Proceedings of the Thirteenth China-Japan International Workshop on Information Technology and Control Applications

Edited by

Luefeng Chen, Weihua Cao, Jian Sun, Xianbo Sun and Jinhua She

Enshi, Hubei, China

26–28 September 2020

China University of Geosciences

Hubei Minzu University

Proceedings of the Thirteenth China-Japan International Workshop on Information Technology and Control Applications

Enshi, Hubei, China

26–28 September 2020

Edited by

Luefeng Chen, Weihua Cao, Jian Sun, Xianbo Sun and Jinhua She

Organized by

China University of Geosciences **Tokyo University of Technology** **Beijing Institute of Technology**

Hosted by

Sponsored by

湖北省科学技术协会
Hubei Association For Science & Technology

富士技術出版株式会社
FUJI TECHNOLOGY PRESS LTD.

The Thirteenth China-Japan International Workshop on Information Technology and Control Applications

26–28 September 2020

Enshi, Hubei, China

Organizing Institutes

China University of Geosciences, China

Beijing Institute of Technology, China

Tokyo University of Technology, Japan

Hosts

China University of Geosciences, China

Hubei Minzu University, China

Co-Hosts

Advanced Control and Intelligent Automation for Complex Systems Overseas Expertise Introduction Center for Discipline Innovation

Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems

Engineering Research Center of Intelligent Technology for Geoexploration, Ministry of Education

Hubei Engineering Research Center of Intelligent Geological Equipment

Hubei Association of Automation

Sponsors

Hubei Association For Science & Technology

Fuji Technology Press Ltd.

Joseph Spencer, UK
Yang Shi, Canada
Chunyi Su, Canada
Takao Terano, Japan
Meiling Wang, China
Junzheng Wang, China
Qinghe Wu, China
Yuanqing Xia, China
Xin Xin, Japan
Yonghua Xiong, China
Li Xu, Japan
Ryuichi Yokoyama, Japan
Changfan Zhang, China
Xiaofeng Zong, China
Xianbo Sun, China

Organizing Committee Chairs

Luefeng Chen, China
Bin Xin, China
Hiroyuki Kameda, Japan
Kunwu Xie, China

CONTENTS

Session B-2: Image, Acoustic, Speech & Signal Processing

Zoom ID: 85797790328 Password: itca20 Date/Time: 2020.09.28 / 15:50 - 18:10

Session C-2: Distributed Control Methods

Zoom ID: 79562906419 Password: itca20 Date/Time: 2020.09.28 / 15:50 - 18:10

Multiview Rule-Based Modeling and Granular Aggregation

Witold Pedrycz*
* University of Alberta, Canada

Abstract: Multiview models, as the name stipulates, are models describing a real-world system observed from different points of view. In establishing such individual perspectives, we typically engage locally available features (attributes, input variables). When considered together, a collection of multiview models has to be aggregated. Multiview models also arise in the presence of data coming with a massive number of variables when building a monolithic model involving all attributes is neither feasible nor computationally sound.

We formulate and discuss these two categories of scenarios by focusing on fuzzy rule-based architectures. An important task when building an aggregate of multiview models is to endow the overall global model with a sound measure of quality using which one can efficiently assess the relevance of the individual results produced by the rule-based models as well as establish the quality of the overall fusion. We advocate that the quality of the results can be quantified in terms of some information granule. In the two scenarios outlined above, the family of multiview models is aggregated with the use of the augmented principle of justifiable granularity -one of the fundamentals of Granular Computing. The related optimization criteria of coverage and specificity are discussed along with the associated optimization process. The emergence of type-2 information granules is also shown.

Witold Pedrycz (IEEE Fellow, 1998) is Professor and Canada Research Chair (CRC) in Computational Intelligence in the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Canada. He is also with the Systems Research Institute of the Polish Academy of Sciences, Warsaw, Poland. In 2009 Dr. Pedrycz was elected a foreign member of the Polish Academy of Sciences. In 2012 he was elected a Fellow of the Royal Society of Canada. In 2007 he received a prestigious Norbert Wiener award from the IEEE Systems, Man, and Cybernetics Society. He is a recipient of the IEEE Canada Computer Engineering Medal, a Cajastur Prize for Soft Computing from the European Centre for Soft Computing, a Killam Prize, a Fuzzy Pioneer Award from the IEEE Computational Intelligence Society, and 2019 Meritorious Service Award from the IEEE Systems Man and Cybernetics Society.

His main research directions involve Computational Intelligence, fuzzy modeling and Granular Computing, knowledge discovery and data science, pattern recognition, data science, knowledge-based neural networks, and control engineering. He has published numerous papers in these areas; the current h-index is 114 (Google Scholar) and 87 on the list *top-h scientists for computer science and electronics* http://www.guide2research.com/scientists. He is also an author of 21 research monographs and edited volumes covering various aspects of Computational Intelligence, data mining, and Software Engineering.

Dr. Pedrycz is vigorously involved in editorial activities. He is an Editor-in-Chief of ***Information Sciences***, Editor-in-Chief of ***WIREs Data Mining and Knowledge Discovery*** (Wiley), and Co-editor-in-Chief of ***Int. J. of Granular Computing*** (Springer) and ***J. of Data Information and Management*** (Springer). He serves on an Advisory Board of ***IEEE Transactions on Fuzzy Systems*** and is a member of a number of editorial boards of international journals.

Modeling and Identification of Strain Wave Gearing for Motion Control Applications to Precision Positioning Devices

Makoto Iwasaki*
* Nagoya Institute of Technology, Japan

Abstract: The invited speech presents a practical motion controller design technique for precision positioning devices including strain wave gearing, e.g. industrial multi-axis robots. Since HarmonicDrive® gears (HDGs), a typical strain wave gearing, inherently possess nonlinear properties, such as Angular Transmission Errors (ATEs), nonlinear stiffness, nonlinear friction etc., due to structural errors and flexibility in the mechanisms, the ideal positioning accuracy corresponding to the apparent resolution cannot be essentially attained at the output of gearing in the devices. In addition, mechanisms with HDGs generally excite resonant vibrations due to the periodical disturbance by ATEs, especially in the condition that the frequency of synchronous components of ATE corresponds to the critical mechanical resonant frequency. This speech, therefore, focuses on the accurate settling and vibration suppression in positioning, in order to improve the performance deteriorations by applying several control approaches, e.g. a model-based feedforward control manner, and adaptive/robust feedback control manners. In the speech, at first, an approach toward modeling and identification of the ATEs is presented to achieve and apply the precise mathematical models to the motion controller design. Following to the accurate mathematical modeling, several motion controller design approaches are practically presented: a model-based feedforward compensation to improve the settling performance in the positioning, an adaptive variable notch filter design as a feedback compensation to effectively suppress the mechanical vibration, and an H_inf controller design to boost the robust capability. The proposed approaches have been applied to precision motion control of actual devices as servo actuators, and verified through numerical simulations and experiments.

Makoto Iwasaki received the B.S., M.S., and Dr. Eng. degrees in electrical and computer engineering from Nagoya Institute of Technology, Nagoya, Japan, in 1986, 1988, and 1991, respectively. Since 1991, he has been with the Department of Computer Science and Engineering, Nagoya Institute of Technology, where he is currently a Professor at the Department of Electrical and Mechanical Engineering. As professional contributions of the IEEE, he has been an AdCom member of IES in term of 2010 to 2019, a Technical Editor for IEEE/ASME TMech from 2010 to 2014, an Associate Editor for IEEE TIE since 2014, a Management Committee member of IEEE/ASME TMech (Secretary in 2016 and Treasurer in 2017), a Co-Editors-in-Chief for IEEE TIE since 2016, a Vice President for Planning and Development in term of 2018 to 2021, respectively. He is IEEE fellow class 2015 for "contributions to fast and precise positioning in motion controller design". He has received the Best Paper Award of Trans of IEE Japan in 2013, the Best Paper Award of Fanuc FA Robot Foundation in 2011, the Technical Development Award of IEE Japan in 2017, the Nagamori Awards in 2017, the Ichimura Prize in Industry for Excellent Achievement of Ichimura Foundation for New Technology in 2018, the Technology Award of the Japan Society for Precision Engineering in 2018, and the Commendation for Science and Technology by the Japanese Minister of Education in 2019, respectively. His current research interests are the applications of control theories to linear/nonlinear modeling and precision positioning, through various collaborative research activities with industries.

Development Direction of Human Coexistence Robot Partner Based on Smart Device

Jinseok WOO*
* Tokyo University of Technology, Japan

Abstract: Social isolation can cause problems on human both mentally and physically. In particular, the isolation of the elderly causes serious problems such as lonely death. However, social participation can lead to healthier lives by reducing isolation. Many technologies are being developed to reduce social isolation and loneliness. Accordingly, we considered to use robot partner in order to prevent social isolation. In this paper, we explain the current state about our development of robot partner, and we discuss development direction of robot. First, we show our robot partner system. Next, we explain the modular structured systems in order to develop robot partner system. Finally, we show several examples of the robot system, and discuss the applicability of proposed system.

Jinseok WOO graduated from the Kumoh National Institute of Technology, Republic of Korea in 2009, and received degree of master of engineering from Tokyo Metropolitan University, Japan in 2011. He has been received degree of doctor of engineering from Tokyo Metropolitan University, Japan in 2017. After he worked for Tokyo Metropolitan University, Japan as a research assistant professor, he joined the School of Engineering, Tokyo University of Technology, as an assistant professor at the Department of Mechanical Engineering in 2019. He has received the Excellent Paper Award of 22th Intelligent System Symposium, Fuzzy, Artificial Intelligence, Neural Networks and Computational Intelligence, Japan in 2012, the Best Presentation Award, International Conference on Fuzzy Theory and Its Applications, Republic of Korea in 2018. His current interests are in the fields of mechatronics, computational intelligence, and human-friendly robot partner.

Accelerated First-Order Distributed Method for Nash Equilibria of Convex-Concave Bilinear Two-Network Zero-Sum Game

Xianlin Zeng*
* Beijing Institute of Technology, China

Abstract: Due to the big data/scale of many modern applications and rise of distributed optimization/game, simple first-order gradient descent algorithms have become dominant and have been particularly revisited because gradients are often computed in a cheap and distributed way. Many researchers focus on the development of accelerated first-order algorithms. However, most existing works focus on unconstrained optimization problems. Convex-concave bilinear two-network zero-sum games have applications in several domains such as machine learning, optimization, and game theory. This talk will introduce a recent result on a dynamical accelerated first-order method for Nash equilibria seeking of a class of two-network zero-sum games. The proposed method achieves an O(1/t2) convergence rate in some sense, which is consistent to Nesterov's accelerated method for unconstrained optimization problems.

Xianlin Zeng received the B.E. and M.S. degrees in Control Science and Engineering from the Harbin Institute of Technology, Harbin, China, in 2009 and 2011, respectively, and the Ph.D. degree in Mechanical Engineering from the Texas Tech University in 2015. He worked as a postdoctoral researcher in National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing, China, from 08/2015 to 07/2017. After that, he worked as a postdoctoral researcher in Beijing Institute of Technology, Beijing, China, from 08/2017 to 08/2019. He joined the School of Automation, Beijing Institute of Technology, as an associate professor in 09/2019. He is with research specialization in the areas of nonsmooth cooperative control and decision of multi-agent systems. His research interests include distributed algorithms for nonsmooth optimization and games, optimization-based and game theory-based control of unmanned autonomous systems, and distributed computation of matrix equations and inequalities. He has published 5 first-author papers in top control science and engineering journals IEEE Transactions on Automatic Control, Automatica, and SIAM Journal on Control and Optimization.

Technology and Application of Intelligent Humanoid Robot System

Xin Chen*
* China University of Geosciences, China

Abstract: After recent advancement of computing and robotics technologies, intelligent robot system spread widely all walks of life. However, these robots face several typical engineering challenges. First of all, the intelligent perception of the environment is the basis of robot operation. Secondly, the robot has real-time decision-making ability to external information. Finally, high precision control can be performed to achieve the target operation. Two typical robot systems, live working robot system and dulcimer intelligent playing robot system, are introduced in detail in this report. A complex workspace environment perception technology put forward by combining broad-spectrum light source and binocular vision to realize high-precision positioning of different sizes of operation targets. A humanoid behavior learning model is proposed, which can realize intelligent real-time decision-making of dual-arm performance by learning the behavior habits of human professional players. A humanoid trajectory planning method is introduced to realize the anthropomorphic motion of two arms with joint constraint. Some preliminary results are demonstrated via video.

Xin Chen received his B.S. and M.S. degrees in engineering from Central South University, Changsha, China, in 1999 and 2002, respectively, and the Ph.D. degree in engineering from University of Macau, China, in 2007. He finished his postdoctoral research on control science and engineering at Central South University. He is currently a Professor with the School of Automation, China University of Geosciences. He won the Science Fund for Distinguished Young Scholars of Hubei Province, and the Distinguished Professor of “Chutian Scholar Program” of Hubei Province. He was a visiting professor with the Department of Electrical and Computer Engineering of the University of Alberta from December 2018 to December 2019.

His research interests include intelligent control, multi-agent systems, robotics and process control. He is the member of Technical Committee of Control Theory, Education Works Committee and Youth Works Committee of Chinese Association of Automation, also the member of Society of Intelligent Aerospace Systems of Chinese Association for Artificial Intelligence. He served as the program committee chair of 37th Chinese Control Conference (CCC 2018).

Vibration Suppression Based on Input Shaping and Adaptive Model Following Control

Lulu Wu*, **, Jinhua She*,**,***, Chuanke Zhang*,**, Zhentao Liu*,**
* School of Automation, China University of Geosciences, Wuhan 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
***School of Engineering, Tokyo University of Technology, Tokyo 192-0982, Japan

Abstract

Vibration suppression is of great significance to improve the positioning accuracy of servo systems. This paper develops a vibration suppression method based on input shaping and adaptive model following control. First, a zero vibration input shaper is used to suppress the vibration caused by an elastic load. This helps to obtain an ideal position output of the load based on a nominal model. Then, a compensation controller in model following control is designed to suppress the vibration caused by parameter changes. Finally, through analyzing the percentage residual vibration, the sum of squared position error is employed to optimize the parameters of the compensation controller. This ensures the method is adaptive to different system parameter changes. Comparisons with other input shaper methods show the effectiveness and superiority of the developed method.

Keywords: Vibration suppression, input shaping, model following control, parameter optimization.

1. INTRODUCTION

A servo system has been widely used in industrial applications due to its fast response and strong tracking ability. In order to achieve high load ratio performance, some elastic connecting devices, such as harmonic gears and speed bumps, are often used in servo systems. However, these elastic devices cause hysteresis errors in position transmission. It will affect the positioning performance of a servo system, in severe cases, it will cause the system unstable. Therefore, it is significant to suppress vibration to enhance the positioning accuracy of servo systems.

There are basically two types of vibration suppression methods: hardware-based and software-based. Furthermore, the software-based methods are divided into passive suppression methods and active suppression methods. The passive ones mainly include a filter methog [1], a H_∞ control method [2], and an iterative learning control method [3]. While the active ones mainly include an input shaping method, an observer method [4], a trajectory planning method [5].

Among these methods, input shaping technology has been widely applied in vibration suppression since it is simple and easy to implement. It was first known as posicast control that is proposed by Smith [6]. Then, Singer and Seering continued to improve it and proposed a complete input shaping theory. ZV (zero vibration) input shaper is the first one that has been systematically researched and put into use. It is simple, however, it is very sensitive to system modeling errors at the same time [7]. In order to overcome this drawback, many methods have been devised to enhance the robustness of a ZV input shaper. One approach is to add additional constraints to improve built-in robustness. For example, a ZVD (zero vibration and derivation) input shaper, a ZVDD (zero vibration and double derivation) input shaper, an EI (extra-insensitive) input shaper, and a SI (specified-insensitivity) input shaper [8]. Unfortunately, the robustness of the above shapers comes at the expense of the response time. Moreover, as the constraints increase, it is more difficult to calculate parameters of the input shapers. Another approach is to adapt shaper parameters either directly or indirectly to improve adaptive robustness. The indirect method focuses on updating the shaper parameters by identifying vibration information of a system from the time domain or the frequency domain [9, 10]. The direct method is to construct an adaptive algorithm from the output information of a system to adjust shaper parameters [11, 12].

The above methods based on input shaping are effective to suppress vibration of a servo system, but the response time of the system is longer due to a constant parameter adjustment of an input shaper. Moreover, it is difficult to suppress vibration when system parameters change greatly. A model following control system consists of a nominal model, an actual model, and a compensation controller. It is noteworthy that the error caused by parameter variation is quickly suppressed by designing a proper compensation controller. This method has good effects and a considerable scope of application [13,14].

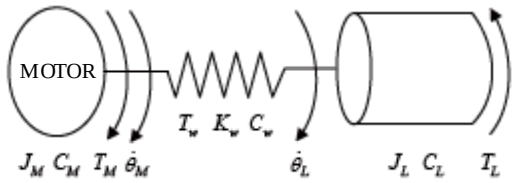


Fig.1 Typical two-inertia mechanical transmission model.

Therefore, combining input shaping with model following control is a good way to obtain satisfactory vibration suppression.

This paper presents a vibration suppression method that is based on input shaping and adaptive model following control to achieve fast and accurate vibration suppression under the situation of parameter changes. An ZV input shaper is first used to suppress the vibration caused by an elastic load and obtain an ideal position output of the load based on a nominal model. Then, a compensation controller is designed to suppress the vibration caused by parameter changes. After that, choosing the sum of squared position error as an objective function to optimize the parameters of the compensation controller to enhance suppression effect. Finally, simulations show the validity of the developed method.

2. ANALYSIS OF MATHEMATICAL MODEL

In practical industry applications, elastic gear devices are often used in order to achieve high load ratio performance. Generally, such a system is regarded as a typical two-inertia mechanical transmission system (Fig.1) for analysis.

Since the damping coefficient of motor and load are very small, their impact on the system is neglected to simplify the analysis. According to Fig.1, the following differential equations in the s domain is established

$$\begin{aligned} J_M\theta_M s^2 &= T_M - T_w \\ J_L\theta_L s^2 &= T_w - T_L \\ T_w &= C_w s(\theta_M - \theta_L) + K_w(\theta_M - \theta_L) \\ \omega_M &= \theta_M s \\ \omega_L &= \theta_L s \end{aligned} \tag{1}$$

Thus

$$\begin{aligned} \frac{\omega_M}{T_M} &= \frac{J_L s^2 + C_w s + K_w}{J_M J_L s^3 + (J_M + J_L)C_w s^2 + K_w(J_M + J_L)s} \\ \frac{\omega_L}{T_M} &= \frac{C_w s + K_w}{J_M J_L s^3 + (J_M + J_L)C_w s^2 + K_w(J_M + J_L)s} \\ \frac{\theta_L}{\theta_M} &= \frac{\omega_L}{\omega_M} = \frac{C_w s + K_w}{J_L s^2 + C_w s + K_w} \end{aligned} \tag{2}$$

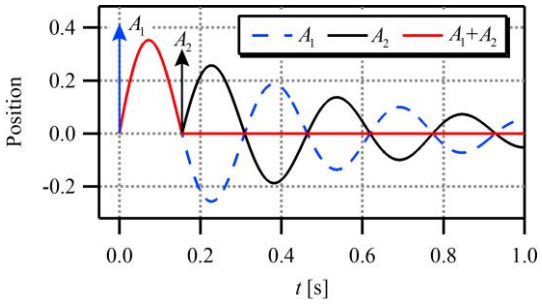


Fig.2 Vibration suppression using two impulse.

In (2), the variables and parameters are defined as follows:

ω_M [rad/s]	Angular velocity (motor)
ω_L [rad/s]	Angular velocity (load)
θ_M [rad]	Motor position
θ_L [rad]	Load position
T_M [Nm]	Electromagnetic torque
T_L [Nm]	Load torque
T_w [Nm]	Torsional torque Torsional torque
C_w [N·m·s/rad]	Spring damping coefficient
K_w [N·m/rad]	Spring stiffness coefficient
J_M [kg·m^2]	The moment of inertia (motor)
J_L [kg·m^2]	The moment of inertia (load)

According to the analysis of a second-order system, the natural frequency and damping ratio are $\omega_n = \sqrt{K_w/J_L}$ and $\xi = \sqrt{C_w^2/4J_L K_w}$, respectively. It is clear that when the system parameters, such as J_L, K_w, and C_w, have variations, the vibration frequency and damping ratio of the system will change at the same time.

3. VIBRATION SUPPRESSION BASED ON INPUT SHAPING AND ADAPTIVE MODEL FOLLOWING CONTROL

This section first explains the design of vibration suppression based on input shaping and model following control. Then, the relationship among the PRV (percentage residual vibration), vibration frequency, and damping ratio is analyzed to obtain an objective function to optimize the compensation controller.

3.1 Input Shaping

As a first step to understand how this approach suppress vibration, it is helpful to start with the simplest input signal - an impulse (Fig.2).

The pulse amplitude and delay time of a conventional two-pulse ZV input shaper are

$$[A_i] = \begin{bmatrix} \frac{K}{1+K} & \frac{1}{1+K} \end{bmatrix}, \quad [t_i] = \begin{bmatrix} 0 & \frac{\pi}{\omega_d} \end{bmatrix} \tag{3}$$

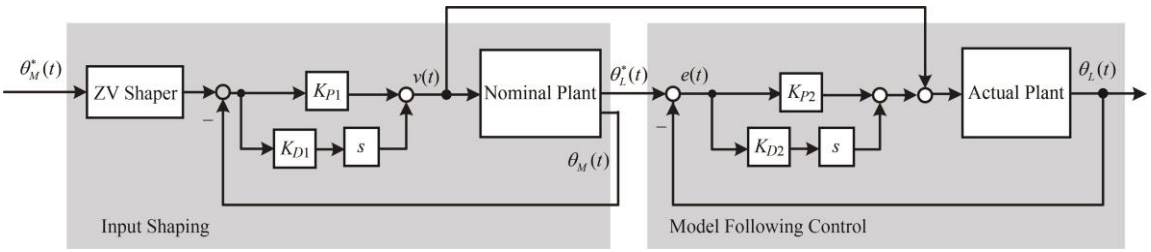


Fig.3 Vibration suppression based on input shaping and model following control.

where

$$K = e^{\frac{-\zeta\pi}{\sqrt{1-\zeta^2}}},\ \omega_d = \omega_n\sqrt{1-\zeta^2} \tag{4}$$

By setting the derivative, with respect to frequency, to equal zero, we get a ZVD shaper. Furthermore, replacing the constraint of zero vibration at the modeling frequency with a constraint that merely limited the vibration to a small value, an EI shaper is obtained. The expression of the two shapers are described in [15].

From the above equations, it is clear that a considerable change in system parameters will result in poor vibration suppression performance of an input shaper. Vaughan *et al*. had compared the sensitivity of different input shapers to parameter modeling errors in [15].

3.2 Model Following Control

As described in Section 3.1, input shaping is a simple feedforward method to suppress vibration, however, when system parameters have changes, an input shaper cannot achieve zero residual vibration. Therefore, this paper develops a vibration method that combines input shaping and model following control to achieve fast suppression speed and high suppression accuracy.

In order to guarantee a simple control structure and a fast vibration suppression speed, this paper employs a ZV shaper and a PD compensation controller. As shown in the block diagram of vibration suppression based on input shaping and model following control (Fig.3), a ZV shaper is first used to suppress the vibration caused by an elastic load. Then, an ideal load position output without vibration is obtained. After that, the ideal output is set as the reference signal of load position input for the actual model. Finally, a compensation controller is designed to improve the vibration suppression effect.

3.3 Parameter optimization of compensation controller

The vibration caused by parameter changes can be effectively suppressed by designing a proper compensation controller. However, if the compensation controller employs fixed parameters, it cannot approach satisfactory effect when system parameters change in a wide range. Taking this problem into consideration, adjusting parameters of the compensation controller during operation time helps to achieve a better vibration suppression performance.

For an underdamped second-order system, the impulse response is

$$y_0(\mathrm{t}) = \frac{A_0\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n(t-t_0)} \sin\left(\omega_n\sqrt{1-\zeta^2}\,(t-t_0)\right) \tag{5}$$

where A_0 is the amplitude of the pulse, t_0 is the time the impulse is applied.

Using the vibration frequency, $\omega_d = \omega_n\sqrt{1-\zeta^2}$, to get the simplification

$$y_0(\mathrm{t}) = \frac{A_0\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n(t-t_0)} \sin\left(\omega_d\,(t-t_0)\right) \tag{6}$$

The amplitude of residual vibration from a single unity-magnitude impulse applied at time zero is

$$A_{\uparrow} = \frac{\omega_n}{\sqrt{1-\zeta^2}} \tag{7}$$

The amplitude of residual vibration from a sequence of impulses after the last impulse applied at $t = t_n$ is

$$A_{\Sigma} = \frac{\omega_n}{\sqrt{1-\zeta^2}} e^{-\zeta\omega_n t_n} \sqrt{\left[C(\omega_n,\zeta)\right]^2 + \left[S(\omega_n,\zeta)\right]^2} \tag{8}$$

where

$$\begin{aligned} C(\omega_n,\zeta) &= \sum_{i=1}^{n} A_i e^{\zeta\omega_n t_i}\cos(\omega_d t_i) \\ S(\omega_n,\zeta) &= \sum_{i=1}^{n} A_i e^{\zeta\omega_n t_i}\sin(\omega_d t_i) \end{aligned} \tag{9}$$

Dividing (8) by (7) yields the PRV equation

$$V(\omega_n,\zeta) = \frac{A_{\Sigma}}{A_{\uparrow}} = e^{-\zeta\omega_n t_n}\sqrt{\left[C(\omega_n,\zeta)\right]^2 + \left[S(\omega_n,\zeta)\right]^2} \tag{10}$$

$$V(\omega_n',\zeta') = e^{-\zeta'\omega_n' t_2}\sqrt{\left[C(\omega_n',\zeta')\right]^2+\left[S(\omega_n',\zeta')\right]^2}$$
$$=e^{-\zeta'\omega_n'\frac{\pi}{\omega_d'}}\sqrt{\left[\frac{1}{1+K}e^{-\zeta'\omega_n'\frac{\pi}{\omega_d'}}+\frac{K}{1+K}e^{\zeta'\omega_n'\frac{\pi}{\omega_d'}}\cos\left(\omega_d'\frac{\pi}{\omega_d}\right)\right]^2+\left[\frac{K}{1+K}e^{\zeta'\omega_n'\frac{\pi}{\omega_d'}}\sin\left(\omega_d'\frac{\pi}{\omega_d}\right)\right]^2} \quad (11)$$
$$=e^{-bc}\sqrt{\left[\frac{1}{1+K}e^{-bc}+\frac{K}{1+K}e^{bc}\cos(c\pi)\right]^2+\left[\frac{K}{1+K}e^{bc}\sin(c\pi)\right]^2}$$

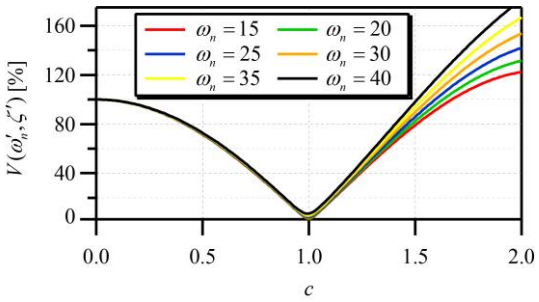


Fig.4 Relationship among PRV, ω_n and ζ.

Table 1. Parameters of the two-inertia system.

Parameter	Value	Parameter	Value
K_e [N·m/A]	1.5	K_t [N·m/A]	1.5
J_L [kg·m^2]	0.008	J_M [kg·m^2]	0.006
C_w [N·m·s/rad]	0.01	K_w [N·m/rad]	5

However, for a two-inertia system, when system parameters change, both natural frequency and damping ratio will change too. In particular, there is a relationship between them ($\zeta=a\omega_n$, $a=C_w/(2K_w)$). Under the case of both ω_n and ζ have changes, the PRV equation is shown in (11), where $b=\zeta'\pi/\left(\sqrt{1-\zeta^2}\right)$, $c=\omega_n'/\omega_n$. According to the above expression, the relationship among PRV, ω_n and ζ is obtained (Fig.4).

When the PRV exceeds a desired range ε, it is considered that there is a parameter change, resulting in the vibration that cannot be suppressed satisfactory by a ZV shaper. Assuming that the system cycle time is T and the sampling time is T_s, then the number of samples per cycle is $N=T/T_s$. The sum of squared position error is

$$S=\sum_{k=1}^{N}\left[V(\omega_n',\zeta')\times A_{\uparrow}(k)\right]^2$$

where $A_{\uparrow}(k)$ is the amplitude of residual vibration without input shaping at the *k-th* sample time. In fact, when system parameters have different changes, the value of the sum of squared position error is different too. This paper selects this value as an objective function to optimize the compensation controller.

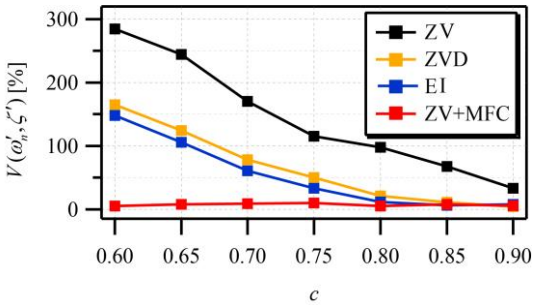


Fig.5 Results of vibration suppression of different methods.

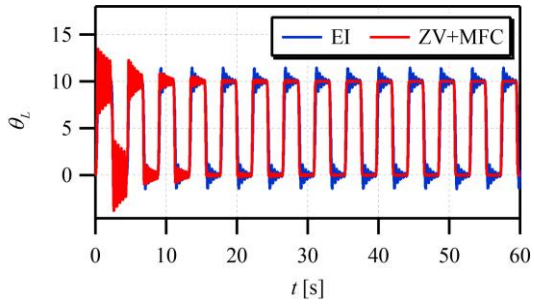


Fig.6 Results of load position curve for change in J_L.

4. NUMERICAL VERIFICATION

Simulations were carried out for a servo system (Table 1) to demonstrate the effectiveness of the developed method. Vibration suppression was carried out and compared for the developed method and three input shaper methods for different change values in the moment of inertia (load).

For changes in J_L varies within 100%-278%, corresponding to c varies within 60%-100%, a comparison among different input shapers (the ZV shaper, the ZVD shaper, and the EI shaper) and the developed method (Fig.~¥ref{Fig:5}) shows that the developed method suppresses the vibration below a desired value in a wide range of frequency change. When the system frequency changes within 80%-90%, the ZV shaper, the EI shaper, and the developed method almost achieve the same effect of vibration suppression. However, when the system frequency changes within 60%-80%, the PRV of the developed method is below 10% while the one of other input shaper methods is beyond 20%. A typical result of the EI shaper and the developed method

($c = 0.6$) is presented in Fig.6. For the EI shaper and the developed method, the peak value of vibration is 1.478 (147.8%) and 0.065 (6.50%), respectively. This presents that the developed method has a strong adaptability to changes of vibration frequency.

5. CONCLUSION

This paper presented a vibration suppression method based on input shaping and adaptive model following control. This method achieves a fast and accurate effect of vibration suppression when system parameters change in a wide range. An ZV input shaper was used to suppress the vibration caused by an elastic load and a compensation controller was designed to suppress the vibration caused by parameter changes. Furthermore, through analyzing the PRV, the sum of squared position error was chosen to optimize the parameters of the compensation controller. Simulations demonstrate the effectiveness of the method and its superior to other input shaper methods.

References

[1] L. Biagiotti, C. Melchiorri, and L. Moriello, "Optimal trajectories for vibration reduction based on exponential filters", IEEE Transactions on Control Systems Technology, Vol. 24, No. 2, 2016, pp. 609-622.

[2] W. Sun, H. Gao, and B. Yao, "Adaptive robust vibration control of full-car active suspensions with electrohydraulic actuators", IEEE Transactions on Control Systems Technology, Vol. 21, No. 6, 2013, pp. 2417-2422.

[3] S. Wang, J. Zhao, and J. Wang, "Open-closed-loop iterative learning control for hydraulically driven fatigue test machine of insulators", Journal of Vibration and Control, Vol. 21, No. 12, 2015, pp. 2291-2305.

[4] W. He and S. Ge, "Vibration control of a nonuniform wind turbine tower via disturbance observer", IEEE/ASME Transactions on mechatronics, Vol. 20, No. 1, 2014, pp. 237-244.

[5] J. Lou, Y. Wei, G. Li, Y. Yang, and F. Xie, "Optimal trajectory planning and linear velocity feedback control of a flexible piezoelectric manipulator for vibration suppression", Shock and Vibration, Vol. 2015, 2015, pp. 1-11.

[6] O. Smith, "Posicast control of damped oscillatory systems", Proceedings of the IRE, Vol. 45, No. 9, 1957, pp. 1249-1255.

[7] W. Singhose, "Command shaping for flexible systems: A review of the first 50 years", International journal of precision engineering and manufacturing, Vol. 10, No. 4, 2009, pp. 153-168.

[8] J. Kim and E. Croft, "Preshaping input trajectories of industrial robots for vibration suppression", Robotics and Computer-Integrated Manufacturing, Vol. 54, 2018, pp. 35-44.

[9] A. Tzes and S. Yurkovich, "An adaptive input shaping control scheme for vibration suppression in slewing flexible structures", IEEE Transactions on Control Systems Technology, Vol. 1, No. 2, 1993, pp. 114-121.

[10] E. Pereira, J. Trapero, I. Díaz, and V. Feliu, "Adaptive input shaping for manoeuvring flexible structures using an algebraic identification technique", Automatica, Vol. 45, No. 4, 2009, pp. 1046-1051.

[11] T. Chen and J. Shan, "Fixed-time consensus control of multiagent systems using input shaping", IEEE Transactions on Industrial Electronics, Vol. 66, No. 9, 2018, pp. 7433-7441.

[12] M. Cole, "A discrete-time approach to impulse-based adaptive input shaping for motion control without residual vibration", Automatica, Vol. 47, No. 11, 2011, pp. 2504-2510.

[13] M. Pai, "Dynamic output feedback rbf neural network sliding mode control for robust tracking and model following", Nonlinear Dynamics, Vol. 79, No. 2, 2015, pp. 1023-1033.

[14] S. Mobayen, "Finite-time robust-tracking and model-following controller for uncertain dynamical systems", Journal of Vibration and Control, Vol. 22, No. 4, 2016, pp. 1117-1127.

[15] J. Vaughan, A. Yano, and W. Singhose, "Comparison of robust input shapers", Journal of Sound and Vibration, Vol. 315, 2008, pp. 797-815.

Trajectory Azimuth Control Based on Equivalent-Input-Disturbance Approach for Directional Drilling Process

Zhen Cai*,**, Xuzhi Lai*,**,†, Min Wu*,**, Chengda Lu*,**, Luefeng Chen*,**
* School of Automation, China University of Geosciences, Wuhan 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
† Corresponding author: Xuzhi Lai, E-mail: laixz@cug.edu.cn

Abstract

This paper is concerned with trajectory azimuth control in directional drilling. The motion process of the drill bit and a series of stabilizers are described, and a state-space model of the trajectory azimuth is built. The scheme of the trajectory azimuth control system is designed based on the equivalent input disturbance approach. An internal model is inserted to represent the drill bit for improving the tracking performance, and a state observer is combined with a low-pass filter to estimate the trajectory azimuth by measuring the azimuth of bottom hole assembly. The control parameters can be obtained by the condition of the system stability, which are derived in terms of linear matrix inequalities. A typical case is used to illustrate the validity and robustness of our approach.

Keywords: directional drilling process, trajectory azimuth, tracking control, equivalent input disturbance.

1. INTRODUCTION

Directional drilling can drill the curved boreholes with the downhole robot system in the complex geological environments. The downhole robotic actuators of the Rotary Steering System (RSS) are often used in the actual project [1]. The operation engineers can control the bit orientation by sending operational commands to the RSS actuators, which can achieve the control of the drilling trajectory.

At present, many researchers have done many works on the trajectory control. Panchal et al. [2] proposed a control strategy based on the borehole propagation process, this control strategy ignores the transient behavior of physical borehole propagation in physics. Bayliss et al. [3] derived the state-space model for drilling expansion and designed the controller based on the proposed model, but it can not capture the basic delay characteristics of borehole propagation dynamics. He also proposed some approaches to improve the control effect. Actually, the effects of time-delay and disturbance on the system are not considered in the above control studies.

In the actual project, the drilling trajectory consists of the trajectory inclination and trajectory azimuth. Since the trajectory azimuth denotes the turning trend of the drilling trajectory, it is essential to the trajectory control in the drilling process. However, few scholars study the azimuth control, and our previous work [4] was presented in only for the drilling attitude control but not the trajectory azimuth control. The article mainly uses the trajectory azimuth as the entry point to study the control problem of the drilling trajectory.

Strictly speaking, the trajectory control system is nonlinear during the drilling process. Considering that the trajectory evolution delay and the coupling between the trajectory inclination and trajectory azimuth, the trajectory system is uncertain and nonlinear. This style system is difficulty to the control of the drilling trajectory. To achieve the control effectively, it is necessary to develop some control strategies to deal with these situations. The disturbance observer-based control method and the sliding-mode control method are proposed to deal with the uncertainties and nonlinearities conditions [5,6]. However, they increase the complexity of the model and reduce system control performance.

To solve the above situation effectively, She et al. [7] proposed a disturbance suppression approach based on the concept of an equivalent input disturbance (EID). The method suppresses the exogenous disturbances without requiring a piece of prior information on the disturbance. Since the EID system configuration is simple, and the parameters are easy to design, this approach has widely applied to the linear systems, the uncertain systems, the nonlinear systems, and the time-delay systems [8,9]. Notably, for the delay-time and nonlinearities of the drill string system, the EID approach is used to improve the control performance of the drilling-string system [10]. Therefore, the EID method provides a fine scheme for us to solve the trajectory azimuth control for directional drilling process.

The paper proposes a trajectory reference-tracking and disturbance rejection control strategy for the trajectory azimuth. The model of the trajectory azimuth is

established according to the evolution prcess of the trajectory turn. The structure of the trajectory control system is built based on the EID, and the robustness of the control system is analyzed in detail. According to the Lyapunov stability theory, a sufficient condition that formed a set of linear matrix inequalities (LMIs) is derived to guarantee the asymptotic stability. Finally, a typical case illustrates the validity of the method.

2. TRAJECTORY AZIMUTH MODEL AND PROBLEM FORMULATION

In this work, the trajectory evolution process is considered in a vertical plane. The section mainly explains the trajectory azimuth model and problem formulation.

1.1 Trajectory Azimuth Model

The geometric description of the drilling trajectory is shown in Fig. 1. The base coordinate of drilling trajectory is $[e_x, e_y, e_z]^T$, where e_z is the gravity direction, e_x and e_z are vertical, and e_x and e_y are orthogonal, respectively. The process of trajectory formation is also considered as the trajectory evolution. Due to the complexity of the drilling trajectory evolution, only the motion process of the trajectory azimuth is studied.

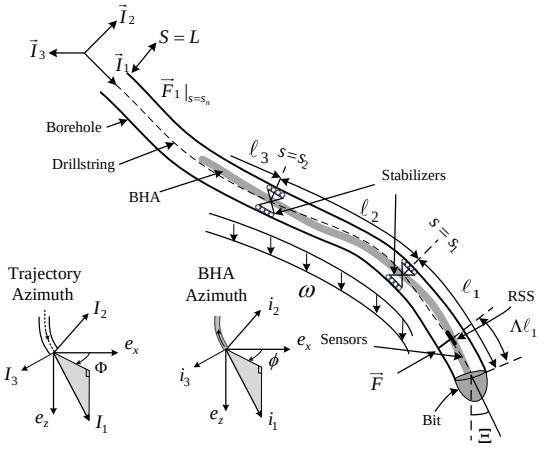


Fig. 1 The trajectory evlution description

The length of drilling trajectory S is described by the trajectory orientation $\Xi(S)$ with respect to the vertical line, where $S \in [0, L]$, L denotes the trajectory length; S indicates a curvilinear coordinate position measured along drilling trajectory. The dimensionless length of drilling trajectory ξ is defined as $\xi = L/\ell_1$, where ℓ_1 indicates the distance between the bit and the first stabilizer, ξ is the independent variable for the dynamics question of the trajectory propagation. Each section between two stabilizers is regarded as an Euler-Bernoulli beam. However, the section $\Lambda\ell_1$ between the bit and the RSS steering device is different from the section between the two stabilizers ℓ_i, $i = 2, \cdots, n$.

The two nonlinear delay differential equations can represente the evolution process of the trajectory azimuth Φ, and the equations also describe the relationship between the BHA azimuth and trajectory azimuth. To simplify the description of the trajectory evolution, the equations are given directly as follows [11]:

$$\eta\Pi(\phi - \Phi) = \mathcal{F}_b(\phi - \langle\Phi\rangle_1) + \mathcal{F}_r\Gamma + \sum_{i=1}^{n-1}\mathcal{F}_i(\langle\Phi\rangle_i - \langle\Phi\rangle_{i+1}), \quad (1a)$$

$$-\chi\Pi\dot{\phi} = \mathcal{M}_b(\phi - \langle\Phi\rangle_1) + \mathcal{M}_r\frac{\Gamma}{\sin\Theta} + \sum_{i=1}^{n-1}\mathcal{M}_i(\langle\Phi\rangle_i - \langle\Phi\rangle_{i+1}), \quad (1b)$$

where η denotes the lateral steering resistance, χ denotes the angular steering resistance, Π denotes the scale active weight-on-bit, ϕ denotes the BHA azimuth, Φ denotes the trajectory azimuth, Γ_Φ denotes the RSS forces on the Φ, respectively. The variable $\langle\Phi\rangle_1$ denotes the average values of trajectory azimuth between the drill bit and the first stabilizer; the variable $\langle\Phi\rangle_2$ denotes the average values of trajectory azimuth from the first stabilizer to the second stabilizer. They describe to

$$\langle\dot{\Phi}\rangle_\iota = [\Phi(\xi_i) - \Phi(\xi_{i-1})]/\varkappa_i, \quad (2)$$

where $\xi_i = \xi - \sum_{j=1}^{i}\varkappa_j$ and $\xi_{i-1} = \xi - \sum_{j=1}^{i-1}\varkappa_j$, $j = 1, \cdots, i$, $i = 1, 2$.

