unknown dynamics and position parameters, which leads to the system dynamics function partially unknown. In this paper, we aim to develop the human-multi-robot shared control with optimized interaction performance, in which the adaptive dynamic programming method using neural networks (NN) is employed to solve the optimized control problem subject to unknown system dynamics Liu et al. (2013). In addition, to achieve coordination of the robot agents in human-multi-robot collaboration, multi-agent reinforcement learning method has been utilized to achieve the coordination of the robot agents, or the robot agent and the human worker Li et al. (2015a), and will be utilized in this paper. The obtained optimized interaction controller is different from the traditional impedance controller, which is deemed as one kind of dynamic behavior control.

The rest of this paper is organized as follows. In Section II, the dynamic behavior control framework of the human-multi-robot interaction is presented and described. In Section III, optimized shared control strategy for human-multi-robot collaboration is proposed. In Section IV, the implementation procedure of the presented method is given. Simulation studies of the above human-multi-robot collaboration control problems are made in Section V. Then, the conclusion of this paper is given in Section VI.

2. HUMAN-MULTI-ROBOT COLLABORATION CONTROL

2.1 System Description

When a human worker and N robots interact with each other, as shown in Fig. 1, the description of the human-multi-robot collaboration system is given as below.

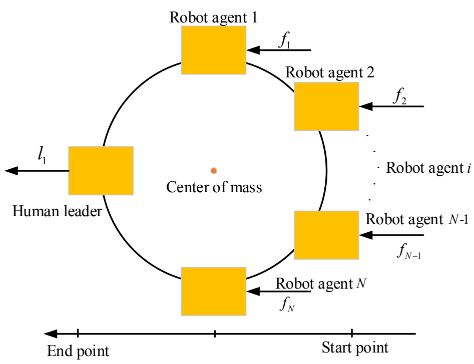

Fig. 1. Human-multi-robot collaborative manipulation for moving the object.

The collaborative task we consider is a 1DOF point-to-point task where a human worker and N robot agents collaboratively move a circular object with mass m by applying forces l_1 and $f_i, i = 1, 2, \cdots, N$ on the object, just as shown in Fig. 1. The position of the object is denoted as x, and x_i and x_f are the start and end positions of the object, respectively. To facilitate analysis, the torques generated during the carrying process are not considered. During the manipulation process, all robot agents can communicate with each other and exchange information, to achieve the coordination of the robot agents.

2.2 Dynamic Behavior Control Framework

To achieve optimized control for human-multi-robot collaboration, a dynamic behavior control framework is adopted, which is shown in Fig. 2.

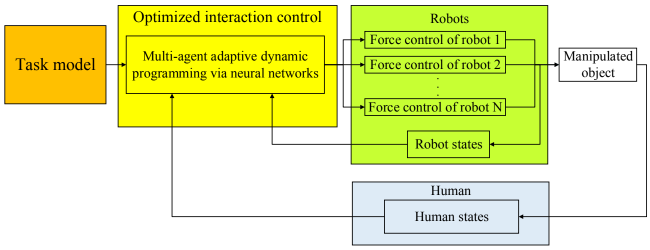

Fig. 2. Dynamic behavior control for physical human-multi-robot interaction.

The proposed framework mainly consists of four parts, namely the task model part, the optimized interaction control part, the robot model part, and the human model part. The task model part specifies tasks to the human worker as well as the robot agents according to the task management system. The optimized interaction control part receives task information from the task model part, the robot state information from the robot model part, and the human state information from the human model part. After receiving the above information, the optimized interaction control part will generate the optimized control and send to the robot model part. The robot force control loop will be realized in this part, after which the robot states will be generated and fed back to the optimized interaction control part. Further, for human-multi-robot collaboration system, the robot model and the human model will interact with each other to generate the system states, which will all be fed back to the optimized interaction control part. It is noteworthy that the human model is also considered and the robot model and the human model form a complete system model to achieve optimized collaboration control performance.

The aim of the human-multi-robot interaction control algorithm is to reduce the conflicts between different robot agents, which can therefore reduce the waste of energy. Further, the more important aim is to achieve optimized interaction control of the robot agents, which will be discussed in the following sections.

