Reinforcement Q-Learning and ILC with Self-Tuning Learning Rate for Contour Following Accuracy Improvement of Biaxial Motion Stage

Wei-Liang Kuo[1], Ming-Yang Cheng[2], Hong-Xian Lin[3]

Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan

Abstract**—***Biaxial motion stages are commonly used in precision motion applications. However, the contour following accuracy of a biaxial motion stage often suffers from system nonlinearities and external disturbances. To deal with the above-mentioned problem, a control scheme consisting of a reinforcement Q-learning controller with a self-tuning learning rate and two iterative learning controllers is proposed in this paper. In particular, the reinforcement Q-learning controller is used to compensate for friction and also cope with the problem of dynamics mismatch between different axes. In addition, one of the two iterative learning controllers is used to suppress periodic external disturbances, while the other one is employed to adjust the learning rate of the reinforcement Q-learning controller. Results of contour following experiments indicate that the proposed approach is feasible.*

Keywords—Reinforcement Learning, Q-Learning, Iterative Learning Control (ILC), Contour Following.

I. INTRODUCTION

Contour following is commonly seen in industrial processes such as machining, cutting, polishing, deburring, painting and welding. In these industrial processes, product quality depends on contour following accuracy. Generally speaking, better contour following accuracy can be achieved by reducing tracking errors and/or contour error [14]. As a matter of fact, tracking error reduction is one of the most important research topics in the contour following problems of multi-axis motion stage [1]-[4]. Due to factors such as external disturbance, system nonlinearity, servo lag and mismatch in axis dynamics, contour following accuracy of the multi-axis motion stage may not be able to meet the accuracy requirements [5]-[7].

There are many existing approaches that can be used in practice to reduce tracking error of a multi-axis motion stage [9]-[12]. For example, the commonly used multi-loop feedback control scheme with command feedforward is very effective in reducing tracking error caused by the servo lag phenomenon [8]. In addition, advanced control schemes such as sliding mode control and adaptive control can be used to reduce tracking error as well. Recently, the number of studies exploiting the paradigm of artificial neural network to improve contour following accuracy of multi-axis motion stage has risen steadily [13]-[22]. For instance, Wen and Cheng [13] proposed a fuzzy CMAC with a critic-based learning mechanism to cope with external disturbance and nonlinearity so as to reduce tracking error. Later on, Wen and Cheng [15] further proposed a recurrent fuzzy cerebellar model articulation controller with a self-tuning learning rate to improve contour following accuracy for a piezoelectric actuated dual-axis micro motion stage. In addition to tracking error reduction, the paradigm of artificial neural network has been applied to different fields such as wind power generation [24], the game of Go [22], and object grasping using robots [25]. Generally, a neural network needs to be trained before it can be used to solve a particular problem. Among different training mechanisms for neural networks, reinforcement learning is the one that has received a lot of attention recently [21]. In this paper, a control scheme consisting of a reinforcement Q-learning controller with an adjustable learning rate and two iterative learning controllers (ILC) is proposed to improve contour following accuracy of a bi-axial motion stage. In the proposed approach, the reinforcement Q-learning controller is responsible for friction compensation and also deals with the dynamics mismatch between different axes. In addition, one of the two ILCs is exploited to deal with the adverse effects due to periodic external disturbances from repetitive motions, while the other ILC is exploited to tune the learning rate of Q-learning based on current tracking error so as to further improve contour following accuracy.

The remainder of the paper is organized as follows. Section 2 gives a brief review on reinforcement learning and iterative learning control. Section 3 introduces the proposed control scheme. Experimental results and conclusions are provided in Section 4 and 5, respectively.

II. BRIEF REVIEW ON REINFORCEMENT LEARNING AND ITERATIVE LEARNING CONTROL

Since the proposed control scheme exploits the idea of reinforcement learning and iterative learning control, brief reviews on these two research topics will be provided in this section.

2.1 Reinforcement Learning and Q-Learning

In general, learning mechanisms of artificial neural networks can be divided into three types: supervised learning, unsupervised learning and reinforcement learning. Unlike the other two types of learning which either need training pairs or expected final outcome, reinforcement learning "learns" proper actions by maximizing the reward simply based on the reward/penalty resulting from previous action and current environment. Fig. 1 illustrates a typical control block diagram that employs reinforcement Q-learning.