Moreover, the parameter Λ is a scale factor and the parameters $\varkappa_i = \ell_i/\ell_1$, $i = 1,2$, is the dimensionless position length of the i-th segment of BHA. Here, given with two stabilizers (i.e., $n = 2$). the coefficients $\mathcal{F}$ and $\mathcal{M}$ denote the dimensionless influence coefficients, and they are provided by

$$\begin{aligned}
\mathcal{F}_1 &= 6/(3+4\varkappa_2), \quad \mathcal{M}_1 = -2/(3+4\varkappa_2),\\
\mathcal{F}_b &= -(6+4\varkappa_2)/(3+4\varkappa_2),\\
\mathcal{M}_b &= 4(1+\varkappa_2)/(3+4\varkappa_2),\\
\mathcal{F}_w &= (6+10\varkappa_2-3\varkappa_2^3)/(12+16\varkappa_2),\\
\mathcal{M}_w &= (-1-2\varkappa_2+\varkappa_2^3)/(12+16\varkappa_2),\\
\mathcal{F}_r &= [-3-4\varkappa_2+\Lambda^2(9+6\varkappa_2)-2\Lambda^3(3+\varkappa_2)]\\
&\qquad /(3+4\varkappa_2),\\
\mathcal{M}_r &= \big[\Lambda(1-\Lambda)\big(3+4\varkappa_2-\Lambda(3+2\varkappa_2)\big)\big]/(3+4\varkappa_2).
\end{aligned}$$

Assume that the desired trajectory azimuth $\Phi_r(\xi)$ is continuously differentiable for the complex stratum. The BHA has infinite stiffness, and two stabilizers are considered, i.e., $n = 2$.

Combining with (1a) and (1b), the differential equation of the trajectory azimuth is as follows:

$$\chi\Pi\dot{\Phi} = \mathcal{M}_b(\langle\Phi\rangle_1 - \Phi) + \frac{\chi}{\eta}\mathcal{F}_b(\Phi - \Phi_1) + \left(\frac{\mathcal{F}_b\mathcal{M}_1 - \mathcal{M}_b\mathcal{F}_1}{\eta\Pi} - \mathcal{M}_1\right)(\langle\Phi\rangle_1 - \langle\Phi\rangle_2) \quad (3)$$

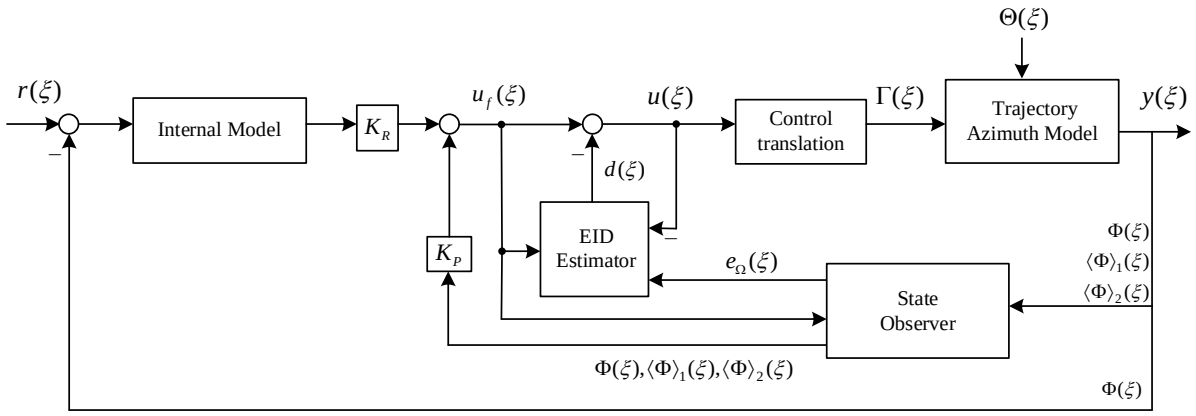


Fig.2 The structure of drilling control system

$$-\frac{\chi}{\eta}\mathcal{F}_1\left(\frac{\Phi-\Phi_1}{\varkappa_1}-\frac{\Phi_1-\Phi_2}{\varkappa_2}\right)+\frac{\chi}{\eta}\frac{\mathcal{F}_r\dot{\Theta}\cos\Theta}{(\sin\Theta)^2}\Gamma$$
$$+\left(\frac{\mathcal{F}_b\,\mathcal{M}_r-\mathcal{M}_b\mathcal{F}_r}{\eta\Pi}-\mathcal{M}_r\right)\frac{\Gamma}{\sin\Theta}-\frac{\chi}{\eta}\frac{\mathcal{F}_r}{\sin\Theta}\dot{\Gamma}.$$

The equation (3) with distributed delays is transformed into a first-order state-space formulation with pointwise delays by considering the average trajectory azimuth $\langle\Phi\rangle_i$.

Define $\dot{z}(\xi):=\mathrm{col}(\Phi(\xi),\langle\Phi\rangle_1(\xi),\langle\Phi\rangle_2(\xi))$, the state-space equation is as follow:

$$\dot{z}(\xi)=A_0z(\xi)+A_1z(\xi-\tau_1)+A_2z(\xi-\tau_2)+B_0\Gamma(\xi)+B_1\dot{\Gamma}(\xi), \quad (4)$$

where $\tau_1=\varkappa_1,\ \tau_2=\varkappa_1+\varkappa_2$, and the system matrices

$$A_0=\frac{1}{\chi\Pi}\begin{bmatrix}-\mathcal{M}_b+\frac{\chi}{\eta}(\mathcal{F}_b-\frac{\mathcal{F}_1}{\varkappa_1}) & \mathcal{M}_b+G & -G\\ \frac{\chi\Pi}{\varkappa_1} & 0 & 0\\ 0 & 0 & 0\end{bmatrix},$$

$$A_1=\frac{1}{\chi\Pi}\begin{bmatrix}\frac{\chi}{\eta}(\frac{\mathcal{F}_1}{\varkappa_1}+\frac{\mathcal{F}_1}{\varkappa_2}-\mathcal{F}_b) & 0 & 0\\ -\frac{\chi\Pi}{\varkappa_1} & 0 & 0\\ \frac{\chi\Pi}{\varkappa_2} & 0 & 0\end{bmatrix},$$

$$A_2=\frac{1}{\chi\Pi}\begin{bmatrix}-\frac{\chi}{\eta}\frac{\mathcal{F}_1}{\varkappa_2} & 0 & 0\\ 0 & 0 & 0\\ -\frac{\chi\Pi}{\varkappa_2} & 0 & 0\end{bmatrix},$$

$$B_0=\frac{1}{\chi\Pi}\begin{bmatrix}\frac{\chi}{\eta}\frac{\mathcal{F}_r\dot{\Theta}\cos\Theta}{(\sin\Theta)^2}+\frac{M}{\sin\Theta} & 0 & 0\end{bmatrix}^T,$$

$$B_1=\frac{1}{\chi\Pi}\begin{bmatrix}-\frac{\chi}{\eta}\frac{\mathcal{F}_r}{\sin\Theta} & 0 & 0\end{bmatrix}^T,$$

$$G=\frac{\mathcal{F}_b\,\mathcal{M}_1-\mathcal{M}_b\mathcal{F}_1}{\eta\Pi}-\mathcal{M}_1,\ M=\frac{\mathcal{F}_b\,\mathcal{M}_r-\mathcal{M}_b\mathcal{F}_r}{\eta\Pi}-\mathcal{M}_r.$$

To simplify the control input in the model (4), the control input $u(\xi)$ are defined as: $Bu(\xi)=B_0\Gamma(\xi)+B_1\dot{\Gamma}(\xi)$. (4) is rewritten as follow: $\dot{z}(\xi)=A_0z(\xi)+A_1z(\xi-\tau_1)+A_2z(\xi-\tau_1)+Bu(\xi)$. Due to the input matrices B_0 and B_1 related to the trajectory azimuth Θ, this simplification will have an impact on the system.

Remark 1 There is the coupling between the trajectory inclination and trajectory inclination in (4). The change of the trajectory inclination has influences for the trajectory azimuth control. Moreover, the delay characteristic of the model itself also affects the control effect. Since the influence of these factors needs to be eliminated, the EID method is used in the control system.

1.2 Problem Formulation

The structure of EID-based control system is shown in Fig. 2, which consists of the plant, an internal model of the reference input, the state observer, the control translation, and an EID estimator. The inner loop contains the state observer and the EID estimator, and the outer one provides the internal model and the output feedback.

There are the trajectory evolution delay and coupling with the trajectory inclination in the drilling process. The influence of the two factors can be regarded as the system disturbance $\Psi(\xi)$. So, the control input $u(\xi)$ should contain two parts: $u_f(\xi)$ for the tracking of the reference input and $\tilde{d}(\xi)$ for the compensation for the system disturbance $\Psi(\xi)$, that is,

$$u(\xi)=u_f(\xi)-\tilde{d}(\xi). \quad (5)$$

Since the term $\Psi(\xi)$ can be treated as a state-dependent EID, letting the EID be $\tilde{d}(\xi)$ to describe the plant as

$$\dot{z}(\xi)=A_0z(\xi)+A_1z(\xi-\tau_1)+A_2z(\xi-\tau_2)+B\left[u(\xi)+\tilde{d}(\xi)\right], \quad (6)$$
$$y(\xi)=Cz(\xi),$$

where $\tau_1=\varkappa_1,\ \tau_2=\varkappa_1+\varkappa_2$.

The state observer is constructed as

$$\dot{\hat{z}}(\xi)=A_0\hat{z}(\xi)+A_1\hat{z}(\xi-\tau_1)+A_2\hat{z}(\xi-\tau_2)+B\left[u(\xi)+\tilde{d}(\xi)\right]+L[y(\xi)-\hat{y}(\xi)], \quad (7)$$
$$\hat{y}(\xi)=C\hat{z}(\xi),$$

where L is the observer gains.

Define the state estimation error to be $e_\Omega(\xi) \coloneqq \mathrm{col}\{e(\xi), e(\xi-\tau_1), e(\xi-\tau_2)\}$ for the state variables $z(\xi)$, $z(\xi-\tau_1)$ and $z(\xi-\tau_2)$, and it is written to be

$$e(\xi-\tau_n) = z(\xi-\tau_n) - \hat{z}(\xi-\tau_n), \tag{8}$$

where $n = 0,1,2$, $\tau_0 = 0$, and $\tau_1 < \tau_2$.

Substituting (8) into (6) yields

$$\begin{aligned}\dot{\hat{z}}(\xi) = {} & A_0\hat{z}(\xi) + A_1\hat{z}(\xi-\tau_1) + A_2\hat{z}(\xi-\tau_2) \\ & + Bu(\xi) + [A_0e(\xi) + A_1e(\xi-\tau_1) \\ & + A_2e(\xi-\tau_2) - \dot{e}(\xi) + B\tilde{d}(\xi)].\end{aligned} \tag{9}$$

To reveal the relationship between (7) and (9), suppose that there exist control inputs $\Delta d(\xi)$ satisfying

$$\begin{aligned}B\Delta d(\xi) = {} & A_0e(\xi) + A_1e(\xi-\tau_1) \\ & + A_2e(\xi-\tau_2) - \dot{e}(\xi).\end{aligned} \tag{10}$$

Substituting (9) into (8) yields

$$\begin{aligned}\dot{\hat{z}}(\xi) = {} & A_0\hat{z}(\xi) + A_1\hat{z}(\xi-\tau_1) + A_2\hat{z}(\xi-\tau_2) \\ & + B[u(\xi) + \hat{d}(\xi)],\end{aligned} \tag{11}$$

where

$$\hat{d}(\xi) = \tilde{d}(\xi) + \Delta d(\xi). \tag{12}$$

Combining (5), (7) and (11) leads to

$$B[\hat{d}(\xi) + u(\xi) - u_f(\xi)] = LCe(\xi). \tag{13}$$

Thus, the least square solution of $\hat{d}(\xi)$ is

$$\hat{d}(\xi) = B^+LCe(\xi) + u_f(\xi) - u(\xi), \tag{14}$$

where $B^+ = (B^TB)^{-1}B^T$ is a generalized inverse of B.

Due to the reference trajectory azimuth tracked precisely, the following internal model is inserted to track the movement of the drill bit in the control loop

$$\dot{z}_R(\xi) = A_Rz_R(\xi) + B_R[r(\xi) - Ey(\xi)], \tag{15}$$

where $E = [1,0,0]$ is used to extract $\Phi(\xi)$ from $y(\xi)$, and A_R, and B_R are constants selected according to on-site requirements. The internal model (15) gives

$$u_f(\xi) = K_P\hat{z}(\xi) + K_Rz_R(\xi), \tag{16}$$

where K_P and K_R denote the control gains, respectively.

Noting (12), an EID estimator is constructed in the form of a low-pass filter to estimate $\tilde{d}(\xi)$ from $\hat{d}(\xi)$

$$\dot{z}_F(\xi) = A_Fz_F(\xi) + B_F\hat{d}(\xi), \tag{17}$$

$$\tilde{d}(\xi) = C_Fz_F(\xi), \tag{18}$$

where A_F, B_F, C_F are selected matrices of appropriate dimensions according to the actual drilling situation. (17) and (18) are used to select the angular frequency bandwidth for the EID estimation. It satisfies

$$F(jw) \approx 1, \forall\omega \in [0, \omega_r], \tag{19}$$

where ω_r is the highest angular frequency for disturbance estimation. The cut-off frequency of the EID estimator is larger than ω_r, and it is suggested to be ten times that of the reference.

Then, let the external signals be zero, i.e. $r(\xi) = 0$. Notice that

$$y(\xi) = Cz(\xi), \tag{20}$$

$$y(\xi) - \hat{y}(\xi) = Ce(\xi). \tag{21}$$

Substituting (20) into (15) gives

$$\dot{z}_R(\xi) = -B_REC\hat{z}(\xi) - B_RECe(\xi) + A_Rz_R(\xi). \tag{22}$$

Substituting (21) into (7) yields

$$\begin{aligned}\dot{\hat{z}}(\xi) = {} & A_0\hat{z}(\xi) + A_1\hat{z}(\xi-\tau_1) + A_2\hat{z}(\xi-\tau_2) \\ & + Bu_f(\xi) + LCe(\xi).\end{aligned} \tag{23}$$

Subtracting (4) by (7) yields

$$\begin{aligned}\dot{e}(\xi) = {} & (A_0 - LC)e(\xi) + A_1e(\xi-\tau_1) + A_2e(\xi-\tau_2) \\ & + Bu(\xi) - Bu_f(\xi),\end{aligned} \tag{24}$$

substituting (5), (8), (17) and (18) into (24) yields

$$\begin{aligned}\dot{e}(\xi) = {} & (A_0 - LC)e(\xi) + A_1e(\xi-\tau_1) + A_2e(\xi-\tau_2) \\ & - BC_Fz_F(\xi).\end{aligned} \tag{25}$$

Then, substituting (5) and (14) into (17) leads to

$$\dot{z}_F(\xi) = (A_F + B_FC_F)z_F(\xi) + B_FB^+LCe(\xi). \tag{26}$$

Denoting $\delta(\xi) \coloneqq \mathrm{col}\{\hat{z}(\xi), e(\xi), z_F(\xi), z_R(\xi)\}$ for the system state variables. Combining (22), (23), (25) and (26), the state-space equations of the closed-loop system is

$$\begin{aligned}\dot{\delta}(\xi) = {} & \bar{A}_0\delta(\xi) + \bar{A}_1\delta(\xi-\tau_1) + \bar{A}_2\delta(\xi-\tau_2) \\ & + \bar{B}u_f(\xi),\end{aligned} \tag{27}$$

where the system matrix

$$\bar{A}_0 = \begin{bmatrix} A_0 & LC & 0 & 0 \\ 0 & A_0 - LC & -BC_F & 0 \\ 0 & B_FB^+LC & A_F + B_FC_F & 0 \\ -B_REC & -B_REC & 0 & A_R \end{bmatrix},$$

$$\bar{A}_1 = \begin{bmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},\ \bar{A}_2 = \begin{bmatrix} A_2 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$\bar{B} = [B^T \quad 0 \quad 0 \quad 0]^{\mathrm{T}}$.

The control law is

$$u_f(\xi) = \bar{K}\delta(\xi), \tag{28}$$

where $\bar{K}$ is the control gain, and $\bar{K} = [K_P, 0, 0, K_R]$.

Substituting (28) into (27), the closed-loop system is normalized as

$$\dot{\delta}(\xi) = \mathcal{A}_0\delta(\xi) + \mathcal{A}_1\delta(\xi-\tau_1) + \mathcal{A}_2\delta(\xi-\tau_2), \tag{29}$$

where $\mathcal{A}_0 = \bar{A}_0 + \bar{B}\bar{K}$, $\mathcal{A}_1 = \bar{A}_1$, $\mathcal{A}_2 = \bar{A}_2$.

The objective of the paper can be formulated in detail as follows: Design a control law to track the desired trajectory azimuth $\Phi_r(\xi)$, by using the EID estimator obtaining the control input $u(\xi)$, while the system state tends to be stable quickly.

3. CONTROLLER DESIGN

The section mainly describes the system stability analysis and controller design.

Lemma 1 *The closed-loop system (29) with the control*

law (28) is asymptotically stable, if there are symmetric positive definite matrices $\mathcal{X}_i$, $\mathcal{Y}_i$, $\mathcal{J}_i$ $(i = 1,2,3,4)$, $\mathcal{X}_{11}$ *and* $\mathcal{X}_{22}$, *appropriate matrix* $\mathcal{W}_1$, $\mathcal{W}_2$, *which make the following feasible*

$$\begin{bmatrix} \psi_{11} & \psi_{12} & \psi_{13} & \psi_{14} & \psi_{15} \\ * & -\psi_{22} & 0 & 0 & 0 \\ * & * & -\psi_{33} & 0 & 0 \\ * & * & * & -\psi_{44} & 0 \\ * & * & * & * & -\psi_{55} \end{bmatrix} < 0, \qquad (30)$$

where

$$\psi_{11} = \begin{bmatrix} \varphi_{11} & LC\mathcal{X}_2 & 0 & \varphi_{14} \\ * & \varphi_{22} & \varphi_{23} & -\mathcal{X}_2 C^T E B_R^T \\ * & * & \varphi_{33} & 0 \\ * & * & * & \varphi_{44} \end{bmatrix},$$

$\varphi_{11} = A_0\mathcal{X}_1 + \mathcal{X}_1 A_0^T + B\mathcal{W}_1 + \mathcal{W}_1^T B^T,$
$\varphi_{44} = A_R\mathcal{X}_4 + \mathcal{X}_4 A_R^T,$
$\varphi_{14} = B\mathcal{W}_2 - \mathcal{X}_1 C^T E B_R^T,$
$\varphi_{22} = A_0\mathcal{X}_2 + \mathcal{X}_2 A_0^T - (LC\mathcal{X}_2 + \mathcal{X}_2 C^T L^T),$
$\varphi_{23} = -BC_F\mathcal{X}_3 + \mathcal{X}_2 C^T L^T B^{+T} B_F^T,$
$\varphi_{33} = (A_F + B_F C_F)\mathcal{X}_3 + \mathcal{X}_3(A_F^T + C_F^T B_F^T),$
$\psi_{12} = \text{diag}\{A_1\mathcal{Y}_1, A_1\mathcal{Y}_2, 0, 0\},$
$\psi_{13} = \text{diag}\{A_2\mathcal{J}_1, A_2\mathcal{J}_2, 0, 0\},$
$\psi_{14} = \psi_{15} = \text{diag}\{\mathcal{X}_1, \mathcal{X}_2, \mathcal{X}_3, \mathcal{X}_4\},$
$\psi_{22} = \psi_{44} = \text{diag}\{\mathcal{Y}_1, \mathcal{Y}_2, \mathcal{Y}_3, \mathcal{Y}_4\},$
$\psi_{33} = \psi_{55} = \text{diag}\{\mathcal{J}_1, \mathcal{J}_2, \mathcal{J}_3, \mathcal{J}_4\}.$

Then, the controller gains are as follows:

$K_P = \mathcal{W}_1\mathcal{X}_1^{-1},\ K_R = \mathcal{W}_2\mathcal{X}_4^{-1}.$

Proof. Choose the following Lyapunov functional of (29) as follows:

$$V(\delta(\xi)) = \delta^T(\xi)\mathcal{P}\delta(\xi) + \int_{\xi-\tau_1}^{\xi} \delta^T(\xi)\mathcal{R}\delta(\xi)ds + \int_{\xi-\tau_2}^{\xi} \delta^T(\xi)\mathcal{Q}\delta(\xi)ds, \qquad (31)$$

where $\mathcal{P}$, $\mathcal{R}$, $\mathcal{Q}$ are positive definite and diagonal matrices with compatible dimensions.

The derivative of (31) is as follows:

$$\begin{aligned} \dot{V}(\delta(\xi)) &= 2\delta^T(\xi)\mathcal{P}\dot{\delta}(\xi) + \delta^T(\xi)\mathcal{R}\delta(\xi) \\ &\quad + \delta^T(\xi)\mathcal{Q}\delta(\xi) - \delta^T(\xi-\tau_1)\mathcal{R}\delta(\xi-\tau_1) \\ &\quad - \delta^T(\xi-\tau_2)\mathcal{Q}\delta(\xi-\tau_2) \\ &= \mathcal{Z}^T(\xi)\Omega\mathcal{Z}(\xi) \end{aligned} \qquad (32)$$

where $\mathcal{Z}(\xi) = [\delta^T(\xi)\ \ \delta^T(\xi-\tau_1)\ \ \delta^T(\xi-\tau_2)]^T$ and

$$\Omega = \begin{bmatrix} \mathcal{P}\mathcal{A}_0 + \mathcal{A}_0^T\mathcal{P} + \mathcal{R} + \mathcal{Q} & \mathcal{P}\mathcal{A}_1 & \mathcal{P}\mathcal{A}_2 \\ * & -\mathcal{R} & 0 \\ * & * & -\mathcal{Q} \end{bmatrix}.$$

If $\Omega < 0$, then there exits a sufficiently small positive scale ε such that $\dot{V}(\delta(\xi)) \le -\varepsilon\|\delta(\xi)\|^T$, the closed-loop system (29) is asymptotically stable. By the Schur complement, Ω is equivalent to be

$$\begin{bmatrix} \mathcal{P}\mathcal{A}_0 + \mathcal{A}_0^T\mathcal{P} & \mathcal{P}\mathcal{A}_1 & \mathcal{P}\mathcal{A}_2 & I & I \\ * & -\mathcal{R} & 0 & 0 & 0 \\ * & * & -\mathcal{Q} & 0 & 0 \\ * & * & * & -\mathcal{R}^{-1} & 0 \\ * & * & * & * & -\mathcal{Q}^{-1} \end{bmatrix} < 0 \qquad (33)$$

Let

$\mathcal{X}_i = \mathcal{P}_i^{-1},\ \mathcal{Y}_i = \mathcal{R}_i^{-1},\ \mathcal{J}_i = \mathcal{Q}_i^{-1},\ i = 1,2,3,4,$

and

$K_P\mathcal{X}_1 = \mathcal{W}_1,\ K_R\mathcal{X}_4 = \mathcal{W}_2.$ (34)

Then, left- and Right-multiplying (33) by the block matrix $\text{diag}\{\mathcal{X}, \mathcal{Y}, \mathcal{J}, I, I\}$, combining with (34), the LMI (30) is obtained. Meanwhile, the control gains are given.

4. SIMULATION RESULTS AND ANALYSIS

A benchmark system is set with the two stabilizers. It is considered that the BHA is composed of a series of steel pipes. There are some geometry parameters of the BHA, such as Young's modulus E_y, density ρ of the BHA pipe, the inner radius I_r, the outer radius O_r, cross-sectional area $A = \pi(O_r^2 - I_r^2)$ and second moment of inertia $I = (O_r^4 - I_r^4)\pi/4$. Moreover, the BHA inherent property parameters are set as follows: the Young's modulus $E_y = 2e11\ \text{N/m}^2$.

The system parameters are selected according to the design of directional drilling system by Marck et al. [12]. These parameters are given in Table I.

Table 1. The system drilling parameters.

Parameter	Value	Paramter	Value
ℓ_1	3.66m	$\varkappa_1$	1
ℓ_2	6.10m	$\varkappa_2$	1.67
Λ	0.167	$\Lambda\ell_1$	0.61m
I_r	0.053m	O_r	0.086m
χ	0.1	η	30
γ	0.0024	Π	0.0087

The parameters of the inner model are selected as follow: $A_R = -0.001,\ B_R = 1$. To ensure the feasibility of LMI (30), A_R is selected as -0.001 instead of 0. Although this processing will produce a minimal error, it is reasonable for the efficient calculation. For satisfying the condition of (19), the state-space parameters of a low-pass filter are $A_F = -100,\ \ B_F = 100,\ C_F = 1.$

According to the setting of the above system parameters, combining the stability analysis and controller design, the control gains are obtained, that is, $K_P = [-2146 -$

1050 524], $K_R = 2547$. These values are rounded for reducing the system calculation.

The designed trajectory azimuth comes from actual engineering. The length of the drilling trajectory ξ is 400 (the corresponding dimensionless range is approximately 1464 m). The desired trajectory azimuth is shown as follows:

$$\Phi_r(\xi) = \begin{cases} 95^\circ, & \xi \in [0,100] \\ 95^\circ - 0.5^\circ\xi, & \xi \in [100,200] \\ 45^\circ, & \xi \in [200,250] \\ 45^\circ - 0.5^\circ\xi, & \xi \in [250,300] \\ 0^\circ, & \xi \in [300,400] \end{cases}$$

In the actual drilling, the initial value of the trajectory azimuth suddenly changed to 95°, which is not suitable for control the steering device. To solve the question, let $\Phi_r(\xi) - 95^\circ$ without changing the desired trajectory.

To illustrate the effectiveness of the proposed control strategy, the PI control method is used to compare with our method. Their distinct is that the PI controller replaces the EID estimator. In comparison, the control parameters are the same.

The unknown disturbances are added to the system at $\xi = 150$ and $\xi = 270$ in Fig. 3, respectively. The unknown disturbance is set as a step function and its amplitude changes to 8°. Such disturbance is considered as a big disturbance in practical engineering, which is an excellent challenge to the drilling control.

Actually, it is shown in Fig. 3 that the trajectory azimuth is variable in the region where the unknown disturbance occurs, and the system tends to the steady-state quickly. Compared with PI control, our method can converge rapidly and has little change.

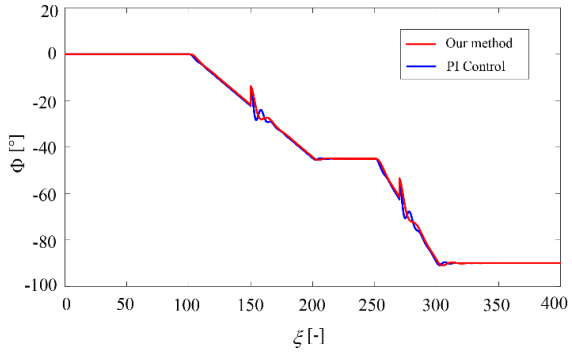


Fig. 3 The azimuth response with unknown disturbance

Fig. 4 shows the error contrast effect of the trajectory azimuth. It can be seen that there are large fluctuations when the disturbance occurs, but our method can quickly converge to 0, and the change is small. Compared with PI control, our method reduces the azimuth error by 2° at the maximum amplitude, which is very important in practical projects.

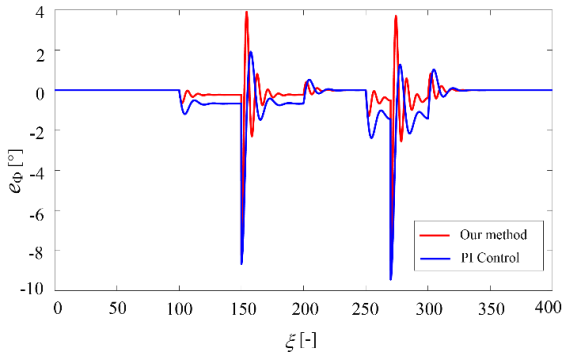


Fig. 4 The azimuth error with unknown disturbance

5. CONCLUSION

The control of the trajectory azimuth with the evolution delay and the angle coupling has been investigated. By analyzing the moment characteristics of the trajectory azimuth, the dynamic model has been derived from the trajectory evolution equations. Then, an EID-based control system has developed two feedback loops: the inner one estimates the trajectory azimuth, and the outer one realizes the reference tracking. The closed-loop control system is designed by some novel condition in terms of linear-matrix-inequalities, and the system stability is proved. The simulation results illustrated the validity of our approach.

Acknowledgement

This work was supported by the National Key R&D Program of China under Grant 2018YFC0603405, the National Natural Science Foundation of China under Grant 61733016, the Hubei Provincial Technical Innovation Major Project under Grant 2018AAA035, the 111 project under Grant B17040, and the Fundamental Research Funds for the Central Universities under Grant CUGCJ1812.

References

[1] N. Demirer, U. Zalluhoglu, J. Marck, R. Darbe and M. Morari, "Autonomous directional drilling with rotary steerable systems", In Proc. 2019 Amer. Control Conf. (ACC), Philadelphia, PA, USA, pp. 5203-5208, 2019.

[2] N. Panchal, M. T. Bayliss, and J. F. Whidborne, "Vector-based kinematic closed-loop attitude control-system for directional drilling", IFAC Proc. Vol., vol. 45, no. 8, pp. 78-83, 2012.

[3] M. T. Bayliss and J. F. Whidborne, "Mixed uncertainty analysis of pole placement and H_∞ controllers for directional drilling attitude tracking", J. Dyn. Sys., Meas., Control, vol. 137, no. 121008, pp. 1-8, Dec. 2015.

[4] Z. Cai, X. Z. Lai, and M. Wu, L. F. Chen, and C. D. Lu, "Compensation control for tool attitude in directional drilling systems", 2019 12th Asian Control Conference (ASCC), Kitakyusyu International Conference Center, Japan, June 9-12, 2019, 376-380.

[5] M. F. Shakib, E. Detournay, and N. V. D. Wouw, "Nonlinear dynamic modeling and analysis of borehole propagation for directional drilling", Int. J. Nonlin. Mech., vol. 113, pp. 178-201, Jul. 2019.

[6] N. A. H. Kremers, E. Detournay, and N. V. D. Wouw, "Model-based robust control of directional drilling systems", IEEE Trans. Control Syst. Technol., vol. 24, no. 1, pp. 226-239, Jan. 2016.

[7] J. H. She, M. Fang, Y. Ohyama, H. Hashimoto, and M. Wu, "Improving disturbance rejection performance based on an equivalent input disturbance approach", IEEE Trans. Ind. Electron., vol. 55, no. 1, pp. 380-389, Jan. 2008.

[8] F. Gao, M. Wu, J.-H. She, and W. Cao, "Disturbance rejection in nonlinear systems based on equivalent-input-disturbance approach", Appl. Math. Comput., vol. 282, pp. 244-253, May. 2016.

[9] Z. Yan, X. Z. Lai, Q. Meng, and M. Wu, "A novel robust control method for motion control of uncertain Single-Link FLexible-Joint manipulator", IEEE Trans. Syst., Man, Cybern., Syst., to be published. doi: 10.1109/TSMC.2019.2900502.

[10] C. Lu, M. Wu, X. Chen, W. Cao, C. Gan, and J.-H. She, "Torsional vibration control of drill-string systems with time-varying measurement delays", Inf. Sci., vol. 467, pp. 528-548, Oct. 2018.

[11] L. Perneder, "A three-dimensional mathematical model of directonal drilling[thesis]", Minneapolis, MN: Faculty of the Graduate School, University of Minnesota, 2013.

[12] J. Mark, and E. Detournay, "Analysis of spiraled-borehole data by use of a noval directional-drilling model", SPE Drill. Complet., vol. 29, no. 3, pp. 267-278, 2014.

Position Control of Machine Tool Moving Axis Based on Sliding Mode Control

Sanqiu LIU*, Wangyong HE*, Haogui LI*
* School of Automation, China University of Geosciences
Wuhan, Hubei 430074, P. R. China
* Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems
Wuhan, Hubei 430074, P. R. China

Abstract

Aiming at high-precision tracking performance requirements of machine tool moving axis control, this paper establishes system mathematical model considering the elastic deformation of ball screw, and designs a sliding mode controller to suppress the influence of uncertainty on the control performance. Then an extended state observer is designed to observe the system state and disturbance, and feedback to the sliding mode controller for position control. Finally, the correctness of the designed sliding mode control and extended state observer are proved by MATLAB simulation analysis.

Keywords: AC Servo System, Machine Tool Moving Axis, Tracking Control, Sliding Mode Control, Extended State Observer.

1. INTRODUCTION

PMSM (Permanent Magnet Synchronous Motor) has the advantages of small size, reliable structure, high efficiency and simple control, is widely used in CNC machine tools, robots, aerospace and other fields [1]. Motion axis servo system drives the worktable to move according to the given input through PMSM, which is common in the position control system of machine tools. This requires the control system to have good dynamic performance and steady-state accuracy. However, servo system of machine tool, including electromechanics, mechanical dynamics, control science and so on, is a complex control system. Among many influencing factors, the disturbance and uncertainty make it difficult to establish an accurate mathematical model. The disturbance and uncertainty of system include unmodeled dynamics, parameter uncertainty and external disturbance [2]. Unmodeled dynamics is mainly determined by its complex internal structure; parameter uncertainty is due to the time-varying mechanical and electrical parameters of the system; external disturbance as the most serious factor affecting the system, mainly includes load force and friction. In view of these disturbances and uncertainties, although the conventional PID (proportion–integral–derivative) control is widely used in the field of industrial control because of its simple control structure, easy implementation, and mathematical model independent of the plant. But when the internal parameters and external disturbance change greatly, PID control can't meet the system performance requirements. In order to improve the robustness of the system and obtain satisfactory control performance, it is necessary to design advanced control method to suppress the disturbance and uncertainty.

In recent decades, based on the idea of PID feedback control, some advanced control strategies have been applied in various control fields. For servo system, there are mainly active disturbance rejection control [3], adaptive control [4], robust control [5], sliding mode control [6], etc. Moreover, with the deepening of research, some advanced control strategies are combined with each other to form new control strategies with complementary advantages, such as adaptive sliding mode control [7], neural network sliding mode control [8]. Among them, sliding mode control is widely used in high-performance servo motor control because of its insensitivity to uncertainty, and its easy implementation [9].

In this paper, a sliding mode control method is designed to track the position of the motion axis of machine tool under the condition of disturbance and uncertainty. Firstly, the system plant is modeled, and then the sliding mode controller is designed for the plant. At the same time, in order to obtain the internal states and disturbance information of the system, the extended state observer is designed. Finally, the simulation analysis of the above design verifies the correctness of the design.

2. MODELING OF MOTION AXIS SERVO SYSTEM

In the position control of machine tools, system modeling has the most important influence on control performance. The motion axis servo system mainly includes motor drive unit, mechanical transmission unit and signal detection unit. Motor drive unit drives the servo motor to rotate, mechanical drive unit converts the rotation motion

of the motor into linear motion of workbench, and signal detection unit detects the working state of workbench or motor during the working process and feeds it back to controller. The system structure is shown in the Fig.1, where v is the speed of the workbench, f is the friction force and F_d is the external disturbance of the workbench.

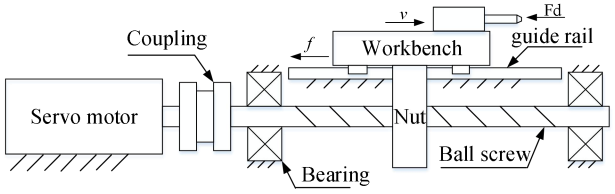


Fig.1 Schematic diagram of system structure.

For the surface-mount PMSM, in the *d-q* rotating coordinate system, when $i_d = 0$, the mechanical equation of PMSM is

$$\begin{cases} T_e - T_L = J_m\ddot{\theta}_m + B_m\dot{\theta}_m \\ T_e = k_i i_q \end{cases} \tag{1}$$

Where T_e is output electromagnetic torque of the motor, T_L is load torque of motor, J_m is motor shaft moment of inertia, B_m is motor damping coefficient, θ_m is output angle of motor shaft, i_a is current of motor q axis, k_i is the motor torque constant.

For the schematic diagram of the system structure, there are two degrees of freedom: the rotation of motor shaft and the movement of workbench. Lagrange equation is established for the whole controlled plant

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} + \frac{\partial D}{\partial \dot{q}_j} = Q_j \qquad j = 1,2 \tag{2}$$

Where q is generalized coordinate of the system

$$q = \begin{bmatrix} \theta_m & x_l \end{bmatrix}^T \tag{3}$$

Where x_l is the displacement of workbench.

L is lagrange function, which is the difference between kinetic energy T and potential energy V of the system. The potential energy is mainly the elastic potential energy produced by the torsion deformation of ball screw

$$L = T - V = \frac{1}{2}J_m\dot{\theta}_m^2 + \frac{1}{2}M\dot{x}_l^2 - \frac{1}{2}K_T\left(\frac{x_l}{i} - \theta_m\right)^2 \tag{4}$$

Where M is total mass converted to the workbench, K_T is torsional stiffness of ball screw, and i is screw transmission ratio.

D is dissipation function of the system, mainly including the dissipation energy produced by viscous damping in the system

$$D = \frac{1}{2}B_m\dot{\theta}_m^2 + \frac{1}{2}B\dot{x}_l^2 \tag{5}$$

Where B is viscous damping coefficient converted from mechanical transmission unit to workbench.

Q is generalized force of the system, mainly the electromagnetic torque T_e of the motor and the external disturbance F_d of the workbench

$$Q = \begin{bmatrix} T_e & F_d \end{bmatrix}^T \tag{6}$$

Substituting equation (3) ~ (6) into equation (2) gives

$$\begin{cases} J_m\ddot{\theta}_m + K_T\left(\theta_m - \dfrac{x_l}{i}\right) + B_m\dot{\theta}_m = T_e \\ M\ddot{x}_l - \dfrac{1}{i}K_T\left(\theta_m - \dfrac{x_L}{i}\right) + B\dot{x}_l = F_d \end{cases} \tag{7}$$

Then the torque balance equation of the servo motor is

$$T_e = k_i i_q = J_m\ddot{\theta}_m + B_m\dot{\theta}_m + i\left(M\ddot{x}_l + B\dot{x}_l - F_d\right) \tag{8}$$

Let the input $u=i_q$. The state vector $\boldsymbol{x}=[x_1\ x_2\ x_3\ x_4]^T=[\ x_l\ \ \theta_m\ \ \dot{x}_l\ \ \dot{\theta}_m\ ]$ are the output displacement of the workbench, the output angle of the motor shaft, the linear speed of the workbench and the angular velocity of the motor shaft respectively. Then the equation of state space is written as follows

$$\begin{bmatrix} \dot{x}_l \\ \dot{\theta}_m \\ \ddot{x}_l \\ \ddot{\theta}_m \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \dfrac{-K_T}{i^2M} & \dfrac{K_T}{iM} & \dfrac{-B}{M} & 0 \\ \dfrac{K_T}{iJ_m} & \dfrac{-K_T}{J_m} & 0 & \dfrac{B_m}{J_m} \end{bmatrix} \begin{bmatrix} x_l \\ \theta_m \\ \dot{x}_l \\ \dot{\theta}_m \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \dfrac{k_i}{J_m} \end{bmatrix} u + \begin{bmatrix} 0 \\ 0 \\ \dfrac{1}{M} \\ 0 \end{bmatrix} F_d \tag{9}$$

In servo motor control, it is often used as a three loop structure, that are current loop, speed loop and position loop in turn from inside to outside. The output of the position and speed controller can only output current i_q by the current loop, and then produce electromagnetic torque. Since the electromagnetic time constant is far less than mechanical time constant, the response speed of current loop is much faster than that of speed loop and position loop. Therefore, it is assumed that the current loop is a proportional link with a coefficient of 1. The output of the outer loop controller is proportional to the electromagnetic torque [10].

3. SYSTEM CONTROLLER DESIGN

3.1 Design of Sliding Mode Controller

Because sliding mode control is not sensitive to system disturbance and parameter perturbation, and easy to implement, the sliding mode controller is selected to control the system. Sliding mode control generally

consists of the following two steps.

(1) Determine the switching function $s(x)$, that is, determine the sliding mode switching surface.
(2) Determined corresponding control function u so that the sliding mode of the system exists, and the system can stabilize and reach switching surface within a limited time.

According to the state space equation of the system mentioned above, the sliding surface in sliding mode control can be designed as follows, where $\boldsymbol{C} = [c_1\ c_2\ c_3\ 1]$.

$$s(\boldsymbol{x}) = \boldsymbol{C}^T \boldsymbol{x} = \sum_{i=1}^{4} c_i x_i = \sum_{i=1}^{3} c_i x_i + x_4 \tag{10}$$

The position tracking problem of servo system is essentially to control the error signal. The error signal is defined and the sliding mode function is designed as follows

$$\begin{cases} e_1 = x_1^* - x_1 = x_l^* - x_l \\ e_2 = x_2^* - x_2 = \theta_m^* - \theta_m \\ e_3 = x_3^* - x_3 = \dot{x}_l^* - \dot{x}_l \\ e_4 = x_4^* - x_4 = \dot{\theta}_m^* - \dot{\theta}_m \end{cases} \tag{11}$$

$$s = c_1 e_1 + c_2 e_2 + c_3 e_3 + e_4 \tag{12}$$

Where C_1, C_2 and C_3 must satisfy Hurwitz condition, so that, let $C_1 = \lambda^3$, $C_2 = 3\lambda^2$, $C_3 = 3\lambda$, $\lambda > 0$.