3. OPTIMIZED CONTROL FOR HUMAN-MULTI-ROBOT COLLABORATION

3.1 System Model and Problem Description

Consider the system composed of N robots collaborating with a human worker to move a rigid object, whose dynamics are modeled as follows:

$$m\ddot{x} = l_1 + \sum_{i=1}^{N} f_i, \tag{1}$$

where $x \in \mathbb{R}$ is the position of the manipulated object, $m \in \mathbb{R}$ is the mass of the object, $l_1 \in \mathbb{R}$ denotes the interaction force exerted by the human worker, $f_i \in \mathbb{R}$ denotes the interaction force exerted by the i-th robot agent.

Assume the control input of the human worker l_1 is modeled as follows,

$$l_1 = K_p(x_d - x) + K_d(\dot{x}_d - \dot{x}), \tag{2}$$

where $K_p \in \mathbb{R}$ and $K_d \in \mathbb{R}$ denote the stiffness and damping coefficients, $x_d \in \mathbb{R}$ and $\dot{x}_d \in \mathbb{R}$ denote the desired position and velocity of the human worker, x and $\dot{x}$ denote the actual position and velocity of the manipulated object. Because the object is assumed to be rigid, the position and velocity of the object is considered equivalent to the positions and velocities of the human worker and the robot agents.

In particular, the desired trajectory x_d of the human worker is generated by a given system

$$\begin{cases} \dot{w} = U_1 w, \\ x_d = V_1 w, \end{cases} \tag{3}$$

where $w \in \mathbb{R}$ is an auxiliary state, and U_1 and V_1 are two matrices designed to generate x_d.

Similarly, the desired trajectory of the robot agents $x_{d1} \in \mathbb{R}$ is generated by a given system

$$\begin{cases} \dot{w}_1 = U_2 w_1, \\ x_{d1} = V_2 w_1, \end{cases} \tag{4}$$

where $w_1 \in \mathbb{R}$ is an auxiliary state, and U_2 and V_2 are two matrices designed to generate x_{d1}.

Then, by denoting $z = [\dot{x}\ x\ w\ w_1]^T$, we have the complete system description

$$\dot{z} = Az + \sum_{i=1}^{N} B_i f_i, y = Cz, \tag{5}$$

where $A = m^{-1}\begin{bmatrix} -K_d & -K_p & K_pV_2 + K_dV_2U_2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & U_1 & 0 \\ 0 & 0 & 0 & U_2 \end{bmatrix}$, $B_i = m^{-1}[1\ 0\ 0\ 0]^T$. According to the definition of the complete state z, the state x_d, i.e., the desired trajectory of the human worker is unknown to the robot agents. Therefore, the output feedback variable y is defined as $y = [\dot{x}\ x\ w_1]^T$, and $C = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

We consider that the system's control objective is to minimize the infinite-horizon cost function

$$\Gamma = \int_0^\infty c(t)\mathrm{d}t,$$
$$c(t) = (x - x_{d1})^T Q_1 (x - x_{d1}) + \dot{x}^T Q_2 \dot{x} + \sum_{i=1}^{N} f_i^T R_i f_i, \tag{6}$$

where the weights Q_1, $Q_2 \in \mathbb{R} \geq 0$, and $R_i \in \mathbb{R} > 0, i = 1, 2, \cdots, N$. The first term of the above cost function penalizes the error between the actual and desired positions of the robot agents. The third term determines the contribution of the interaction forces of the robot agents.

According to the definition of the output feedback variable y, the cost function (6) can be rewritten as

$$\Gamma = \int_0^\infty [y^T Q y + \sum_{i=1}^{N} f_i^T R_i f_i]\mathrm{d}t, \tag{7}$$

where $Q = \begin{bmatrix} Q_1 & 0 & 0 \\ 0 & Q_2 & -Q_2V_2 \\ 0 & -V_2^T Q_2 & V_2^T Q_2 V_2 \end{bmatrix}$.

3.2 Multi-Agent Dynamic Programming

Human-multi-robot shared collaboration control can be investigated on the basis of game theory. In particular, the robot agents are involved in a common game and have individual objectives. There are different solutions to the game that will result in different multi-agent behaviors Li et al. (2015a), and Nash equilibrium is considered in this paper to achieve the coordination of the multi-robots to collaborate with the human worker.