In general, the reinforcement Q-learning controller can be expressed as:

$$Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha_t[r_t + \gamma \max_{a_i} Q_{t+1}(s_{t+1}, a_i) - Q_t(s_t, a_i)] \tag{1}$$

Where $Q_t(s_t, a_t)$: the Q value corresponding to the state s_t and action a_t in the Q-table; s: state; a_i: action; i: action index in action space; α_t: learning rate; r: reward; γ:discount factor. $\max_{a_i} Q_{t+1}(s_{t+1}, a_i)$ is the maximum value of Q corresponding to state $s_{t+1}$and action a_i; t: time variable.

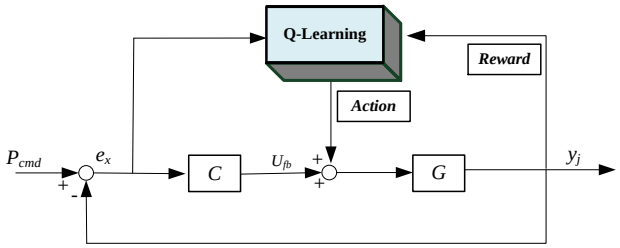

FIG. 1 A TYPICAL CONTROL BLOCK DIAGRAM THAT EMPLOYS REINFORCEMENT BASED Q-LEARNING

The probability of selecting action a_i in Q-learning is described by (2).

$$P(s, a_i) = \frac{e^{Q(s,a_i)}}{\sum e^{Q(s,a_i)}} \tag{2}$$

2.2 Iterative Learning Controller

As reported in many previous studies, ILC is effective in suppressing periodic disturbances caused by repetitive motions [17-19]. Fig. 2 is the block diagram for a control scheme consisting of a feedback controller and a control law based ILC [16]. In Fig. 2, U_{ilc} is the control force generated by ILC, L is the learning function and F is a low-pass filter. All the tracking error e_x and the control force U_{ilc} generated by ILC in the previous iteration are stored and used to update U_{ilc} in the current iteration. The total control force U_j in the jth iteration to the plant G is the sum of U_{ilc} and the feedback controller output U_{fb}.

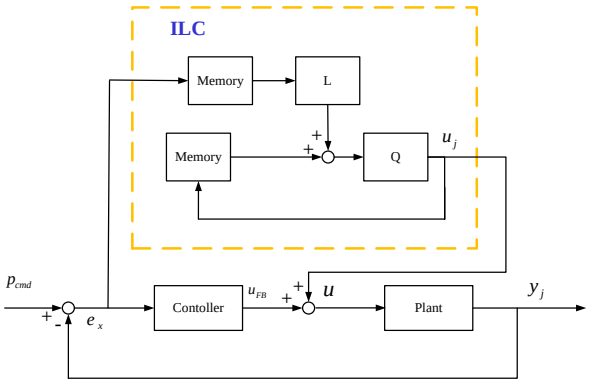

FIG. 2 BLOCK DIAGRAM FOR A CONTROL SCHEME CONSISTING OF A FEEDBACK CONTROLLER AND A CONTROL LAW BASED ILC

Based on Fig.2, the relationship between output y_j in the jth iteration and input P_{cmd} can be described as:

$$y_j = (1+GC)^{-1}G \cdot U_j + (1+GC)^{-1}GC \cdot P_{cmd} \tag{3}$$

where the total control force U_{j+1} in the j+1th iteration is updated using Eq. (4)

$$U_{j+1} = F(U_j + Le_j) + Ce_{j+1} \tag{4}$$

where e_j, and u_j are the tracking error and total control force in the jth iteration, respectively. Note that in Eq. (4), Ce_{j+1} can be regarded as the feedback control force. The control force U_{ilc} generated by ILC aims at reducing the tracking error. Namely, better performance of ILC leads to smaller tracking error so that feedback control force decreases as well.