The Lyapunov function is defined

$$V_1 = \frac{1}{2} s^2 \tag{13}$$

Then

$$\begin{aligned} \dot{s} &= c_1 \dot{e}_1 + c_2 \dot{e}_2 + c_3 \dot{e}_3 + \dot{e}_4 \\ &= c_1 \left(\dot{x}_l^* - \dot{x}_l \right) + c_2 \left(\dot{\theta}_m^* - \dot{\theta}_m \right) + \\ &c_3 \left[\ddot{x}_l^* - \left(\frac{K_T}{iM} \theta_m - \frac{K_T}{i^2 M} x_l - \frac{B}{M} \dot{x}_l + \frac{1}{M} F_d \right) \right] + \\ &\left[\ddot{\theta}_m^* - \left(-\frac{K_T}{J_m} \theta_m - \frac{B_m}{J_m} \dot{\theta}_m + \frac{K_T}{i J_m} x_l + \frac{k_i}{J_m} u \right) \right] \end{aligned} \tag{14}$$

In order to reduce the chattering effect of sliding mode control, the exponential reaching law is adopted

$$\dot{s} = -\varepsilon \operatorname{sgn} s - ks \qquad s > 0, k > 0 \tag{15}$$

By substituting the above formula into equation (14), the sliding mode control rate is obtained as follows

$$\begin{aligned} u = \frac{J_m}{k_i} &\left\{ \begin{array}{l} \varepsilon \operatorname{sgn} s + ks + \ddot{\theta}_m^* + c_1 \left(\dot{x}_l^* - \dot{x}_l \right) + c_2 \left(\dot{\theta}_m^* - \dot{\theta}_m \right) + \\ c_3 \left[\ddot{x}_l^* - \left(\frac{K_T}{iM} \theta_m - \frac{K_T}{i^2 M} x_l - \frac{B}{M} \dot{x}_l + \frac{1}{M} F_d \right) \right] \end{array} \right\} \\ &+ \frac{K_T}{k_i} \theta_m + \frac{B_M}{k_i} \dot{\theta}_m - \frac{K_T}{i k_i} x_l \end{aligned} \tag{16}$$

In the above control rate, the control variable u cannot be realized because the disturbance F_d is unknown. When the upper and lower bounds of F_d are known, the control rate can be designed by the bounds of F_d

$$\begin{aligned} u = \frac{J_m}{k_i} &\left\{ \begin{array}{l} \varepsilon \operatorname{sgn} s + ks + \ddot{\theta}_m^* + c_1 \left(\dot{x}_l^* - \dot{x}_l \right) + c_2 \left(\dot{\theta}_m^* - \dot{\theta}_m \right) \\ + c_3 \left[\ddot{x}_l^* - \left(\frac{K_T}{iM} \theta_m - \frac{K_T}{i^2 M} x_l - \frac{B}{M} \dot{x}_l \right) \right] \end{array} \right\} \\ &+ \frac{K_T}{k_i} \theta_m + \frac{B_m}{k_i} \dot{\theta}_m - \frac{K_T}{i k_i} x_l - F_{de} \end{aligned} \tag{17}$$

Where

$$F_{de} = \frac{J_m c_3}{M k_i} F_d \qquad (F_{dL} \le F_{de} = \frac{J_M c_3}{M k_i} F_d \le F_{dU}) \tag{18}$$

In order to satisfy the stability condition, let

$$F_{de} = \frac{F_{dU} + F_{dL}}{2} - \frac{F_{dU} - F_{dL}}{2} \operatorname{sgn} s \tag{19}$$

3.2 Design of Extended State Observer

The state observer can observe the internal state of the plant according to its input and output, this makes it easy to get some state variables that are difficult to measure in the system, such as velocity and acceleration, for the sliding mode control rate of the system, the speed of the workbench and the angular speed of the motor can be obtained by the observer. However, the control rate also contains the unknown disturbance information of the system, so the extended state observer can be designed to observe the system state and unknown disturbance. The extended state observer takes the disturbance as the expanded state, and feeds back the observed value to the controller for control compensation.

The extended state observer is designed as follows

$$\begin{cases} \dot{\hat{x}}_l = \dot{\hat{x}}_l + \beta_{01} (x_l - \hat{x}_l) \\ \dot{\hat{\theta}}_m = \dot{\hat{\theta}}_m + \beta_{02} (x_l - \hat{x}_l) \\ \ddot{\hat{x}}_l = -\frac{K_T}{i^2 M} \hat{x}_l + \frac{K_T}{iM} \hat{\theta}_m - \frac{B}{M} \dot{\hat{x}}_l + \frac{1}{M} \hat{F}_d + \beta_{03} (x_l - \hat{x}_l) \\ \ddot{\hat{\theta}}_m = \frac{K_T}{i J_m} \hat{x}_l - \frac{K_T}{J_m} \hat{\theta}_m + \frac{B_m}{J_m} \dot{\hat{\theta}}_m + \frac{k_i}{J_m} u + \beta_{04} (x_l - \hat{x}_l) \\ \dot{\hat{F}}_d = \beta_{05} (x_l - \hat{x}_l) \end{cases} \tag{20}$$

Where $\hat{x}_l, \hat{\theta}_m, \hat{F}_d$ are the observed values of x_l, θ_m, F_d, $\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{05}$ are observer gains, and ensure the observer is stable.

Then

$$u=\frac{J_m}{k_i}\left\{\begin{array}{l}\varepsilon \mathrm{sgn}s+ks+\ddot{\theta}_m^*+c_1\left(\dot{x}_l^*-\dot{\hat{x}}_l\right)+c_2\left(\dot{\theta}_m^*-\dot{\hat{\theta}}_m\right)\\ +c_3\left[\ddot{x}_l^*-\left(\frac{K_T}{iM}\hat{\theta}_m-\frac{K_T}{i^2M}\hat{x}_l-\frac{B}{M}\dot{\hat{x}}_l+\frac{1}{M}\hat{F}_d\right)\right]\end{array}\right\}$$
$$+\frac{K_T}{k_i}\hat{\theta}_m+\frac{B_m}{k_i}\dot{\hat{\theta}}_m-\frac{K_T}{ik_i}\hat{x}_l \tag{21}$$

The control block diagram of sliding mode control based on extended state observer is shown in Fig.2.

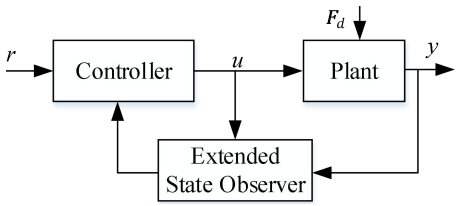


Fig.2 Block diagram of sliding mode control based on extended state observer

4. SIMULATION AND ANALYSIS

Simulink simulation analysis is performed to verify the correctness of the design. The parameters of the controlled plant in the system are shown in Table 1. The controller parameters and signal input of the system are set in Table 2. The given input signal are 0.02t, the disturbance F_d is a constant value disturbance. Let the root of the observer's characteristic equation be -30, -40, -50, -60, -70, then the observer gains can be obtained.

Table 1. Model parameters of controlled plant.

Parameter	Value	Unit	Meaning
K_T	1000	*N*m/rad*	Torsional stiffness of ball screw
B	1.6	*N*s/m*	Damping coefficient converted to workbench
B_m	0	*N*s/m*	Damping coefficient of motor
J_m	5	*Kg*m²*	Moment of inertia of motor shaft
M	100	*kg*	Mass converted to workbench
k_i	1	-	the motor torque constant
i	0.002	*m/rad*	Screw transmission ratio

Table 2. Controller parameters and input signals.

Parameter	Value	Remarks
λ	20	Then c_1=8000, c_2=1200, c_3=60
ε	5	Coefficient of exponential reaching law
k	100	Coefficient of exponential reaching law
r	0.02t	Given input
F_d	1000	t<10s，F_d=0； t≥10s，F_d=1000N

Firstly, the observation effect of the extended state observer is tested. When t=10s, a disturbance force of 1000N is applied to the plant. The observed disturbance force and actual force are shown in the Fig.3. It can be seen from the diagram that the extended state observer has good observation effect.

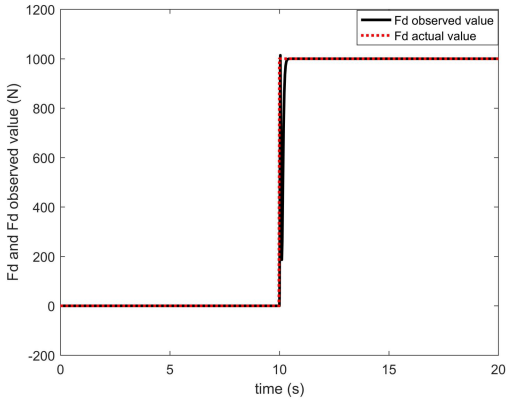


Fig.3 Comparison of disturbance observation value and actual value based on extended state observer.

Then, assuming that the value of disturbance is known, the F_d is 1000N, the sliding mode control based on the boundary of disturbance and sliding mode control based on extended state observer are respectively adopted for simulation analysis. For the given position of 0.02t, the results are shown in the fig.4. It can be seen from the figure that the sliding mode control based on the boundary of disturbance can't estimate the real-time situation of the disturbance, so it can only consider that the disturbance force of 1000N acts on the plant in the whole process, which results in the greater tracking error of the first half part and the phenomenon of jitter. The sliding mode control based on the extended state observer can observe the disturbance in real time and carry out compensation control, the control effect is better and the jitter is not obvious.

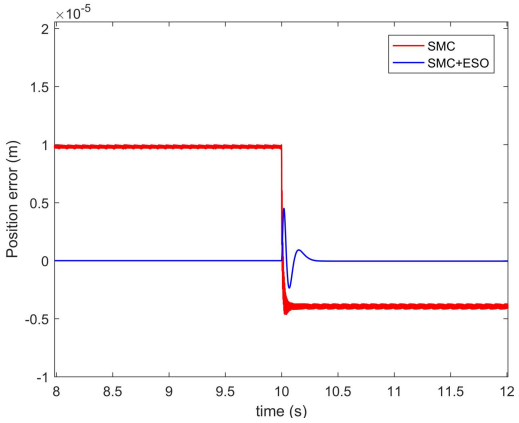


Fig.4 Position errors of sliding mode control based on the boundary of disturbance (SMC) and extended state observer (SMC+ESO)

.

For unknown disturbance, using sliding mode control rate based on the boundary of disturbance requires guessing the bounds of F_d. Fig.5 and Fig.6 show the position

tracking error of sliding mode control based on the speculative boundary of disturbance. In Fig.5 the speculative boundary of F_d is 500, it is less than the actual value of F_d. In Fig.6 the speculative boundary of F_d is 2000, it is greater than the actual value of F_d. It can be seen from Fig. 5 and Fig. 6 that when the disturbance boundary set is inconsistent with the actual value, the position tracking error is also different. Therefore, the sliding mode control based on disturbance boundary can't eliminate the influence of disturbance well in the case of unknown disturbance.

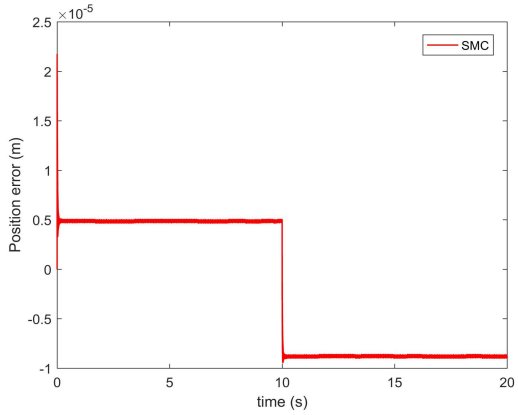


Fig.5 Position error of sliding mode control based on the boundary of disturbance (Supposing F_d is 500N).

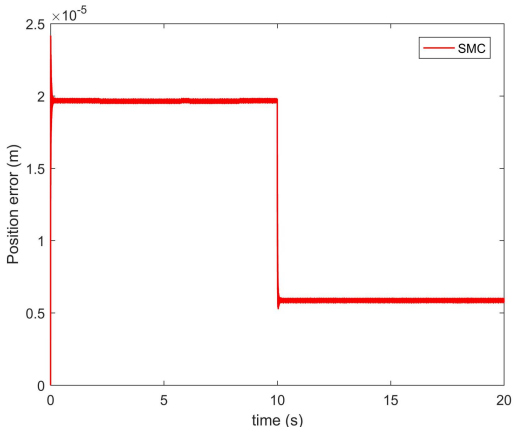


Fig.6 Position error of sliding mode control based on the boundary of disturbance (Supposing F_d is 2000N).

5. CONCLUSION

Aiming at the position tracking control of machine tool moving axis servo system, this paper establishes system model with considering the elastic deformation of ball screw, and obtains the fourth order mathematical model of the system. Then, designs a sliding mode controller, in order to overcome the problem of unknown system disturbance, an extended state observer is designed to observe the internal state and disturbance of the system. Finally, two kinds of sliding mode control methods are simulated, and the comparison shows that the sliding mode control based on the extended state observer can observe the disturbance and suppress disturbance well.

References

[1] Z. M. Zhou, B. Zhang and D. P. Mao, Robust Sliding Mode Control of PMSM Based on Rapid Nonlinear Tracking Differentiator and Disturbance Observer, Sensors, Vol. 18, No. 4, 2018, pp. 1031-1049.

[2] Y. Yan, J. Yang and Z. Sun, et al. Robust Speed Regulation for PMSM Servo System with Multiple Sources of Disturbances via An Augmented Disturbance Observer, IEEE/ASME Transactions on Mechatronics, Vol. 23, No. 2, 2018, pp. 769-780.

[3] S. Hebertt, L. Jesus and G. Carlos, et al. On the Control of the Permanent Magnet Synchronous Motor: An Active Disturbance Rejection Control Approach, IEEE Transactions on Control Systems Technology, Vol. 22, No. 5, 2014, pp. 2056-2063.

[4] J. Hu, Y. Qiu and H. Lu. Adaptive robust nonlinear feedback control of chaos in PMSM system with modeling uncertainty, Applied Mathematical Modelling, Vol. 40, No. 19, 2016, pp. 8265-8275.

[5] S. Kim, J. Lee and K. Lee. Robust speed control algorithm with disturbance observer for uncertain PMSM, International Journal of Electronics: Theoretical & Experimental, Vol. 105, No. 8, 2018, pp. 1300-1318.

[6] X. G. Zhang, L. Sun and K. Zhao, et al. Nonlinear Speed Control for PMSM System Using Sliding-Mode Control and Disturbance Compensation Techniques, IEEE Transactions on Power Electronics, Vol. 28, No. 3, 2013, pp. 1358-1365.

[7] B. Zhang, H. Y. Li. A PMSM Sliding Mode Control System Based on Model Reference Adaptive Control, International Power Electronics & Motion Control Conference. IEEE, 2002, pp. 336-341.

[8] M. S. Wang, T. M. Tsai. Sliding Mode and Neural Network Control of Sensorless PMSM Controlled System for Power Consumption and Performance Improvement, Energies, Vol. 10, No. 11, 2017, pp. 1780-1794.

[9] S. H. Chang, P. Y. Chen and Y. H. Ting, et al. Robust current control-based sliding mode control with simple uncertainties estimation in permanent magnet synchronous motor drive systems, IET Electric Power Applications, Vol. 4, No. 6, 2010, pp. 441-450.

[10] H. Liu, S. Li. Speed Control for PMSM Servo System Using Predictive Functional Control and Extended State Observer, IEEE Transactions on Industrial Electronics, Vol. 59, No. 2, 2012, pp. 1171-1183.

Output Stabilization for Wind Power System Using Equivalent-Input-Disturbance Approach

Junyang Shen*, Jinhua She*
* Graduate School of Engineering, Tokyo University of Technology
Katakuramachi 1404-1, Hachioji, Tokyo 192-0982, Japan

Abstract

Large fluctuations in wind speed cause big changes in the power output of a wind generator and may even destroy the stability of the system. This paper applies the equivalent-input-disturbance (EID) approach to suppress the changes in the power output so as to improve the quality of the power supply. An output-stabilization control system is constructed by combining an EID estimator and a PI controller. The system takes wind fluctuations as an EID and uses the EID estimator to estimate it. Then, the system incorporates the estimate into the PI control law to yield high-quality power output.

Keywords: output fluctuation, pitch-angle control, equivalent input disturbance (EID), wind-power generator, PID control.

1. INTRODUCTION

With the depletion of fossil fuels and the attention to environmental issues, countries around the world have begun to pay attention to renewable clean energy. As explained in a statistical review of world energy provided by BP plc [1], Renewable energy (including biofuels) posted a record increase in consumption in energy terms (3.2 EJ). This was also the largest increment for any source of energy in 2019. By energy source, wind generation provided the largest contribution to growth (1.4 EJ) followed closely by solar (1.2 EJ). Wind installed capacity was 620 GW in 2019 a net increase of 58 GW. Another report given by Institute for Sustainable Energy Policies showed that the annual share of renewable energy to total power generation (including self-consumption) in Japan in 2019 was estimated to have increased to 18.5% from 17.4% in the previous year and wind energy accounted for 0.76% [2].

In vast flatlands in Europe, the United States, and China, there is little turbulence in wind speed. However, 73% of the terrain in Japan is a mountain country. Thus, wind farms in Japan suffer large wind speed turbulence and there are large fluctuations in power outputs if proper control systems are not used.

In this paper, we explain an output-stabilization control system that combines the equivalent-input-disturbance (EID) approach with a PI controller to suppress the influences in wind speed.

2. WIND POWER SYSTEM

2.1 Configuration of wind power generation system

The wind power generation system (Fig. 1) contains a pitch-angle control system, a hydraulic servo system, and a windmill and generator [3]. A PI controller is used in the pitch-angle control system.

The working principle of the system is as follows. First, find the tracking error, e (t), between the reference input, P_{g0} (t), and the generator output, P_g (t). Next, calculate the pitch-angle command value, β_{CMD} (t), from the pitch-angle control system and send it to the hydraulic servo system. Then, the servo system sends the pitch angle, β (t), to the windmill to adjust the power output of an induction generator, P_g (t).

2.2 Pitch angle control system

The condition for the control of the pitch angle is divided into the four ranges [4]:

(1) $V_w < 5$ m/s : The pitch angle is set to be 90° to stop rotating.

(2) $5\text{ m/s} \le V_w < 13\text{ m/s}$: Since wind power is weak, the pitch angle is set to 10° to receive the wind force as much as possible.

(3) $13\text{ m/s} \le V_w < 24\text{ m/s}$: The pitch angle is controlled maintain stable output. The relationship between a change in the ouput, ΔP (t), and that in the pitch angle, $\Delta\beta$ (t), is given by

$$G(\beta(t)) = \frac{\Delta\beta(t)}{\Delta P(t)} = \frac{1}{Ab_1 + Ab_2 V_w^{\,2}} \qquad (1)$$

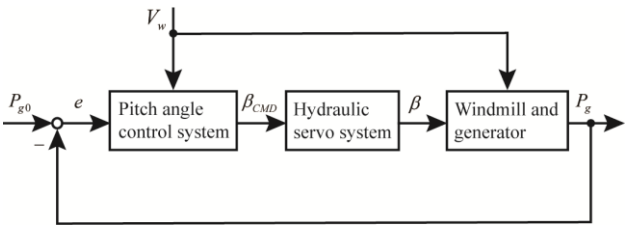


Fig. 1. Wind power generation system.

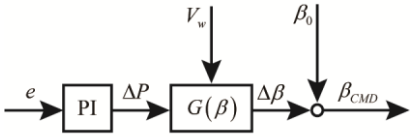


Fig. 2. Pitch-angle control system.

where

$$\begin{cases} Ab_1 = a_{12} + 2a_{13}\beta + 3a_{14}\beta^2 , \\ Ab_2 = a_{22} + 2a_{23}\beta + 3a_{24}\beta^2, \end{cases} \quad (2)$$

where a_{12}, a_{13}, a_{14}, a_{22}, a_{23}, and a_{24} are constants.

(4) $V_w \geq 24$ m/s: In order to protect the power system, the pitch angle is set to be 90° and the rotation is stopped.

Fig. 2 shows the pitch-angle control system.

2.3 Hydraulic servo system

The system is nonlinear. A first-order linear approximate model is [5] [6]

$$Q(s) = \frac{1}{T_c s + 1}, \quad (3)$$

where T_c is the time constant of the system.

A limiter is used to restrict the pitch-angle command, β_{CMD} (t), in the range [10°, 90°] (Fig. 3).

2.4 Windmill and generator

Fig. 4 shows the windmill and generator. The output of the windmill, P_w (t), is

$$P_w(t) = \frac{C_p(\lambda,\beta) V_w{}^3 \rho A}{2}, \quad (4)$$

where ρ is the density of air, A is the swept area of the rotor, and V_w is the wind speed. The power efficiency of the wind turbine $C_p(\lambda,\beta)$ is a function of the blade pitch angle, θ, and the tip-speed ratio, λ. The relationship between them is usually obtained through experiments. It can be approximated by [6]

$$C_p(\lambda(t),\beta(t)) = c_1(\beta)\lambda^2 + c_2(\beta)\lambda^3 + c_3(\beta)\lambda^4, \quad (5)$$

$$\begin{cases} c_1(\beta) = c_{10} + c_{11}\beta + c_{12}\beta^2 + c_{13}\beta^3 + c_{14}\beta^4 , \\ c_2(\beta) = c_{20} + c_{21}\beta + c_{22}\beta^2 + c_{23}\beta^3 + c_{24}\beta^4, \\ c_3(\beta) = c_{30} + c_{31}\beta + c_{32}\beta^2 + c_{33}\beta^3 + c_{34}\beta^4, \end{cases} \quad (6)$$

where $c_{10} - c_{34}$ are constants that represent the characteristics of the system. And

$$\lambda(t) = \frac{R\omega}{V_w}, \quad (7)$$

where $\omega(t)$ is the rotational speed of the rotor and R is the radius of the windmill rotor.

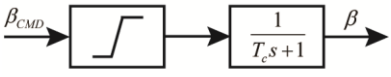


Fig. 3. Hydraulic servo system.

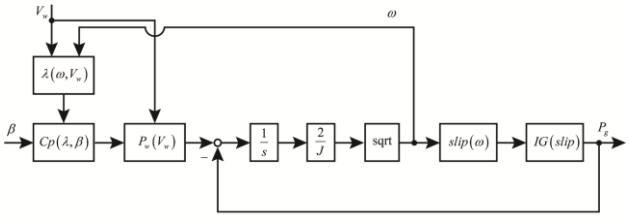


Fig. 4. Windmill and generator.

In this study, a squirrel-cage induction generator is used as a wind power generator. Since the wind power generator is connected to the wind turbine with large inertia, the electrical transient phenomenon can be ignored and the output of the generator $P_g(t)$ is given by

$$P_g(t) = \frac{-3V^2 S(S+1)R_2}{(R_2 - SR_1)^2 + S^2(X_1 + X_2)^2}, \quad (8)$$

where X_1 and X_2, and R_1 and R_2 are the reactance and resistance of the stator and rotor of the motor, respectively; V is the phase voltage; and $S(t)$ is the slip that is given by

$$S(t) = \frac{\omega_0 - \omega}{\omega_0}, \quad (9)$$

where ω_0 is the synchronous speed of windmill.

The rotating speed of the windmill rotor, $\omega(t)$, is given by

$$\omega^2 = \int_0^t \frac{2}{J}(P_w - P_g)dt, (t > 0), \quad (10)$$

where J is the moment of inertia of the windmill.

2.5 Plant modelling

Since the model contains nonlinearities, we use linear approximation to derive a linear model.

A linear relationship between $P_w(t)$ and $\beta(t)$ is

$$\begin{aligned} \Delta P_w &= \frac{dP_w}{d\beta}\Delta\beta \\ &= \frac{\rho A V_{w0}{}^3}{2}\left[c_1'(\beta_0)\lambda_0{}^2 + c_2'(\beta_0)\lambda_0{}^3 + c_3'(\beta_0)\lambda_0{}^4\right]\Delta\beta \\ &= K_{pw}\Delta\beta \end{aligned} \quad (11)$$

A linear relationship between $P_w(t)$ and $\beta(t)$ is

$$\begin{aligned} \Delta P_g &= \frac{dP_g}{dS}\Delta S \\ &= -3V^2 R_2 \left\{ \frac{(2S_0 - 1)\left\{(R_2 - SR_1)^2 + S_0{}^2(X_1 + X_2)^2\right\}}{\left\{(R_2 - SR_1)^2 + S_0{}^2(X_1 + X_2)^2\right\}^2} \right. \\ &\left. + \frac{(2S_0{}^2 + 2S_0)\left\{R_1(R_2 - SR_1) - S_0(X_1 + X_2)^2\right\}}{\left\{(R_2 - SR_1)^2 + S_0{}^2(X_1 + X_2)^2\right\}^2} \right\}\Delta S \\ &= K_{pg}\Delta S \end{aligned} \quad (12)$$

And a linear relationship between $S(t)$ and $\omega(t)$ is

$$\Delta S = \frac{dS}{d\omega}\Delta\omega = \frac{d}{d\omega}\left(1 - \frac{\omega}{\omega_0}\right)\Delta\omega = -\frac{1}{\omega_0}\Delta\omega \quad (13)$$

Defining

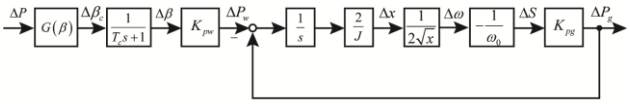


Fig. 5. Linear model of control system.

$$x = \omega^2 \tag{14}$$

yields

$$\frac{d\omega}{dx} = \frac{d}{dx}\sqrt{x} = \frac{1}{2\sqrt{x}}. \tag{15}$$

Thus,

$$\Delta\omega = \frac{d\omega}{dx}\Delta x = \frac{1}{2\sqrt{x}}\Delta x. \tag{16}$$

Summarizing the above explanation gives the linear model of the control system (Fig. 5).

3. CONTROLLER DESIGN

We employ the EID approach [7] in this study to suppress the fluctuations in the output. The control system is constructed by combining the EID compensator with a PI controller (Fig. 6). A state observer estimates the state of the plant, and an EID estimator uses the state estimate to produce an EID estimate. We used the pole-placement method to design the PI controller and the state observer gain, L.

The state observer is constructed as

$$\dot{\hat{x}} = A\hat{x}(t) + Bu_f(t) + LC\left[x(t) - \hat{x}(t)\right]. \tag{17}$$

As explained in [12], an estimate of the EID is given by

$$\hat{d}(t) = B^{+}LC\left[x(t) - \hat{x}(t)\right] + u_f(t) - u(t), \tag{18}$$

where

$$B^{+} \coloneqq \left(B^T B\right)^{-1} B^T. \tag{19}$$

$\hat{d}(t)$ is filtered by a low-pass filter $F(s)$, which selects the angular-frequency band for disturbance estimation and rejection. Thus, the filtered disturbance estimate $\tilde{d}(t)$ is given by

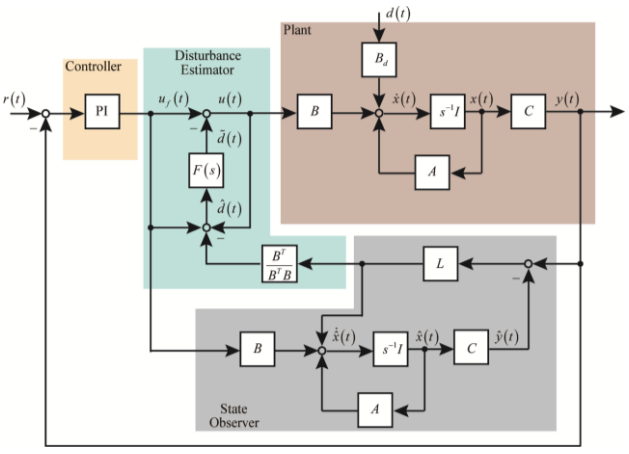


Fig. 6. Configuration of output-stabilization control system.

Table I, WIND TURBINE PARAMETERS

R [m]	J [kg·m^2]	ρ [kg/m^3]	P_g [kW]
14	62993	1.225	320
X_1 [Ω]	X_2 [Ω]	V [V]	p
0.00397	0.0534	400√3	2
q	f [Hz]	R_1 [Ω]	R_2 [Ω]
34.3	50	0.0376	0.00443

And $F(s)$ is chosen to be a first-order one for simplicity

$$\tilde{D}(s) = F(s)\hat{D}(s). \tag{20}$$

$$F(s) = \frac{1}{Ts+1}, \tag{21}$$

where T is the time constant of the filter.

4. SIMULATIONS AND RESULT ANALSIS

The physical parameters of the windmill, wind turbine, and induction motor are shown in Table I and according to Subsection 2.5, we obtain

$$\begin{aligned} &G(\beta) = -0.0365,\ \mathrm{K}_{pw} = -25613,\ \frac{2}{J} = 0.00003175,\\ &\frac{1}{2\sqrt{x}} = 0.1082,\ -\frac{1}{\omega_0} = -0.2163,\ \mathrm{K}_{pg} = 30631000, \end{aligned} \tag{22}$$

and the transfer function of the plant be

$$P(s) = \frac{53197}{s^2 + 25.26s + 56.9}. \tag{23}$$

The state-space expression of the model is

$$\begin{cases} \dot{x}(t) = \begin{bmatrix} -25.26 & -56.9 \\ 1 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(t) \\ y(t) = \begin{bmatrix} 0 & 53197 \end{bmatrix} x(t) \end{cases}. \tag{24}$$

For the PI controller

$$C(s) = K_p + \frac{K_i}{s}, \tag{25}$$

the gains were chosen to be

$$K_p = 0.011,\ K_i = 0.003. \tag{26}$$

Selecting the poles of the state observer to be $-100 \pm 10j$ yields the gain of the state observer

$$L = \begin{bmatrix} 0.1058 \\ 0.0033 \end{bmatrix}. \tag{27}$$

The time constant of the low-pass filter, $F(s)$, is selected to be $T = 0.01$ s.

4.3. Results and analysis

We used the nonlinear model in simulations to verify the effectiveness of our method and set the initial condition to be

$$P_{g0} = 320 \text{ kW},\ V_w = 18 \text{ m/s},\ \beta_0 = 22^{\circ}, \tag{28}$$

The wind speed is shown in Fig. 7.

We tested control performance for the control system

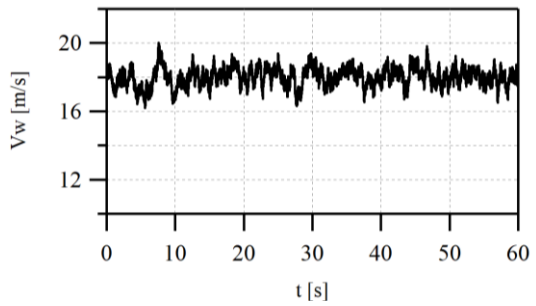


Fig. 7. Wind speed.

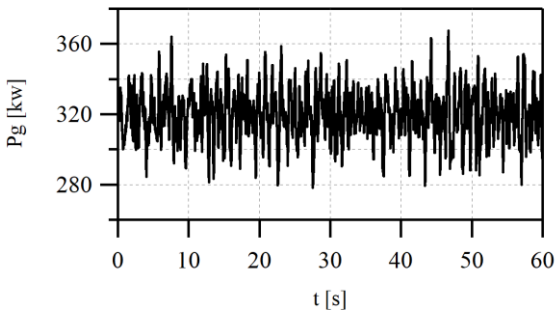


Fig. 8. Output for PI control.

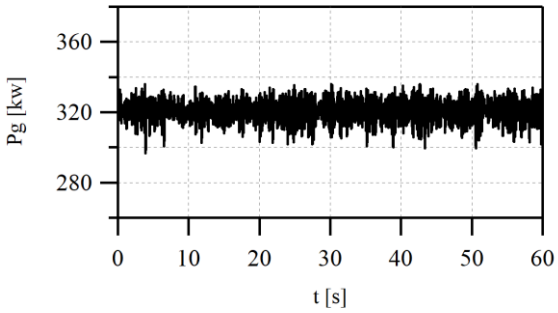


Fig. 9. Output for PI + EID control.

with and without the EID compensator. Fig. 8 shows the output for the system uses only PI control, and Fig. 9 shows the output for the system using the combination of the PI and EID control.

It is clear that the fluctuations in P_g is much smaller when the EID compensator is incorporated. More specifically, P_g varied between 280 kW and 365 kW for the PI control system without incorporating the EID compensator, and it varied between 310 kW and 340 kW for the PI control system that incorporated the EID compensator. The fluctuations in P_g was reduced by about 56%.

From Figs. 10 and 11, we can see that the pitch actuator works more frequently for PI + EID control. The pitch angle was adjusted more often to suppress the wind turbulence, the range is larger, and the response is much faster for PI + EID control than for PI control.

On the other hand, frequently changing the pitch angle may shorten the lifetime of mechanical parts and increase the risk of failure. So, we need to consider a trade-off

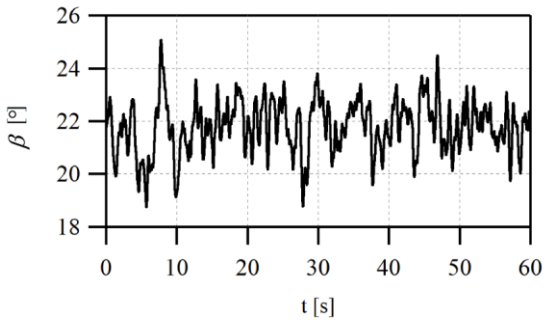


Fig. 10. Pitch angle for PI control.

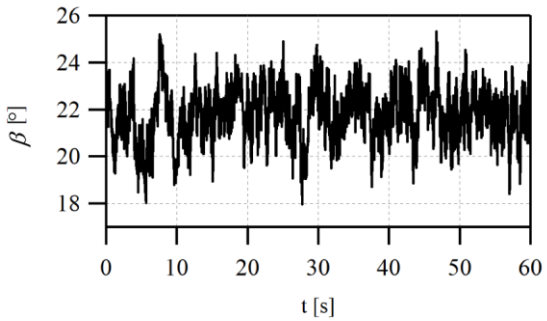


Fig. 11. Pitch angle for PI + EID control.

between the control performance and the lifetime of the system in the design of the pitch control system.

5. CONCLUSION

This paper presented a control system for a wind power generator that combines the PI control the EID approach. Simulations show that the output power variation is much smaller for our method than for conventional PI control. Our method effectively suppressed the influences in the wind speed. It has advantages:

(1) The control system does not change the hardware structure of the conventional PI control system.

(2) Parameters of the PI controller and the EID compensator are designed independently.

References

[1] Statistical Review of World Energy 2020 (June 2020), available: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2020-full-report.pdf

[2] Share of renewable energy electricity in Japan, 2019, available: https://www.isep.or.jp/en/879/

[3] T. Senjyu, R. Sakamoto, N. Urasaki, H. Higa, K. Uezato, and T. Funabashi, "Output power control of wind turbine generator by pitch angle control using minimum variance control," Electrical Engineering in Japan, Vol. 154, No. 2, 2006, pp. 10-18.

[4] T. Matsuzaka and K. Tuchiya, "Study on Stabilization of a Wind Generator Power Fluctuation," IEEJ Transactions on Power and Energy, Vol. 117, No. 5,

1997, pp. 625-633.

[5] N. Kodama, T. Matsuzaka and S. Yamada, "Modeling and analysis of the NEDO 500-kW wind generator," Electrical Engineering in Japan, Vol. 135, No. 3, 2001, pp. 37-47.

[6] T. Matsuzaka and K. Tuchiya, "Study on Stabilization of a Wind Generator Power Fluctuation," IEEJ Transactions on Power and Energy, Vol. 117, No. 5, 1997, pp. 625-633.

[7] J. H. She, M. X. Fang, Y. Ohyama, H. Hashimoto, and M. Wu, "Improving disturbance-rejection performance based on an equivalent-input-disturbance approach," IEEE Transactions on Industrial Electronics, Vol. 55, No. 1, 2008, pp. 380–389.

[8] J. She, A. Zhang, X. Lai, and M. Wu, "Global stabilization of 2-DOF underactuated mechanical systems – an equivalent-input-disturbance approach –," Nonlinear Dynamics, Vol. 69, 2012, pp. 495–509.

Asymptotic Stabilization for a Class of Linear Fractional-Order Composite Systems

Zhe ZHANG*, Toshimitsu USHIO**, Jing ZHANG*, Feng LIU***, Can DING*,
*College of Electrical and Information Engineering, Hunan University, 410082, China
E-mail: qyzz@hnu.edu.cn
**Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka, Japan
E-mail: ushio@sys.es.osaka-u.ac.jp
***School of Automation, China University of Geosciences, Wuhan , 430074, China
E-mail:fliu@cug.edu.cn

Abstract

In this paper, we designs a decentralized control method consisting of a series of local state feedback controllers for a class of linear fractional composite systems. In addition, the corresponding asymptotic stabilization criterion is derived. First, we design local state feedback controllers for each fractional subsystem of a class of linear fractional composite systems. Then, according to the method of vector Lyapunov function, we combine the above series of local state feedback controllers into a decentralized controller. Through this controller, we propose the asymptotic stabilization criterion for a class of linear fractional composite systems. Finally, the numerical simulation of a class of linear fractional composite system verifies the correctness and effectiveness of the decentralized control method.

Keywords: Fractional-order Composite system, Control Theory, Linear System, Stabilization Control.

1. INTRODUCTION

In recent years, the research of fractional-order systems is mainly divided into two parts. The first part is to study the characteristics of the fractional-order system itself, including stability and bifurcation, etc. The other part is to control the unstable fractional-order system via some different control methods so as to make the fractional-order system reach the target state. On one side, many researchers have studied the stability analysis of particular fractional-order systems in recent years, such as neural network systems [1,2], gene regulation network systems [3-5], HIV systems [6-8], and Lorenz dynamical systems [9-11], etc. On the other hand, in recent years, many researchers have also studied how to apply appropriate controllers to the fractional-order system so that the fractional-order system can achieve asymptotic stability. In [12], Zhang et al. have controlled the fractional-order Newton-Leipnik system via linear feedback controllers.

But, for the composite system where each subsystem is modeled by an integer-order system, many decentralized control methods have been proposed. The introduction of an M-matrix was first proposed in [16]. Compared with negative definite matrices, the M-matrix has lower conservatism and less restriction. In [16], the M-matrix is mainly applied to the stability analysis and control of integer-order systems. However, this paper generalizes the M-matrix to the stabilization control of fractional-order systems for the first time. In addition, Fukuda and Ushio proposed decentralized event-triggered control of the integer-order composite systems in [13].

The rest of the paper is organized as follows. In Section 2, some definitions and lemmas that need to be used in this paper are reviewed. In Section 3, the main results in this paper are shown. In Section 4, numerical simulations are represented to illustrate the correctness and feasibility of the proposed decentralized control. Finally, Section 5 concludes the paper.

2. PREPARATION

Definition 1 [14] The Caputo fractional-order function with the fractional-order operator parameter α is defined by

$$ {}^{C}D^{\alpha} f(t) = \frac{1}{\Gamma(n-\alpha)} \int_0^t (t-\tau)^{n-\alpha-1} f^{(n)}(\tau) d\tau, \tag{2.1} $$

where $f(t)$ is an arbitrary integrable function. As a special condition, if $\alpha \in (0,1)$ and $n=1$, then we obtain the Caputo fractional-order function in its most common form as follows.

$$ {}^{C}D^{\alpha} f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} f^{(1)}(\tau) d\tau. \tag{2.2} $$

We consider the following fractional-order system.

$$ {}^{C}D^{\alpha} x(t) = f(x(t)), \tag{2.3} $$

where the $x(t) \in \mathbb{R}^n$ is the state of the fractional-order

system (2.3) and $f:\mathbb{R}^n \to \mathbb{R}^n$ satisfies the following two conditions:
(1) The origin is an equilibrium point of (2.3), that is, $f(0)=0$;
(2) The function $f(\cdot)$ is locally Lipschiz continuous in a neighborhood of the origin.

Lemma 1[15] Let $x=0$ be an equilibrium point for (2.3) and a Lyapunov function $V:\mathbb{R}^n \to \mathbb{R}$ of (2.3) be a convex and continuously differentiable function such that $V(0)=0$.

Then, the following inequality holds for all $t \geq 0$.

$$ {}^{C}D^{\alpha}V(x) \leq \frac{\partial V(x)}{\partial x}\,{}^{C}D^{\alpha}x, \tag{2.4}$$

where the fractional-order operator $\alpha \in (0,1)$.

Definition 2[15] The solution $x(t)$ of (2.3) is called stable if, for any $\varepsilon>0$, there exists $\delta=\delta(\varepsilon)>0$ such that, for every $\|x_0\|<\delta$ with $x_0=x(t_0)$, where t_0 represents the initial time, we have

$$\|x(t)\|<\varepsilon, \text{ for any } t \geq t_0. \tag{2.5}$$

The solution of (2.3) is called asymptotically stable if it is stable and there exists $\hat{\delta}>0$ such that $\lim_{t\to\infty} x(t)=0$ whenever $\|x_0\|<\hat{\delta}$.

Lemma 2[14] If the convex and continuously differentiable Lyapunov function $V:\mathbb{R}^n \to \mathbb{R}$ of (2.3) satisfies

$$\vartheta_1(\|x\|) \leq V(x) \leq \vartheta_2(\|x\|), \tag{2.6}$$

$$ {}^{C}D^{\alpha}V(x) \leq -\vartheta_3(\|x\|), \tag{2.7}$$

$\forall t \geq 0$, where $\vartheta_i (i=1,2,3)$ are class-κ functions and the fractional-order operator is $\alpha \in (0,1)$, then the equilibrium point $x=0$ of the fractional-order system (2.3) is asymptotically stable.

Definition 3[16] A real $n\times n$ matrix $W=[w_{ii}]$ is an M-matrix if the element $w_{ii} \leq 0$, for $i \neq j$, and if its all principal minor determinants are positive.

Lemma 3[16] If $W=[w_{ii}]$ is an M-matrix , there exists a diagonal matrix $P=diag\{p_1,p_2,\ldots p_N\}$ with elements $p_i>0$, $i \in N$, such that the matrix

$$C=W^{T}P+PW, \tag{2.8}$$

3. MAIN RESULT

In this section, we will propose a design method of the decentralied control for a class of linear fractional-order composite systems.

The Fig. 1 illustrates a decentralized control of a composite system with three fractional-order subsystems.

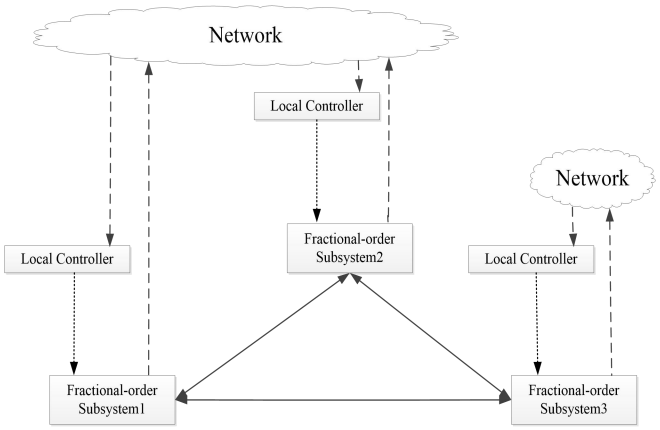


Fig. 1 Illustration of the decentralized control of the fractional-order composite system in the case of three fractional-order subsystems.

There exist communications among several subsystems. The local controller of subsystem 1 can receive the states of subsystem 2 by the network and determines the local input to subsystem 1 using both states of subsystem 1 and 2. While the local controller of subsystem 3 determines the local input using only its state since it does not receive states of the other subsystems.