Then, the Nash equilibrium can be achieved by the optimal control

$$f_i^* = -\frac{1}{2} R_i^{-1} B_i P z^*, \tag{8}$$

where P is obtained by solving the following well-known Riccati equation

$$A^T P + PA + Q - \sum_{i=1}^{N} P B_i R_i^{-1} B_i^T P = 0. \tag{9}$$

Remark 1: If the control coefficients K_p, K_d, the human desired trajectory x_d, and the mass of the object m are all known to the robot agents, it is obvious that the system matrices A and B_i are known to the robot agents. Therefore, the optimal control forces $f_i, i = 1, 2, \cdots, N$ can be easily computed via the coupled algebraic Riccati equation (ARE) (9). In this paper we assume that the control coefficients K_p, K_d, and the mass of the object m, are unknown to the robot agents. Thus, the system matrices A and $B_i, i = 1, 2, \cdots, N$ are unknown to the robot agents. For this case, the adaptive dynamic programming method with neural networks will be utilized in this paper.

3.3 Adaptive Dynamic Programming via Neural Networks

For the optimized interaction control problem with unknown dynamics, the adaptive dynamic programming method developed in Liu et al. (2013) will be utilized, to address the optimization problem discussed above.

In particular, the optimal control f_i^* that minimizes the cost function (6) subject to the system function is given by

$$f_i^*(z) = -\frac{1}{2}R_i^{-1}B_i^T(z)\Gamma_z^*, \tag{10}$$

where $\Gamma_z^* = \frac{\partial \Gamma^*}{\partial z}$ and $\Gamma^* = \min_{f_1(z),f_2(z),\cdots,f_N(z)}\Gamma$ is the optimal cost function. However, as A, B_i^T and Γ^* are unknown due to unknown dynamics of the human and the mass of the object, the above optimal control needs to be obtained by approximation.

First, an observer has been designed to identify the system with unknown dynamics, as below

$$\dot{\hat{z}} = A_0\hat{z} + \hat{G}(\hat{z}, f_1, f_2, \cdots, f_N) + L(y - C\hat{z}), \tag{11}$$

where A_0 is a given Hurwitz matrix, $\hat{G}(z, f_1, f_2, \cdots, f_N)$ is the estimate of $G = Az + \sum_{i=1}^{N} B_i f_i - A_0\hat{z}$, $\hat{z}$ is the output of the observer, and L is chosen such that $A - LC$ is also a Hurwitz matrix. $\hat{G}(\hat{z}, f_1, f_2, \cdots, f_N)$ is approximated using a three-layer neural networks (NN), which is expressed as below

$$\hat{G}(\xi) = \hat{W}_1 S_1(\hat{V}_1\xi), \tag{12}$$

where $\xi = \begin{bmatrix}\hat{z}^T & f_1 & f_2 & \cdots & f_N\end{bmatrix}^T$ is the NN input, $p1 \in \mathbb{R}$ denotes the number of neural nodes in the hidden layer, $\hat{V}_1 \in \mathbb{R}^{p1\times(4+N)}$ and $\hat{W}_1 \in \mathbb{R}^{4\times p1}$ are the estimates of the ideal weights $V_1 \in \mathbb{R}^{p1\times(4+N)}$ and $W_1 \in \mathbb{R}^{4\times p1}$, respectively. $\hat{W}_1$ is updated according to the following weight update law

$$\begin{aligned}\dot{\hat{W}}_1 =& -k_1 A_c^{-T}C^T \parallel \tilde{y} \parallel S^T(\hat{V}_1\xi) - k_2 \parallel \tilde{y} \parallel \hat{W}_1,\\ \dot{\hat{V}}_1 =& -k_3^T \parallel \tilde{y} \parallel CA_c^{-1}\hat{W}_1(I_{p1} - \mathrm{diag}(\mathrm{S}_1(\hat{\mathrm{V}}_1\xi)))^{\mathrm{T}}\mathrm{sgn}(\xi)\\ &- k_4 \parallel \tilde{y} \parallel \hat{V}_1,\end{aligned} \tag{13}$$