III. THE PROPOSED REINFORCEMENT Q-LEARNING CONTROLLER WITH AN ADJUSTABLE LEARNING RATE

Fig. 3 illustrates the block diagram for the control scheme consisting of a feedback controller, a control law based ILC, and the proposed reinforcement Q-learning controller with an adjustable rate.

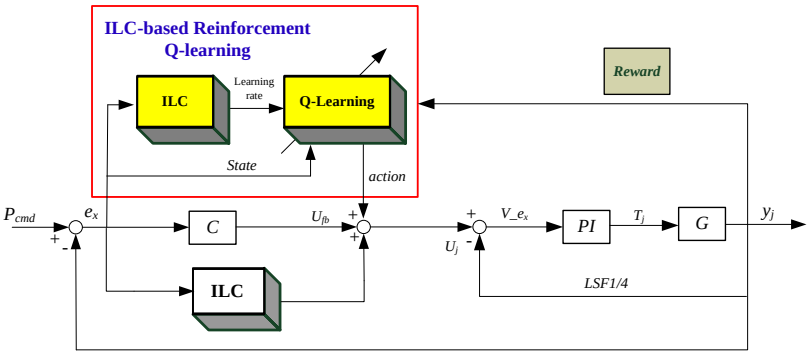

FIG. 3 BLOCK DIAGRAM FOR THE CONTROL SCHEME CONSISTING OF A FEEDBACK CONTROLLER, A CONTROL LAW BASED ILC, AND THE PROPOSED REINFORCEMENT Q-LEARNING CONTROLLER WITH AN ADJUSTABLE RATE.

In the proposed approach, the reinforcement Q-learning controller is modified as:

$$Q_{t+1}(e_t, a_t) = Q_t(\mathrm{e}_t, a_t) + L_{ilc}[\mathrm{r}_t + \gamma \max_{a_i} Q_{t+1}(\mathrm{e}_{t+1}, \mathrm{a}_i) - Q_t(\mathrm{e}_t, \mathrm{a}_i)] \tag{5}$$

Compared with Eq. (1), the learning rate α_t in the conventional Q-learning algorithm is replaced by L_{ilc} in Eq. (5), where L_{ilc} is updated using Eq. (6)

$$L_{ilc_{j+1}} = F[L_{ilc_j} + Le_j] \tag{6}$$

where L_{ilc_j+1} is the learning rate for the reinforcement Q-learning controller in the j+1th iteration. Note that in this paper, the value of L_{ilc} is constrained to be between zero and one. Moreover, in this paper, the aim is to reduce the tracking error of a multi-axis motion stage. As a result, the state s in Eq. (1) is replaced by tracking error e in Eq. (5). In addition, three possible actions — accelerate, decelerate and maintain constant velocity, and can be selected for a_t in Eq. (5) to adjust the velocity command for the motion stage. The probability of selecting action a_i in the Q-learning algorithm is rewritten as:

$$P(\mathrm{e}, a_{i,}) = \frac{e^{Q(\mathrm{e},a_i)}}{\sum e^{Q(\mathrm{e},a_i)}} \tag{7}$$

where the state is the tracking error e.

In this paper, the action space A consists of three actions

$$A = \{a_1, a_2, a_3\} \tag{8}$$

where a_1: accelerate: a_2: decelerate; a_3: maintain constant speed.

In this paper, the reward is designed to reduce tracking error. In particular, the reward is determined using Eq. (9).

$$r_{t+1} = \begin{cases} 100, & \text{if} \quad e_t - e_{t+1} < 0 \\ 0, & \text{if} \quad e_t - e_{t+1} = 0 \\ -1, & \text{if} \quad e_t - e_{t+1} > 0 \end{cases} \tag{9}$$

Fig. 4 illustrates the block diagram of the motion control scheme for a bi-axial motion stage proposed in this paper. In Fig.4, the velocity command ω_{cmd_x} for the *x*-axis consists of the control force U_{ilc_x} generated by ILC, the feedback control force U_{fb_x}, and the control force U_{RL_x} generated by the reinforcement Q-learning controller. It can be expressed as:

$$\omega_{cmd\_x} = U_{fb\_x} + U_{ilc\_x} + U_{RL\_x} \tag{10}$$

where

$$U_{fb\_x} = C \cdot e_{j+1} \tag{11}$$

$$U_{ilc\_x} = F(u_j + L \cdot e_j) \tag{12}$$

$$U_{RL\_x} = v + \Delta v \tag{13}$$

Note that the velocity command for the *y*-axis is designed similarly.