We assume that there are no communication delays among the local controllers.

Then, we consider the following linear fractional-order composite system where each subsystem is described by a fractional-order linear differential equation.

$$ {}^{C}D^{\alpha}x_i(t)=A_{ii}x_i+B_iu_i+\sum_{j\in D_i}A_{ij}x_j,\ i,j\in\{1,2,...,N\}, \tag{3.1}$$

where $x_i \in \mathbb{R}^{n_i}$ and $u_i \in \mathbb{R}^{m_i}$ are the state and the local input of the fractional-order composite system i , and $A_{ii} \in \mathbb{R}^{n_i\times n_j}$, $B_i \in \mathbb{R}^{n_i\times m_i}$ are constant matrices of appropriate dimensions. Then, we design the following local state feedback controller for each fractional-order subsystem

$$u_i(t)=-K_{ii}x_i-\sum_{j\in Z_i}f_{ij}K_{ij}x_j, \qquad i\in \mathrm{N}, \tag{3.2}$$

where the gain matrices $K_{ii} \in \mathbb{R}^{n_i\times n_j}$ are constant . $F=(f_{ii})$ is a binary $N\times N$ matrix which distributes over the constant gain matrices K_{ii} of the fractional-order linear composite systems (3.1). Then, the fractional-order linear composite system (3.1) can be rewritten as follows

$$ {}^{C}D^{\alpha}x_i(t)=\hat{A}_ix_i+\sum_{j\in D_i}A_{ij}x_j-B_i\sum_{j\in Z_i}f_{ij}K_{ij}x_j, \quad i,j\in\{1,2,...,N\}, \tag{3.3}$$

with

$$\hat{A}_i=A_{ii}-B_iK_{ii}, \tag{3.4}$$

we assume that (3.4) is Hurwitz.

Assumption 1 We assume a positive definite Lyapunov function $V_i(x_i(t))$ of each subsystem, a series of class-κ functions $\beta_{ki}(x), k=1,2,3$, and constant $\mu_i > 0$ such that

$$\begin{cases} \beta_{1i}\left(\|x_i\|\right) \le V_i(x_i(t)) \le \beta_{2i}\left(\|x_i\|\right), \\ \dfrac{\partial V_i(x_i)}{\partial x_i}\widehat{A}_i \le -\mu_i \beta_{3i}\left(\|x_i\|\right), \\ i = 1,2,...N, \end{cases} \tag{3.5}$$

Assumption 2 There exist a non-negative real numbers $\kappa_{ij}(i,j=1,2,...,N)$ with $\kappa_{ij} \ge 0$ for $i \ne j$ such that

$$\begin{aligned} &\left\| \frac{\partial V_i(x_i)}{\partial x_i}\left(\sum_{j\in D_i} A_{ij}x_j - B_i \sum_{j\in Z_i} f_{ij}K_{ij}x_j \right) \right\| \\ &\le \sqrt{\beta_{3i}(\|x_i\|)} \sum_{j\in D_i \cup Z_i} \kappa_{ij} \sqrt{\beta_{3j}(\|x_j\|)}, \end{aligned} \tag{3.6}$$

Theorem 1 Assume that the fractional-order composite system (3.1) satisfies Assumption 1 and Assumption 2, and the following matrix W is an M-matrix , then the fractional-order composite system (3.1) is asymptotic stable.

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1j} \\ w_{21} & w_{22} & \cdots & w_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ w_{i1} & w_{i2} & \cdots & w_{ij} \end{bmatrix}_{i\times j}, \qquad w_{ij} = \begin{cases} \mu_i - \kappa_{ii}, & \text{if } i = j, \\ -\kappa_{ij}, & \text{otherwise}, \end{cases} \tag{3.7}$$

Proof. We choose the following vector Lyapunov function which is convex and continuously differentiable of fractional-order composite system (3.1).

$$V(x(t)) = \sum_{i=1}^{N} p_i V_i(x_i(t)), \tag{3.8}$$

then, we perform a fractional derivation of (3.8), and according to Lemma 1 we have

$$\begin{aligned} &{}^{C}D^{\alpha}V(x(t)) \\ &\le \sum_{i=1}^{N} p_i \frac{\partial V_i(x_i)}{\partial x_i} {}^{C}D^{\alpha}x_i(t) \\ &= \sum_{i=1}^{N} p_i \frac{\partial V_i(x_i)}{\partial x_i}\left(A_{ii}x_i + B_i u_i + \sum_{j\in D_i} A_{ij}x_j \right) \\ &= \sum_{i=1}^{N} p_i \frac{\partial V_i(x_i)}{\partial x_i}\left(\widehat{A}_{ii}x_i + \sum_{j\in D_i} A_{ij}x_j - B_i \sum_{j\in Z_i} f_{ij}K_{ij}x_j \right) \\ &\le \sum_{i=1}^{N} p_i \left\{ -\mu_i \beta_{3i}\left(\|x_i(t)\|\right) + \beta_{3i}^{1/2}(\|x_i\|) \sum_{j\in D_i\cup Z_i} \kappa_{ij}\beta_{3j}^{1/2}(\|x_j\|) \right\} \\ &\le -\frac{1}{2}\beta_3^T\left(\|x(t)\|\right)\left(W^TP + PW\right)\beta_3\left(\|x(t)\|\right), \end{aligned} \tag{3.9}$$

where

$$\beta_3\left(\|x(t)\|\right) = \left[\beta_{31}^{1/2}\left(\|x_1\|\right), \beta_{32}^{1/2}\left(\|x_2\|\right), \ldots, \beta_{3N}^{1/2}\left(\|x_N\|\right) \right]^t. \tag{3.10}$$

according to Lemma 3, we can select $P = diag\{p_1,...,p_N\}$ such that $C = W^TP + PW$ is positive definite. Then, we have

$$\begin{aligned} {}^{C}D^{\alpha}V(x(t)) &\le -\frac{1}{2}\beta_3^T\left(\|x(t)\|\right) C \beta_3\left(\|x(t)\|\right) \\ &< 0, \qquad \text{if } x(t) \ne 0. \end{aligned} \tag{3.11}$$

Then, by Lemma 2, the equilibrium of the fractional-order composite system (3.1) is asymptotically stable, which completes the proof.

4. NUMERICAL SIMULATION

In order to verify the effectiveness of the controller we designed in the previous section, we present an example of numerical simulation in this section.

Example We consider the following two-dimensional linear fractional-order composite system.

$$\begin{cases} D^{\alpha}x(t) = Ax(t) + By(t), \\ D^{\alpha}y(t) = Cy(t) + Dx(t), \end{cases} \tag{4.1}$$

where

$$\begin{cases} A = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.8 \end{bmatrix}, B = \begin{bmatrix} 0.3 & 0 \\ 0 & 0.5 \end{bmatrix}, \\ C = \begin{bmatrix} 0.9 & 0 \\ 0 & 0.7 \end{bmatrix}, D = \begin{bmatrix} 0.2 & 0 \\ 0 & 0.4 \end{bmatrix}, \\ x = \left[x_1, x_2\right]^T, \\ y = \left[y_1, y_2\right]^T, \end{cases} \tag{4.2}$$

then, we give the time responses of the system (4.1) without input with different fractional-order operator and the initial states in Fig. 2

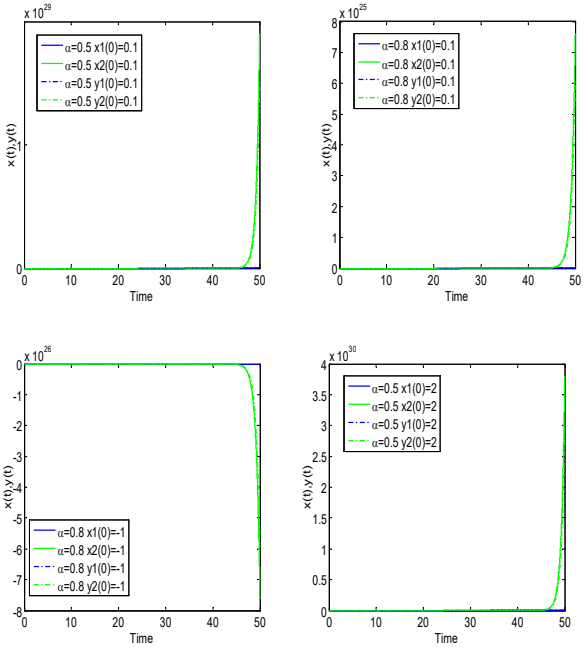


Fig. 2 The time responses of the system (4.1) without inputs

we can clearly see that the uncontrolled system is unstable. Then, we give the following inputs as the local control law with constant gain matrices that satisfy the Theorem 1.

$$\begin{cases} D^{\alpha}x(t) = Ax(t) + By(t) + u_1, \\ D^{\alpha}y(t) = Cy(t) + Dx(t) + u_2, \end{cases} \tag{4.3}$$

where

$$\begin{aligned} u_1 &= \begin{bmatrix} -1.4 & 0 \\ 0 & -1.6 \end{bmatrix} x + \begin{bmatrix} -0.2 & 0 \\ 0 & -0.3 \end{bmatrix} y, \\ u_2 &= \begin{bmatrix} -1.8 & 0 \\ 0 & -1.4 \end{bmatrix} x + \begin{bmatrix} -0.1 & 0 \\ 0 & -0.1 \end{bmatrix} y. \end{aligned} \tag{4.4}$$

Then, we have

$$\begin{cases} D^{\alpha}x(t) = A^{*}x(t) + B^{*}y(t), \\ D^{\alpha}y(t) = C^{*}y(t) + D^{*}x(t), \end{cases} \tag{4.5}$$

where

$$\begin{aligned} A^{*} &= \begin{bmatrix} -0.9 & 0 \\ 0 & -0.8 \end{bmatrix}, B^{*} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}, \\ C^{*} &= \begin{bmatrix} -0.9 & 0 \\ 0 & -0.7 \end{bmatrix}, D^{*} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.3 \end{bmatrix}. \end{aligned} \tag{4.6}$$

We select a vector Lyapunov function as follows.

$$\begin{aligned} V(x(t), y(t)) &= V(x(t)) + V(y(t)) \\ &= x^{T}(t)x(t) + y^{T}(t)y(t). \end{aligned} \tag{4.7}$$

Then, we have

$$\begin{cases} \dfrac{\partial V(x(t))}{\partial x_1(t)} x_1(t) \le -0.9\|x_1(t)\|^2, \\ \dfrac{\partial V(x(t))}{\partial x_2(t)} x_2(t) \le -0.8\|x_2(t)\|^2, \\ \dfrac{\partial V(y(t))}{\partial y_1(t)} y_1(t) \le -0.9\|y_1(t)\|^2, \\ \dfrac{\partial V(y(t))}{\partial y_2(t)} y_2(t) \le -0.7\|x_2(t)\|^2, \\ \dfrac{\partial V(x(t))}{\partial x_1(t)} y_1(t) \le 0.1\|x_1(t)\|^2\|y_1(t)\|^2, \\ \dfrac{\partial V(x(t))}{\partial x_2(t)} y_2(t) \le 0.2\|x_2(t)\|^2\|y_2(t)\|^2, \\ \dfrac{\partial V(y(t))}{\partial y_1(t)} x_1(t) \le 0.1\|y_1(t)\|^2\|x_1(t)\|^2, \\ \dfrac{\partial V(y(t))}{\partial y_2(t)} x_2(t) \le 0.3\|y_2(t)\|^2\|x_2(t)\|^2, \end{cases} \tag{4.8}$$

and we can get that

$$W = \begin{bmatrix} 0.8 & 0 & -0.1 & 0 \\ 0 & 0.6 & 0 & -0.2 \\ -0.1 & 0 & 0.8 & 0 \\ 0 & -0.3 & 0 & 0.4 \end{bmatrix}. \tag{4.9}$$

It is obvious that the matrix W is an M-matrix, which satisfies the conditions of Theorem 1. Then, we get the time response of the system (4.1) controlled by Eq.(4.4) with different α and initial states. It is clear from the time responses that the controlled system (4.1) is asymptotically stable when the control law satisfies the conditions of Theorem 1.

5. CONCLUSION

In this paper, we effectively extend the decentralized control to a class of linear fractional-order composite system. We showed that the sufficient conditions of asymptotic stabilization for the class of linear fractional-order composite system. Moreover, we design the rule of the control laws via a vector Lypunov function and an M-matrix. By simulation, we compare time responses of the fractional-order composite system with and without control in all cases, which shows the correctness and usefulness of the proposed decentralized control method.

We dealt with the case where the fractional-operator lies in (0,1), and the proposition of this paper has been well verified through the simulation. Its extension to all fractional-operators is future work.

References

[1] Messadi M, Mellit A. Control of chaos in an induction motor system with LMI predictive control and experimental circuit validation[J]. Chaos Solitons & Fractals, 2017, 97:51-58.

[2] Liu H, Li S, Li G, et al. Robust adaptive control for fractional-order financial chaotic systems with system uncertainties and external disturbances[J]. Information Technology & Control, 2017, 46(2).

[3] Almeida R. A Caputo fractional derivative of a function with respect to another function[J]. Communications in Nonlinear Science & Numerical Simulation, 2017, 44:460-481.

[4] Baleanu D, Wu G, Zeng S. Chaos analysis and asymptotic stability of generalized Caputo fractional differential equations[J]. Chaos Solitons & Fractals, 2017.

[5] Bao B, Wang N, Chen M, et al. Inductor-free simplified Chua ' s circuit only using two-op-amp-based realization[J]. Nonlinear Dynamics, 2016, 84(2):511-525.

[6] Rocha R, Ruthiramoorthy J, Kathamuthu T. Memristive oscillator based on Chua ' s circuit: stability analysis and hidden dynamics[J]. Nonlinear Dynamics, 2017, 88(4):2577-2587.

[7] Kengne J. On the Dynamics of Chua ' s oscillator with a smooth cubic nonlinearity: occurrence of multiple attractors[J]. Nonlinear Dynamics, 2016:1-13.

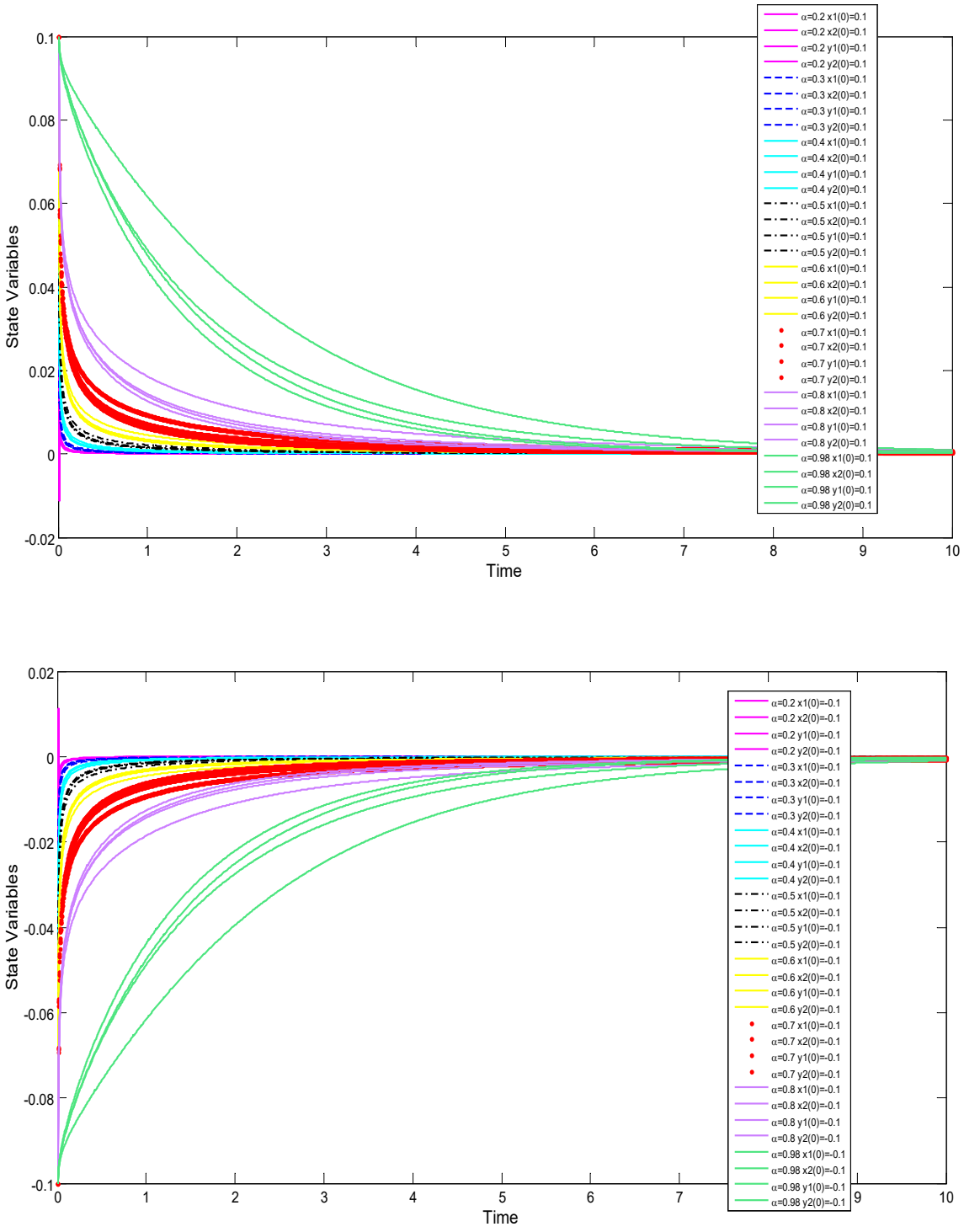


Fig. 3 The time responses of the system (4.1) with inputs.

[8] Zhang Z, Zhang J, Ai Z， A novel stability criterion of the time-lag fractional-order gene regulatory network system for stability analysis. Communications in Nonlinear Science & Numerical Simulation, 2019, 66, 96-108.

[9] Guner O. Exp-Function Method and Fractional Complex Transform for Space-Time Fractional KP-BBM Equation[J]. Communications in Theoretical Physics, 2017, 68(8):149-154.

[10] Chung W S, Zare S, Hassanabadi H. Investigation of Conformable Fractional Schr ¨odinger Equation in Presence of Killingbeck and Hyperbolic Potentials[J]. Communications in Theoretical Physics, 2017, 67(3):250-254.

[11] Pahnehkolaei S M A, Alfi A, Machado J A T. Dynamic stability analysis of fractional order leaky

integrator echo state neural networks. Communications in Nonlinear Science & Numerical Simulation, 2017, 47:328-337.
[12] Fernandez-Anaya G, Nava-Antonio G, Jamous-Galante J, et al. Lyapunov functions for a class of nonlinear systems using Caputo derivative. Communications in Nonlinear Science & Numerical Simulation, 2017, 43:91-99.
[13] Fukuda K, Ushio T (2018) Decentralized Event-Triggered Control of Composite Systems Using M-Matrices. IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences, 101(8): 1156-1161.
[14] Li Y, Chen Y Q, Podlubny I (2009) Mittag-Leffler stability of fractional order nonlinear dynamic systems. Automatica, 45:1965-1969.
[15] Tuan H T, Trinh H (2017) Stability of fractional-order nonlinear systems by Lyapunov direct method. IET Control Theory & Applications, 12:2417-2422.
[16] Siljak D D (2012) Decentralized control of complex systems. Academic Press.

Speed-Sensorless Control of IPMSM based on Novel Nonsingular Fast Terminal Sliding Mode Observer and Fractional-Order Software Phase-Locked Loop

Kaihui ZHAO*, Ruirui ZHOU*, Jinhua SHE **, Aojie LENG*, Wangke DAI*, Gang HUANG***
* College of Electrical and Information Engineering, Hunan University of Technology
No. 88, Taishan Road, Tianyuan District, Zhuzhou, Hunan Province, 412007, China
** School of Engineering, Tokyo University of Technology
1404-1 Katakura, Hachioji, Tokyo 192-0982, JAPAN.
*** College of Traffic Engineering, Hunan University of Technology
No. 88, Taishan Road, Tianyuan District, Zhuzhou, Hunan Province, 412007, China

Abstract

This paper presents a novel method of improving the speed-sensorless control performance of the interior permanent magnet synchronous motor (PMSM) based on the nonsingular fast terminal sliding mode observer (NFTSMO) and fractional-order software phase-locked loop. First, the system of interior PMSM is described. Then, the NFTSMO is constructed to estimate the *d-q* -axis back electromotive force (EMF). Moreover, the speed and position of the rotor are tracked accurately by the fractional-order software phase-locked loop. The effectiveness and the feasibility are verified by simulation in Matlab/Simulink. The result shows excellent performances in spite of speed fluctuation, torque ripple.

Keywords: Interior permanent magnet synchronous motor (IPMSM), Speed-sensorless control, Nonsingular fast terminal sliding mode observer (NFTSMO), Fractional-order software phase-locked loop (FO-SPLL).

1. INTRODUCTION

Permanent Magnet Synchronous Motor (PMSM) is developed and applied in industrial drives, railway transportation, and generators in renewable energy power plants [1]. PMSM has many advantages, such as high-energy efficiency, rugged construction, reliable operation, high power factor [2, 3].

The speed-sensorless control can enhance the control accuracy, improve the reliability, and reduce the volume, the cost of the motor drive system. Therefore, the speed-sensorless control of PMSM is attractive in many industrial applications.

The speed-sensorless control method can be divided into three classes [4]. The first one is based on the back electromotive force (EMF) for the medium and the high speed [5]. The second one is based on the estimated flux [6]. The third one is the signal injection estimation for low-speed operation [7, 8].

The back EMF-based method gets a lot of attention, because it has a good control performance at the medium and high speed. But the method is mostly on the α-β -axis back EMF [9]. A speed-sensorless control method is proposed for interior PMSM (IPMSM) based on an adaptive super-twisting sliding-mode observer (AST-SMO) and improved SPLL in α-β stationary reference frame [5]. A sensorless control of PMSM is proposed to estimate the speed and stator resistance using a full-order sliding mode observer (FO-SMO) in α-β stationary reference frame [2]. To achieve high-performance speed-sensorless control for interior PMSM, a robust backstepping controller with terminal SMO is proposed [7]. A sensorless drive scheme is proposed based on the virtual third harmonic back EMF and fractional-N phase-soft locked loop (FN-SPLL) [10]. A sensorless control system of interior PMSM is presented based on the voltage injection with a pulse-width modulation (PWM) carrier and extended EMF [11]. To achieve low-speed and zero operation of Interior PMSM, a scheme is proposed based on low-frequency voltage injection and an enhanced vector-tracking observer [12]. A sensorless control method of PMSM is proposed for wide-speed-range drives and energy-efficient, and it is verified by extensive experiments [13]. However, hardly any work on *d-q*-axis back EMF can be found [8].

In this paper, a novel control strategy is implemented in the *d-q* axis reference frame based on nonsingular fast terminal sliding mode observer (NFTSMO) and fractional-order software phase-locked loop (FO-SPLL). The *d-q* axis back EMF is estimated by the NFTSMO. The rotor speed and position are tracked accurately by using the FO-SPLL to improve the speed-sensorless control performance of interior PMSM.

The remainder is organized as follows: The *d-q* axis mathematical model of interior PMSM is introduced in Section 2. Then, an NFTSMO is designed to estimate the

d-q-axis back EMF, and the rotor speed and position is tracked accurately using FO-SPLL in Section 3. The overall system is tested by Simulink/Matlab simulation in Section 4. The conclusions are given in Section 5.

2. SYSTEM DESCRIPTION

The *d-q* axis voltage equations of PMSM are [14]

$$\begin{cases} u_d = R_s i_d + \dfrac{d\psi_d}{dt} - \omega_e \psi_q \\ u_q = R_s i_q + \dfrac{d\psi_q}{dt} + \omega_e \psi_d \end{cases} \tag{1}$$

where R_s is stator resistance, u_d, u_a are the *d-q*-axis voltages, i_d, i_a are the *d-q*-axis stator currents, ψ_d, ψ_a are *d-q*-axis stator flux linkage, ω_e is motor electrical angular velocity, respectively.

The *d-q* axis stator flux linkage of interior PMSM are

$$\begin{cases} \psi_d = L_d i_d + \psi_r \\ \psi_q = L_q i_q \end{cases} \tag{2}$$

where ψ_r is the rotor magnets flux linkage, L_d, L_a are the *d-q* axis inductances.

From (1), (2), the *d-q*-axis voltage equations of interior PMSM can be rewrote as

$$\begin{cases} u_d = R_s i_d + L_d \dfrac{di_d}{dt} - \omega_e L_q i_q \\ u_q = R_s i_q + L_q \dfrac{di_q}{dt} + \omega_e L_d i_d + \omega_e \psi_r \end{cases} \tag{3}$$

From (3), the *d-q*-axis current equations of interior PMSM can be expressed as

$$\begin{cases} \dfrac{di_d}{dt} = -\dfrac{R_s}{L_d} i_d + \omega_e \dfrac{L_q}{L_d} i_q + \dfrac{1}{L_d} u_d + \dfrac{1}{L_d} e_d \\ \dfrac{di_q}{dt} = -\dfrac{R_s}{L_q} i_q - \omega_e \dfrac{L_d}{L_q} i_d + \dfrac{1}{L_q} u_q + \dfrac{1}{L_q} e_q \end{cases} \tag{4}$$

where e_d and e_a are the back EMF in *d-q*-axis frame, and $e_a = -\omega_e \psi_r$, $e_d = 0$.

According to (4), state equation of interior PMSM can be summarized as follows

$$\begin{cases} \dot{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{B}\boldsymbol{d} \\ \boldsymbol{y} = \boldsymbol{C}\boldsymbol{x} \end{cases} \tag{5}$$

where $\boldsymbol{x} = \begin{bmatrix} i_d & i_a \end{bmatrix}^T$ are state vector, $\boldsymbol{u} = \begin{bmatrix} u_d & u_a \end{bmatrix}^T$ are inputs vector, $\boldsymbol{d} = \begin{bmatrix} e_d & e_a \end{bmatrix}^T$ are reconfigurable vector, $\boldsymbol{y} = \begin{bmatrix} i_d & i_a \end{bmatrix}^T$ are output vector, and

$$\boldsymbol{A} = \begin{bmatrix} -\dfrac{R_s}{L_d} & \omega_e \dfrac{L_q}{L_d} \\ -\omega_e \dfrac{L_d}{L_q} & -\dfrac{R_s}{L_q} \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \dfrac{1}{L_d} & 0 \\ 0 & \dfrac{1}{L_q} \end{bmatrix}, \quad \boldsymbol{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

3. SENSORLESS CONTROL OF INTERIOR PMSM USING NFTSMO AND FO-SPLL

3.1 Design of the NFTSMO

To improve the precision and eliminate the chattering of the observer, a novel nonsingular fast terminal sliding mode (NFTSM) observer is proposed to estimate the back EMF in *d-q*-axis reference frame.

According to (5), a NFTSMO can be designed as follows:

$$\dot{\hat{\boldsymbol{x}}} = \boldsymbol{A}\hat{\boldsymbol{x}} + \boldsymbol{B}\boldsymbol{u} + \boldsymbol{v} \tag{6}$$

where $\hat{\boldsymbol{x}} = \begin{bmatrix} \hat{i}_d & \hat{i}_a \end{bmatrix}^T$, $\wedge$ denotes the estimated values; $\boldsymbol{v} = \begin{bmatrix} v_d & v_a \end{bmatrix}^T$ is the control input vector of the NFTSMO.

From (6) and (5), the error equation can be obtained

$$\dot{\boldsymbol{e}} = \boldsymbol{A}\boldsymbol{e} + \boldsymbol{B}\boldsymbol{d} - \boldsymbol{v} \tag{7}$$

where $\boldsymbol{e} = \boldsymbol{x} - \hat{\boldsymbol{x}} = \begin{bmatrix} e_1 & e_2 \end{bmatrix}^T = \begin{bmatrix} x_1 - \hat{x}_1 & x_2 - \hat{x}_2 \end{bmatrix}^T$, $\boldsymbol{e}$ is the error vector.

The sliding mode manifold can be defined as follows:

$$\boldsymbol{s} = \begin{bmatrix} s_1 & s_2 \end{bmatrix}^T = \boldsymbol{e} = \begin{bmatrix} e_1 & e_2 \end{bmatrix}^T \tag{8}$$

Feng et al. [15] presents the traditional nonsingular terminal sliding mode (NTSM) manifold

$$\boldsymbol{l} = \boldsymbol{s} + \boldsymbol{\beta}\dot{\boldsymbol{s}}^{p/q} \tag{9}$$

where $\boldsymbol{l} \in R^2$, $\boldsymbol{s} = \begin{bmatrix} s_1 & s_2 \end{bmatrix}^T$, $\dot{\boldsymbol{s}}^{p/q} = \begin{bmatrix} \dot{s}_1^{p/q} & \dot{s}_2^{p/q} \end{bmatrix}^T$, $\boldsymbol{\beta} = diag(\beta_1, \beta_2)$, $\beta_1 > 0$, $\beta_2 > 0$, $1 < p/q < 2$, $p > 0$, $q > 0$, p and q are odd.

The novel NFTSM manifold is presented [16, 17]

$$\boldsymbol{l} = \boldsymbol{s} + \boldsymbol{\alpha}\left|\boldsymbol{s}\right|^{g/h} sgn\boldsymbol{s} + \boldsymbol{\beta}\left|\dot{\boldsymbol{s}}\right|^{p/q} sgn\dot{\boldsymbol{s}} \tag{10}$$

where $\boldsymbol{l} \in R^2$, $\boldsymbol{s} = \begin{bmatrix} s_1 & s_2 \end{bmatrix}^T$, $\dot{\boldsymbol{s}}^{p/q} = \begin{bmatrix} \dot{s}_1^{p/q} & \dot{s}_2^{p/q} \end{bmatrix}^T$, $\boldsymbol{\alpha} = diag(\alpha_1, \alpha_2)$, $\boldsymbol{\beta} = diag(\beta_1, \beta_2)$, $\alpha_1 > 0$, $\alpha_2 > 0$, $\beta_1 > 0$, $\beta_2 > 0$, $1 < p/q < 2$, $g/h > p/q$, $p > 0$, $q > 0$, p and q are odd.

Remark 1: When the state $\boldsymbol{s}$ is close to the system equilibrium, ignoring the higher order terms $\boldsymbol{s}$, the NFTSM manifold (10) is equivalent to the NTSM manifold (9). So, the convergence speed of NFTSM is approximately equal to NTSM (9).

Remark 2: When the state is far from the system equilibrium, the high term of $\boldsymbol{s}$ in the right of equation (10) plays an important role. The convergence speed of NFTSM is faster than NTSM (9).

3.2 Stability Analysis of the NFTSMO

The control law of the NFTSMO is designed according

to Theorem 1.

Theorem 1: The error equation (7) is asymptotically stable, if the manifold (10) is chosen, and the control law (11) of the NFTSMO is designed as

$$\boldsymbol{v} = \boldsymbol{v}_{ea} + \boldsymbol{v}_n \tag{11}$$

where

$$\boldsymbol{v}_{ea} = \boldsymbol{A}\boldsymbol{e} \tag{12}$$

$$\boldsymbol{v}_n = \int_0^t \left[\frac{(1+\alpha \frac{g}{h}|\boldsymbol{s}|^{\frac{g}{h}-1})}{\beta \frac{p}{q}} |\dot{\boldsymbol{s}}|^{2-\frac{p}{q}} \mathrm{sgn}\dot{\boldsymbol{s}} + (k+\eta)\mathrm{sgn}\boldsymbol{l} + \mu \boldsymbol{l} \right] d\tau \tag{13}$$

where, $k > \max\left(\boldsymbol{B}\|\dot{\boldsymbol{d}}\|\right)$, $\mathrm{sgn}(\boldsymbol{l}) = \left[\mathrm{sgn}(l_1) \quad \mathrm{sgn}(l_2)\right]^T$, $k>0$, $\eta>0$, $\mu>0$ are the designed parameters.

Proof: From (10), it gets

$$\dot{\boldsymbol{l}} = \dot{\boldsymbol{s}} + \frac{g}{h}\alpha |\boldsymbol{s}|^{\frac{g}{h}-1}\mathrm{sgn}\dot{\boldsymbol{s}} + \frac{p}{q}\beta |\dot{\boldsymbol{s}}|^{\frac{p}{q}-1}\mathrm{sgn}\ddot{\boldsymbol{s}} \tag{14}$$

Define the following Lyapunov function

$$V(t) = \frac{1}{2}\boldsymbol{l}^T\boldsymbol{l} \tag{15}$$

Substituting (14) into (15) yields

$$\begin{aligned} \dot{V}(t) &= \boldsymbol{l}^T\dot{\boldsymbol{l}} \\ &= \boldsymbol{l}^T\left[\dot{\boldsymbol{s}} + \frac{g}{h}\alpha|\boldsymbol{s}|^{\frac{g}{h}-1}\mathrm{sgn}\dot{\boldsymbol{s}} + \frac{p}{q}\beta|\dot{\boldsymbol{s}}|^{\frac{p}{q}-1} + \ddot{\boldsymbol{s}}\right] \\ &= \boldsymbol{l}^T\frac{p}{q}\beta|\dot{\boldsymbol{s}}|^{\frac{p}{q}-1}\left[\frac{(1+\alpha\frac{g}{h}|\boldsymbol{s}|^{\frac{g}{h}-1})}{\beta\frac{p}{q}}|\dot{\boldsymbol{s}}|^{2-\frac{p}{q}}\mathrm{sgn}\dot{\boldsymbol{s}} + \ddot{\boldsymbol{s}}\right] \end{aligned} \tag{16}$$

From the stator current error system (7), one obtains

$$\dot{\boldsymbol{e}} = \boldsymbol{A}\boldsymbol{e} + \boldsymbol{B}\boldsymbol{d} - \boldsymbol{v} = \boldsymbol{B}\boldsymbol{d} - \boldsymbol{v}_n \tag{17}$$

From (8) and (17), (16) can be rewritten as

$$\begin{aligned} \dot{V}(t) &= \boldsymbol{l}^T\frac{p}{q}\beta|\dot{\boldsymbol{s}}|^{\frac{p}{q}-1}\left[\frac{(1+\alpha\frac{g}{h}|\boldsymbol{s}|^{\frac{g}{h}-1})}{\beta\frac{p}{q}}|\dot{\boldsymbol{s}}|^{2-\frac{p}{q}}\mathrm{sgn}\dot{\boldsymbol{s}} + \boldsymbol{B}\dot{\boldsymbol{d}} - \dot{\boldsymbol{v}}_n\right] \\ &= \boldsymbol{l}^T\frac{p}{q}\beta|\dot{\boldsymbol{s}}|^{\frac{p}{q}-1}\left[\boldsymbol{B}\dot{\boldsymbol{d}} - (k+\eta)\mathrm{sgn}(\boldsymbol{l}) - \mu\boldsymbol{l}\right] \end{aligned} \tag{18}$$

Since the parameter k satisfies $k > \max\left(\boldsymbol{B}\|\dot{\boldsymbol{d}}\|\right)$, the following inequation can be obtained

$$\begin{aligned} \dot{V}(t) &\le -\boldsymbol{l}^T\frac{p}{q}\beta\dot{\boldsymbol{s}}^{\frac{p}{q}-1}\left[\eta\mathrm{sgn}(\boldsymbol{l}) + \mu\boldsymbol{l}\right] \\ &= -\frac{p}{q}\min_{i=1,2}\left(\beta_i\dot{s}_i^{\frac{p}{q}-1}\right)\left[\eta\|\boldsymbol{l}\| + \mu\|\boldsymbol{l}\|^2\right] \end{aligned} \tag{19}$$

Since p and q are all odds and $1 < p/q < 2$, e.g. $q = 2m+1$, $p = 2m+3$, $m \in N$, the following equation can be obtained

$$\dot{s}_i^{p/q-1} = \dot{s}_i^{(p-q)/q} = \left(\dot{s}_i^2\right)^{(p-q)/(2q)} = \left(\dot{s}_i^2\right)^{1/(2m+1)} \ge 0$$

Then, the term $\dot{s}_i^{p/q-1}$ has two possible cases:

$$\begin{cases} \dot{s}_i^{p/q-1} > 0 & for \ \dot{s} \ne 0 \\ \dot{s}_i^{p/q-1} = 0 & for \ \dot{s} = 0 \end{cases} \tag{20}$$

Substituting (20) into (19), one obtains $\dot{V}(t) \le 0$.

The error equation (7) will converge to zero asymptotically [15]. □

Remark 3: The convergence speed of l can be regulated by selecting α, β, p, q, g and h.

Remark 4: When the system state reaches and stays on the NFTSM manifold $\boldsymbol{l} = 0$, the equivalent control principle of sliding mode ensures $\boldsymbol{s} = \dot{\boldsymbol{s}} = 0$.

According to the error equation (7), it gets

$$\boldsymbol{B}\boldsymbol{d} = \boldsymbol{v} \tag{21}$$

One can get the estimated back EMF $\hat{e}_d$, $\hat{e}_q$ in *d*-*q*-axis

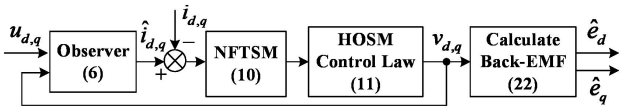


Fig. 1 Principle diagram of NFTSMO of back EMF.

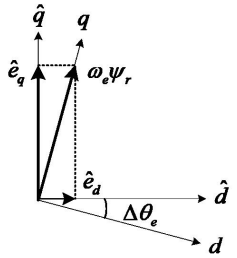


Fig. 2: Assumed $\hat{d}-\hat{q}$ and actual *d*-*q*-axis reference frame.

reference frame.

$$\begin{cases} \hat{e}_d = L_d v_d \\ \hat{e}_q = L_q v_q \end{cases} \tag{22}$$

Remark 5: From (11)-(13), it can get that the control law $\boldsymbol{v}$ in (11) is continuous and smooth. The control law $\boldsymbol{v}$ can be used to estimate back EMF directly. Then, the estimated back EMF can be calculated from(22).

The proposed NFTSMO can be described in Fig.1.

3.3 Design the FO-SPLL to Estimate Rotor Position and Speed

The *d*-*q*-axis back EMF is input to the FO-SPLL. FO-SPLL can track the rotor speed and position

accurately.

Software phase locked-loop (SPLL) is an adaptive closed-loop system. SPLL can track the frequency and phase real-time. It also has good tracking performance even if the phase angle of the voltage in imbalance conditions, and on large harmonic, relatively. This paper estimated rotor speed and rotor position utilizing FO-SPLL.

In sensorless control of interior PMSM, since the rotor speed is not known, the $\hat{d}-\hat{q}$ frames are assumed to estimate the rotor speed. The relationship between the *d*-*q*-axis and estimated $\hat{d}-\hat{q}$ axis is shown in Fig.2 [18].

The phase error $\Delta\theta_e$ is defined as

$$\Delta\theta_e = \hat{\theta}_e - \theta_e = \arctan\frac{\hat{e}_d}{\hat{e}_q} \tag{23}$$

where $\hat{\theta}_e$ is estimated position of the rotor, θ_e is the real position of the rotor.

In Fig.2, the geometrical relationship is shown as following

$$\begin{cases} \hat{e}_q = \omega_e \psi_r \cos\Delta\theta_e \\ \hat{e}_d = \omega_e \psi_r \sin\Delta\theta_e \end{cases} \tag{24}$$

When the estimated rotor position $\hat{\theta}_e$ follows the real position θ_e, one can get $\Delta\theta_e = \hat{\theta}_e - \theta_e = 0$ by FO-SPLL.

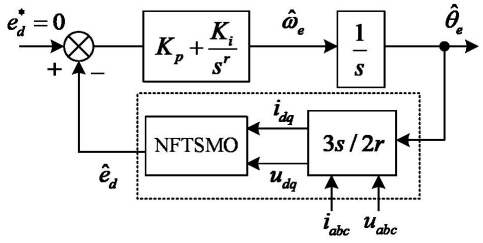


Fig. 3: Transfer function of FO-PLL.

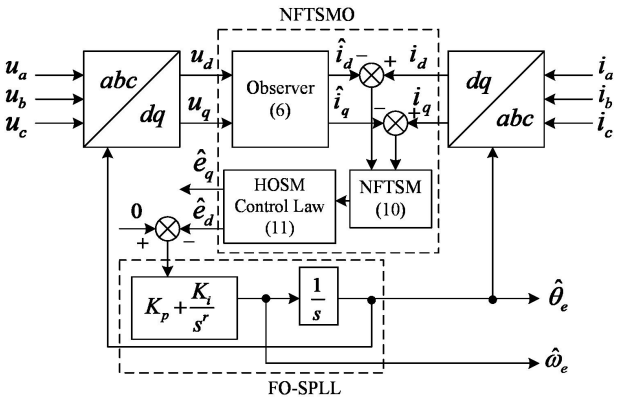


Fig. 4: Principle diagram of speed-sensorless control based on NFTSMO and FO-SPLL.

This gives

$$\begin{cases} \hat{e}_q = \omega_e \psi_r \cos\Delta\theta_e = \omega_e \psi_r \\ \hat{e}_d = \omega_e \psi_r \sin\Delta\theta_e = \omega_e \psi_r (\hat{\theta}_e - \theta_e) = 0 \end{cases} \tag{25}$$

Fig.3 shows the diagram of FO-SPLL. Therefore, the closed -loop transfer function of FO-SPLL is

$$G(s) = \frac{\hat{\theta}_e(s)}{\theta_e(s)} = \frac{e_q K_P s^r + e_q K_I}{s^{r+1} + e_q K_P s^r + e_q K_I} \tag{26}$$

The error transfer function of FO-SPLL can be obtained as

$$G_E(s) = \frac{\Delta\theta_e(s)}{\theta_e(s)} = \frac{s^{r+1}}{s^{r+1} + e_q K_P s^r + e_q K_I} \tag{27}$$

where $0 < r \le 1$, r is called the fractional-orders of the FO-SPLL.

With constant speed command, the rotor position $\hat{\theta}_e(t)$ is a ramp function, and the steady-state error of FO-SPLL is given by

$$\begin{aligned} \Delta\theta_e(\infty) &= \lim_{s\to 0} s \cdot \Delta\theta_e(s) \\ &= \lim_{s\to 0} \frac{s^r}{s^{r+1} + e_q K_p s^r + e_q K_i} = 0 \end{aligned} \tag{28}$$

Therefore, the rotor position can be estimated by FO-SPLL.