where $\tilde{y} = y - C\hat{z}$, $k_1, k_2, k_3, k_4 \in \mathbb{R}$ are positive scalars, $A_c = A_0 - LC$, and L is selected such that $A_0 - LC$ is a Hurwitz matrix. With the above approximation of the system dynamics, the control force f_i, which is the estimate for the optimal control f_i^*, can be achieved as

$$f_i = -\frac{1}{2}R_i^{-1}\left(\frac{\partial\hat{G}(\hat{z}, f_1, f_2, \cdots, f_N)}{\partial f_i}\right)^T\hat{\Gamma}_{\hat{z}}, \tag{14}$$

where $\frac{\partial\hat{G}(\hat{z}, f_1, f_2, \cdots, f_N)}{\partial f_i}$ is

$$\begin{aligned}\frac{\partial\hat{G}(\hat{z}, f_1, f_2, \cdots, f_N)}{\partial f_i} =& \hat{W}_1\frac{\partial S_1(\hat{V}_1\xi)}{\partial\hat{V}_1\xi}\hat{V}_1\\ &[0_{4\times1}\ \lambda_1\ \lambda_2\ \cdots\ \lambda_i\ \cdots\ \lambda_N]^T,\end{aligned} \tag{15}$$

where $\lambda_k, k = 1, 2, \cdots, N$ denote the coefficients for the control forces of different robot agents, $\lambda_k = 1$ for $k = i$, and $\lambda_k = 0$ otherwise.

Second, another critic NN is used to approximate $\Gamma(\hat{z})$, which is shown as below

$$\hat{\Gamma}(\hat{z}) = \hat{W}_2^T S_2(\hat{z}), \tag{16}$$

where $\hat{W}_2 \in \mathbb{R}^{p2\times1}$ is the estimate of the ideal weight $W_2 \in \mathbb{R}^{p2\times1}$, and $S_2(\hat{z}) \in \mathbb{R}^{p2\times1}$ denotes the activation function, $p2 \in \mathbb{R}$ is the number of hidden neurons that is similar as in (12). $\hat{W}_2$ is online updated according to the following weight update law

$$\dot{\hat{W}}_2 = -k_5\frac{\sigma}{(\sigma^T\sigma + 1)^2}(\sigma^T\hat{W}_2 + \hat{z}^TQ\hat{z} + \sum_{i=1}^{N} f_i^T R_i f_i) \tag{17}$$

where k_5 is a positive scalar and

$$\sigma = [\frac{\partial S_2^T(\hat{z})}{\partial\hat{z}}]^T\dot{\hat{z}}. \tag{18}$$

Thus, $\hat{\Gamma}_{\hat{z}}$ is obtained as

$$\hat{\Gamma}_{\hat{z}} = \frac{\partial\hat{\Gamma}(\hat{z})}{\partial\hat{z}} = \hat{W}_2\frac{\partial S_2^T(\hat{z})}{\partial\hat{z}}. \tag{19}$$

Remark 2: In summary, the control input in (10) can be obtained with the two NN approximation equations (12) and (16), with the weight updating law in equations (13) and (17), respectively. The specific implementation procedure of the proposed algorithm will be given in the following section.

4. IMPLEMENTATION PROCEDURE OF THE PROPOSED ALGORITHM

The implementation procedure of the proposed algorithm is as follows.

Step 1: The human-multi-robot collaboration control system is initialized. The parameters Q_1, Q_2, R_1, and R_2 in the cost function are set. The desired trajectory of the robot agents is determined. The initial values of the NN parameters, such as the weights of the two NNs, W_1, V_1, W_2, and the learning parameters k_1, k_2, k_3, k_4, k_5, are predetermined.

Step 2: When the collaboration system starts running, the initial interaction control f_i in (14), is utilized. Collect the motion and force data and feed back to the presented adaptive dynamic programming algorithm, in which the weight values of the two NNs, W_1, V_1, and W_2, are updated according to (13) and (17), respectively.

Step 3: The updated weight values of the two NNs are utilized in (14) for the next time moment.

5. SIMULATION STUDIES

5.1 Simulation Conditions

Without loss of generality, here we consider two robots collaborating with a human worker to move the object, and the method proposed in this paper can be extended to N robots ($N > 2$) collaborating with human worker, easily.