In Fig. 4, the reinforcement Q-learning controller with adjustable learning rate is responsible for friction compensation and also deals with the dynamics mismatch between the *x*-axis and *y*-axis. Fig. 4 also shows that two ILCs are employed in the proposed motion control scheme. In particular, one ILC is exploited to adjust the learning rate of the reinforcement Q-learning controller, while the other ILC is exploited to deal with the adverse effects due to periodic external disturbances so as to further reduce tracking error.

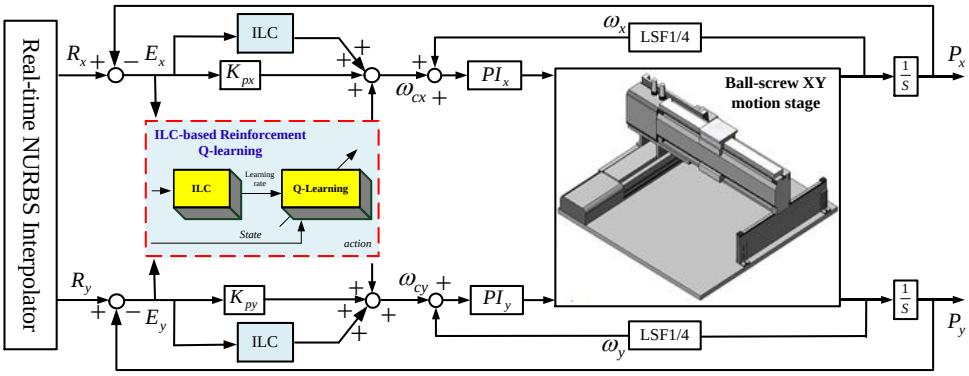

FIG. 4 THE BLOCK DIAGRAM OF THE PROPOSED MOTION CONTROL SCHEME FOR A BI-AXIAL MOTION STAGE.

IV. EXPERIMENTAL RESULTS

Fig. 5 shows a the photograph of the bi-axial motion stage used to assess the effectiveness of the proposed approach. Fig. 6 (a) shows the circle-shaped contour represented in NURBS form used in the contour following experiment. Under S-curve acceleration/deceleration motion planning, a NURBS interpolator [23] is employed to convert the circle-shaped contour into the position commands for the *x*-axis (Fig. 6(b)) and *y*-axis (Fig. 6(c)). The duration time for each circle following is 9.5 seconds. In each experiment, circle following will be performed seven times (i.e. seven iterations). In total, four different control schemes are tested in the contour following experiments. They are:

Control scheme #1: PI type feedback controller combined with an ILC.

Control scheme #2: PI type feedback controller combined with a reinforcement Q-learning controller with a fixed learning rate.

Control scheme #3: PI type feedback controller combined with an ILC and a reinforcement Q-learning controller with a fixed learning rate.

Control scheme #4: PI type feedback controller combined with an ILC and a reinforcement Q-learning controller with adjustable learning rate.

Due to the limitations in the paper length, only the experimental results of the tracking error in the *x*-axis for these four tested control schemes are shown in Fig. 7. In addition, performance indices in terms of root mean square of tracking error (RMS), average of integral of absolute tracking error (AIAE), and maximum tracking error (MAX) are listed in TABLE 1.

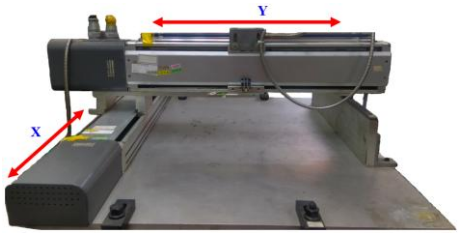

FIG. 5 PHOTOGRAPH OF THE BI-AXIAL MOTION STAGE USED IN THIS PAPER.