Remark 6: When the fractional order r decreases, the bandwidth of FO-SPLL increases, and the phase delay increases [19].

Remark 7: Obviously, when $r = 1$, the FO-SPLL becomes the conventional SPLL. Because the adjustable range of r is wide, the performance of FO-SPLL is better than traditional SPLL.

The diagram of interior PMSM speed-sensorless control based on NFTSMO and FO-SPLL is shown in Fig.4.

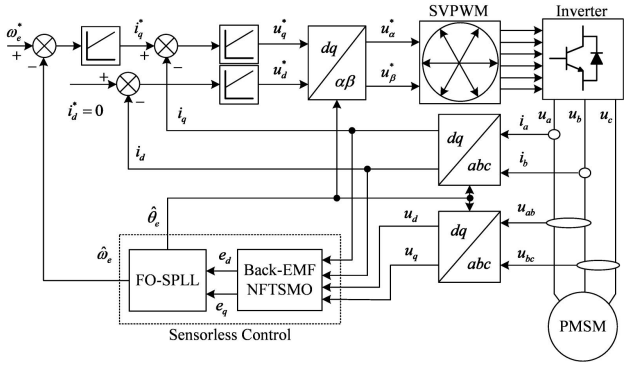


Fig. 5 The block of speed-sensorless control system for interior PMSM.

Table 1. Parameters of interior PMSM

Parameters	Unit	Values
Rated voltage (U_N)	V	380
Rated current (I_N)	A	3.5
Rated speed (n_N)	r/min	1000
Stator resistance (R_s)	Ω	2.875
q axis inductance (L_q)	H	0.0075
d axis inductance (L_d)	H	0.0025
Rotational Inertia (J)	$kg.m^2$	0.0008
Rotor flux (ψ_r)	Wb	0.175
Number of pole pairs (n_p)	$paires$	4

4. SIMULATIONS

The overall block of the interior PMSM speed-sensorless control system is shown in Fig.5. The system consists of a speed PI regulator, *d*-*q*-axis current PI regulators, a speed and rotor position observer implemented by NFTSMO with FO-SPLL, etc.

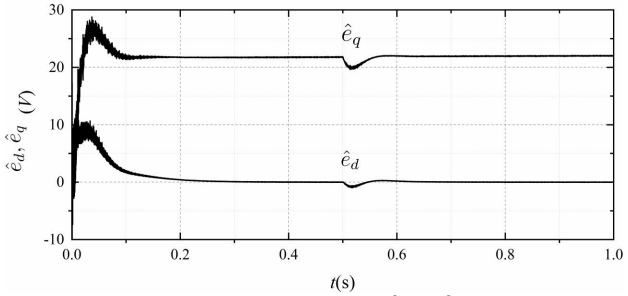
The i_d=0 control mode is carried out. The parameters of interior PMSM is listed in Table 1.

The proposed NFTSMO in *d*-*q*-axis frame is designed to observe the back EMF. According to Theorem 1, the parameters of the observer are chosen as $p=7$, $q=5$, $g=5$, $h=3$, $\alpha_1=\alpha_2=\beta_1=\beta_2=1$, $k+\eta=3000$, $\mu=2000$. The parameters of FO-SPLL are chosen as $K_p=100$, $K_i=500$, $r=0.9$.

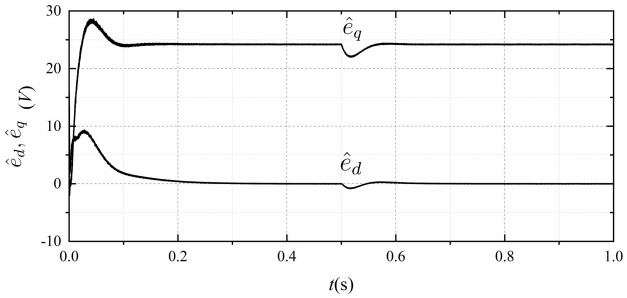
The initial speed is set to 1000 r/min. The initial load torque is 0 Nm, and is increased to 2 Nm at 0.5s. The simulation results are shown in Fig. 6-8.

Fig.6 shows the simulation results of the observed *d*-*q*-axis back EMF by SMO and NFTSMO. The estimated back EMF e_d, e_q by SMO are shown in Fig.6 (a), and the estimated back EMF e_d, e_q by NFTSMO are shown in Fig.6 (b). It is clear from Fig.6 that, the torque also has an impact on the speed at the 0.5s.

Fig.7 shows the simulation results of the actual, estimated and errors of rotor speed by SMO and NFTSMO. The actual, estimated and errors of rotor speed by SMO are shown in Fig.7 (a), and the actual, estimated and errors of rotor speed by NFTSMO are shown in Fig.7(b). It is clear



(a) The estimated back EMF $\hat{e}_d$, $\hat{e}_a$ by SMO.



(b) The estimated back EMF $\hat{e}_d$, $\hat{e}_q$ by NFTSMO.

Figure 6: The *d*-*q*-axis back EMF $\hat{e}_d$, $\hat{e}_q$.

from Fig.7 that, the *d*-*q*-axis back EMF is affected by the torque change at the 0.5 s.

Fig.8 shows the simulation results of the actual, estimated and errors of rotor position.

It can be drawn from the simulation results:

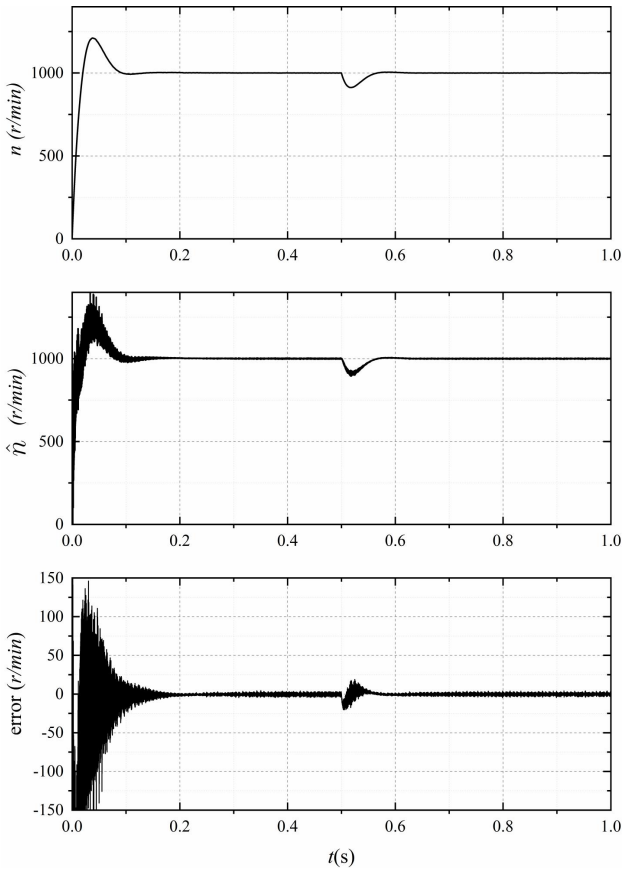
(1) The *d*-*q*-axis back EMF and the estimated rotor position are affected by the torque change at the 0.5 s.
(2) The torque also has an impact on the speed at the 0.5 s. The estimated speed is very close to the actual speed, and the estimated errors are small.
(3) It can be seen from simulation results that the SMO and NTSMO all operate well. When the load torque are changing at 0.5 s, the SMO has the chattering phenomenon, and the NTSMO almost has no chattering.
(4) The simulation results show that the proposed method can reduce or eliminate the chattering, and improve the dynamic response and precision.

5. CONCLUSION

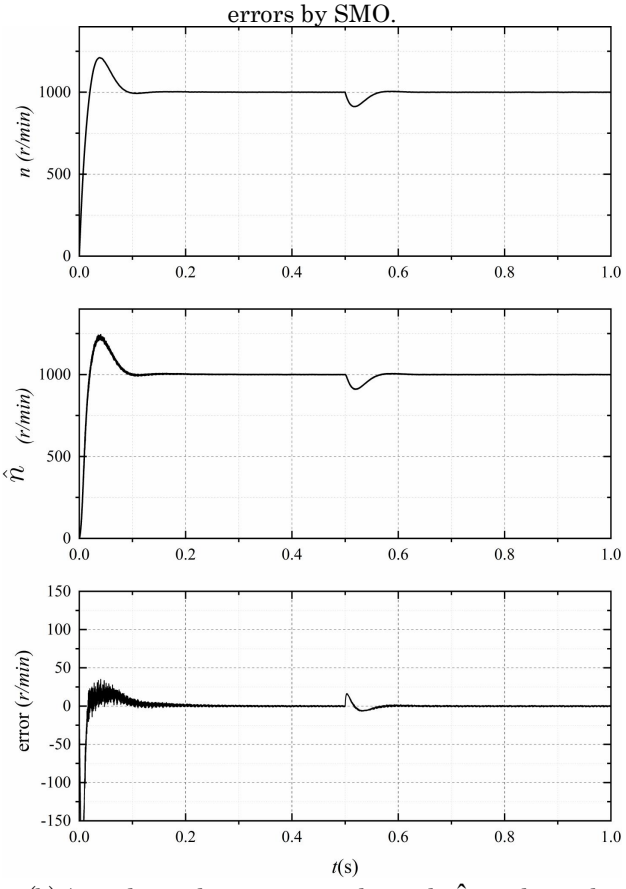
The robust speed-sensorless control scheme for interior PMSM is presented combining a NFTSMO and FO-SPLL. The NFTSMO is constructed to estimate the *d*-*q*-axis back EMF. The fractional-order phase-locked loop (FO-PLL) is designed to track the speed and position of rotor accurately. Furthermore, the system was implemented in Matlab/Simulink. The simulation results verify the efficiency of the proposed scheme have good robustness, good performance, good convergence.

Acknowledgements

This work was supported in part by the program of JSPS (Japan Society for the Promotion of Science) International Research Fellows under Grant 19F19703; the National Key Research and Development Program of China under Grants 2017YFB1300900 and 2018YFD0400705; the Natural Science Foundation of China under Grants 61773159 and 61873348; the Hunan Provincial Natural Science Foundation of China under

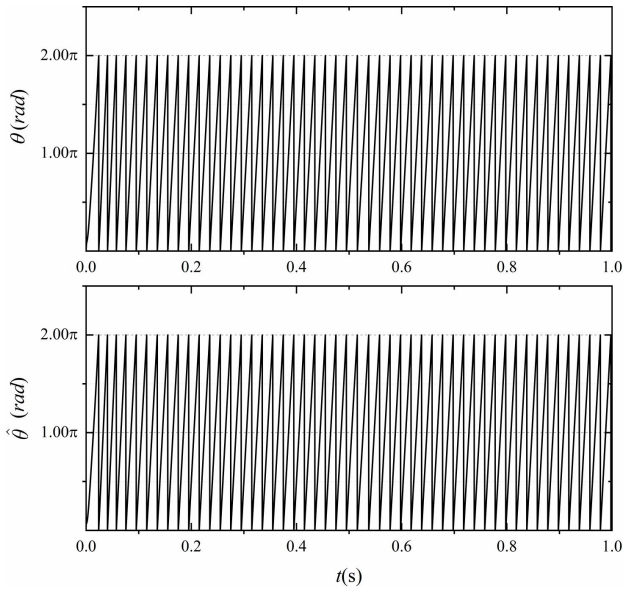


(a) Actual speed n, estimated speed $\hat{n}$, and speed errors by SMO.

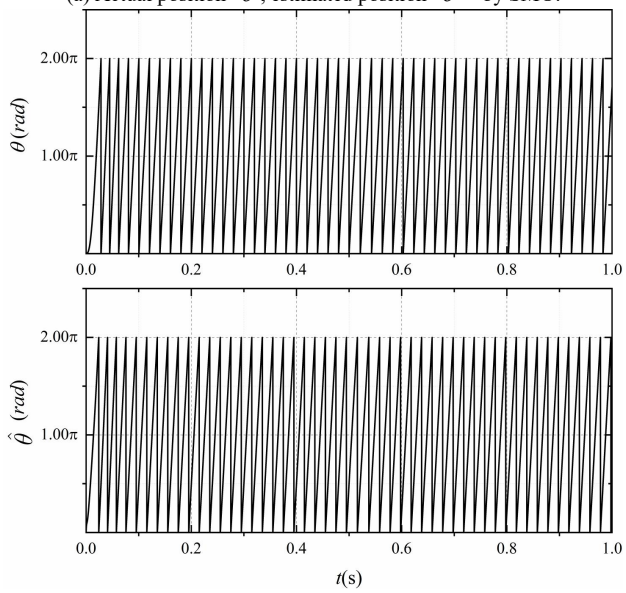


(b) Actual speed n, estimated speed $\hat{n}$, and speed errors by NFTSMO.

Fig. 7 The actual speed n, estimated speed $\hat{n}$, and speed errors.



(a) Actual position θ, estimated position $\hat{\theta}$ by SMO.



(b) Actual position θ, estimated position $\hat{\theta}$ by NFTSMO.

Fig. 8 The actual position θ, estimated position $\hat{\theta}$.

Grants 2020JJ6083, 2019JJ40072, and 2018JJ4066; the Teaching Reform Research Project of Hunan Provincial Education Department of China (Hunan Education Notice [2019] No.291) under grant no. 543; the Degree & Postgraduate Education Reform Project of Hunan Province under grant no. 2019JGZD068; and the Hunan Postgraduate Research Innovation Project under Grant CX20190861.

References

[1] C. F. Zhang, G. P. Wu, R. Fei, J. H. Feng, L. Jia, J. He, and S. D. Huang, "Robust Fault-Tolerant Predictive Current Control for Permanent Magnet Synchronous Motors Considering Demagnetization Fault," IEEE Transactions on Industrial Electronics,

Vol. 65, No. 7, 2018, pp. 5324–5334.

[2] O. Saadaoui, A. Khlaief, M. Abassi, I. Tlili, A. Chaari, and M. Boussak, “A New Full-Order Sliding Mode Observer Based Rotor Speed and Stator Resistance Estimation for Sensorless Vector Controlled PMSM drives,” Asian Journal of Control, Vol. 21, No. 3, 2019, pp. 1318–1327.

[3] K. Zhao, T. Yin, C. Zhang, J. He, X. Li, Y. Chen, R. Zhou, and A. Leng, “Robust Model-Free Nonsingular Terminal Sliding Mode Control for PMSM Demagnetization Fault,” IEEE Access, Vol. 7, 2019, pp. 15 737–15 748.

[4] K. Zhao, P. Li, J. She, C. Zhang, and J. He, “Sensorless Control for IPMSM Based on NFTSMO and FOPLL,” In The 6th International Workshop on Advanced Computational Intelligence and Intelligent Informatics (IWACIII 2019), Chengdu, China, 2019, pp. 1–5.

[5] S. Chen, X. Zhang, X. Wu, G. Tan, and X. Chen, “Sensorless Control for IPMSM Based on Adaptive Super-Twisting Sliding-Mode Observer and Improved Phase-Locked Loop,” Energies, Vol. 12, No. 7, 2019.

[6] E. G. Shehata, “Speed Sensorless Torque Control of an IPMSM Drive with Online Stator Resistance Estimation Using Reduced Order EKF,” International Journal of Electrical Power & Energy Systems, Vol. 47, 2013, pp. 378–386.

[7] S. Wu and J. Zhang, “A Terminal Sliding Mode Observer Based Robust Backstepping Sensorless Speed Control for Interior Permanent Magnet Synchronous Motor,” International Journal of Control Automation and Systems, Vol. 16, No. 6, 2018, pp. 2743–2753.

[8] G. Foo and M. F. Rahman, “Sensorless Sliding-Mode MTPA Control of an IPM Synchronous Motor Drive Using a Sliding-Mode Observer and HF Signal Injection,” IEEE Transactions on Industrial Electronics, Vol. 57, No. 4, 2010, pp. 1270–1278.

[9] H. Zhan, Z. Q. Zhu, M. Odavic, and Y. Li, “A Novel Zero-Sequence Model-Based Sensorless Method for Open-Winding PMSM with Common DC Bus,” IEEE Transactions on Industrial Electronics, Vol. 63, No. 11, 2016, pp. 6777–6789.

[10] X. Song, B. Han, S. Zheng, and S. Chen, “A Novel Sensorless Rotor Position Detection Method for High-Speed Surface PM Motors in a Wide Speed Range,” IEEE Transactions on Power Electronics, Vol. 33, No. 8, 2018, pp. 7083–7093.

[11] T. Yokoyama, K. Uchida, and H. Kubota, “Position Sensorless Control for IPMSM Based on Extended EMF and Voltage Injection Synchronized with Pulse-Width Modulation Carrier,” Electrical Engineering in Japan, Vol. 199, No. 1, 2017, pp. 28–39.

[12] G. Wang, D. Xiao, N. Zhao, X. Zhang, W. Wang, and D. Xu, “Low-Frequency Pulse Voltage Injection Scheme-Based Sensorless Control of IPMSM Drives for Audible Noise Reduction,” IEEE Transactions on Industrial Electronics, Vol. 64, No. 11, 2017, pp. 8415–8426.

[13] S. Shinnaka and Y. Amano, “Elliptical Trajectory-Oriented Vector Control for Energy-Efficient/Wide-Speed-Range Drives of Sensorless PMSM,” IEEE Transactions on Industry Applications, Vol. 51, No. 4, 2015, pp. 3169–3177.

[14] K. H. Zhao, T.-F. Chen, C.-F. Zhang, J. He, and G. Huang, “Online Fault Detection of Permanent Magnet Demagnetization for IPMSMS by Nonsingular Fast Terminal-Sliding-Mode Observer,” Sensors, Vol. 14, No. 12, 2014, pp. 23119–23136.

[15] Y. Feng, X. H. Yu, and Z. H. Man, “Non-singular Terminal Sliding Mode Control of Rigid Manipulators,” Automatica, Vol. 38, No. 12, 2002, pp. 2159–2167.

[16] S. Chen, W. Liu, and H. Huang, “Nonsingular Fast Terminal Sliding Mode Tracking Control for a Class of Uncertain Nonlinear Systems,” Journal of Control Science and Engineering, Vol. 2019, 2019, p. 8146901.

[17] W. Liu, S. Chen, and H. Huang, “Adaptive Nonsingular Fast Terminal Sliding Mode Control for Permanent Magnet Synchronous Motor Based on Disturbance Observer,” IEEE Access, Vol. 7, 2019, pp. 153791–153798.

[18] P. Kshirsagar, R. P. Burgos, J. Jihoon, A. Lidozzi, W. Fei, D. Boroyevich, and S. Seung-Ki, “Implementation and Sensorless Vector-Control Design and Tuning Strategy for SMPM Machines in Fan-Type Applications,” IEEE Transactions on Industry Applications, Vol. 48, No. 6, 2012, pp. 2402–2413.

[19] J. Huang, H. Li, F. Teng, and D. Liu, “Fractional Order Sliding Mode Controller for the Speed Control of a Permanent Magnet Synchronous Motor,” in 24th Chinese Control and Decision Conference (CCDC), 2012, pp. 1203–1208.

A Tomato Disease Recognition System Based on Image Enhancement and Deep Learning

Yong-Hua XIONG*, Long-Fei LIANG*
* School of Automation, China University of Geosciences
Wuhan, 430074, China

Leaf diseases are a major problem in the agricultural sector that affecting crop yields and economic benefits. Using deep learning methods to identify tomato leaf diseases has become a current research hotspot. However, recent studies have shown that changes in the light environment may cause the wrong result of deep learning models in practical applications. To solve this problem, we propose a deep learning method using the image enhancement algorithm with color restoration (MSRCR) to identify tomato leaf diseases in this paper. Compared with other image enhancement algorithms, the MSRCR algorithm enhances the image brightness while retaining the color information of the original image. We apply the MSRCR algorithm to the disease identification of tomato crops, which enabled the recognition rate of the deep learning model to reach 82% in low brightness environments. Compared with using unprocessed low brightness environment images, the MSRCR algorithm improves the recognition accuracy of the deep learning model by more than 65%. We select the EfficientNet Convolutional Neural Network (CNN) model as the deep learning model to identify tomato leaf diseases and achieve a recognition accuracy rate of 93.9% on our test set.

Keywords: Deep Learning, Color Restoration, Disease Recognition, Tomato Leaves, Convolutional Neural Network.

1. INTRODUCTION

At present, the production of tomato in China ranks first in the world, and it is one of the most important economic crops in China [1]. However, diseases in tomato leaves can cause major production and economic losses every year, as well as a decline in the quality and quantity of output in the vegetable industry [2]. By identifying these diseases, huge losses in production can be reduced. Similarly, in terms of quality and quantity, the final agricultural products can also be improved.

Traditional crop disease detection efficiency and reliability are poor because they mainly rely on manual observation. And farmers lack professional knowledge and agricultural experts cannot always serve the field, so they can easily miss the best time for disease prevention [3]. Researchers have developed many recognition systems based on images of plant diseases. For example, The author in [4] proposed an apple fruit disease classification method based on image processing technology. The author in [5] proposed an algorithm based on color and regional growth information for identifying leaf diseases in greenhouse vegetables.

In recent years, the deep CNN model has been widely used in crop disease image recognition. The author in [6] applied deep learning methods in their work to develop smartphone-assisted disease diagnosis systems. They used CNN to train the model by using a dataset of 54306 images of healthy and infected plant leaves. The model received training in using images to recognize 14 crops and 26 diseases. They evaluated the applicability of CNN on the classification of crop diseases and made their model achieve 99.35% accuracy. Although their models produced the latest results, they performed poorly when tested on image sets taken under different conditions.

The author in [7] used tomato leaf color image samples from the PlantVillage dataset to train the Alexnet, GoogleNet, and VGGNet models, with test accuracy rates of 91.52%, 89.68%, and 95.25%, respectively. The author in [8] used the same data set to train and compare the results of popular CNN models: VGG16, VGG19, Inception-V3, and ResNet50. They conducted experiments on different models, and the VGG16 model achieved the best test accuracy of 90.4%.

Although it seems that the CNN model performs well in crop disease recognition, there are still some problems waiting for researchers to solve. First, due to the uncertainty of the external environment, the trained model lacks a certain generalization ability. The second is that the computational load included in the CNN model is still very large, which requires a long time to train and evaluate. The third is that research at this stage is still in the experimental stage and lacks a certain application ability.

In this study, we propose to add the MSRCR algorithm before using the CNN model to help improve the

generalization ability of the model. The main research results of this article are as follows.

- We propose to apply the MSRCR algorithm to the disease recognition of tomato crops, which improves the recognition accuracy of the model in a low brightness environment by 65%.
- We train and use a relatively lightweight and novel efficientnet model, which finally achieved a 93.9% correct rate of tomato disease recognition.

The rest of the paper is organized as follows: Section 2 introduces the research methods used in this paper, including the MSRCR algorithm and the CNN model. In Section 3 we evaluate the performance of the method. Finally, Section 4 presents some conclusions and future directions.

2. MATERIALS AND METHODS

In this study, we use the public dataset of tomato disease images to train and test the EfficientNet model, and use the MSRCR algorithm to improve the model's recognition ability in low brightness environments and achieve a good application result.

2.1 Dataset Description

For the training of CNN models, large image datasets are considered necessary. Researchers usually use existing public datasets of crop disease images to train and evaluate their models to achieve the purpose of identifying specific crop diseases. The dataset that we use comes from the public dataset PlantVillage which has been applied to the study of plant health assessment many times, such as Ramcharan [9], Ferentinos [10].

Table 1. Information of the database images.

Label	Category	Quantity
1	Healthy (TH)	1591
2	Bacterial Spot (TBS)	2127
3	Early Blight (TEB)	1000
4	Late Blight (TLB)	1909
5	Leaf Mold (TLM)	952
6	Leaf Spot (TLS)	1771
7	Target Spot (TTS)	1404
8	Mosaic Virus (TMV)	5357

We select 16111 images from PlantVillage to form the dataset we need to assist our research. The dataset includes 8 different diseases of tomato crops, as shown in Table 1. It is worth noting that the dataset also contains 1591 images of healthy tomato crop leaves. Figure 1 shows us some samples in this dataset.

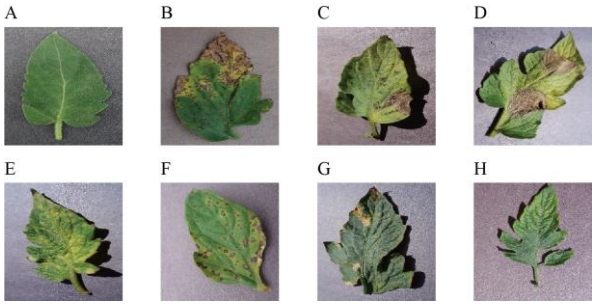


Fig.1 Examples of images from 8 classes in the tomato dataset. (A) Healthy (TH), (B) Bacterial Spot (TBS), (C) Early Blight (TEB), (D) Late Blight (TLB), (E) Leaf Mold (TLM), (F) Leaf Spot (TLS), (G) Target Spot (TTS), (H) Mosaic Virus (TMV).

To facilitate the reading and processing of images by the CNN model, we adjust the size of all images to 224*224 pixels. We also divide the dataset into a training set, a verification set and a test set according to the ratio of 8:1:1 for the training and testing of models. Many researchers have also proved the scientificity of this method.

2.2 The MSRCR Algorithm

Color constancy means that the color of the same object under different light or light sources is constant. The color of the object is determined by its reflective properties, regardless of the external incident light source. The ability of the human visual system to discern the color of an object is only determined by the reflective properties of the surface of the object, and has nothing to do with the process of receiving incident light. Retinex theory is about how the human visual system adjusts the brightness and color of the sensed objects. It explains the mechanism that the same object can maintain constant color under the illumination of different light or light sources.

The expression of the Multi-scale Retinex algorithm(MSR) is:

$$R_i(x,y) = \sum_{k=1}^{K} W_k \left(\log I_i(x,y) - \log[F_k(x,y) * I_i(x,y)]\right) \quad (1)$$

In the formula: $i = (i = 1, 2, 3, ..., N)$ means the band number. For grayscale images, $N = 1$; for color images, $N = 3$, respectively corresponding to the R, G, and B components of the color image, and also corresponding to the long, medium, and short waves of the spectrum. $R_i(x,y)$ represents the Retinex enhanced image of the i-th band, $I_i(x,y)$ represents the original image of the i-th band, and $F_k(x,y)$ is the k-th Gaussian function. '*' represents the convolution operation, W_k is the weighting factor of the k-th scale.

The MSR algorithm is a synthesis of the Single-scale Retinex algorithm at multiple scales, which makes the image have a good balance between dynamic range compression and color rendering. But the experiment found that the image processed by the MSR algorithm still has the phenomenon of color distortion and image lightening. The reason for this phenomenon is that its RGB components have changed after the image enhanced by the MSR algorithm. The RGB component of the image processed by the MSRCR has almost no change compared to the original image.

The expression of the MSRCR algorithm is:

$$R_{MSRCR_i}(x,y) = \alpha_i(x,y) \cdot R_{MSR_i}(x,y) \quad (2)$$

$$\alpha_i(x,y) = \log(\beta \frac{I_i(x,y)}{\sum_{n=1}^{N} I_n(x,y)}) \quad (3)$$

N is the number of bands, I_i is the image of the i-th band, and β is the adjustment parameter. Generally, β can be 125. It can be seen from formula (2) that the MSRCR algorithm is based on the MSR algorithm and is obtained by adding an adjustment factor $\alpha_i(x,y)$, which takes into account the ratio of RGB components in the original image. By this factor, the MSR enhanced image is compensated, so that the MSRCR output image can better maintain the original color without distortion.
There is a limit of 10 pages for each paper in the Proceedings. Be sure to fill out your Conference Registration Form and your Speaker's Biographical Sketch and send them with one copy of your camera-ready manuscript.

2.3 CNN Model

The development of CNN models is usually carried out under limited resources and then expands it to larger computing resources to obtain better accuracy when conditions permit. In this study, we chose the EfficientNet model to identify and classify our dataset. Generally speaking, the more complex the CNN model structure, the more deep-level features of the dataset can be mined, and the final recognition effect will be better. However, as the structure of the model becomes more complex, the amount of calculation of the model will also increase exponentially.

Unlike other CNN models, EfficientNet improves performance by scaling in three dimensions of width, depth, and resolution using a set of fixed scaling factors that meet certain constraints. EfficientNet's network structure borrows from the MnasNet model and adopts the method of optimizing accuracy (ACC) and computational load (FLOPS) simultaneously. The specific network structure of EfficientNet is shown in Table 2.

Table 2. EfficientNet baseline network.

Stage	Operator	Resolution	Channels	Layers
1	Conv3x3	224x224	32	1
2	MBConv1,k3x3	112x112	16	1
3	MBConv6,k3x3	112x112	24	2
4	MBConv6,k5x5	56x56	40	2
5	MBConv6,k3x3	28x28	80	3
6	MBConv6,k5x5	14x14	112	3
7	MBConv6,k5x5	14x14	192	4
8	MBConv6,k3x3	7x7	320	1
9	Conv1x1/Pooling/FC	7x7	1280	1

In this work, we use the transfer learning method to retrain the existing weights of the last layer parameters of the EfficientNet model, which has learned a lot of prior visual knowledge from the imagenet database, to classify our dataset.

3. RESULTS

The computing hardware foundation of this experiment is the NVIDIA MX250 graphics processor to assist the calculation processing, and Tensorflow, OpenCV, Keras, CUDA are also used for the realization of the software part.

3.1 Comparative Experimental Design and Results

In this study, we select a more lightweight EfficientNet model to identify tomato crop diseases. Table 3 shows us the parameter information of the model. In order to prevent the model from overfitting, in this article we limit the number of model iterations to 1,000. Each time the model randomly reads 32 images from the training set for calculation and adjusting the internal parameters of the model.

Figure 2 shows us the entire training process of the EfficientNet model. Figure 2(A) shows the change curve of the model's correct rate during the training process. Figure 2(B) shows the transformation curve of the model cross entropy loss during the training process. Also, Table 4 shows us the performance of EfficientNet on each dataset at the end of the training.

Table 3. The EfficientNet model parameter settings.

Label	Parameter	Value
1	Batches/Epoch	1000
2	Learning rate	0.005
3	Train batch size	32
4	Validation batch size	-1

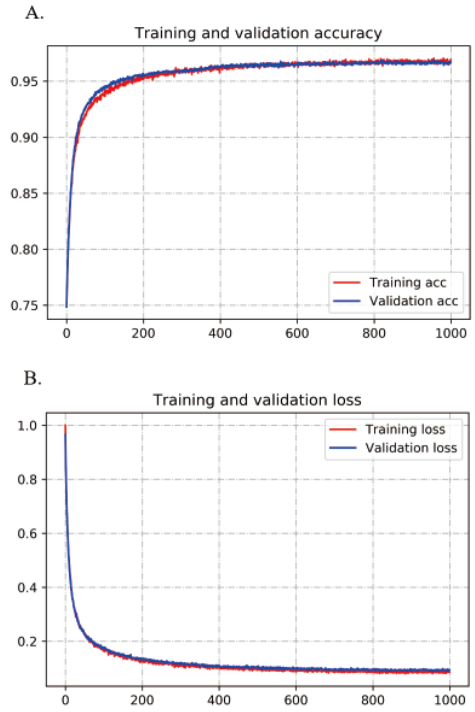


Fig.2 Train and validation accuracy and cross entropy of the EfficientNet model on dataset, during training.

Combining Figure 2 and Table 4, it is not difficult to find that the EfficientNet model is very outstanding in accuracy and cross entropy. From Figure 2(A), we can see that the change rate of the correct rate began to be flat after 600 iterations, and it was in a small-scale floating state after 800 iterations. Figure 2(B) shows the change rule is almost the same as Figure 2(A). It should be noted here that although it seems that the two curves are close to coincide in Figure 2, this does not mean that our model has overfitting. From the results in Table 4, we can see that the final model accuracy rate on the training set is 96.9%, while the model accuracy rate on the validation set is 94.5%. At the same time, the cross loss value of the model on the validation set is also 0.1 greater than that on the training set.

Table 4. Performance of EfficientNet on the dataset.

Label	Dataset	Correct Rate	Loss
1	Train	96.9%	0.084
2	Verification	94.5%	0.094
3	Test	93.9%	-

In the end, the accuracy of the model on our test set reached 93.9%. Although this result is lower than the accuracy of the model on the validation set, it also proves that our model does not appear to be overfitting.

3.2 Performance Analysis of The MSRCR Algorithm

In this section, in order to scientifically make the contrast effect more obvious, we randomly select 50 images from the test set as experimental observation objects. First, we adjust the exposure coefficient of 50 randomly selected images to -6 to simulate a realistic low brightness environment. We use the MSRCR algorithms to restore the brightness and color of the processed 50 images to observe the actual performance.

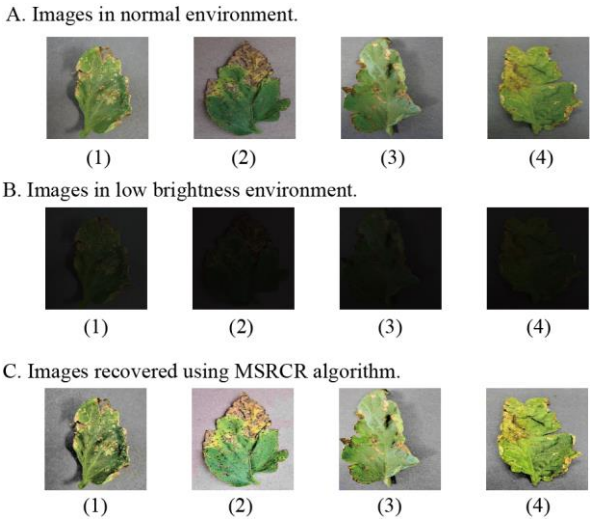


Fig.3 Comparison of tomato leaf images before and after enhancement.

Due to space limitations, here we randomly select 4 of these 50 independent experiments as the result display, as shown in Figure 3. From Figure 3(B), we can see that the image adjusted by the exposure coefficient has reached a level that is invisible to the human eye. Comparing Figure 3(A), Figure 3(B) and Figure 3(C), we can see that the image processed by MSRCR not only restores the brightness information of the original image, but also retains the color information of the original image.

Table 5. The accuracy of the EfficientNet model on different datasets.

Label	Category	Correct Rate
1	Original images	94%
2	Low brightness images	14%
3	MSRCR	82%

To observe the influence of the images processed by the MSRCR algorithm on the CNN model, we use the trained EfficientNet to identify the above image data, and the results are shown in Table 5. From the results in Table 5, we can see that the EfficientNet model is almost unrecognizable for images in low brightness environment, and its accuracy is only 14%. However, after the MSRCR processing of the low brightness image, the accuracy of the model recognition increased to 82%. Although the image processed by MSRCR still has a 12% gap compared with the original image, it is increased by 68% compared to the image with low brightness environment. The disease information contained in the images in the

low brightness environment is greatly reduced, which leads to the incorrect classification of the model.

4. CONCLUSION

How to make the crop disease identification model adapt to different outdoor environment to achieve better disease identification effect, has been one of the main research contents of crop disease identification. In this study, we tried to integrate the image enhancement idea with color restoration to solve the problem that the CNN model could not correctly identify the tomato diseases under low brightness environment, and achieved good results. This research has two main contributions:

- An image enhancement algorithm with color restoration is applied to the disease identification of tomato crops, which improves the system's recognition accuracy rate from 14% to 82% in low brightness environments.
- The efficientnet model we train and use to recognize tomato disease images has a correct rate of 93.9%.

Although experimental results have shown the effectiveness of our method, this research requires further efforts, including how to correctly identify different stages development of the same disease.

References

[1] Durmus H , Gunes E O , Kirci M . Disease detection on the leaves of the tomato plants by using deep learning[C]// 2017 6th International Conference on Agro-Geoinformatics. 2017.

[2] Karthik R, Hariharan M, Anand S, et al. Attention embedded residual CNN for disease detection in tomato leaves[J]. Applied Soft Computing, 2020, 86: 105933.

[3] Kamilaris A, Prenafeta-Boldú F X. Deep learning in agriculture: A survey[J]. Computers and electronics in agriculture, 2018, 147: 70-90.

[4] Dubey S R, Jalal A S. Apple disease classification using color, texture and shape features from images[J]. Signal, Image and Video Processing, 2016, 10(5): 819-826.

[5] Ma J, Du K, Zhang L, et al. A segmentation method for greenhouse vegetable foliar disease spots images using color information and region growing[J]. Computers and Electronics in Agriculture, 2017, 142: 110-117.

[6] Mohanty S P, Hughes D P, Salathé M. Using deep learning for image-based plant disease detection[J]. Frontiers in plant science, 2016, 7: 1419.

[7] Suryawati E, Sustika R, Yuwana R S, et al. Deep Structured Convolutional Neural Network for Tomato Diseases Detection[C]//2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). IEEE, 2018: 385-390.

[8] Wang G, Sun Y, Wang J. Automatic image-based plant disease severity estimation using deep learning[J]. Computational intelligence and neuroscience, 2017, 2017.

[9] Ramcharan A, Baranowski K, McCloskey P, et al. Deep learning for image-based cassava disease detection[J]. Frontiers in plant science, 2017, 8: 1852.

[10] Ferentinos K P. Deep learning models for plant disease detection and diagnosis[J]. Computers and Electronics in Agriculture, 2018, 145: 311-318.

Speech Emotion Recognition Based on Improved Synthetic Minority Over-Sampling Technique

Zhen-Tao Liu**, Bao-Han Wu**, Peng Xiao **, Jin-Meng Xu **
* School of Automation, China University of Geosciences,
Wuhan, Hubei 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems
Wuhan, Hubei 430074, China

Abstract

In small sample environment for speech emotion recognition, the problem of data imbalance may occur in the data preprocessing, which impacts the learning of different emotional categories in the decision space by the classifiers. To solve the problem, a data imbalance processing method based on improved synthetic minority over-sampling technique (ISMOTE) is proposed to reduce the impact of sample imbalance on emotion recognition results. Experiments on two databases (i.e., Emo-DB, SAVEE) are performed, from which the experimental results show that our method obtains average recognition accuracy of 82.61% (Emo-DB) and 72.92% (SAVEE) for speaker-dependent SER.

Keywords: Speech emotion recognition, Data imbalance processing, SMOTE, Small sample environment.

1. INTRODUCTION

With the rapid development of theories and technologies in human-computer interaction, intelligent service systems with emotions have become a hotspot. Since speech is the main way of human communication and the important medium for emotional expression, speech emotion recognition is of great significance to the development of artificial intelligence.

It is widely accepted that speech conveys not only the semantic meaning but also the emotional information of speakers [1, 2]. Both speaker-dependent (SD) and speaker-independent (SI) SER for small sample environment have attracted much attention [3–5]. However, it is time and cost demanding to prepare a certain amount of training data for SD SER, and even more severe in SI SER [6]. Thus, we mainly focus on SD SER in this paper.

In the actual application environment, it is difficult to obtain standard speech emotion data, and data imbalance of each emotion category often appears [7]. In view of the above problems, the solutions can be divided into two categories: data and algorithm [8, 9]. The data-level approach is to target the training set samples, which changes their sample distribution through corresponding strategies to reduce the degree of data imbalance. There are two main processing strategies, i.e., subsampling and oversampling [10]. The subsampling method is generally applied to the case where the data imbalance is small and the majority of the samples are sufficient, but it causes a certain degree of emotional in-formation loss. Oversampling can reduce the degree of data imbalance at the data level by constructing new samples, but artificially synthesized minority samples may increase the risk of overlearning in minority samples [11]. In addition, the algorithm-level method is to target the learning algorithms, which appropriately modifies the algorithm to adapt to the imbalanced classification problem according to the defects of the algorithm when solving the imbalanced problem. There are two commonly used strategies, i.e., classifier integration method and cost-sensitive method [12]. The classifier integration method is mainly to improve the learning ability of the learner through the method of ensemble learning to reduce the impact of imbalanced data on the classifier. Cost-sensitive learning is to assign different misclassification costs to each class during the training process of the learner [13]. In SER, unbalanced emotional speech samples often exhibit multi-category, small-scale, and high degree of emotional confusion [14]. At present, only a few data imbalance processing methods have been studied for SER.

The problem with SER in small sample environment is that data imbalance always exists in the emotional corpora, which impacts the learning of different emotional categories in the decision space by the classifiers [15]. To solve the problem, an improved synthetic minority over-sampling technique (ISMOTE) is proposed to reduce the impact of data imbalance, in which the decision space of minority class is expanded as much as possible while reducing the influence of the synthetic samples on the decision space of majority class. Emo-DB (German sentiment corpus) and SAVEE (English sentiment corpus) are used in the comparative experiment to validate the effectiveness of our method. The experimental results demonstrate the feasibility of our method.

The remainder of paper is organized as follows. Data imbalance processing in SER is presented in Section 2. Experiments on SER and discussion are given in Section 3.

2. DATA IMBALANCE PROCESSING IN SER

2.1 Framework of the SER

Flowchart of the proposed SER model is shown in Fig. 1. After the originally extracted features are obtained from the pre-processed speech samples based on the low-level emotion descriptor (LLED) [16–18], the rest of the process is composed of data imbalance processing and emotion classification. The degree of data imbalance in the sample can be measured by the imbalance rate, i.e., the ratio of the difference between the majority sample size and minority sample size to the corresponding minority sample size. If the imbalance rate is greater than or equal to 100%, the emotion classifier will affect classification because of the data imbalance. Therefore, it is necessary to determine whether building a classification model.

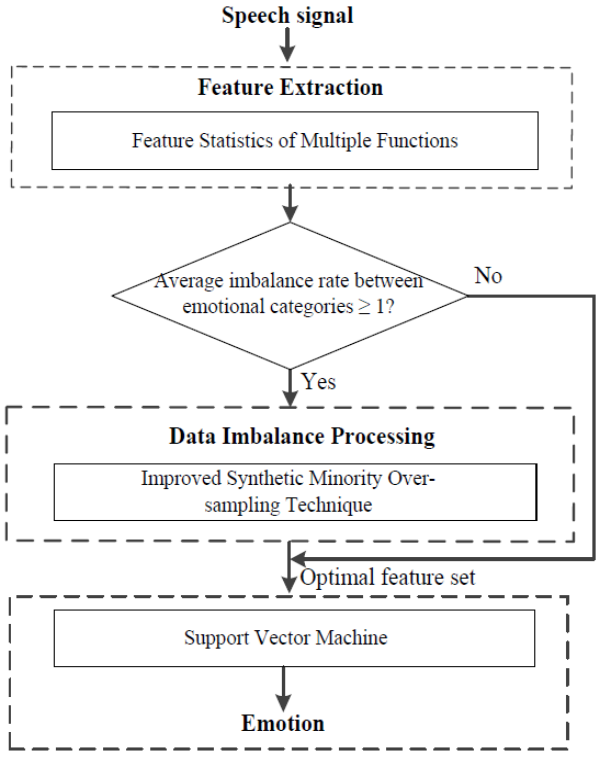


Fig.1 Flowchart of the proposed speech emotion recognition.