The parameters used in the simulation of the optimized control for human-multi-robot collaboration system are as follows,

$m = 1$kg, $K_p = 900$, $K_d = 60$, $U_1 =$ -1, $V_1 = 1$, $U_2 =$ -0.8, $V_2 = 1$, $x_i = 0.2$m. Then,

$$A = \begin{bmatrix}-60 & -900 & 852 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -0.8\end{bmatrix},\ B_i = \begin{bmatrix}1\\0\\0\\0\end{bmatrix},\ C = \begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix},$$

$i = 1, 2.$

For the cost function parameters chosen for simulation, $Q_1 = 1$, $Q_2 = 1000$, $R_1 = 0.01$, $R_2 = 0.001$.

It is noteworthy that the parameters in the system function, K_p, K_d, U_1, V_1, and the mass of the object m, are unknown to the two robot agents. These parameters are just set for simulation.

The matrices A_0 and L are selected as follows.

$$A_0 = \begin{bmatrix} -50 & -500 & 500 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & -0.8 & 0 \\ 0 & 0 & 0 & -0.8 \end{bmatrix}, \ L = \begin{bmatrix} 10 & 10 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The observer NN is a three-layer NN with one hidden layer to approximate the unknown system dynamics. The input layer involves 6 neurons and the output layer contains 4 neurons, and the numbers of neural nodes in the hidden layer, $p1$, is set as 15. The activation function $S_1(\cdot)$ is selected as hyperbolic tangent function $\tanh(\cdot)$. Let the learning rates be $k_1 = k_3 = 10$ and the parameters be $k_2 = k_4 = 10$. Additionally, the initial weights of W_1 and V_1 are all set to be random within [0, 0.2]. Then, we can complete the design of the NN observer for the unknown system dynamics.

Then, based on the observed states, a feedforward neuro-controller is constructed via the ADP method to obtain the optimized control of the human-multi-robot collaboration system Liu et al. (2013), in which a critic NN is built to approximate the cost function. Since the cost function is a quadratic function of the system state variables, the activation functions of the critic NN are chosen from the second-order series expansion of the value function, which is shown as below.

$$S_2(\hat{z}) = [\hat{z}^2(1), \hat{z}(1)\hat{z}(2), \hat{z}(1)\hat{z}(3), \hat{z}(1)\hat{z}(4), \hat{z}^2(2), \\ \hat{z}(2)\hat{z}(3), \hat{z}(2)\hat{z}(4), \hat{z}^2(3), \hat{z}(3)\hat{z}(4), \hat{z}^2(4)]. \quad (20)$$

It is obvious that the number of neurons in the hidden layer is $p_2 = 10$. In addition, the initial weights of $W_2 \in \mathbb{R}^{10\times 1}$ are set as $[0, 0, \ldots, 0]T$, and the learning rate $k_5 = 1.5$.

5.2 Simulation Results

For the given set of cost function parameters, using the presented algorithm, after simulation, the convergence results of the NN weight parameters W_1, V_1, and W_2 are shown in Figs. 3, 4, respectively.

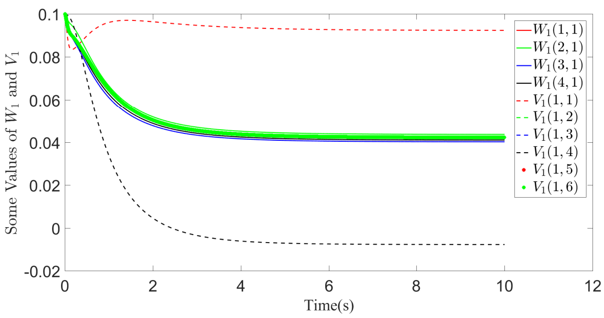

Fig. 3. The convergence results of W_1 and V_1 parameters.

The resulted system states and the estimation results of the human-multi-robot collaboration system states are shown in Fig. 5.