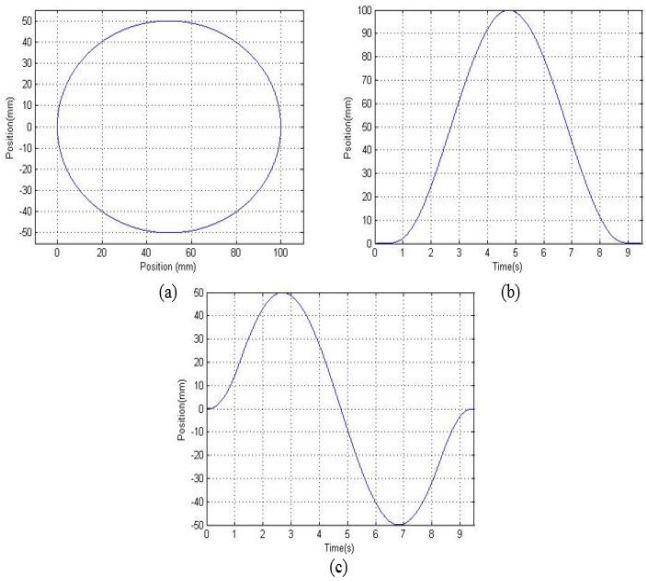

FIG. 6 (A) CIRCLE SHAPED CONTOUR REPRESENTED IN NURBS FORM (B) POSITION COMMANDS FOR X-AXIS, (C) POSITION COMMANDS FOR Y-AXIS.

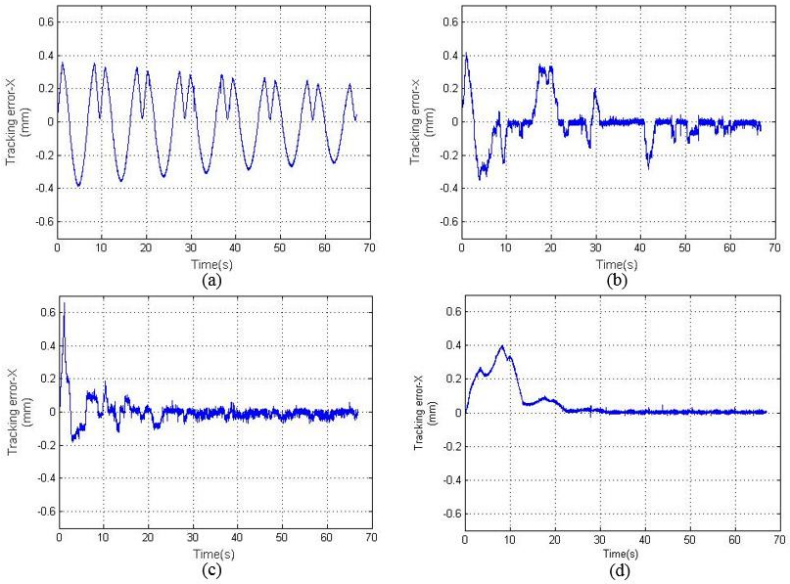

FIG. 7 EXPERIMENTAL RESULTS OF TRACKING ERROR IN THE X-AXIS (A) CONTROL SCHEME #1 (B) CONTROL SCHEME #2 (C) CONTROL SCHEME #3 (D) CONTROL SCHEME #4

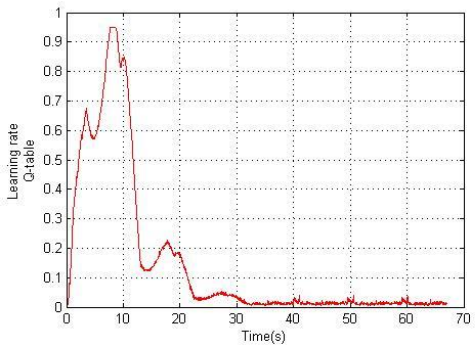

FIG. 8 THE VALUES OF THE LEARNING RATE OF THE REINFORCEMENT Q-LEARNING CONTROLLER VARIES WITH RESPECT TO TIME.