In data imbalance processing, an improved synthetic minority over-sampling technique (ISMOTE) is proposed, in which the problem of unbalanced data appearing in emotional classification is solved. Then, Support Vector Machine is adopted to classify the emotion categories such as neutral, happy, sad, surprise and angry.

2.2 Imbalance data processing based on ISMOTE

Emotional corpora that are commonly used to train SER models are standard, but it is difficult to collect data-balanced speech samples in real environments. For imbalanced data sets, samples of minority class are sparsely distributed in sample space compared with the overwhelming amount of majority class, which the recognition process is more difficult. Therefore, it is very important to improve the recognition rate of minority class samples.

Synthetic minority over-sampling technique (SMOTE) is a classical oversampling algorithm that constructs corresponding new samples from minority class information obtained by neighbor relations, which is a scheme based on random oversampling algorithm [19]. The implementation of SMOTE is mainly to find k nearest neighbors by Euclidean distance for each sample x_i in minority classes, and generate a random number between (0,1) to multiply the difference between x and x_i. The new synthesized sample $x_{new} = x + rand(0, 1) * (x_i - x)$.

When SMOTE is applied to emotional speech sample processing, it needs to be improved to overcome these shortcomings. Firstly, it uses all the minority samples in the sampling without considering whether there will be noise data in these samples [20]. Although the sampling space can be expanded to increase the recognition rate of the minority class after completing the sampling process, it will affect the decision space and recognition rate of majority class. Secondly, it is considered as an interpolation method. If the feature dimension of the sample is two, the new sample x_{ij} synthesized by the algorithm is limited to the line x_i is connected to its neighbor point x_{nn}. This interpolation method is limited in the way of extension of minority samples.

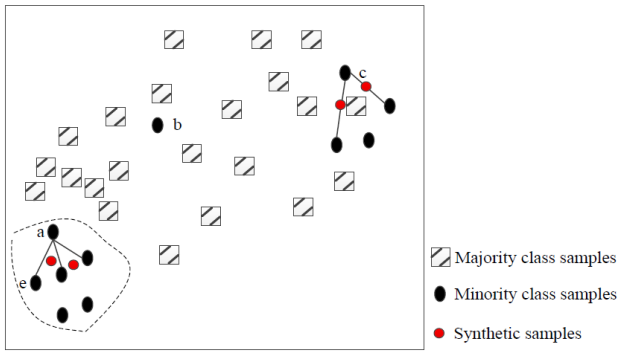


Fig.2 Diagram of the ISMOTE algorithm.

In view of above analysis, an improved synthetic minority over-sampling technique (ISMOTE) is proposed to solve the problem of data imbalance in SER, for which not all the minority samples need to be upsampled, but only the corresponding target points are interpolated. What's more, the interpolation method is different from SMOTE, by

which the decision space of minority class is expanded as much as possible while reducing the influence of the synthetic samples on the decision space of majority class. Fig. 2 is the schematic diagram of ISMOTE algorithm for two-dimensional feature set.

Combined with the schematic diagram in Fig. 2, the ISMOTE can be divided into the following steps.

Step 1: For each speech sample of minority emotional classes x_i, k nearest neighbors in the speech samples of minority emotional class are calculated based on Euclidean distance and the set of neighbors is denoted as S_{i1}. Besides, k nearest neighbors in all speech samples for x_i are obtained and its neighbor set is denoted as S_{i2}.

Step 2: Let $num_i = count(S_{i1} \cap S_{i2})$. If $num_i = 0$, x_i is marked as a noise point, eliminate it without participating in any subsequent sampling operations. If $0 < num_i <= k/2 + t$, t is the regulatory factor, x_i is marked as a target point. If $k/2 + t < num_i <= k$, x_i is marked as a nontarget point, and mark the num_i neighbors of x_i as points in the security domain Q.

Step 3: Count the number of target points n, and determine the sampling magnification of the algorithm N according to the imbalance ratio of the emotional sample and the number of target points.

Step 4: Interpolate all target points to construct a new sample. If the target point x_o belongs to the security domain, i.e., $x_o \in Q$, randomly select two neighbors points $\tilde{x}_1$ and $\tilde{x}_2$ from the set of neighbors of x_o in minority class samples for one interpolation, i.e., $x_{new} = x_o + rand(0, 1)*(\tilde{x}_1 - x_o) + rand(0, 1)*(\tilde{x}_2 - x_o)$, as shown in the dotted line area in Fig. 3. On the other hand, randomly select one neighbor point $\tilde{x}_1$ from the set of neighbors of x_o in minority class samples for one interpolation, i.e., $x_{new} = x_o + rand(0, 1) * (\tilde{x}_1 - xo)$.

As shown in Fig. 2, sample point "a" chooses the nearest neighbor points to interpolate and construct a few new samples. Sample point "b" is directly judged as noise sample points and no longer participates in any postorder interpolation operations. This helps avoid the influence on decision space of most speech samples. In view of the different neighbor distributions of sample points "a" and "c", the interpolation space between sample point "a" and nearest neighbor points is no longer available in two-dimensional speech feature space. Limited to the connection line, the region expands to a triangular region, while the sample point "c" still adopts linear interpolation method, i.e., interpolation in the connection area between the sample point "c" and its nearest neighbors due to the small number of neighbors.

3. EXPERIMENTS

The experiment was designed with the following steps. First of all, openSMILE toolkit and MATLAB R2012b were used to extract of speech emotional features, in which multidimensional features are extracted separately. And the pro-posed ISMOTE was carried out in data imbalance processing, in which the unbalanced emotional data reaches equilibrium. Then, Support Vector Machine (SVM) [21] was adopted for speech emotion classification an a Radial Basis Function (RBF) was used as kernel function, in which penalty coefficient C and kernel parameter gamma are obtained based on grid search.
Two sets of experiments rely on the extracted speech emotional features were conducted respectively. Firstly, using the same initial feature set, the experiments by different data imbalance processing methods were performed to verify the effectiveness of ISMOTE on Emo-DB and SAVEE. Then, speaker-dependent SER on SAVEE and Emo-DB databases was carried out, in which our method is compared with some state-of-the-arts works.

3.1 Speech database

3.1.1 Surrey Audio-Visual Expressed Emotion (SAVEE) Database

It consists of 480 short English utterances recorded by four speakers in seven basic emotions (i.e., angry, fear, dis-gust, surprise, happy, sad, and neutral), in which the speech samples are picked from the standard TIMIT corpus and each emotion is phonetically-balanced [23].

3.1.2 Berlin Database of Emotional Speech (Emo-DB)

It is a German emotional speech database recorded by the Technical University of Berlin, by the 10 actors (5 males and 5 females) of 10 statements (5 long 5 short) of seven emotions (i.e., happy, angry, anxious, fearful, bored, disgusted, and neutral) simulation, contains a total of 800 sentence corpus, sampling rate of 48kHz (16kHz, 16bit after compression) [22]. The speech recorded in a professional studio, requirements in the interpretation of a particular emotional actor before through the memories of their true experience or experience of mood brewing, to enhance the sense of reality of emotions.

3.2 Emotional data imbalance processing

The ISMOTE was tested using standard feature set IN-TERSPECH 2010 [18] on the Emo-DB and SAVEE database because 2-5 times imbalance in the data of each emotional category exist, in which five sets of experiments are per-formed based on different kinds of data imbalance processing method in total (i.e., None, Subsampling, Random Oversampling, SMOTE, and

ISMOTE). The original sample feature set and the feature set processed by different kinds of data imbalance processing method are used in SD SER experiment respectively. SVM is used for emotion classification.

All samples of each individual are used, in which 70%samples are randomly used for training and the remaining 30% samples were used for testing. The experiment is di-vided into two groups based on different databases. The first group, i.e., 535 emotional speech samples in Emo-DB were randomly divided into training set and testing set in proportion (7:3), in which the emotional samples in training set are divided into 90 “anger” samples, 58 “boredom” samples, 30 “disgust” samples, 56 “anxiety” samples, 56 “happiness” samples, 36 “sadness” samples, and 36 “neutral” samples. After these unbalanced training samples were processed in different ways and training the SVM classifier, 161 testing sets were used to test the classifier. The second group, i.e, 480 emotional speech samples from SAVEE were randomly divided into 336 speech samples in training set and 144 speech samples in testing set, in which the training set consists of 43 “anger” samples, 44 “disgust” samples, 38 “fear” samples, 42 “happiness” samples, 79 “neutral” samples, 48 “sadness” samples, and 42 “surprise” samples. Table 1 shows the contrast results in the initial samples and the samples after using ISMOTE on Emo-DB. Table 2 gives the contrast results in the initial samples and the samples after using ISMOTE on SAVEE. As shown in Table 1, through the unbalanced data processing, the accuracy and recall rate of the category are 0.76 and 0.95, respectively. The former represents 76% of all the samples identified by the learner as the category, 24% is actually other categories; the latter shows that 95% of the test samples in this category are correctly classified, and 5% of the samples are misclassified into other emotional categories.

Table 1. Contrast results in the initial samples and the samples after using ISMOTE on Emo-DB.

Category	None			ISMOTE			Number
	Precision	Rec-all	F1	Precision	Recall	F1	
Anger	0.76	0.95	0.84	0.76	0.95	0.84	37
Boredom	0.70	1.00	0.82	0.77	1.00	0.87	23
Disgust	0.92	0.75	0.83	1.00	0.75	0.86	16
Anxiety	0.64	0.69	0.67	0.75	0.69	0.72	13
Happiness	0.87	0.52	0.65	0.94	0.64	0.76	25
Sadness	0.86	0.83	0.84	0.84	0.91	0.87	23
Neutral	0.83	0.62	0.71	0.89	0.71	0.79	24
Avg/Total	0.80	0.78	0.78	0.84	0.83	0.82	161

Table 2. Contrast results in the initial samples and the samples after using ISMOTE on SAVEE.

Category	None			ISMOTE			Number
	Precision	Rec-all	F1	Precision	Recall	F1	
Anger	0.50	0.65	0.56	0.57	0.71	0.63	17
Boredom	1.00	0.31	0.48	1.00	0.38	0.55	16
Disgust	0.82	0.41	0.55	0.81	0.59	0.68	22
Anxiety	0.50	0.50	0.50	0.71	0.56	0.63	18
Happiness	0.74	0.98	0.84	0.82	0.98	0.89	41
Sadness	0.69	0.75	0.72	0.60	0.75	0.67	12
Neutral	0.62	0.72	0.67	0.65	0.83	0.73	16
Avg/Total	0.70	0.67	0.65	0.76	0.73	0.72	144

The Emo-DB’s 374-sentence training set samples were processed in five different ways for unbalanced emotional speech. After using up-sampling method, the training set samples included 210 sentences. The training set samples were expanded to 630 sentences after other over-sampling methods and the speech samples was balanced among emotional classes. In the same way, SAVEE’s 336 speech samples were processed by different methods. The training set samples were reduced to 266 sentences after using up-sampling method and the training set samples were expanded to 553 sentences after other over-sampling methods.

The classification model was trained using the processed data, and the test was performed using the same testing set. Contrast results using different data imbalance methods on Emo-DB and SAVEE are shown in Table 3.

Table 3. Contrast results using different imbalance processing methods on Emo-DB and SAVEE

Database	Recognition Rate(%)				
	None	Subsampling	Random Over-sampling	SMOTE	ISMOTE
Emo-DB	78.26%	74.53%	81.75%	81.99%	82.61%
SAVEE	66.67%	57.64%	70.14%	72.22%	72.92%

Table 1 shows the emotion recognition results when no data imbalance is processed including the accuracy of each emotion category in the two sets of data. Both the precision and the recall are greatly offset, and the corresponding F1 value is lower, e.g., the precision of “happy” category on Emo-DB is 0.87, which represents 87% of the results are correctly classified, and the recall rate is only 0.52, which means that the classifier only

classifies 52% of the test samples of the category correctly, which results in an overall F1 value is only 0.62. This shows that the data imbalance between the categories of speech data extremely affects the learning of sentiment classifiers. Excessive attention to most types of speech samples leads to higher recall rates and relatively lower accuracy in most emotional categories, such as “angry” and “neutral” in Table 1, while the under-learning of a few categories led to a lower recall rate and a higher ac-curacy rate, which affects the overall emotional recognition accuracy.

As shown in Table 1 and Table 2, the precision and the re-call rate of each emotional category is lower than that of the data unbalanced processing, and the F1 values of each category are given, from which the data imbalance processing method improves recognition results obviously. The degree of improvement indicates that data for different emotion categories in the training set is balanced and supplemented, and the learner’s degree of over-learning for most types of emotion categories and the degree of under-learning for a few classes are greatly reduced.

At the same time, the recognition rate of the learning model by different data imbalance processing methods in Table 3 demonstrates the effectiveness of the ISMOTE algorithm in the unbalanced emotional speech processing com-pared with other methods. The emotional data imbalance processing method can extend the decision space of a few sentiment categories to achieve the inter-class balance while reducing the influence of the synthesis of minority speech samples on the decision space of most emotional classes.

4. CONCLUSION

In this paper, a new SER method based on ISMOTE algorithm was put forward. The effectiveness of the proposal was validated in multiple contrast experiments with different experimental conditions. The ISMOTE was demonstrated to be more suitable for solving data imbalance in speech emotional recognition than the traditional SMOTE.

In future work, this model will be further optimized. For example,feature selection can eliminate emotion-independent features in the initial feature set and minimize the number of emotional redundant features, therefore adding feature selection can further improve the accuracy of recognition.

References

[1] R. A. Calix, S. A. Mallepudi, B. Chen, et al., Emotion Recognition in Text for 3-D Facial Expression Rendering, IEEE Transactions on Multimedia, 12(6): 544-551, 2010.

[2] F. Tao, G. Liu, Q. Zhao, An ensemble framework of voicebased emotion recognition system for films and TV programs, IEEE International Conference on Acoustics, Speech and Signal Processing, 2018.

[3] Y. Wang, L. Guan, A. N. Venetsanopoulos, Kernel Cross-Modal Factor Analysis for Information Fusion With Application to Bimodal Emotion Recognition, IEEE Transactions on Multimedia, 14(3): 597-607, 2012.

[4] El. M. Ayadi, M. S. Kamel, F. Karray, Survey on speech emotion recognition: features, classification schemes, and databases, Pattern Recognition, 44(3): 572-587, 2011.

[5] J. Rybka, A. Janicki, I. Giannoukos, Comparison of Speaker Dependent and Speaker Independent Emotion Recognition, International Journal of Applied Mathematics and Computer Science, 23(4): 797-808, 2013.

[6] Z. T. Liu, F. F. Pan, M. Wu, A multimodal emotional communication based humans-robots interaction system, The 35th Chinese Control Conference, 6363-6368, 2016.

[7] C. N. Anagnostopoulos, T. Iliou, I. Giannoukos, Features and classifiers for emotion recognition from speech: a survey from 2000 to 2011, Artificial Intelligence Review, 43(2): 155-177, 2015.

[8] W. W. Ng, J. Hu, D. S. Yeung, et al., Diversified Sensitivity-Based Undersampling for Imbalance Classification Problems, IEEE Transactions on Cybernetics, 45(11): 2402-2412, 2017.

[9] P. D. Gutierrez, M. Lastra, J. M. Benitez, et al., SMOTE-GPU: Big Data preprocessing on commodity hardware for imbalanced classification, Progress in Artificial Intelligence, 6(4):347- 354, 2017.

[10] A. Anand, G. Pugalenthi, G. B, et al., An approach for classification of highly imbalanced data using weighting and under-sampling. Amino Acids, 39(5):1385-1391, 2010.

[11] M. Buda, A. Maki, Mazurowski, et al., A systematic study of the class imbalance problem in convolutional neural networks, Neural Networks, 106: 249-259, 2018.

[12] C. Bianchi, Nicolo, Re, et al., Synergy of multi-label hierarchical ensembles, data fusion, and cost-sensitive methods for gene functional inference, Machine Learning, 88(1-2):209-241, 2012.

[13] M. Liu, C. Xu, Y. Luo, et al., Cost-Sensitive Feature Selection by Optimizing F-Measures, IEEE Transactions on Image Processing, PP(99):1-1, 2017.

[14] L. Chen, S. K. Zheng, Speech Emotion Recognition: Features and Classification Models, Digital Signal Processing, 22(6): 1154-1160, 2012.

[15] O. Loyola-Gonzalez, J. F. Martinez-Trinidad, J. A. Carrasco-Ochoa, et al., Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases, Neurocomputing, 175: 935-957, 2016.

[16] F. Eyben, M. Wollmer, and B. Schuller, openEAR: Introducing the munich open-source emotion and affect recognition toolkit, International Conference on Affective Computing & Intelligent Interaction & Workshops. IEEE, 576-581, 2009.

[17] B. W. Schuller, S. Steidl, A. Batliner, The INTERSPEECH 2009 Emotion Challenge, INTERSPEECH, 312-315, 2009.

[18] B. Schuller, S. Steidl, A. Batliner, et al., The INTERSPEECH 2010 paralinguistic challenge, INTERSPEECH, 2794-2797, 2010.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, et al., SMOTE: synthetic minority over-sampling technique, Journal of Artificial Intelligence Research, 16(1): 321-357, 2012.

[20] H. Han, W. Y. Wang, B. H. Mao, Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning, International Conference on Advances in Intelligent Computing, 2005.

[21] C. C. Chang, C. J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol,2: 1-27, 2011.

[22] F. Burkhardt, A. Paeschke, M. Rolfes, W. Sendlmeier and B. Weiss, A database of german emotional speech, Proc. of Inter-speech, 1517-1520, 2005.

[23] S. Haq, P. J. B. Jackson, and J. D. Edge, Audio-Visual feature selection and reduction for emotion classification, Proc. International Conf. on Auditory-Visual Speech Processing, 185-190, 2008.

Feasibility Architecture for Processing Multimodal Signal for a Robot Control System

Motohiro Akikawa*, Masayuki Yamamura*
* Department of Computer Science, School of Computing, Tokyo Institute of Technology
Yokohama, Kanagawa 226-8502, Japan
m_akikawa@ali.c.titech.ac.jp, my@c.titech.ac.jp
+81-45-924-5211

Abstract

In recent years, many systems embedding deep learning in robots have been developed. Some use multimodal information to achieve higher accuracy. In this paper, we point out three problems of such systems, high cost, robustness, and system optimization. First, the optimization of huge architectures using real environments is computationally expensive, so it is not easy to develop such architectures. Second, in a real environment, noise such as changes in lighting is often contained in inputs, so the architecture should be robust against noise. Finally, a system composed of individually optimized modules may find it difficult to coordinate them, so the system is better optimized as one architecture. To address these problems, a simple and highly robust architecture named Memorizing and Associating Converted Multimodal Signal Architecture (MACMSA) is proposed. Verification experiments are conducted and the potential of the proposed architecture is also discussed. The results of the experiments show that MACMSA decreases the effects of noise and obtains much better robustness than a simple autoencoder. MACMSA takes us a step closer to building robots that can truly interact with humans.

Keywords: Deep-learning architecture, Robot control, Multimodal input, Feasibility architecture, Robust architecture.

1. Introduction

The purpose of this paper is proposing a feasibility architecture for a robot, such as humanoid robots, control system. As robotics has progressed, robots that have the ability to interact with human have been developed. ASIMO [1]and Pepper [2] are examples of such robots. However, these robots interact with humans only under limited conditions. There is no robot that interacts with humans perfectly under any conditions.

Traditional mainstream controlling systems for robots consist of classical programs that are adopted and extended by programmers. However, these systems are not always able to adapt to well-discussed problems such as the frame problem nor can they take complex actions [3].

In recent years, system design using machine learning such as deep-learning has become mainstream [4-6]. Using deep-learning and machine learning techniques, highly accurate recognition has become possible in computer vision [5-12]. Given the achievements in computer vision, these techniques have also been adapted to robot systems [13-16]. In addition, there are some systems that use multimodal information, e.g., a combination of picture and audio, as input to obtain higher accuracy [16-19]. However, the studies that use deep learning with multimodal information as input have three big challenges. The first is the increasing cost of optimizing networks. This cost is not only a computational cost but also a cost to design an architecture. If multimodal information is simply used as input to one neural network, bloated network structures cause computational costs to increase [6, 7, 16]. On some systems, restricted network structures are used, but it is not obvious how networks should be restricted [15]. Thus, it is a burden on the system designer to restrict the network structure. In addition, it is important that how networks connect to the other networks when an architecture consisted of some types of networks are designed. Ramachandram and Taylor [5] mentioned "architecture design has been more an art than a science." because the designers have to pay so much attentions to the structure. However, the architecture design should be a science because, without reasons, the architecture is no able to be expanded and adapted to other tasks. The second is robustness. There is a system that processes time-series data as input by treating several frames of information as one input [17]. However, the system is unstable at the beginning of a motion or until all inputs include environmental changes such as lighting changes that should be treated as the noise. The methods in the state of the art only use the function approximating of neural network and outputs based on the features. These methods solved the problem to earn high accuracy for the specified tasks in experiments. However, it is more important for communicating with humans that the researches pay attentions to the way to reduce the noise in real environment and adapt unknown situations. The third is system optimization. On some system, deep learning is

used only for object recognition and the dedicated software for a specific robot is used to control it [18]. If sensor and drive modules are optimized individually and either modules' specifications are changed, the system immediately becomes unstable. It is also obvious that if sensor modules are optimized but drive modules are not, the robots may not take the action that the programmer expected. Therefore, it would be best for a robotics system to be optimized as one architecture that includes all modules.

In this research, considering three problems above (cost, robustness, and system optimization), the Memorizing and Associating Converted Multimodal Signal Architecture (MACMSA) is proposed. The proposed architecture is new in terms of the combination of components. The experiments are conducted to focus on investigating the performances and the limits of the proposed architecture. The validity of MACMSA is also discussed with respect to the three points above when the architecture is used on robots.

2. Materials and Methods

2.1 Outline of the Architecture

The architecture is composed of three modules (Figure 1). The associator, that is most important module, is a memory device that behaves as an associative memory. By employing a sparse Hopfield network [19] as the associator, the robots store and recall strongly denoised signals. The examples of previous works related the implemented associative memory on robots are [20-21]. The encoders are address translators that convert a real number, that is an input from sensors, to a binary value. The reason why the encoders output a binary value is that Hopfield network obtains a binary value as an input and processes a binary value. The decoders are also address translators that convert a binary value, which is an output from Hopfield network, into a real number. The reason why the encoders output a real number is that the inputs for many types of sensors are real numbers. The encoder and the decoder are independently implemented as three-layer feedforward neural network for each mode of information. In this study, it is assumed that three modes of information, visual information, audio information that has been transformed using a Fourier transform, and actuator states including angle, angular velocity, and angular acceleration are fused.

2.2 Optimization Procedure

MACMSA is optimized by the following three steps. First, the encoder and decoder are optimized. Second, all training data are encoded into the binary patterns by the optimized encoders. Third, the associator stores all the encoded training data. The details of each step are below.

Step 1 is optimization of the encoder and decoder. The encoder and decoder of each mode of information are optimized as a sparse autoencoder [22]. There are many systems implemented autoencoders in robotics [16][17][23]. It is typical for only half of the network structure to be used to detect the features of the data or compress the dimensions of the data after optimizing the autoencoder [24]. In this study, half of the network structure is used as an encoder and the other half of the network structure is used as the decoder. The output layer of the encoder and the input layer of the decoder share the same layer. Thus, the encoder and decoder can be optimized as five-layer autoencoder (Figure 2). All layers use the sigmoid function as the activation function and have the same value of gain except for the output layer of the encoder. It is necessary for the output of the encoder to approximate a binary value because the Hopfield network, which is used as the associator, processes binary values. Backpropagation with stochastic gradient descent optimizes the autoencoder. The cost function used to optimize the sparse autoencoder which is given by Equation 1 is the squared error with a sparse regularization term, which is Kullback–Leibler divergence (KL), added [25]. $E(w)$ is the cost function for weights, w. n is an index of data. y_n is nth output from the autoencoder. t_n is an answer signal for nth input. β is the hyperparameter to control the effect of the sparse regularization term on the cost function. j is the index of neuron composing the network. ρ is the target value for the sparseness and $\hat{\rho}$ is the actual measured value.

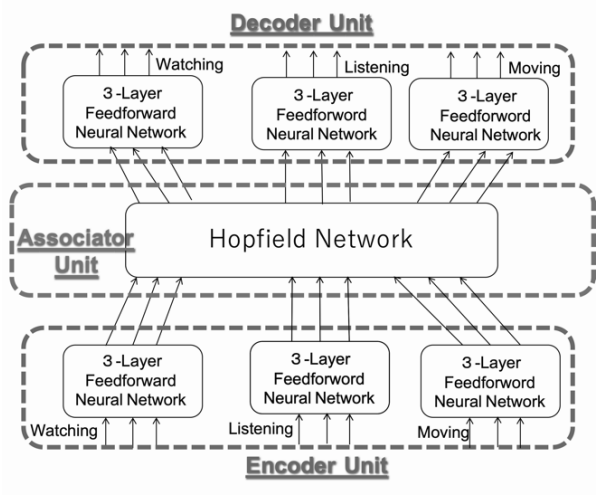


Figure 1: The architecture is composed of three modules, the encoders, the associator and the decoders. Each mode information is input to each encoder. Then, the encoders output information converted to binary. The outputs from the encoders are concatenated to one vector to be the input for associator unit. The vector obtained as the output from associator are divided to be the inputs to decoders. The system outputs each mode information again at each decoder.

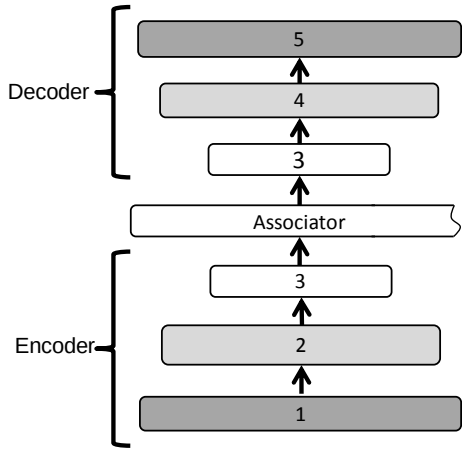


Figure 2: One autoencoder are divided at layer 3 to be encoder and decoder. At output layer of the encoder, a part of vector to be an input of associator is obtained as an output from the encoder. The input layer of the decoder obtains a part of vector as an input.

$$E(w) = \sum_n \|\boldsymbol{y}_n - \boldsymbol{t}_n\|^2 + \beta \sum_j KL(\rho||\widehat{\rho_j}) \quad (1)$$

As a result, the derivative form for updating the weights becomes Equation 2 [25].

$$\frac{\partial E}{\partial u_j^{(l)}} = \left\{ \sum_k \frac{\partial E}{\partial u_k^{(l+1)}} w_{kj}^{(l+1)} + \beta \left(-\frac{\rho}{\widehat{\rho_j}} + \frac{1-\rho}{1-\widehat{\rho_j}} \right) \right\} f'\left(u_j^{(l)}\right) \quad (2)$$

j and k are an index of neuron. Thus, u_k is the error from kth neuron and u_j is also the error from jth neuron. l is the index of layer. f' is the derivative of the activation function.

As previously noted, the gain at output layer of the encoder is increased. To increase the value of the gain, the optimization and setting the gain are repeated followed the flowchart indicated Figure 3.

Step 2 is generation of the encoded patterns for associator. The encoders optimized in step 1 are given all the training data as input and output the encoded patterns. These patterns are approximately binary but are still real numbers at this point. Therefore, by the operation indicated Figure 4, all patterns stored by associator are replaced to binary pattern. The system adopts this replacement even when the system is operating.

Step 3 is storing the patterns on associator. Each weight for Hopfield network is determined by Equation 3 as J_{ij} [19]. Here, M is the number of stored patterns,

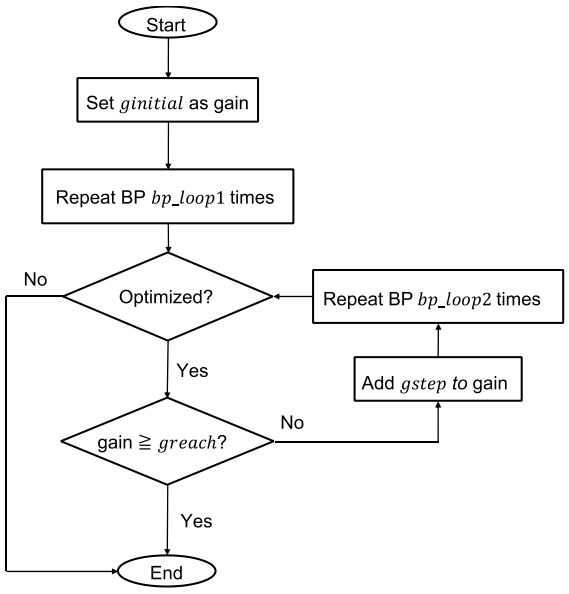


Figure 3: First, the encoders adopt a low initial gain ($ginitial$) and update the weight for bp_loop1 iterations using backpropagation. Then, the encoders adopt a new gain by adding $gstep$ to the old gain and update the weights for bp_loop2 iterations. Hereafter, the encoders increase the gain by $gstep$ every bp_loop2 iterations and update weight until the gain reach $greach$.

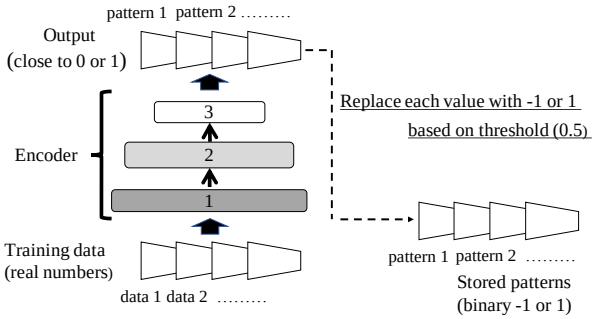


Figure 4: Firstly, all training data are encoded by encoder. Then, all outputs are replaced with 0s or 1s using 0.5 as the threshold. In addition, 0 is replaced with −1 as a matter of programming convenience. This patterns with -1 or 1 are stored at associator.

and μ is an index of stored patterns. a is an average pattern. ζ_i is a signal for ith neuron and obtained from the output layers of the encoders. Note that these i and j are the index of neuron for Hopfield network and this j is different from j for the autoencoder in step 1.

$$J_{ij} = \frac{1}{M} \sum_{\mu=1}^{M} (\zeta_i^\mu - a)(\zeta_j^\mu - a) \quad (3)$$

For the sparse Hopfield network, the recall function is given by Equation 4 and Equation 5 [19].

$$x_i^{t+1} = F\left(\sum_{j\neq i}^{N} J_{ij}\, x_j^t + h\right) \quad (4)$$

$$F(u) = sgn(u) - b \quad (5)$$

In Equation 4, N is the number of neurons. x_j^t is the state of each neuron when the time is t. h describes the shift, and the signal is maximized when $h = a(1-a^2)$. In Equation 5, b is the bias and cross-talk noise is minimized when $b = a$. When the system works, Hopfield network obtains the input as initial state from the encoders and output a pattern as steady state by following Equation 4. In addition, the theoretical storage capacity is calculated by Equation 6 and approximated using signal-to-noise (S/N) analysis [19].

$$\alpha_C(a) = \frac{\alpha_C(0)}{(1-a^2)} \quad (6)$$

Here, $\boldsymbol{\alpha_C(0) = 0.138N}$ and $\boldsymbol{N}$ is the number of neurons.

2.3 Experiments conditions

Three experiments were conducted, the verification of concept, the observation of performances and limits and a comparison with a simple autoencoder. In the verification of concept, four experiments with pseudo data were conducted to show MACMSA performs in accordance with the concept. In the observation of performances and limits, three experiments were conducted, in which the sizes of the encoder, associator, and decoder directly affecting the storage capacity were varied, to indicate the relationship between accuracy and storage capacity. In a comparison with a simple autoencoder, a comparison with the simple autoencoder which processes three modalities concurrently was conducted.

2.3.1The verification of concept

Experiment 1-1: This test demonstrates the optimization of the three autoencoders. In this experiment, to prove that the proposed architecture is able to treat different dimensions data, the input layer size of the largest encoder and the input layer size of the smallest encoder were set to 1,000 and 100, respectively. One function of decoders is the dimension reduction. Thus, the size of hidden layers may be smaller than the size of input layers. However, the sum of neurons in all hidden layer 3s is the number of neurons in Hopfield network and the number of neurons directly affects the storage capacity of Hopfield network. Therefore, not to consider the storage capacity, the size of hidden layer 3s should not be too small in this experiment. Considering these reasons, the number of neurons in hidden layers 3s was half of input layers and the number of neurons in hidden layers 2s was three quarter of input layers. The input layer size of the middle size network was set to 500 which is almost intermediate size. To indicate that the encoders and the decoders have the ability to process data with various averages and variances, the training data are pseudo data composed of a combination of three averages, 0.25, 0.5, 0.75, and three standard deviations, 0.25, 0.28, 0.31. Hence, 729 combinations are possible, all of which were tested. Three average and three standard deviations reproduces the case that each encoder obtains inputs from different type of sensor which outputs data having different averages and standard deviations. Ten pseudo data were generated from normally distributed random numbers. Note that the range of possible values was [0.0,1.0]. Not to consider the storage capacity of Hopfield network, the number of pseudo data should be small because the number of training data is the number of patterns to be stored. For each combination, 10 trials were tested with different seeds. Thus, 7,920 conditions were tested. Table 2 shows the number of neurons in the autoencoder. All the weighs of the three autoencoders were initialized with normally distributed random number with an average of 0.0 and a standard deviation of 0.3. $ginitial$ was 1.0 to make the activation function the standard sigmoid function. $greach$ was 10 so that the encoders output approximately binary values. Setting high $gstep$ causes an optimization failure because the network structure is changed dramatically. bp_loop1 and bp_loop2 should be set with α, learning rare, to optimize the autoencoder properly. Sparseness of patterns for associator is able to be controlled by ρ. If β is big, such as 3, the autoencoder extracts the input data directly as feature. However, if β is too small, such as 0.001, the autoencoder expresses all data by one feature [46]. Thus, β should be set carefully. Considering reasons, $gstep$ bp_loop1, bp_loop2, α, ρ, and β were fine-tuned by supplementary experiments. Table 3 shows the hyperparameters used to optimize the three autoencoders.

Experiment 1-2: The three autoencoders optimized in Experiment 1-1 were given input with noise and the amount of error in the output was measured. The input consisted of training data with noise levels of 0% to 50% in steps of 2%. The noise level of 50% indicates half of dimensions of all training data is changed and it is hard to assume that half of input data are changed in the situation exactly same. Thus, in this experiment, the maximum noise level was set to 50%. In this experiment, it should be observed that how outputs will be changed against tiny changes of inputs. Therefore, 2% was set as the small change. The noise was added by replacing a portion of the training data with a uniform random number from the range $[-1.0,1.0]$ for each mode of information. The reason why the minimum value of noise was -1.0 was that some types of sensors work as analog devices and they should output the values outside the defined range as noises. It seems that the probability distribution of data and the probability distribution of noises are different in real environment. Therefore, the noises were generated with a uniform random number. If the noise level of 25% is set, all training data are set to baseline and 25% of all

Table 1: Number of neurons in each autoencoder and loop values.

Autoencoder structure	Input layer	Hidden layer 1	Hidden layer 2	Hidden layer 3	Output layer	bp_loop1	bp_loop2
Structure 1	1,000	750	500	750	1000	15,000	7,500
Structure 2	500	375	250	375	250	10,000	5,000
Structure 3	100	75	50	75	100	7,500	3,750

Table 2: Hyperparameters and their settings

$ginitial$	$greach$	$gstep$	α	ρ	β
1.0	10.0	0.5	0.001	0.4	0.2

training data of all modality are replaced to uniform random numbers. For instance, for the first modality, 250 dimensions which is 25% of training data, are chosen at random for each training data, and then replaced to uniformly distributed random number. Slimily, for second modality, 125 dimensions are chosen at random for each training data, and replaced to uniformly distributed random numbers. For third modality, 25 dimensions are chosen at random for each training data, and replaced to a uniformly distributed random number. When the noise is 0%, the input is exactly same as the training data. Each combination of training data and noise rates were tested 10 times. Thus, 72,900 conditions were tested in one noise level and 1,895,400 conditions was tested in total because there are 26 steps noise level.

Experiment 1-3: To test robustness, the encoders and associator were given input with noise and how much the associator can recall correctly was measured. The autoencoders optimized in Experiment 1-1 were used as the encoders. The associator was initialized using the patterns generated by the methods in steps 2 and 3 in Section 2.2. The encoders were given the input with noise levels ranging from 0% to 50% in intervals of 2%, just as in experiment 1-2. This test was also performed 10 times for all combinations and with all noise rate. Thus, 1,895,400 conditions were tested in total.

Experiment 1-4: To test how much the system was robust against noise, the whole system was tested in this experiment. The encoders optimized in Experiment1-1 encoded the input with the noise just as in the other experiments. The associator was given the pattern from the encoders then recalled the output pattern. The decoders decoded the pattern to the original formatted data. If the system operates properly, even if the input contains noise, the output from system should be close to the original training data. Just as in the other experiments, the system obtained the input with noise levels from 0% to 50% in increments of 2% and was tested 10 times for each noise level and all training data. Thus, 1,895,400 conditions were tested in total just as in the other experiments.

Figure 5 shows the networks focused on each experiment and what to measure.

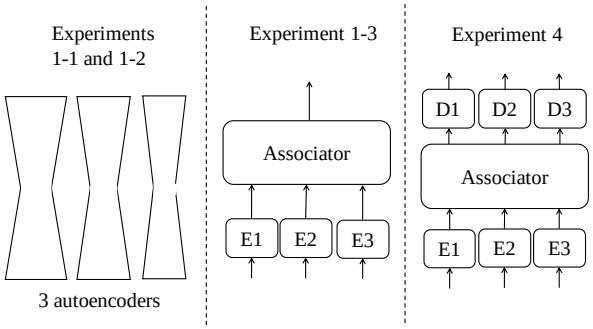


Figure 5: Experiment 1-1 evaluates the performance of three autoencoders for the optimization of the encoders and decoders. Experiment1-2 measures the error of the three autoencoders. Experiment1-3 tests the associator with input from the encoder. Experiment1-4 measures the error of the system.

2.3.2 The observation of performances and limits

Three experiments that correspond to Experiments 1-2, 1-3, and 1-4 in the verification of concept were conducted. In these experiments, four architecture sizes were tested for a fixed number of training data of 10. It seems that 10 data may be few to be training data but the smallest architecture which is the minimum size of proposed architecture is able to store up to only fourteen patterns when patterns are not sparse patterns because of the storage capacity. Experiment 2-1 was an operating experiment with simple three autoencoders, Experiment 2-2 was an associator experiment with the encoders and associator, and Experiment 2-3 was an operating experiment with the total system. Table 3 shows the number of neurons for each architecture as well as the values for bp_loop1 and bp_loop2. The number of neurons was determined by considering the storage capacity of the Hopfield network and the size of autoencoder that can be optimized. Note that autoencoders smaller than structure 3 of architecture Arch 1 could not be optimized by the proposed method. The number of training data was fixed and the number of neurons for the Hopfield network was decreased to control the loading rate, which is the amount of storage used for patterns with respect to the storage capacity. Three average, 0.25, 0.5,0.75, and one standard deviation, 0.28, were used to generate the training pseudo data. Unlike Experiment 1, a combination of averages was not considered, so training data with the same averages were used for the optimizations of structures 1, 2, and 3. For all

Table 3: the components of each systems

Archit-ecture	Autoencoder Structure	Input layer	Hidden layer 1	Hidden layer 2	Hidden layer 3	Output layer	bp_loop1	bp_loop2
Arch 1	Structure 1	100	75	50	75	100	7,500	3,750
	Structure 2	70	53	35	53	70	7,500	3,750
	Structure 3	40	30	20	30	40	7,500	3,750
Arch 2	Structure 1	180	135	90	135	180	7,500	3,750
	Structure 2	120	90	60	90	120	7,500	3,750
	Structure 3	60	45	30	45	60	7,500	3,750
Arch 3	Structure1	360	270	180	270	360	7,500	3,750
	Strucure2	240	180	120	180	240	7,500	3,750
	Structure3	120	90	60	90	120	7,500	3,750
Arch 4	Structure1	1,000	750	500	750	1,000	15,000	7,500
	Strucure2	500	375	250	375	250	10,000	5,000
	Structure3	100	75	50	75	100	7,500	3,750

average values of the training data, 10 trials were performed with different seeds. Thus, 900 conditions were tested in one noise level and 23,400 conditions was tested in total. The hyperparameters (Table 2) and other conditions were the same as in Experiment 1.

2.3.3 A comparison with a simple autoencoder

The proposed architecture which had the same size of neurons as the verification of concept (Table 2) was compared with a simple autoencoder which processed three modalities concurrently. It is assumed that the three sensors output data which has different averages. An average of the mode 1 was 0.25. an average of the mode 2 was 0.5. an average of the mode 3 was 0.75. All modes of the standard deviations were 0.28. For proposed architecture, the encoder 1 was obtained the mode 1, the encoder 2 was obtained the mode 2 and the encoder 3 was obtained the mode 3. The number of training data was 10. All hyperparameters to optimize the proposed architecture was the same as the hyperparameters in the verification of concept (Table 3). The five layers autoencoder was obtained the simply concatenated three modal information as inputs (Figure 6). The number of neurons in input layer for the autoencoder was 1,600, which was the sum of neurons in input layers in the encoders for the proposed architecture. Similarly, the number of neurons in each layer had the sum of neurons in each layer of the encoders and the decoders for the proposed architecture. Thus, the layer 2 had 1,200 neurons and the layer 3 had 800 neurons. The total number of neurons for the autoencoder were the same as the total number of neurons in the encoders and the decoders for the proposed architecture. The first parts of input layer, which were 1,000 neurons, obtained the mode 1 as inputs. The second parts of input layer, which were 500 neurons, obtained the mode 2. The third parts of input layer, which were 100 neurons, obtained the mode 3. The loss function was simply the sum of squares. The number of backpropagation iterations with SGD was

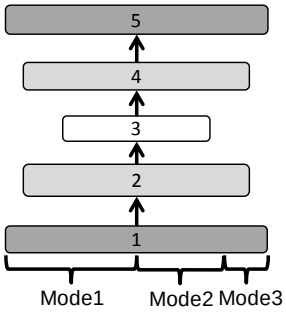


Figure 6: The simple autoencoder consisted of five layers. The autoencoder obtained three modality inputs. The average of mode 1 was 0.25. the average of mode 2 was 0.5. The average of 3 was 0.75. The total number of neurons was the same as the total number of neurons of the encoder and the decoder in the proposed architecture.