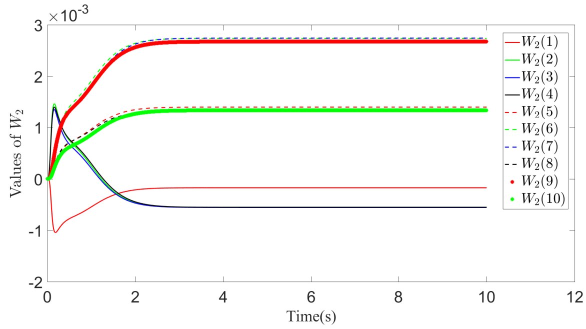

Fig. 4. The convergence results of the W_2 parameters.

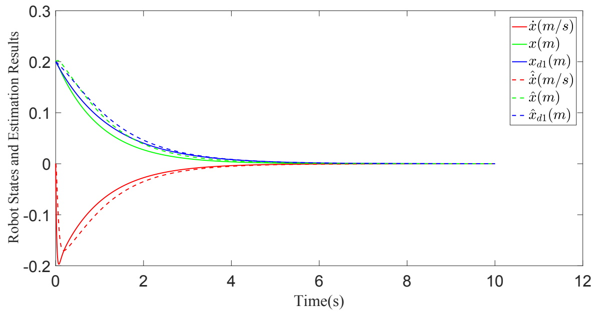

Fig. 5. The actual robot states and the estimation results of the human-multi-robot collaboration system.

The resulted control forces of the two robot agents in the human-multi-robot collaboration is shown in Fig. 6.

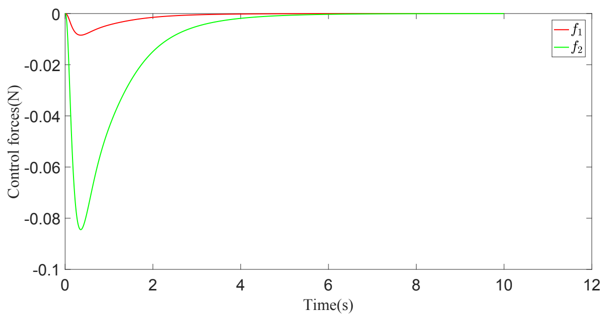

Fig. 6. The control forces of the two robot agents.

5.3 Simulation Results Discussion

From the simulation results in Figs. 3, 4, we can see that the weight matrices W_1, V_1, and W_2 in the two neural networks converge steadily to the optimized solutions. In addition, when the weight parameter for the interaction force of the robot agent 1, R_1, is larger than R_2, the corresponding interaction forces f_1 are smaller than f_2, which is shown in Fig. 6 and very consistent with the theoretical derivation.

It can also be seen from Fig. 5 that the estimation results of the system states $\dot{x}, x$, and x_{d1} quickly and steadily follow the changes of the actual system states. This verifies the stability of the presented algorithm using neural networks.

For the convenience of system modeling and computation, the impedance parameters of the human worker K_p and K_d are assumed to be constant. In practice, this assumption is hard to be satisfied as the human can adaptively regulate these parameters for better working performance or stability. Therefore, in the future work, these two parameters will be assumed to be changeable to make the assumptions more realistic.

In this paper, the human-multi-robot collaborative task is considered as a 1DOF point-to-point task, which is somewhat simplified. In practical applications, the collaborative task should be conducted in three-dimensional space. It is noteworthy that the algorithm presented in this paper can be easily extended to the collaborative scenarios in three-dimensional space. In addition, the torques generated in the human-multi-robot collaborative task is neglected for convenience of analysis, which will be considered in the future work.

For the stability and the boundedness of the approximation error of the two neuron networks used for adaptive dynamic programming, in the previous work Liu et al. (2013), the authors have given the proof. Therefore, in this paper, the neural networks are utilized in the adaptive dynamic programming algorithm for the optimized collaboration control. Also, in the future work, the presented algorithm in this paper will be verified through experimental studies.

6. CONCLUSION

In this paper, an optimized control method for physical human-multi-robot shared collaboration system has been developed. The coordination of the robot agents have been taken into account in the problem formulation. Considering the unknown human dynamics and position parameters during human-multi-robot collaboration, the multi-agent adaptive dynamic programming method has been employed to present an optimized control algorithm subject to unknown system dynamics. The neural network approximation method was utilized in this paper to approximate the unknown system dynamics and the cost function to obtain the optimized control policy. The effectiveness of the presented method has been verified through simulation studies.