TABLE 1
TRACKING ERROR COMPARISON AMONG FOUR TESTED CONTROL SCHEMES

	Tracking error of X axis		
	RMS(μm)	AIAE(μm)	Max(μm)
Scheme #1	19.22	17.01	35.79
Scheme #2	9.48	5.16	35.08
Scheme #3	4.28	2.33	17.63
Scheme #4	2.62	1.51	9.93

Fig. 7(a) shows the tracking error of the circle following experiment using control scheme #1. Since the ILC will be activated after the 1st iteration, the tracking error for the first 9.5 seconds in Fig. 7(a) can be regarded as the results for using the PI feedback control only. Clearly, tracking error gradually decreases after the 2nd iteration, indicating that ILC indeed is effective in suppressing periodic external disturbance. Based on the experimental results shown in Fig. 7(b)~(d), the tracking error for using control schemes #2, #3 or #4 all converges much faster than that for using the control scheme #1. These facts indicate that the reinforcement Q-learning controller is indeed effective in reducing tracking error since control schemes #2, #3 or #4 all include a reinforcement Q-learning controller. In particular, the proposed control scheme (i.e. control scheme) has the best performance among the four tested control schemes.
Fig. 8 shows the values of the learning rate of the reinforcement Q-learning controller in the proposed control scheme varies with respect to time. After three iterations (after 28.5 seconds), the learning rate only changes slightly since the tracking error becomes very small after three iterations.

V. CONCLUSION

This paper has proposed a motion control scheme consisting of two ILCs and one reinforcement Q-learning controller for contour following accuracy improvement. In particular, one ILC is used to tune the learning rate of the reinforcement Q-learning controller that is mainly used to cope with system nonlinearities, while the other ILC is exploited to suppress periodic disturbance during repetitive contour following motions. Results of contour following experiments conducted on a bi-axial motion stage indicate that the proposed control scheme is feasible and outperforms other control schemes also tested in the experiment.

ACKNOWLEDGEMENTS

The authors would like to thank the MOST of Taiwan for their support of this research under Grant MOST 105-2221-E-006-105-MY3.

REFERENCES

[1] F. Hiroshi and T. Tadashi, "High-Precision Control of Ball-Screw-Driven Stage Based on Repetitive Control Using n-Times Learning Filter," IEEE Trans. Ind. Electron, vol. 61, pp. 3694–3703, 2014.

[2] Y. Li and Q. Xu, "Design and Robust Repetitive Control of a New Parallel-Kinematic XY Piezostage for Micro/Nanomanipulation," IEEE/ASME Trans. Mechatronics, vol. 17, pp. 1120–1132, 2012.

[3] C. Hu, B. Yao, Z. Chen and Q. Wang, "Adaptive Robust Repetitive Control of an Industrial Biaxial Precision Gantry for Contouring Tasks," IEEE Trans. Control Systems Technology, vol. 19, pp. 1559–1568, 2011.

[4] K.-C. Yang, and C. Hsieh, "Nanometer Positioning of a Dual-Drive Gantry Table with Precise Yaw Motion Control," J. CSME, vol. 36, no. 2, pp. 107–117, 2015.

[5] M. Tomizuka, "Zero Phase Error Tracking Algorithm for Digital Control," ASME J Dyn Syst Meas Control, vol. 109, pp. 65–68, 1987.

[6] G. Cheng, K. Peng, B.-M. Chen, and T.-H. Lee, "Improving Transient Performance in Tracking General References Using Composite Nonlinear Feedback Control and Its Application to High-Speed XY-Table Positioning Mechanism," IEEE Trans. Ind. Electron, vol. 54, pp. 1039–1051, 2007.

[7] E.C. Park, H. Lim and C.-H. Choi, "Position Control of X-Y Table at Velocity Reversal Using Presliding Friction Characteristics," IEEE Trans. Control Systems Technology, vol. 11, pp. 24–31, 2003.

[8] M. -C. Tsai, I.-F. Chiu and M. -Y. Cheng, "Design and Implementation of Command and Friction Feedforward Control for CNC Motion Controllers," IEE proceedings, Control Theory and Applications, vol. 151, Issue 1, pp. 13–20, Jan. 2004.