75,000 to converge completely the loss function. The gain for sigmoid function was 1.0 to make the activate function the standard sigmoid function. The learning rate α was 0.001, which was the same as α for the proposed architecture. Both the propose architecture and the autoencoder were tested 10 trials with different seeds. 10 times were tested in one noise level for both. Thus, 2,600 times were tested in total.

3. Results

3.1 The verification of concept

Experiment 1-1: The average of evaluation of all autoencoders was shown to converge (Figure 6). The evaluation value degraded periodically because the network was slightly modified by the gain resetting.

Experiment 1-2: As the network structure increases, the error at noise level 0% increased but quite low (Figure 8). At the noise level of 2%, the error for structure 1 was

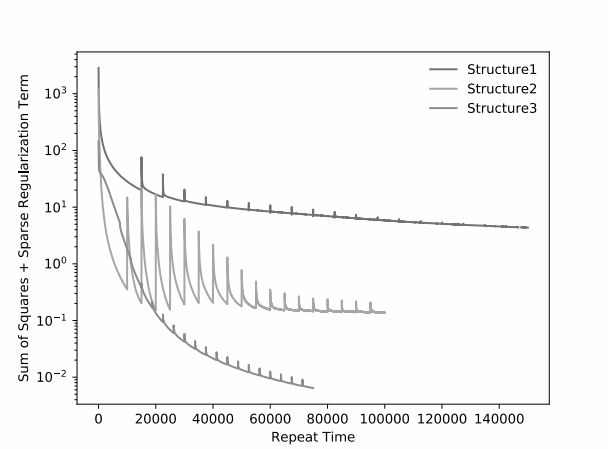


Figure 7: It shows the result of optimizing of each autoencoder. The X-axis is the number of iterations of backpropagation, and the Y-axis is the value of sum of squares + sparse regularization term (Eq. 1).

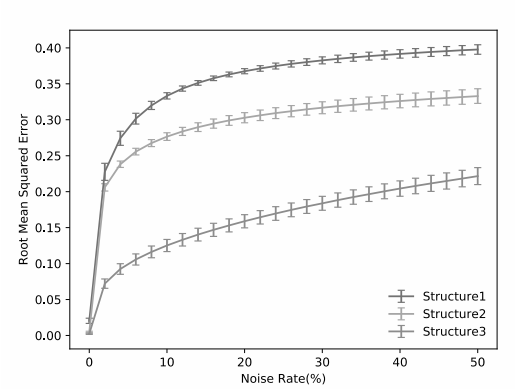


Figure 8: It shows the error on the output from the autoencoders when the input contains noise. The X-axis is the noise rate, the Y-axis is the root mean squared error, and the error bars indicate the standard deviation.

about 0.23, the error for structure 2 was about 0.2, and the error for structure 3 was about 0.07. The error increased sharply up at noise levels of 2%, considering quite low errors at noise level 0%. A tendency that the error increased gradually after the error increased sharply up were regardless of the structure size.

Experiment 1-3: When the noise rates were from 0% to 12%, the accuracy was 100%. It means that the correct pattern was recalled (Figure 9). Thereafter, the percentage of correct outputs gradually decreased. As the noise level increased and the percentage of correct outputs decreased, the standard deviation increased.

If the correct pattern was not recalled, an incorrect pattern was recalled, or the system was in an unstable state. When the level of noise was low and the accuracy was not 100%, it is assumed that the input pattern was not sufficiently close to recall the correct pattern, but a pattern that was relatively similar to the correct pattern

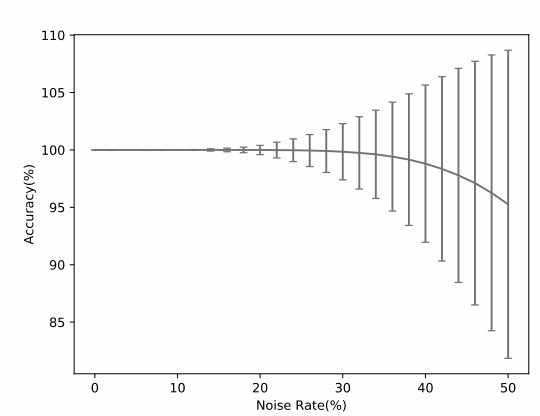


Figure. 9: It show the percentage of correct numbers when the associator is given input including noise from the encoder. The X-axis is the noise rate, the Y-axis is the percentage of correct output measured as the normalized Hamming distance. The error bars indicate the standard deviation.

was recalled because the input was not enough close to recall the correct pattern and that the accuracy was not low indicates that the Hamming distance between the correct pattern and output pattern was short. When the level of noise was high, it is assumed that the input pattern was so far from the correct pattern that a totally different pattern was recalled. Depending on the stored patterns, even if the noise rate was high, there were cases in which the correct pattern was recalled. As a result, when the level of noise was low, the standard deviation was small, and when the level of noise was high, the standard deviation increased.

Experiment 1-4: Up to 30% noise, the system maintained extremely low levels of error (Figure 10). After exceeding a level of 30% noise, the errors from the system increased closely linear. At 50% noise, the error from structure 1, which had the worst accuracy in this experiment, was about 0.1. In contrast, in Experiment 1-2, the error from structure 3, which had the best accuracy at 50% noise rate, was over 0.2.

The results of Experiment 1-2 and this experiment suggest that a system including an associator decreased the error and obtained much better robustness than a simple autoencoder.

3.2 The observation of performances and limits

Experiment 2-1: When the noise level is 0%, the error was quite low (Figure 11). As the noise rate increased, the error also increased. As same as experiment 1-2, there was a tendency that the error sharply increased at first then the error increased gradually as the noise increased. The tendency was regardless of the structure size.

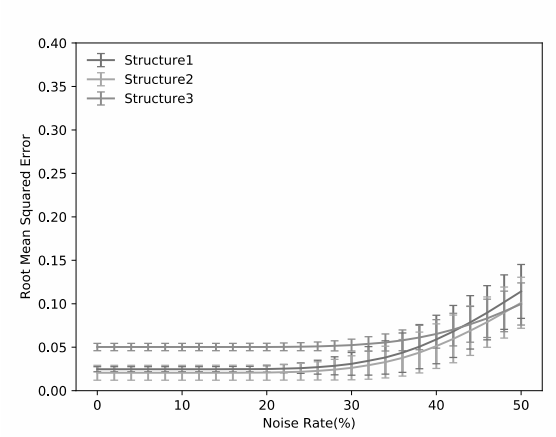


Figure 10: It shows the error on the output from the system when the input contains noise. The X-axis is the noise rate, the Y-axis is the root mean squared error, and the error bars indicate the standard deviation.

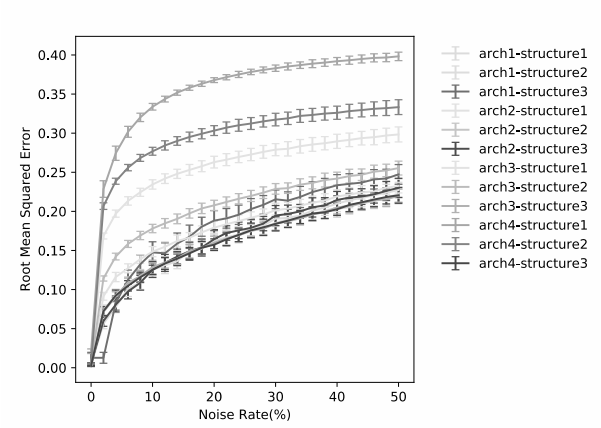


Figure 11: It shows the error on the output from autoencoders each architecture when the input contains noise. The X-axis is the noise rate, and the Y-axis is the root mean squared error. The error bars indicate standard deviations.

Experiment 2-2: Table 4 shows the measured a which was the average of pattern, theoretical storage capacity of the Hopfield network, and loading rate. On Arch 2, Arch 3, and Arch 4, the correct pattern was recalled even if the input to the encoders contained noise (Figure 12). The ability to recall the correct pattern even with high levels of noise depended on the size of the associator. As the associator increases, this ability also increased.

Table 4: Results for a, storage capacity, and loading rate for each architecture.

Architecture	a	Storage capacity	Loading rate
Arch 1	0.211	$0.144N$	66.1%
Arch 2	0.206	$0.144N$	38.6%
Arch 3	0.203	$0.144N$	19.4%
Arch 4	0.205	$0.144N$	8.7%

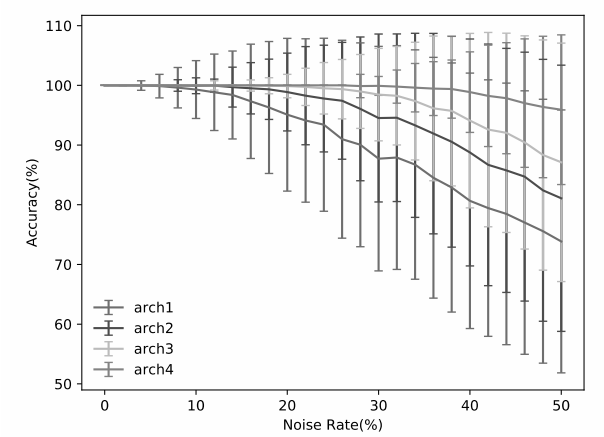


Figure 12: It shows the percentage of correct outputs from the associator of each architecture. The X-axis is the noise rate, and the Y-axis is a percentage of correct number, which is measured using the normalized Hamming distance. The error bars indicate standard deviations.

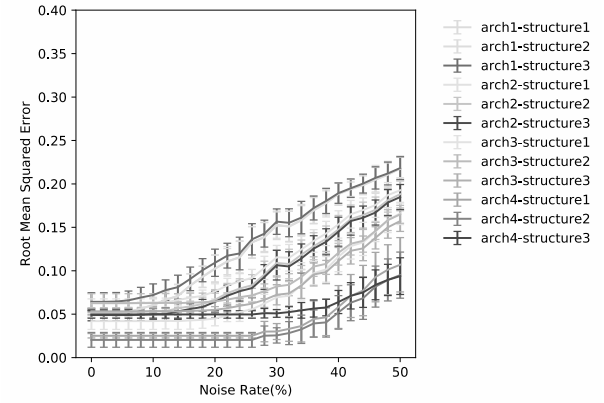


Figure 13: It shows the error on the output from the system of each architecture. The X-axis is the level of noise, and the Y-axis is the root mean squared error. The error bars indicate standard deviations.

Experiment 2-3: The robustness of the total system against the noise increased as the loading rate decreased (Figure 12). In particular, the total system of Arch 4, loaded under 10% of the storage capacity, maintained extremely low error even if the inputs contained up to 30% noise. As the loading rate increased, the total system was not able to keep the error extremely low when the noise also increased. After the point at which the total system was able to keep the error low, the error increased linearly. Storage capacities were about 4% up setting by lower ρ, which was the term about sparse, than 0.5.

3.3 A comparison with a simple autoencoder

At the noise level of 20%, the errors of the proposed architecture were over six times less than the errors of the autoencoder (Figure 14). The shape of increasing the errors for the proposed architecture was the same as the shape in the verification of concept and the architecture output quite low errors up to 30%. The autoencoder which processed three modalities output higher errors when

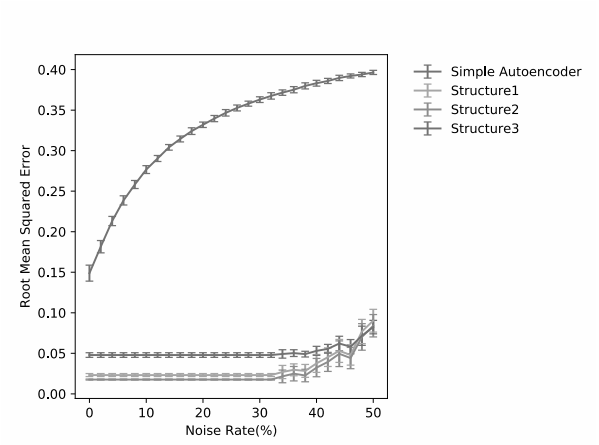


Figure 14: It shows the error on the output from the proposed architecture and an autoencoder when the input contains noise. The X-axis is the noise rate, the Y-axis is the root mean squared error, and the error bars indicate the standard deviation.

inputs contained errors. The shape of increasing the errors for the autoencoder was similar to the shape of logarithmic function.

Remark that it is obviously that the shapes of increasing the errors for the proposed architecture and the autoencoder are different greatly.

4. Discussion

Given these results, the validity of MACMSA as robot system in terms of cost, robustness, and system optimization are discussed, as mentioned in the introduction. Five conclusions can be drawn from results of the four experiments in the verification of concept, three experiments in the observation of performances and limits and the experiment in a comparison with a simple autoencoder: i) The proposed architecture MACMSA performs correctly based on the system design concept. ii) The encoder and decoder can be optimized with data with various averages and standard deviations. iii) The associator is able to store and recall the patterns encoded by the encoder. iv) The system composed of encoders, decoders, and associator can decrease the error better than a simple autoencoder. v) The attributes are maintained even if the module sizes in the system are different.

Cost: If a simple autoencoder processes multimodal information, the number of neurons in the input layer is multiplied dimensions of modalities by the number of modalities. For instance, in the experiment for a comparison with the simple autoencoder, the number of optimized weights between the input layer and the hidden layer 2 is 1,600 ×1,200=1,920,000 and the total number of optimized weights for the autoencoder is 5,760,000. On the other hand, for proposed architecture, the total number of optimized weights for the structure 1 is 2,250,000, the total number of optimized weights for the structure 2 is 562,500, and the total number of optimized weights for the structure 3 is 22,500. Thus, the number of optimized weights for the proposed architecture is 2,835,000. The number of neurons consisted for the simple autoencoder and the proposed architecture is same. However, the total number of weights for the simple autoencoder is twice bigger than the total number of weights for the proposed architecture. According to this difference, the computational cost for the proposed architecture is smaller than the computational cost for the simple autoencoder. The computational cost to store the pattern to Hopfield network is much smaller than the computational cost to optimize the encoders and the decoder. Hence, it is able to ignored. In addition, the computational cost to optimize the network increases exponentially due to the added modalities. For example, to add new modality which has 100 dimensions to the autoencoder in the experiment for a comparison with a simple autoencoder, 247,500 connections are added to between the input layer and the hidden layer 2. On the other hand, to add new modality which has 100 dimensions to the proposed architecture, only 7,500 connections are added to between the input layer and the hidden layer 2 because just one more independent autoencoder is optimized as encoder and decoder. The computational cost is only a constant multiple to add new modality on MACMSA.

In this research, offline learning and batch learning were used to reduce computational cost. Usually, it is considered that online learning is necessary to optimize a robot system because a robot is forced to continue working once it starts working. However, it is considered that offline learning should be sufficiently practical by giving the robots times to optimize the system. Humans also take time to sort out information in the brain. This time should be such as sleeping.

As similar method, there is Deep Boltzmann Machine (DBM) [26]. Srivastave and Salakhutdinov [26] used DBM as a generative model. Boltzmann Machine has potentially a function of denoise. The difference between Boltzmann Machine and Hopfield network is that Boltzmann Machine has a stochastic behavior while Hopfield network has a deterministic behavior. It seems that there is a benefit to use Hopfield network because a cost to implement Hopfield network is smaller than a cost to implement Boltzmann Machine.

Robustness: In an actual environment, it is assumed that noise is often added to the data due to changes in the environment or the state of the sensor. To consider such situations, some of the data were replaced by up to 50% noise in the experiments. The results of Experiment 2-3

show that as the loading rate decreased, the range of noise levels over which extremely low errors can be maintained was wider. Moreover, above that range, the errors increased linearly not non-linearly. It is assumed that the ability to reduce noise comes from the associator because the encoders do not have a strong ability to reduce noise. According to the results of Experiments 1-4 and 2-3, the decoders were able to output values close to those of the training data if input was close to the stored pattern. Hence, if the encoder has a strong ability to reduce noise, the number of errors in Experiments 1-2 and 2-1 should be low. However, in Experiments 1-2 and 2-1, which used only autoencoders, the number of errors was not negligible and increased non-linearly with respect to the noise levels of the input. In addition, at the point of 50% noise, if the network size is small, the errors of the autoencoders and proposed architecture are comparable. In contrast, if the network size is large, errors of the proposed architecture were much lower than those of the autoencoders. In real environments, the number of dimensions of the input from sensors is larger than the number of dimensions of the inputs considered in this study. Thus, it is easier to keep errors low. For example, if the inputs are 64×64-pixel pictures, the network size of the encoder and decoder for these inputs are four times larger than the network size of the largest encoder and decoder used in this study. Focusing on the Hopfield network as an associator, the robustness against noise seems to be lower than the theoretical robustness. It is assumed that the stored patterns are not ideal patterns because they are generated by encoders. Therefore, it is important that the pattern formatted by an encoder be linearly independent. By improving the linear independence of the pattern, the associator can recall the correct pattern at a higher accuracy. Then, it is assumed that the error at the decoder can also be decreased against higher noise rates. In addition, in situations in which the input includes a high level of noise (such as 50%), the input should be treated as another input derived from a different situation because it is seems unlikely that half of the input would be replaced with the noise, even though it is true that the noise often occurs in actual environments. Thus, inputs including high levels of noise should be added to training data as new data. The scale of the Hopfield network used in Experiments 1-1 to 1-4 had the capacity to store about 100 patterns without sparse coding. Moreover, by controlling sparseness precisely, the storage capacity substantially increases so that it is easy to add new training data. Note that ρ was set as the target value of the firing rate for all the autoencoders but the sparseness was not otherwise controlled in this study.

System optimization: MACMSA is a simple architecture. It is composed of three blocks of encoder, associator, and decoder units. The encoders and decoders are optimized simultaneously as a single autoencoder for each mode. Moreover, when the autoencoders have been optimized, the pattern stored by the associator is determined uniquely. In other words, the whole system should be optimized when the autoencoders have been optimized. We have not found published research on the importance of system optimization for robotics control systems. However, things composed of individually best parts are not always the best themselves. Likewise, it is also not always true that a system composed of the individually best modules is the best system. We believe that it is most important to optimize the combination of structures and modules. It is not necessary for structures and module to be state of the art. In fact, the proposed architecture is composed by no state of the art components but the experiments indicates that the proposed architecture withstands in practical use in terms of robustness against noises. System optimization is also important for processing speed under the real world. When a robot works in an actual environment, the system has to output a control signal in real time. Even if modules can process inputs very fast, if a system composed of a combination of these modules takes time to output a control signal for the robot, it is useless. By considering system optimization, it becomes possible to build a robot that works in a real environment and real time.

Finally, if it is necessary to add developmental functions in the next step, the following two functions would be appropriate. The first function is one that outputs a signal for time t + 1 at time t. To control a robot, a system obtains the state at time t as inputs and outputs a signal to move a part of the robot's body to a position at time t+1. Storing the state of time t+1 as a pattern and recalling it using the associator is one potential way to output a signal that specifies the state at time t + 1. The second potential function is precise control of the sparseness. A maximization of the storage capacity is important to cover a wide variety of possible states and allowable actions in a real environment. Therefore, controlling the sparseness at least in an internal hidden layer that is the output layer of an encoder instead of the whole autoencoder is desirable.

The proposed method was implemented using two different approaches. One consisted of writing original programs in C. In the programs, Equation 2 is simply iterated for a fixed number of times to optimize the networks. The other employed Pytorch version 1.4.0, which calculates the gradients from Equation 1 using automatic differentiation. Stable results were obtained by using the C programs, but could not be obtained with Pytorch. Therefore, all the results in this paper were obtained with the programs implemented in C.

5. Conclusion

MACMSA was proposed to consider cost, robustness, and system optimization. MACMSA is intended for robot

control systems that process multimodal information. The robustness of MACMSA was evaluated by adding the noise to the input. The results showed that MACMSA which is designed according to the system optimization is a suitable architecture, even when the input includes high levels of noise. The cost was discussed with a specific example and the system optimization were also discussed. The design cost of the proposed architecture is smaller than the cost of state of the art. The calculation cost of proposed architecture is also smaller than a simple autoencoder which processes multimodal information concurrently. Implementing MACMSA with the functions mentioned in the discussion will take us a step closer to building robots that can truly interact with humans.

References

[1] Sakagami Y, Watanabe R, Aoyam C, Matsunaga S, Higaki N and Fujimura K, "The intelligent ASIMO: System overview and integration", *International Conference on Intelligent Robots and Systems*, 2002, pp. 2478–2483.

[2] Tanaka F, Isshiki K, Takahashi F, Uekusa M, Sei R and Hayashi K, "Pepper learns together with children: Development of an educational application", *International Conference on Humanoid Robots*, 2015, pp. 270–275.

[3] Bekey GA, Autonomous Robots: From Biological Inspiration to Implementation and Control. MIT Press, 2005.

[4] Sünderhauf N, Brock O, Scheirer W, Hadsell R, Fox D, Leitner J, Upcroft B, Abbeel P, Burgard W, Milford M and Corke P, "The limits and potentials of deep learning for robotics", *The International Journal of Robotics Research*, 2018, Vol. 37, Issue.4–5, pp. 405–420.

[5] Ramachandram D and Taylor G, "Deep Multimodal Learning", *IEEE Signal Processing Magazine*, 2017, Vol. 34, No. 6, pp.96-108.

[6] Ngiam J, Khosla A, Kim M, Nam J, Lee H and Ng AY, "Multimodal deep learning", *Proceedings of the International Conference on Machine Learning*, 2011, pp. 689–696.

[7] Chollet F, "Xception: Deep learning with depthwise separable convolutions", *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–258.

[8] Qi CR, Su H, Mo K and Guibas LJ, "PointNet: Deep learning on point sets for 3D classification and segmentation", *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 652–660.

[9] Levine S, Pastor P, Krizhevsky A, Ibarz J and Quillen D, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection", *The International Journal of Robotics Research*, 2017, Vol. 37 Issue 4–5, pp. 421–436.

[10] Akhtar N and Milan A, "Threat of adversarial attacks on deep learning in computer vision: A survey", *IEEE Access* 6, 2018, pp. 14410–14430.

[11] Chen Z, Jacobson A, Sünderhauf N, Upcroft B, Liu L, Shen C, Reid I and Milford M, "Deep learning features at scale for visual place recognition", *IEEE International Conference on Robotics and Automation*, 2017, pp. 3223–3230.

[12] Pierson HA and Gashler MS, "Deep learning in robotics: A review of recent research", *Advanced Robotics*, 2017, Vol. 31, Issue 16, pp. 821–835.

[13] Suzuki K, Mori H and Ogata T, "Motion switching with sensory and instruction signals by designing dynamical systems using deep neural network", *IEEE Robotics and Automation Letters*, 2018, Vol. 3, Issue 4, pp. 3481–3488.

[14] Saito N, Kim K, Murata S, Ogata T and Sugano S, "Tool-use model considering tool selection by a robot using deep learning", *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2018, pp. 270–276.

[15] Sergeant J, Sünderhauf N, Milford M and Upcroft B, "Multimodal deep autoencoders for control of a mobile robot", *Australasian Conference on Robotics and Automation*, 2015, pp. 1–10.

[16] Noda K, Arie H, Suga Y and Ogata T, "Multimodal integration learning of robot behavior using deep neural networks", *Robotics and Autonomous Systems*, 2014, Vol. 62, Issue 6, pp. 721–736.

[17] Lenz I, Lee H and Saxena A, "Deep learning for detecting robotic grasps", *The International Journal of Robotics Research*, 2015, Vol. 34, Issue 4–5, pp. 705–724.

[18] Yang P-C, Sasaki K, Suzuki K, Kase K, Sugano S and Ogata T, "Repeatable folding task by humanoid robot worker using deep learning", *IEEE Robotics and Automation Letters*, 20-17, Vol. 2, Issue 2, pp. 397–403.

[19] Okada M, "Notions of associative memory and sparse coding", *Neural Networks*, 1996, Vol. 9, Issue 8, pp. 1429–1458.

[20] Billard A, Dautenhahn K, Hayes G, "Experiments on human-robot communication with Robota, an imitative learning communicating doll robot", *Socially situated intelligence workshop*, 1998, pp. 4-16

[21] Jockel S, Mendes M, Zhang J, Coimbra A.P, Crisostomo M, "Robot navigation and manipulation based on a predictive associative memory", *IEEE international Conference on Development and Leaning*, 2009, pp. 1-7

[22] Ng A, Sparse Autoencoder, CS294A Lecture notes https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf (accessed 5th May 2020)

[23] Finn C, Tan Y.X, Duan Y, Darrell t, Levine S, Abbeel

P, “Deep Spatial Autoencoder for Visuomotor Learning”, *IEEE International Conference on Robotics and Automation*, 2016, pp.512-519

[24] Asoh H, “Deep Representation Learning by Multi-Layer Neural Networks”, *Journal of the Japanese Society for Artificial Intelligence*, 2013, Vol. 28, Issue 4, pp. 649-659 [in Japanese]

[25] Okatani T, Deep learning, 2015 [in Japanese].

[26] Srivastava N, Salakhutdinov R, “Multimodal Learning with Deep Boltzmann Machines”, *Advances in Neural Information Processing System* 2012, Vol. 25, pp. 2222-2230

Car Body Precision Monitoring and Analysis Based on Big Data

Yixin YANG*,**, Jianjun GAO***, Yiping Feng***,Konghui GUO*,**
* State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body
Changsha, 410082, China
** College of Mechanical and Transportation Engineering, Hunan University
Changsha, 410082, China
*** Beijing Automobile Industrial Company Limited Zhuzhou Branch
Zhuzhou,412007, China

Abstract

In this paper, we present a big data system architecture based on Hadoop. The proposed system integrates various elements such as parts, assemblies, jigs, inspection tools, and various forms of measurement resources, and measurement data into a big data platform. Utilizing the method of big data analysis, we then devise algorithms to improve the efficiency and accuracy of body precision monitoring. Using correlation analysis method, principal component analysis (PCA), and improved PCA method, we developed techniques to analyse complex dimension deviation problems. We further established failure modes and devise monitoring and diagnosis models based on time series analysis.

Keywords: Big Data, PCA, Precision Monitoring, Intelligent Manufacturing, automobile industry.

1. INTRODUCTION

Dimensional precision is an important indicator of body welding quality. It determines the precision and reliability of the vehicle assembly, and further affects the stability and smoothness of the chassis, the handling of the car and the matching quality of the interior and exterior trims. Nevertheless, improvements made on the flexibility of the body production line which enable production of mixed lines of multiple models, cause difficulties in debugging and controlling the dimensional precision.

Usually, the influencing factors of car body dimensional precision include the precision of stamping and small parts, precision and stability of welding fixtures, the trolley of automated production lines, precise positioning of the material frame, personnel operation, and the stability of the robot welding system. Therefore, multiple factors affect the dimensional precision which are also themselves interrelated and influencing each other. Hence it is difficult to analyse and control the cause of deviations.

To investigate this problem, in early 1990s, General Motors, Chrysler, and the University of Michigan jointly launched "2mm Project" [1]. In China, the research and development of automobile dimensional engineering mainly started with the introduction of technology by joint-venture automobile companies. Systematic research on dimension control in body welding is presented in Lin et al. [2]. Jiang et al. also proposes a system optimization method for body parts and fixture design [3]. Yang et al. study an automatic alarm method for body dimension online detection data [4]. The concept of precision quality system for the RPS positioning reference point of car body is proposed by Chen et al. [5].

As it is seen in the above, various aspects of body dimensional precision control are investigated in the current literature. Nevertheless, the complexity of influencing factors of body dimensional precision and their multiple sources causes practical issues in dimensional control. The development of intelligent manufacturing and big data technology provides new methods and means for controlling and improving the body precision. In the paper we establish a big data system and use the methods of big data analysis to carry out body dimensional precision monitoring and analyse the influencing factors.

2. DIMENSION CORRELATION ANALYSIS OF MULTIPLE DEVIATION SOURCES

The change of body dimension generally follows a normal distribution. Hence the dimensional precision control is conventionally built upon the principle of normal distribution. The CMM measuring method is then used to control and improve the dimensional precision according to the actual measurements. Such a method is suitable for simple troubleshooting and improvement. For the multi-factor dimension problem however, it is often difficult to determine the causes of the deviation and to quantitatively analyse the degree of influence of a given deviation source. To address this issue, here we introduce a correlation analysis technique to quantitatively analyse multiple deviation sources and to quickly determine the main factors of deviation sources.

Table 1. Measuring points deviation (n1 and n2) under wearing condition of round pin A. Unit:mm

	1	2	3	4	5	6	7	8	9	10
X1	0.45	0.35	0.25	-0.3	-0.21	0.11	0.22	0.35	-0.36	-0.23
Y1	0.23	0.12	0.22	0.31	-0.15	-0.13	-0.26	0.18	0.25	-0.21
X2	0.42	0.34	0.24	-0.28	-0.22	0.12	0.23	0.33	-0.35	-0.23
Y2	0.01	0.11	-0.05	0.09	0.07	-0.11	-0.06	0.13	0.05	-0.04

Table 2. Measuring points deviation (n1 and n2) under wearing condition of round pin B. Unit:mm

	1	2	3	4	5	6	7	8	9	10
X1	0.12	0.06	-0.1	0.08	0.02	0.11	-0.11	-0.07	0.09	0.08
Y1	0.06	0.08	0.1	-0.06	-0.05	-0.13	0.07	0.1	0.12	0.05
X2	0.1	0.09	-0.08	0.06	-0.12	0.12	-0.05	0.05	0.05	-0.12
Y2	0.36	0.25	0.45	-0.23	-0.19	0.33	-0.48	0.36	0.25	0.33

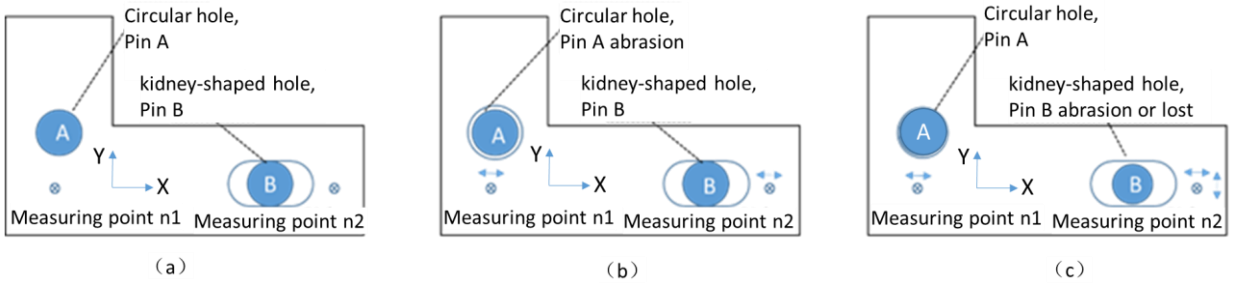


Fig.1 Parts positioning and wear mode of locating pins.

2.1 Principle of dimension correlation analysis

Correlation coefficient quantifies the degree of correlation between variables. Correlation analysis is statistical method to discover if there is a relationship between variables and the level of such a relationship [6]. Covariance is often used to quantify the linear correlation between two random variables. For two variables X and Y the covariance is defined as [7]:

$$\mathrm{cov}(X,Y) = \frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{n-1} \quad \text{..... (1)}$$

The correlation coefficient is defined as:

$$\gamma = \frac{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})}{\sqrt{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})^2 \sum_{i=1}^{n}(x_{2i}-\overline{x_2})^2}} = \frac{\sum_{i=1}^{n}(x_{1i}-\overline{x_1})(x_{2i}-\overline{x_2})}{(n-1)S_1S_2} \quad (2)$$

where x_{1i} is the i-th measurement value of the first variable, x_{2i} is the i-th measurement value of the second variable; $\overline{x_1}$ is the average of the first variable; $\overline{x_2}$ is the average of the second variable, n is the total number of measurements, S_1 is the standard deviation of the first variable and S_2 is the standard deviation of the second variable.

To normalize the correlation coefficient as defined in (2), we further set $Z_{1i} = \frac{x_{1i}-\overline{x_1}}{S_1}$ and $Z_{2i} = \frac{x_{2i}-\overline{x_2}}{S_2}$, therefore the correlation coefficient γ becomes:

$$\gamma = \frac{\sum_{i=1}^{n} Z_{1i}Z_{2i}}{n-1} \quad \text{..... (3)}$$

2.2 Dimension correlation analysis

As shown in Figure 1, the part is positioned with a round and a kidney-shaped hole on the plane. The two measuring points on the part are n_1 and n_2. If n_1 and n_2 move horizontally at the same time and with the same magnitude, it means that the wear of round pin A becomes smaller. Similarly, if the movement amplitudes of n_1 and n_2 are not the same and n_2 moves up and down, then round pin B becomes smaller or gets lost. Here we propose to utilise correlation analysis to determine the type of measurement point fluctuations. In the experiment, CMM is used to mesure the dimension. For convenience of measurement, holes were drilled at n_1 and n_2 positions. The diameter of pins are measured by micrometer. The deviation value is obtained by calculating the deviation between the measured point position and the theoretical position of multiple random clamping measurements.

As shown in Figure 1(b), pin A is worn and the deviations of measuring points n_1 and n_2 are (x_{1i}, y_{1i}) (x_{2i}, y_{2i}). The measuring points are shown in Table 1,2. The correlation coefficient between X1, X2, Y1, Y2 are:

- $\gamma(X1, X2) = 0.999$, $\gamma(X1, Y1) = 0.089$,

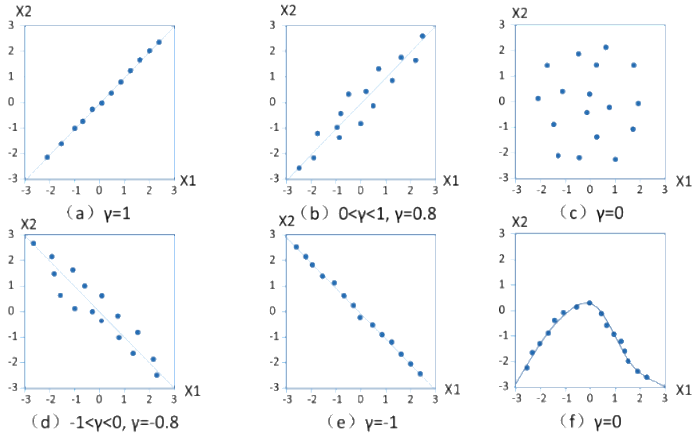


Fig.2 Scatter diagrams of two variables with different correlation coefficients.

- γ(X1, Y2) = -0.055
- γ(X2, Y1) = 0.082, γ(X2, Y2) = -0.007,
- γ(Y2, Y1) = 0.529

Note that the value of the correlation coefficient indicates the strength of the linear correlation between the two variables. As it is seen in Figure 2, γ = 1 indicates that the two variables have a clear linear relationship; 0 < γ < 1, indicating that the two variables are positively correlated; γ = 0, indicating that the two variables are not related in a linear form; -1<γ<0, indicating that the two variables are negatively correlated; γ=-1, indicating that the two variables have a linear relationship with a negative slope.

Obtaining correlation coefficient between multiple variables is usually done through a computer programme. The correlation coefficient matrix corresponding to X1, X2, Y1, Y2 is shown in Table 3. It can be seen from the correlation coefficient matrix that the correlation coefficients of X1 and X2 measurement points are close to 1> this means that the changes in these two measurement are linearly related. when pin A is worn, n1 and n2 move in X direction syn-chronously, and the two points have strong

Table 3. Correlation matrix of measuring points, n1, and n2, under wearing condition of round pin A.

Correlation matrix					
Correlation		X1	Y1	X2	Y2
	X1	1.000	0.089	0.999	-0.055
	Y1	0.089	1.000	0.082	0.529
	X2	0.999	0.082	1.000	-0.07
	Y2	-0.055	0.529	-0.07	1.000

Table 4. Correlation matrix of measuring points, n1, and n2, under wearing condition of round pin B.

Correlation matrix					
Correlation		X1	Y1	X2	Y2
	X1	1.000	-0.412	0.474	0.250
	Y1	-0.412	1.000	-0.141	0.279
	X2	0.474	-0.141	1.000	0.268
	Y2	0.250	0.279	0.268	1.000

correlation in X direction. We also note that correlation coefficients in other directions are very small, due to the limitation of pin B, the movement of n2 is limited ,hence there is no obvious correlation between them. Based on the correlation coefficient and measuring point data, the wear of pin A can be then inferred.

For the cases where the correlation coefficient is γ=0, see Figure 2(f), one may conclude that there is no linear relationship between the two variables. It is however seen that there is a certain relationship between the two variables, which can be described by a non-linear curve.
In Figure 1(c), pin B is worn, and points n1, n2 rotate around pin A. The correlation coefficient matrix is shown in Figure 4. The correlation coefficient is small.There is no obvious correlation in the XY direction. In this case, the correlation analysis cannot effectively indicate the root cause of the changes of the measuring point, and a new methods need to be utilized for analyzing such changes. Instances of such changes are: the rotation mode and multiple parts, multiple processes, and multiple jigs for multiple sources of analysis.

3. DIMENSION ANALYSIS BASED ON PCA AND KPCA

The correlation coefficient analysis is suitable for determining linear correlation between multivariate pairs. For more complex analysis Principal Component Analysis (PCA) and Kernel Principle Component Analysis (KPCA) based on kernel function [8] are often used.

3.1 Analysis of dimension deviation sources based on PCA

For non-random variables, K, Pearson introduces the principal component analysis (PCA). The concept is further extended by H. Hotellin to the random vectors [9].

Table 5. Deviation data of door frame measuring points. Unit: mm

	P1			P2			P3			P4			P5			P6			P7		
	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z	X	Y	Z
1	0.60	0.07	0.24	0.35	-.016	0.27	0.14	0.23	0.19	0.03	0.01	1.12	0.15	0.04	0.99	0.13	-0.11	0.50	0.18	-0.13	-0.35
2	0.25	0.28	0.13	0.03	-0.34	0.09	0.08	0.60	0.05	0.18	0.79	0.56	-0.01	0.41	0.50	-0.07	0.62	0.39	-0.01	0.26	0.03
3	0.30	-0.19	0.20	0.10	0.19	0.08	0.08	0.19	0.15	0.17	0.28	0.50	0.19	-0.14	0.40	0.12	-0.13	0.37	0.23	0.17	-0.08
4	-0.50	0.50	-0.07	-0.43	-0.10	-0.19	-0.23	-0.27	0.06	0.18	0.59	-0.70	-0.35	0.33	-0.56	-0.35	0.01	-0.16	-0.36	0.12	0.64
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
20																					

The PCA is a statistical method for dimensionality reduction. In body dimension analysis, the PCA is usually called principal vector analysis, by finding a combination of the relevant variables, several irrelevant variables, i.e. the principal vector, are derived and represented by Z_i. The principal vector Z_i can be interpreted as a linear combination of N original correlation variables, X_i:

$$\begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_n \end{pmatrix} = \begin{bmatrix} a_{11}a_{12}\cdots a_{1n} \\ a_{21}a_{22}\cdots a_{2n} \\ \vdots \\ a_{n1}a_{n2}\cdots a_{nn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \quad \text{..... (4)}$$

The purpose of principal vector analysis is to determine the eigenvalues and eigenvectors of the correlation coefficient matrix. The eigenvalues are the variance of the principal vectors, Var（Zi）$=\lambda i$. The elements of the eigenvector are, $a_{i1}, a_{i2}, \ldots, a_{in}$. The equation that represents the changes is:

$$[\lambda I - C]V = 0 \quad \text{..... (5)}$$

where λ represents the eigenvalue, i.e., the variance of the principal vector, I is the identity matrix, C is the covariance matrix, and V is the eigenvector. The eigenvectors characterise the fluctuation via the values and signs of their elements. This can be used to find out the source of dimension fluctuation and point out the direction. The original n related variables, $x_1, x_2, \ldots, x_n$, is used with n master vectors $Z_1, Z_2, \ldots, Z_n$, as

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{bmatrix} a_{11}a_{12}\cdots a_{1n} \\ a_{21}a_{22}\cdots a_{2n} \\ \vdots \\ a_{n1}a_{n2}\cdots a_{nn} \end{bmatrix}^{-1} \begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_n \end{pmatrix} =$$

$$\begin{bmatrix} b_{11}b_{12}\cdots b_{1n} \\ b_{21}b_{22}\cdots b_{2n} \\ \vdots \\ b_{n1}b_{n2}\cdots b_{nn} \end{bmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ \cdot \\ \cdot \\ Z_n \end{pmatrix} \quad \text{..... (6)}$$

Therefore, each original variable is described as a linear combination of multiple principal vectors:

$$x_i = b_{i1}Z_1 + b_{i2}Z_2 + \cdots + b_{in}Z_n \quad \text{..... (7)}$$

The variance of the original variable is expressed as：

$$V_{ar}[x_i] = b_{i1}^2 v_{ar}[Z_1] + b_{i1}^2 v_{ar}[Z_1] + \cdots + b_{in}^2 v_{ar}[Z_n] \quad (8)$$

$$V_{ar}[x_i] = b_{i1}^2\lambda_1 + b_{i1}^2\lambda_2 + \cdots + b_{in}^2\lambda_n \quad \text{..... (9)}$$

As mentioned earlier, the correlation matrix is symmetrical, and the number of matrix rows is equal to the number of eigenvalues and eigenvectors.

Here we used PCA to investigate the causes of deviations for analysing the fluctuation of the window frame size during the welding of the door assembly. As shown in Figure 3, gate frame measuring points 1, 2, 3, 13, 14, 15 and 16 fluctuate greatly in X and Z directions, therefore, it is not straightforward to find the rule. It is also difficult to determine the source of deviation of part dimension fluctuation only by using correlation coefficient method.

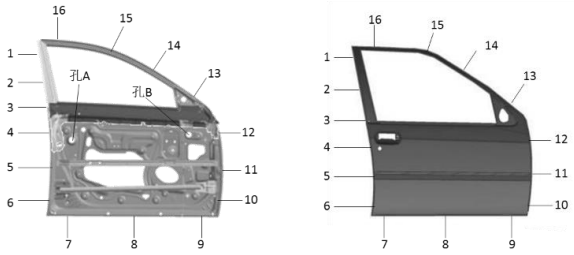


Fig.3 Measuring points of door sub assembly and assembly.