REFERENCES

Albu-Schäffer, A., Ott, C., and Hirzinger, G. (2007). A unified passivity-based control framework for position, torque and impedance control of flexible joint robots. *The International Journal of Robotics Research*, 26(1), 23–39.

Cacace, J., Finzi, A., and Lippiello, V. (2016). Implicit robot selection for human multi-robot interaction in search and rescue missions. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 803–808.

Franklin, D.W., Burdet, E., Tee, K.P., Osu, R., Chew, C.M., Milner, T.E., and Kawato, M. (2008). CNS learns stable, accurate, and efficient movements using a simple algorithm. *Journal of Neuroscience*, 28(44), 11165–73.

Gao, W., Jiang, Y., Jiang, Z.P., and Chai, T. (2016). Output-feedback adaptive optimal control of interconnected systems based on robust adaptive dynamic programming. *Automatica*, 72(8), 37–45.

Ge, S.S., Li, Y., and Wang, C. (2014). Impedance adaptation for optimal robot-environment interaction. *International Journal of Control*, 87(2), 249–263.

Hogan, N. (1985). Impedance control: An approach to manipulation, part i - theory, part ii - implementation, part iii - applications. *ASME Transactions Journal of Dynamic Systems & Measurement Control B*, 107(1), 304–313.

Lawitzky, M., Mrtl, A., and Hirche, S. (2010). Load sharing in human-robot cooperative manipulation. In *19th International Symposium in Robot and Human Interactive Communication*, 185–191.

Li, Y., Ge, S.S., and Yang, C. (2011). Impedance control for multi-point human-robot interaction. In *2011 8th Asian Control Conference (ASCC)*, 1187–1192.

Li, Y., Tee, K.P., Chan, W.L., Yan, R., Chua, Y., and Limbu, D.K. (2015a). Continuous role adaptation for human–robot shared control. *IEEE Transactions on Robotics*, 31(3), 672–681.

Li, Y., Tee, K.P., Yan, R., Limbu, D.K., and Ge, S.S (2015b). Shared control of human and robot by approximate dynamic programming. In *2015 American Control Conference (ACC)*, 1167–1172. IEEE.

Liu, D., Huang, Y., Wang, D., and Wei, Q. (2013). Neural-network-observer-based optimal control for unknown nonlinear systems using adaptive dynamic programming. *International Journal of Control*, 86(9), 1554–1566.

Matinfar, M. and Hashtrudi-Zaad, K. (2005). *Optimization-based Robot Compliance Control: Geometric and Linear Quadratic Approaches.* Sage Publications, Inc.

Moertl, A., Lawitzky, M., Kucukyilmaz, A., Sezgin, M., Basdogan, C., and Hirche, S. (2012). The role of roles: Physical cooperation between humans and robots. *International Journal of Robotics Research*, 31(13), 1656–1674.

Noohi, E., efran, M., and Patton, J.L. (2016). A model for humanhuman collaborative object manipulation and its application to humanrobot interaction. *IEEE Transactions on Robotics*, 32(4), 880–896.

Reed, K.B. and Peshkin, M.A. (2008). Physical collaboration of human-human and human-robot teams. *IEEE Transactions on Haptics*, 1(2), 108–120.

Sklar, E., Parsons, S., Ozgelen, A.T., Schneider, E., and Epstein, S.L. (2013). Hrteam: A framework to support research on human/multi-robot interaction. In *International Conference on Autonomous Agents & Multi-agent Systems.*

Vrabie, D., Pastravanu, O., Abu-Khalaf, M., and Lewis, F.L. (2009). Brief paper: Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica*, 45(2), 477–484.

Wang, C., Li, Y., Ge, S.S., and Lee, T.H. (2015). Optimal critic learning for robot control in time-varying environments. *IEEE Transactions on Neural Networks & Learning Systems*, 26(10), 2301–2310.

Wu, M.H., Ogawa, S., and Konno, A. (2016). Symmetry position/force hybrid control for cooperative object transportation using multiple humanoid robots. *Advanced Robotics*, 30(2), 131–149.