[9] Y. Koren, "Cross-Coupled Biaxial Computer Control for Manufacturing Systems," ASME J Dyn Syst Meas Control, vol. 102, pp. 265–272, 1980.

[10] M.-Y. Cheng and C.-C. Lee, "Motion Controller Design for Contour-Following Tasks Based on Real-Time Contour Error Estimation," IEEE Trans. Ind. Electron, vol. 54, pp. 1686–1695, 2007.

[11] K.-H. Su, and M.-Y. Cheng, "Contouring Accuracy Improvement Using Cross-Coupled Control and Position Error Compensator," International Journal of Machine Tools and Manufacture, vol. 48, pp. 1444–1453, 2008.

[12] H.-R. Chen, M.-Y. Cheng, C.-H. Wu, and K.-H. Su, "Real Time Parameter Based Contour Error Estimation Algorithms for Free Form Contour Following," International Journal of Machine Tools & Manufacture, vol. 102, pp.1–8, 2016.

[13] C.-M. Wen, and M.-Y. Cheng, "Contouring Accuracy Improvement of a Piezo-Actuated Micro Motion Stage Based on Fuzzy Cerebellar Model Articulation Controller," Control Engineering Practice, vol. 20, pp. 1195–1205, 2012.

[14] C.-Y. Chen and M.-Y. Cheng, "Velocity Field Control and Adaptive Virtual Plant Disturbance Compensation for Planar Contour Following Tasks," IET Control Theory & Applications, vol. 6, pp. 1182–1191, 2012.

[15] C.-M. Wen and M.-Y. Cheng, "Development of a Recurrent Fuzzy CMAC with Adjustable Input Space Quantization and Self-Tuning Learning Rate for Control of a Dual-Axis Piezoelectric Actuated Micromotion Stage," IEEE Trans. Ind. Electron, vol. 60, pp. 5105–5115, 2013.

[16] D.A. Bristow, M. Tharayil and A.G. Alleyne, "A Survey of Iterative Learning Control," IEEE Control Systems Magazine, vol. 26, pp. 96–114, 2006.

[17] C.-L. Chen and K.-S. Li, "Observer-Based Robust AILC for Robotic System Tracking Problem," J. CSME, vol. 30, no. 6, pp. 483–491, 2009.

[18] C.-K. Chen, C.-J. Lin, J. Hwang and C.-W. Hung, "The Iterative Learning Control of a Stewart Platform System," J. CSME, vol. 34, no. 1, pp. 21–30, 2013.

[19] K.L. Barton and A.G. Alleyne, "A Cross-Coupled Iterative Learning Control Design for Precision Motion Control," IEEE Trans. Control Systems Technology, vol. 16, pp. 1218–1231, 2008.

[20] M. Wiering and I. Martijn van Otterlo, Reinforcement Learning: State-of-the-Art. Germany: Springer-Verlag Berlin Heidelberg, 2012.

[21] R.S. Sutton and A.G. Barto, Reinforcement Learning : An Introduction. 2nd ed. London, England: in progress, 2012.

[22] D. Silver, A. Huang, C.J. Maddison, A. Guez, L. Sifre, G.V.D. Driessche, J. Schrittwieser, et al. "Mastering the Game of Go with Deep Neural Networks and Tree Search," Nature, vol. 529, pp. 484–489, 2016.

[23] M.-Y. Cheng, M.-C. Tsai and J.-C. Kuo, "Real-Time NURBS Command Generators for CNC Servo Controllers," International Journal of Machine Tools and Manufacture, vol. 42, pp. 801–813, 2002.

[24] Chun Wei, Zhe Zhang, Wei Qiao and Liyan Qu, "Reinforcement-Learning-Based Intelligent Maximum Power Point Tracking Control for Wind Energy Conversion Systems," IEEE Trans. Ind. Electron, vol. 62, no. 10, pp. 6360–6370, 2015.

[25] Sergey Levine, Peter Pastor, Alex Krizhevsky and Deirdre Quillen, "Learning Hand-Eye Coordination for Robotic Grasping with Deep Learning and Large-Scale Data Collection," unpublished.