Table 6. PCA results of the doorframe measuring points deviation.

Component	Component	Component	Component
1	11.537	54.939	52.939
2	2.296	10.934	65.873
3	1.902	9.055	74.928
4	1.287	6.128	81.057
5	1.116	5.316	86.373
6	0.933	4.443	90.815
…	…	…	…

Extraction method: Principal Component Analysis

To address this issue, we use the PCA method to analyse the above measurement points. The measured data of the selected points are presented in Table 5 including 20 groups. The PCA analysis results are shown in Table 6. There are 5 principal vectors greater over 1. The characteristic value of the first principal vector is 11.537, and the

contribution rate is 54.9%. These explain that the source of dimension deviation is relatively single.

In the process of applying the PCA method, the data should be analysed in three directions of xyz. However, due to the high computational complexity in determining the main factors, it is more efficient to analyse the measuring point data with normal or Y direction deviation instead of three-direction. As shown in Figure 3, the door fluctuations in Y direction are unstable. Only Y-direction measuring points deviation and normal measuring point deviation data orthogonal to the outer plate are considered.

Table 7. Comparison of PCA results of three data deviation（a/b/c）.

a. Explanation of deviation variance in Y direction

Initial Eigenvalue			
Component	characteristic value	Contribution rate	Cumulative%
1	6.291	39.321	39.321
2	4.864	30.402	69.724
3	2.107	13.169	82.892
4	1.166	7.286	90.178
5	0.814	5.087	95.265
6	0.295	1.845	97.110
…	…	…	…

b. Explanation of deviation variance in T normal direction

Initial Eigenvalue			
Component	characteristic value	Contribution rate	Cumulative%
1	7.570	47.312	47.312
2	3.667	22.920	70.233
3	1.929	12.057	82.289
4	1.363	8.518	90.808
5	0.604	3.776	94.583
6	0.243	1.519	96.103
…	…	…	…

c. Explanation of total variance of PCA in three directions of XYZ

Initial Eigenvalue			
Component	characteristic value	Contribution rate	Cumulative%
1	5.375	37.092	37.092
2	2.886	19.916	57.008
3	2.237	15.437	72.445
4	1.896	13.084	85.529
5	1.230	8.488	94.017
6	0.243	1.678	95.695
…	…	…	…

We use three methods to measurement and take PCA analysis. The PCA results are presented in Table 7.

Table 7 shows that the first three principal components account for 72.4% of the PCA in three directions of XYZ. The three principal vectors are relatively dispersed, and in the analysis using Y and T normal deviations, the contribution of the first three principal components reached to 82.9%, and 82.3% respectively. The contribution rate of the first principal vector calculated with T normal deviation is 47.3%, which is greater than the contribution rate of the first principal vector calculated in Y where deviation is 39.3%. This indicates that the principal component analysis using T normal deviation is not only efficient in calculation, but also provides a good reference value in determining the principal mode of deviation source.

3.2 Analysis of the dimension deviation source based on KPCA method

The PCA only decouples data thus unable to analyze nonlinear problems. To address this issue, we utilize a kernel function and apply the KPCA to extract data deviation sources with nonlinear features. The KPCA method maps the input space variables from a low-dimensional space to a high-dimensional space through a nonlinear function and then applies the principal component analysis on the variables.

In the KPCA, the key is to transform the inner product of the characteristic space, to the core function of the original space after a non-linear transformation, by introducing the core function. This greatly simplifies the calculation. Here we use the core technology of the SVM support vector machine, to avoid "dimensional disaster". In other words, the inner product operation of samples in the feature space is replaced by a core function satisfying Mercer condition, i.e., the core function must be a semi-positive definite function [10].

The corresponding mapping, $\Phi: x \rightarrow \mathrm{F}$, maps the point x to F, by_. In the corresponding high-dimensional feature space, the variables also meet the condition of de-centralization, i.e.

$$\sum_{\mu=1}^{M} \Phi(x_u) = 0 \quad \text{.....(10)}$$

The covariance matrix of the feature space is then given by

$$\mathrm{C} = \frac{1}{M} \sum_{u=1}^{M} \Phi(x_u) \Phi(x_u)^T \quad \text{.....(11)}$$

and the eigenvalues and eigenvectors of C are obtained from

$$V \epsilon F \backslash \{0\},\ Cv = \lambda v \quad \text{.....(12)}$$

The eigenvector can be expressed as a linear combination of $\Phi(x_1)$, $\Phi(x_2)$, …, $\Phi(x_M)$, where $v = 1,2,\dots,M$. Define an $M \times M$ matrix K, the eigenvalues and eigenvectors are obtained by solving. From this, the projection of the test variables in the new eigenvector space v^k is:

$$\left(v^k \cdot \Phi(x)\right) = \sum_{i=1}^{M} (\alpha_i)^k \left(\Phi(x_i), \Phi(x)\right) \quad (13)$$

Here we then replace the inner product with the kernel function, therefore,

$$\left(v^k \cdot \Phi(x)\right) = \sum_{i=1}^{M} (\alpha_i)^k K(x_i, x) \quad \dots\dots\dots (14)$$

If (14) is not established, then it is adjusted as the following:

$$\Phi(x_u) \to \Phi(x_u) - \frac{1}{M}\sum_{i=1}^{M} \Phi(x_v)\, u = 1,2,\dots,M \quad (15)$$

The kernel matrix is then modified to:

$$K_{uv} \to K_{uv} - \frac{1}{M}\left(\sum_{w=1}^{M} K_{uw} + \sum_{w=1}^{M} K_{wv}\right) +$$

$$\frac{1}{M^2}\sum_{w,\tau=1}^{M} K_{w\tau} \quad \dots\dots\dots (16)$$

Based on the principle of KPCA, the relevant calculation process is as the following.

① The n indexes obtained are written into $m \times n$-dimensional matrix (each index has m samples).

$$\mathrm{A} = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \ddots & \dots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} \quad \dots\dots\dots (17)$$

② Standardizing the matrix, and set $\mathrm{X} = \left(x_{ij}\right)_{m\times n}$

③ Calculating the correlation coefficient matrix, $\mathrm{R} = \frac{1}{m-1} X^T \cdot X = \left(r_{ij}\right)_{n\times n}$

④ Using Jacobian iterative method to obtain the eigenvalues, $\lambda_1, \lambda_2, \dots, \lambda_n$, and the corresponding eigenvectors, $v_1, v_2, \dots, v_n$.

⑤ The strong eigenvalues are then sorted in descending order to get $\lambda_1^{'} > \lambda_2^{'} > \cdots > \lambda_n^{'}$, and the corresponding adjusted eigenvectors are $v_1^{'}, v_2^{'}, \dots v_n^{'}$.

⑥ Unit orthogonalization of the eigenvectors is then conducted using Schmidt orthogonalization to obtain $\alpha_{1,}\alpha_{2,} \dots \alpha_n$.

⑦ The cumulative contribution rate of eigenvalues are then acquired through calculating $\{B_{1,}B_{2,} \dots B_n\}$. We then set the extraction efficiency ρ value if $B_t \geq \rho$ then t principal components $\alpha_{1,}\alpha_{2,} \dots \alpha_t$ are extracted.

⑧ Calculate the projection $Y = X \cdot \alpha$ of the sample variable (X standardized) on the extracted eigenvectors, where $\alpha = \alpha_{1,}\alpha_{2,} \dots \alpha_n$. Y is the dimension-reduced data.

The advantage of KPCA over PCA is that the latter is an algebraic feature analysis method. PCA requires a rather large memory space and its algorithm is computationally complex. For an original space with dimension n, PCA needs to decompose an $n \times n$ non-sparse matrix. As a linear mapping method, the dimension-reduced expression of PCA is generated by a linear mapping. Therefore, it ignores nonlinear relationship in data samples, hence the optimal feature may be overlooked. This is the main reason in some cases the PCA method is not effective [11]. KPCA uses a nonlinear method to extract the principal components, which maps the variables to a high-dimensional space F through a mapping function and then PCA is used to analysis function space F.

Here we take the deviation source analysis of the body side taillight installation area as an example to compare

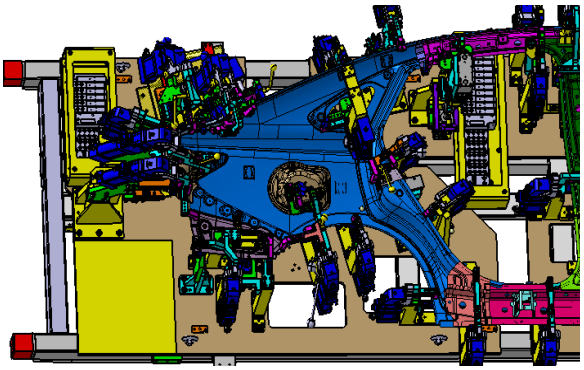


Fig.4 Clamping drawing of the rear area of the body side.

the analysis results of the PCA and KPCA methods. The dimension fluctuation in the mounting area of the side taillight is an issue in body dimension control (see Figure 4) as the matching relationship between this area and the taillights, rear bumper, and rear trunk is rather complicated and the dimension of the rear body side area is the key to the matching quality of this area.

We use both PCA and KPCA methods to analyze the measurement data and compare the results. Several commonly used kernel functions are linear kernel function, P-order polynomial kernel function, Gaussian RBF kernel function, multilayer perceptron kernel function. Here, noting the characteristics of the analysis of the deviation source of the flexible sheet, we adopt a polynomial kernel function as the following

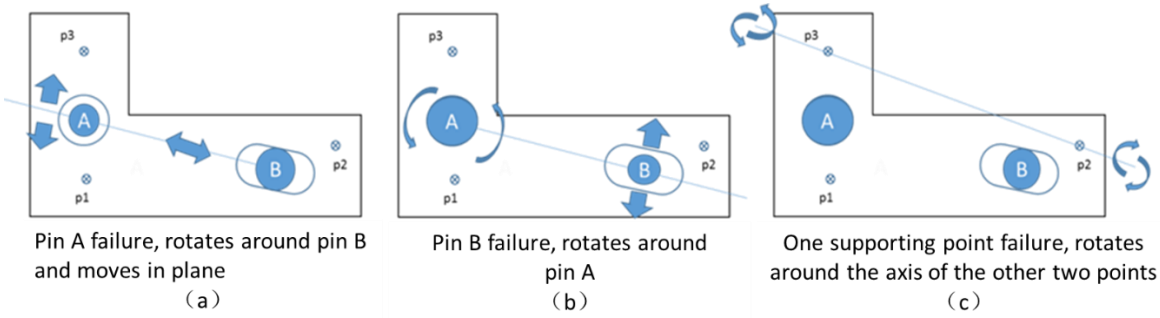
$$K\left(x \cdot x_i\right) = \left[\left(x \cdot x_i\right) + 1\right]^p \quad \dots\dots\dots (18)$$

The comparison of PCA and KPCA analysis is shown in Table 8. The contribution rate of the first three principal components obtained using PCA is 63.9%, and the contribution rate of the first principal component is 26.9%. The contribution rate of the first three principal components in KPCA is 80.7%, of which the contribution rate of the first principal component is 62.5%. As it is seen, the contribution rate of KPCA analysis principal component, especially the first principal component, is higher than PCA. It's easier to determine the main mode of deviation. According to the on-site analysis, due to the loose clamps of the body side tail, and the rebound of the outer plate of the body side tail, the rear light assembly area is rotated

Table 8. Comparison of PCA and KPCA analysis results of rear body side deviation.

PCA				KPCA			
Component	Characteristic value	Contribution rate	Cumulative%	Component	Characteristic value	Contribution rate	Cumulative%
1	2.536	26.921	26.921	1	0.285	62.500	62.500
2	1.809	19.024	46.125	2	0.062	13.596	76.096
3	1.678	17.813	63.938	3	0.021	4.605	80.702
4	1.106	11.741	75.679	4	0.017	3.728	84.430
5	0.812	8.620	84.299	5	0.012	2.632	87.061
6	0.678	7.197	91.497	6	0.008	1.754	88.816
…	…	…	…	…	…	…	…

Fig.5 Three failure modes of fixtures.



around a certain axis. After the clump is repaired, the measuring points in this area become stable and qualified.

4. ANALYSIS BASED ON THE SOUTCE OF MODEL DIMENSION DEVIATION

The deviation of part assemblies is often caused by the deviation of sub-assemblies or single parts. In such cases, we use the data analyzed to recognize the source of the deviation. The data here refers to the corresponding measuring point of the assembly and the sub-assembly or single parts. Here we study two analysis methods, one is based on the failure mode of parts and fixtures, and the other is based on time series to identify the source of dimension deviation.

4.1 Principal vector analysis based on failure mode

Principal component analysis (PCA) is used to extract the principal vectors (deviation mode vectors) p_i and s_j of parts or subassemblies and assemblies. The degree of vector correlation is expressed by the correlation coefficient,

$$\eta_{ij} = \left| \frac{p_i \cdot s_j}{|p_i||s_j|} \right| \quad (19)$$

where η_{ij} is the correlation coefficient between p_i and s_j, p_i is the i-th principal component of the part or sub-assembly, p, and s_j is the j-th principal component on the assembly, s. In order to eliminate the noise interference as much as possible, we set a threshold, V_{comp}, and a mapping relationship is formed between part p and assembly s if the correlation coefficient is greater than V_{comp}. The contribution rate, γ, represents the strength of the mapping relationship:

$$\gamma = \frac{\sum_{j=1}^{q} \lambda_j}{\sum_{m=1}^{m} \lambda_m}, \quad (20)$$

where λ_j is the eigenvalue of s_j corresponding to p_i; q is the number of eigenvectors greater than the threshold V_{comp} (if $q \neq 1$), and m is the total number of extracted eigenvectors. The threshold, V_m, of the first feature principal vector is set to reduce the computational complexity of correlation analysis. If the contribution of the first principal component extracted by a part is less than V_m, then it suggests that the part has no clear impact on the deviation of the assembly. If the contribution of the first principal component is greater than V_m, it is considered to have a significant influence on the deviation of the assembly [12].

In addition to the deviation source of the welding assembly caused by the parts, another important factor is the fixtures. The analysis of the deviation source generated by the fixtures is based on investigating the potential failure modes of the direction and the amplitude of the movement in each measuring point.

According to the 3-2-1 positioning principle for parts, the failure modes of fixture are divided into failure of main positioning pin, failure of secondary positioning pin, and failure of positioning and clamping point. The corresponding parts also appear in three modes: translation

along the AB direction, rotation in the plane around point A, and rotation in space around the axis formed by AB. The failure modes are shown in Figure 5.

For the first failure mode, most of the measuring points in Figure 5(a) move along the kidney-shaped hole and slightly rotate around B. The deviation mainly appears in a specific direction and the deviation is the same. The failure mode in this case can be obtained by unitizing the column vector composed of all measuring points. For Figs. 5(b), (c), the measuring point deviation has mainly a fixed axis for rotating motion. Suppose the vector of the rotation direction of the measuring point is $e_i = (x_i', y_i', z)$, the distance from the measuring point to the center of rotation is d_i', and $e_i \times d_i'$, the column vector is $[e_1.d_1, e_2 \cdot d_2, \dots, e_n \cdot d_n,]^T$. After unitization, $e_i \times d_i'$, is the failure mode under rotation.

For the fixture failure mode, the analysis of the welding assembly deviation source is based on calculating the correlation coefficient, η_{ik}, between the fixture failure mode, a_i, and the main eigenvector, v_k, of the assembly measurement point data:

$$\eta_{ik} = \left| \frac{a_i \cdot s_j}{|a_i||s_j|} \right| \quad (21)$$

Here, again to reduce the amount of calculation, a threshold value, V_{jig}, is considered and the mapping relationship is determined between all the main eigenvectors larger than V_{jig}, and the failure modes. In order to determine the contribution degree of the main eigenvector in the assembly deviation, the deviation contribution coefficient ω is defined as

$$\omega_i = \eta_{ik} \cdot \frac{\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_p} \cdot 100\% \quad (22)$$

4.2 Discrimination of Dimensional Deviation Source Based on Time Series Analysis

Here we first establish a time series AR model to process the continuous measurement data and obtain a stable normal distribution with zero mean. The first step is to extract the trend term to measure the time series data $\{x_t\}, (t = 1,2, \dots, N)$ and to remove the nonstationary part d_t : $y_t = x_t - d_t$ and form the stationary time series, $\{y_t\}$. Then, the time series are zeroed and normalized to obtain the normal distribution $\{x_t\}$, where x_t~N(0,1). The basic expression of AR(n) is:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_n x_{t-n} + a_t \quad a_t \sim NID(0, \sigma_a^2) \quad (23)$$

Among them: $a_t - \{x_t\}(t = 1,2, \dots, N)$ residuals correspond to a normal distribution with variance σ^2. $\varphi_1, \varphi_2, \dots \varphi_n$ and σ_a^2 are the parameters estimated by a certain method for time series $\{x_t\}(t = 1,2, \dots, N)$. Generally, the least square method is used for parameter estimation, which is relatively simple and unbiased. The least square estimation of the model is:

$$\hat{\varphi} = (x^T x)^{-1} x^T y \quad (24)$$

where $\varphi = (\varphi_1, \varphi_2, \dots \varphi_n)^T$, $y = [x_{n+1}, x_{n+2}, \dots, x_N]^T$ and

$$x = \begin{bmatrix} x_n & x_{n-1} & \cdots & x_1 \\ x_{n+1} & x_n & & x_2 \\ \vdots & & \ddots & \vdots \\ x_{N-1} & x_{N-2} & \cdots & x_{N-n} \end{bmatrix}$$

Principal component analysis is used to extract the feature roots $\lambda_1, \lambda_2, \dots, \lambda_m$, and eigenvector V from an n-dimensional pattern vector, $\varphi = (\varphi_1, \varphi_2, \dots \varphi_n)^T$. The vector, V, is a low dimensional eigenvector obtained by PCA reduction of pattern vector, φ:

$$V_{m\times 1} = A_{m\times 1}\varphi_{n\times 1} \quad (25)$$

To construct the discriminant function for a certain part, for K dimension deviation sources, $F_{R1}, F_{R2}, \dots, F_{RK}$, the corresponding K reference states are obtained. The constructed function is the discriminant function. Then classify the part to be inspected mode vector φ_t, analyze and judge its deviation mode.

The discriminant function adopts the distance discriminant function, and its measurement method is geometric Euclidian distance. The geometric distance between φ_t and F_R in an n-dimensional geometric space which is expressed as the sum of squares of the coordinate difference of two points in the space:

$$D^2(X,Y) = \sum_{i=1}^{n}(x_i - y_i)^2 = (X,Y)^T(X,Y), \quad (26)$$

where $X = [x_1 x_2 \dots x_N]^T$ and $Y = [y_1 y_2 \dots y_N]^T$ are arbitrary points in space. The time series data $\{x_t\}_T$ to be checked is formed into a coefficient matrix X_T, which is substituted in the AR model to form the reference deviation:

$$X_T \emptyset_R = a_{RT} \qquad X_T \emptyset_{RT} = a_T \quad (27)$$

where a_{RT} is the calculated residual vector of the coefficient matrix, XT, to be tested and the reference model parameter, $\emptyset_R$, and $X_T \emptyset_{RT} = a_T$ is the AR model to be tested. The geometric distance between a_T and a_{RT} represents the residual offset distance $D^2(a_T, a_{RT})$:

$$D^2(a_T, a_{RT}) = N_T(\emptyset_T - \emptyset_R)^T R_T(\emptyset_T - \emptyset_R) \quad (28)$$

where R_T is the covariance matrix of the test time series data $\{x_t\}_T$, and N_T is the length of the test time series data $\{x_t\}_T$. Let N_T be equal to the length of the reference deviation mode time series data, that is $N_T = N_R = N$.

It can be seen from equation (28) that the residual offset distance is a function of $\emptyset_T$ and $\emptyset_R$, which can be expressed as a weighted residual offset distance measured

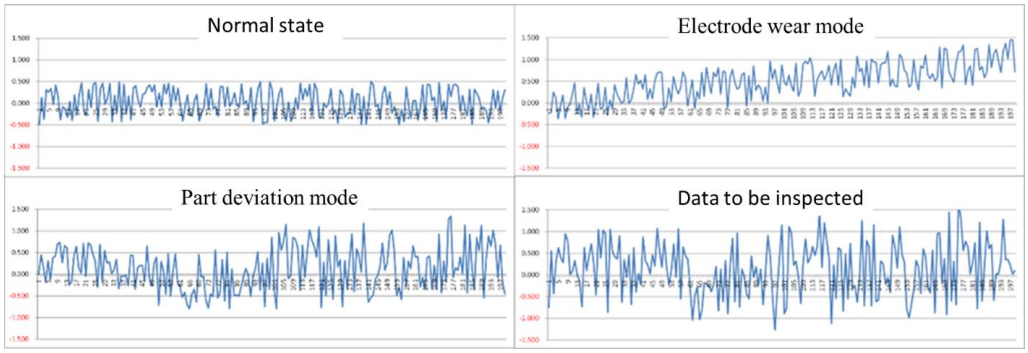


Fig.6 Deviation data of the body top cover measurement points.

Table 9. Euclidean distance and Square Euclidean distance.

	Approximation matrix Euclidean distance				Approximation matrix squared Euclidean distance			
	Normal deviation mode	Part deviation mode	Electrode wear mode	Actual test data	Normal deviation mode	Part deviation mode	Electrode wear mode	Actual test data
Normal deviation mode	0.000	0.300	0.370	0.223	0.000	0.090	0.137	0.050
Part deviation mode	0.300	0.000	0.234	0.083	0.090	0.000	0.055	0.007
Electrode wear mode	0.370	0.234	0.000	0.218	0.137	0.055	0.000	0.047
Actual test data	0.223	0.083	0.218	0.000	0.050	0.007	0.047	0.000

in n+1 dimensional space. The offset distance function related to the residual φ is then derived as

$$D^2(a_T, a_{RT}) = N_T(\varphi_T - \varphi_R)^T r_T(\varphi_T - \varphi_R) \text{ ..... (29)}$$

where r_T is the covariance matrix of the time sequence to be checked $\{x_t\}_T$, which is equivalent to the nth-order sub-matrix of R_T minus the first row and the first column. For the discrimination of K deviation source test modes, we take the reference mode with the smallest residual offset distance value, therefore,

$$D_a{}^2(\varphi_T, \varphi_{R(j)}) = \min\{D_a{}^2(\varphi_T, \varphi_{R(i)})\ (i = 1,2,\dots,K)\}\ \varphi_T \in F_{Rj} \text{ ..... (30)}$$

The reference population where φ_T is located should satisfy the minimum residual offset distance.

Geometric distance calculation and deviation diagnosis are as follows. First, the deviation source is classified. According to the corresponding deviation state of dimension deviation data, an autoregressive model is then established to obtain the mode vector. The principal vector is then extracted to obtain the covariance. Using Eq. (29), (30) we then evaluate the distance and establish the corresponding deviation mode for reference. The next step is to model the inspection data. The geometric distances under different deviation states are also obtained and compared to determine the source of dimension deviation [13].

Here we consider the upper edge of the windshield glass of the car body roof. Here the normal measurement point matching the glass often fluctuates. This is mainly due to the wear of the welding electrode cap, or deformation of the incoming material. Two deviation modes can be established corresponding to the wear of the electrode cap, and deformation of the incoming material. We collect real-time data of the implementation detection data. By calculating, tracking, and comparing the pre-established deviation modes, we automatically identify the deviation modes which eliminate the need on manual inspection.

The data under three states including normal operation, incoming material deformation, and electrode cap wear are collected and shown in Fig. 6. The time series model is then established according to the AR modeling requirements and the feature vector is extracted by the PCA method. The AR model under three states is established as the following:

① Normal:

$$x_t = -0.567x_{t-1} + 0.112x_{t-2} - 0.093x_{t-3} + 0.112x_{t-4} - 0.101x_{t-5} + a_t$$

② Incoming material deformation：

$$x_t = -0.583x_{t-1} + 0.159x_{t-2} - 0.065x_{t-3} - 0.063x_{t-4} + 0.136x_{t-5} + a_t$$

③ Electrode cap wear:

$x_t = -0.505x_{t-1} + 0.012x_{t-2} + 0.062x_{t-3} - 0.148x_{t-4}$ +Based on the measured data of workshop, the AR model is established as follows:

$$x_t = -0.579x_{t-1} + 0.131x_{t-2} - 0.053x_{t-3} - 0.028x_{t-4} + 0.067x_{t-5} + a_t$$

The measured data in the workshop and the residual deviation distance of each failure mode (the Euclidean distance) are presented in Table 9. As it is seen, the deviation mode between the measured data and the reference sample is the smallest, therefore the deviation mode might be the same (or quite similar). Therefore, the deviation mode of the measured data is most likely the deviation mode caused by part fluctuation. Another distance calculation method is Euro-square distance. The calculation results are 0.505 (distance normal mode), 0.007 (distance part deviation mode), and 0.047 (distance electrode wear mode). It can be seen from the data that the Euclidean square distance provides a better and more deviation mode discrimination.

5. CONCLUSION

(1) The dimensional precision is affected by the measurement resources, measurement information, and inspection data. Understanding the corresponding big data improves the efficiency of monitoring and analysis of the body dimensional precision.

(2) Big data analysis enriches the means of analyzing the precision of body dimension. Correlation analysis, PCA, and KPCA analysis method are capable of diagnosis in complex dimension deviation problems. The failure mode-based principal component analysis model and the body dimension deviation source monitoring diagnosis model based on time series analysis were also shown to be able to effectively improve the efficiency of failure mode and deviation source diagnosis.

Data Availability

The data used to support the findings in this paper are available upon request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledge

The study was supported by the China State Key Laboratory of Advanced Design and Manufacturing of Automobile Body Open Fund Project (no. 31515010).

References

[1] D. Tang, “From American Automobile 2mm Program to China Automobile Manufacturing Quality Program,” Development & Innovation of Machinery & Electrical Products, Vol.19, No.1, pp. 26-28, 2006.

[2] Z. Lin, Quality control technology of automobile body manufacturing: China Machine Press, 2005.

[3] H. Xie, C. Jiang, and Z. Zhang, “Interval description of dimensional tolerance and uncertainty optimization,” 2014 China computational mechanics conference, 2014.

[4] D. Yang, “Research on real-time monitoring and fault diagnosis system and its key technologies in BIW Welding Process,”, 2014.

[5] X. Chen, and J. Huang, “Application of RPS in precision designef body,” Automotive technology, No.8, pp. 18-21, 42, 2006.

[6] Y. Tao, H. Shi, B. Song, and S. Tan, “A Novel Dynamic Weight Principal Component Analysis Method and Hierarchical Monitoring Strategy for Process Fault Detection and Diagnosis,” IEEE T IND ELECTRON, Vol.67, No.9, pp. 7994-8004, 2020.

[7] T. Sibillano, A. Ancona, V. Berardi, and P. M. Lugarà, “Correlation analysis in laser welding plasma,” OPT COMMUN, Vol.251, No.1-3, pp. 139-148, 2005.

[8] S. Yi, Z. He, X. Jing, Y. Li, and F. Nie, “Adaptive Weighted Sparse Principal Component Analysis for Robust Unsupervised Feature Selection,” IEEE T NEUR NET LEAR, Vol.31, No.6, pp. 2153-2163, 2019.

[9] H. Caussinus, and A. Ruiz Gazen, Principal Component Analysis, Generalized: John Wiley & Sons, Inc., 2005.

[10] Z. Sun, J. Zhao, and L. Li, “Comparison between PCA and KPCA methods in comprehensive evaluation of robotic kinematic dexterity,” High Technology Letters, Vol.20, No.2, pp. 154-160, 2014.

[11] S. Yang, “Research on BIW size deviation based on wavelet modulus maximum and principal component analysis,”: Chongqing Jiaotong University, 2014.

[12] Y. Yang, S. R. Shen, Y. H. Liu, L. I. Zheng-Ping, and S. Jin, “Diagnosis of body online detecting dimensional deviation based on pattern recognition,” Machinery Design & Manufacture, No.1, pp. 188-190, 2012.

[13] K. Geng, J. He, K. Wang, and Y. Chen, “Gear Fault Diagnosis Based on Pattern Recognition of Acoustic Emission Signals,” Machine Tool & Hydraulics, Vol.47, No.16, pp. 192-196, 208, 2019.

Reconstruction Method for Missing Measurement Data of High-Speed Train Using Generative Adversarial Network

Changfan ZHANG* , Hongrun CHEN* , Jing HE*
* College of Electrical and Information Engineering, Hunan University of Technology
No.89 Taishan Xi Road, Tianyuan District, Zhuzhou, Hunan412007, China

Abstract

Concentrating on the issue of measurement data missing caused by complex and changeable working conditions during the operation of high-speed trains, this paper proposes a frame for reconstruction of missing measurement data based on generative adversarial network. Suitable parameters are set for the frame. Discrete measurement data are taken as the input of the frame for preprocessing of data dimensionality ascending. Then the convolutional neural network learns the correlation between different characteristic values of each device in unsupervised pattern, and constrains and improves the reconstruction accuracy taking advantage of the context similarity of authenticity. When there are measurement data missing to different extents, it is demonstrated by experiments that, the model of the paper can still maintain high reconstruction accuracy. In addition, the reconstruction data also conform well to the distribution law of measurement data.

Keywords: High-speed train， generative adversarial network，data dimensionality ascending，convolutional neural network， reconstruction accuracy

1. INTRODUCTION

High-speed trains may cross through mountain area, continuous tunnel and a variety of complex environments, and there may be network failure, transmission interruption, harmonic interference, etc. As a consequence, there are a lot of data missing, and it is not conducive to train state evaluation and system failure judgment. Safe operation of the train can only be guaranteed by real-time and accurate system monitoring achieved through data reconstruction to the maximum extent for the data missing.

Incomplete data imputation can be realized by parameter estimation result adopting some existing imputation methods such as Maximum likelihood estimation (MLE). However, it is hard to realize proper imputation adopting the method in case of more missing data and noncompliance with observation conditions. Mean imputation[2], regression recovery method[3] and other data imputation methods using mathematical theory are adopted in other papers, but it is hard to realize the imputation of different equipment measurement data of high-speed trains. EM algorithm [4] is used extensively in dealing with missing data, but the calculation speed will drop in case of a lot of missing data in the dataset.

In fact, the correlation between the data measured and the changes under multiple working conditions in high-speed train measurement can become important basis for the reconstruction of missing data. For example, the reconstruction of missing values of voltage and power is realized by Miranda et al.[5] using Autoencoders. However, the method is the shallowest neural network, and no effective description for complex equipment relationship can be carried out. In recent years, with the rapid develop of neural network and the continuous improvement of deep learning technology, there are more and more technical methods for super-resolution reconstruction and missing image repair based on deep learning[6-8]. The missing image repair and resolution reconstruction are essentially similar to the reconstruction of the monitoring missing data of high-speed train for the issues. For all of them, the original historycal data are used to infer the missing part complying with the objective law[9]. Generative Adversarial Networks (GAN) which is very popular has been proposed recently. The data distribution law and characteristics can be learned automatically in unsupervised pattern and the data meet these laws and characteristics are generated in GAN. Thus the reconstruction of missing data is realized.

GAN is widely used in digital image and computer vision. Mudavathu [10] proposed a method to generate a model for generated images by handling auxiliary conditions using labels for the expansion of image dataset. Yao Naiming et al. [11] reconstructs partial missing of face image using GAN, which improves the discrimination of facial expressions. The reconstruction of missing data can be analogous to the issue of imputation in image defects as abovementioned. Yao et al.[12] restore distorted signal by generative adversarial network model so as to improve the signal recognition ability. Vanishing gradient and unable to generate discrete data distribution and other issues occur easily for conventional GAN model when the data similarity is small. Wasserstein GANs (WGAN) method is proposed in reference[13] to improve the stability of the

network. In the method, EM distance is used instead of JS distance. In reference[14] WGAN is introduced into the electric power system for reconstruction of the measurement data with time sequence. However, WGAN still has great limitations in processing 1-D discrete data without distribution law, and it cannot learn accurately data characteristics.

There are many methods for the processing of 1-D discrete data. Kiranyaz et al. [15] design a one-dimensional CNN to detect the damage in a real-time manner. Zhang et al. [16] design a two-dimensional non-negative matrix factorization model to extract bearing signal characteristics. Wen[17] converts the signal into two-dimensional image and extracts the features of converted 2D image. A lot of expert experiences are required to set the parameters when adopting such methods. Therefore, the data preprocessing method not requiring setting any parameters is adopted in the paper.

Inspired by the success processing of discrete data using GAN in the image field, the method for reconstruction of missing measurement data of generative adversarial network is studied in the paper. The main work in the paper is as below: (1) A new data reconstruction frame is proposed based on WGAN. EM distance and context similarity loss between data are used as reconstruction constraints in the model, and the reconstruction data conforming to the complex rules between measurements are generated. (2) Discrete measurement data points are converted into two-dimensional network form, which simplifies effectively the model training time and improves reconstruction accuracy. (3) It is demonstrated by experiments that, when there are a lot of data missing, relatively high reconstruction accuracy can be maintained, and the data reconstructed are consistent with the real data in the distribution law.

2. BRIEF DESCRIPTION OF RELATED WORK

2.1 Generative Adversarial Network

GAN consists of two convolutional neural networks, i.e. generator (G) and discriminator (D). The generator (G) is used to learn data distribution characteristics and is mapped to be new data. Then the discriminator (D) measures the similarity between real data and the data generated, and updates continuously the parameters itself through the relationship established by connecting two seemingly independent deep neural networks based on the training relationship of historical data, to generate the new data which meet the distribution relationship of measurement data[18]. The GAN network structure established by these mappings is shown in Fig. 1 as below.

It may be difficult for the traditional GAN to train. Smooth Wassersterin distance is used to measure the distance between two probability distributions, which can also improves network training stability. Wassersterin distance[19] is defined as below:

$$W(p_r, p_g) = \inf_{\gamma \in \prod(p_r, p_g)} E_{(x,y)\sim\gamma}\left[\|x-y\|\right] \quad (1)$$

Where, $\inf(\bullet)$ refers to the lower bound of expectations of the function; $\prod(p_r, p_g)$ means the set of the joint distribution (γ) of p_r and p_g ; unique Kantotovich-Rubinstein dual form[20] is used since it is impossible to solve directly $\inf_{r\sim\prod(p_r,p_g)}$ in Wassersterin distance:

$$W(p_r, p_g) = \frac{1}{K}\sup_{\|f\|_L \le K} E_{x\sim p_r}[f(x)] - E_{x\sim p_g}[f(x)] \quad (2)$$

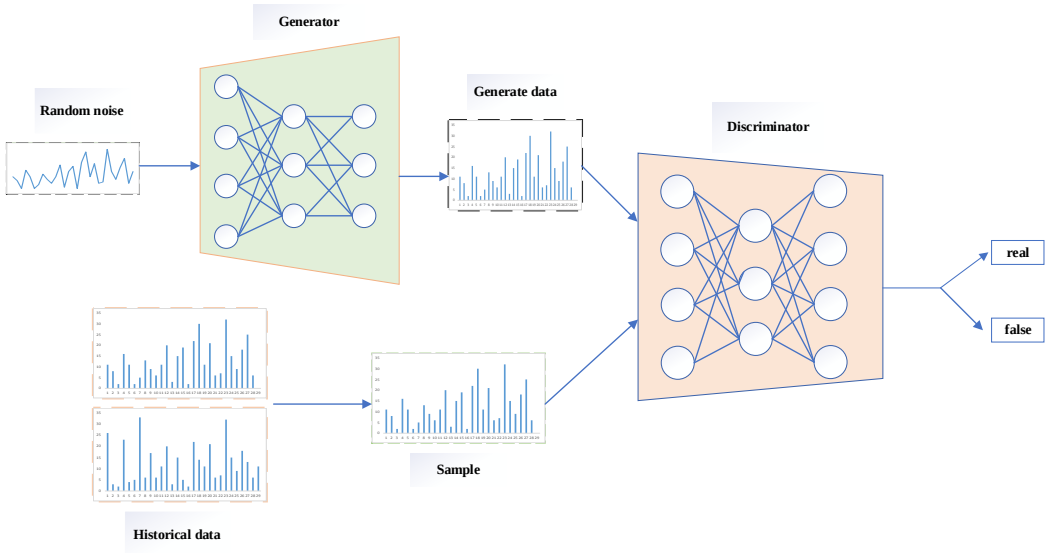


Fig.1 Network Structure of GAN

Gradient clipping optimization is carried out for the parameters of the discriminator so that its parameters are constrained into a relatively small range and the discriminator network can converge rapidly.

3. GENERATIVE ADVERSARIAL NEWORK ORIENTED MISSING DATA RECONSTRUTION

3.1 Generation Adversarial Network Oriented Missing Data Reconstruction Network Structure

Inspired by the work of generative adversarial network in 2D image field, a generative adversarial network frame for repair and completion of missing one-dimensional measurement discrete data is established. Please refer to Fig. 2 for the overall frame for reconstruction of missing data in the paper. Discrete data points acquired are taken as the input of the network. For the reconstruction of discrete data points, the neural network training slows down when there are too many discrete data features. Therefore, 1-D data are converted into 2-D matrix as the input of the generator. Each row of the matrix consists of sampling point coordinates (x, y, z). To improve the training speed of the model, only three convolutional layers and fully connected layers are designed for the module generated in the paper. The first input layer is the high-dimensional hidden variable random noise. The noise (z) is input into the first convolution kernel by converting into $M \times 5$ form linearly using a Linear function for dimensional reduction and add normalized processing for each convolution layer. ReLU is used as the activation function in all layers of the generator for final output of the data except in output layer where Thnh function is used. A data with the number of filter of 5 is finally output, i.e. maximum value and minimum value of AC positioning voltage, the maximum value and minimum value of DC positioning voltage and the average value of DC voltage. The dimensionality of the output is kept consistent with the number of inputs through the adjustment of the parameters of each layer.

It is necessary for the discriminator to determine effectively if it is real data or the data generated in the data reconstruction network. Therefore, LeakyRelu activation function is used for the whole discrimination network to improve the model recognition accuracy. The convolutional layer of the discrimination network is generally consistent with that of the network generated. 0 or 1 is generally output through final sigmoid activation function, in which 0 is false and 1 is true. At the time, the parameters of the discriminator are returned to the model generated again. The parameters are fixed, and the secondary data are generated again, and iterated repeatedly until the data less than EM distance is generated.

When historical measurement data and hidden variable (z) meeting normal distribution are used for training of the generator and discriminator, adam optimizer is used to update network parameters and calculate different adaptive learning rates for different parameters. The parameter update each time will improve the speed of training of the model.

Table 1 Generator Parameter Structure

Description	Parameters	Value
Fully connected layer	Number of neurons	5×128
	Activation function	ReLU
	Number of filters	128
Convolution layer	Size of convolution kernel	3
	Activation function	ReLU
	Number of filters	64
Convolution layer	Size of convolution kernel	3
	Activation function	ReLU
	Number of filters	5
Convolution layer	Size of convolution kernel	2
	Activation function	tanh

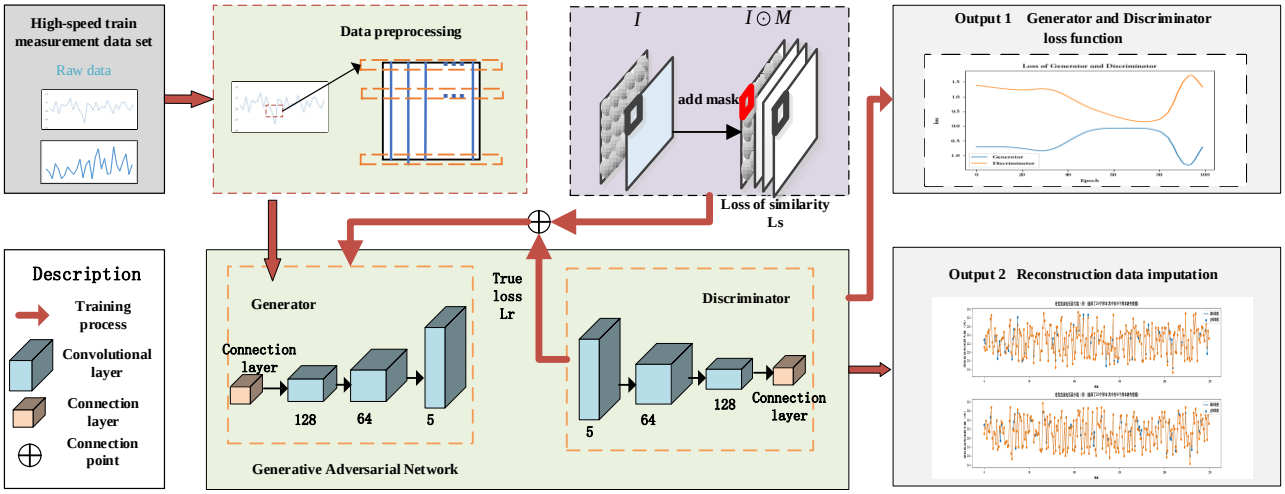


Fig. 2 Reconstruction Frame of High-Speed Train Measurement Data based on WGAN

3.2 Data Preprocessing

Generative adversarial network is a process of continuous data learning, but original data cannot be directly used in numerous data-driven methods. Therefore, the main purpose of data processing method is to extract characteristics from the original data. However, it is a complicated process to extract the characteristics required from one group of data, and the ability for WGAN to learn discrete data is limited. So 1-D form of the original data is converted into 2-D grid matrix not requiring parameter setting in the paper.

The complete sequence of data preprocessing without parameter setting is shown in Fig. 3. To obtain M×M matrix, it is necessary to take one section of 2M long from the original dataset. L(i) ($i=1,\cdots,M^2$) is used to refer to the value of each section of signal, as shown in following equation (6):

$$P(\mathrm{j},\mathrm{k})=\mathrm{round}\left\{\frac{L(\mathrm{j}-1)\times \mathrm{M}+\mathrm{k}-\mathrm{Min(L)}}{Max(L)-Min(\mathrm{L})}\times 255\right\} \tag{3}$$

Where, $round(\bullet)$ is a rounding function. 2×2 filters are used and the whole 2D network matrix is normalized to 255 from 0 in the paper. The value usually selected for M is 2n. The specific selection depends on the data value, e.g. 16, 32, 64, 128, etc.

The advantages of the data processing are as below: The original data can be processed and calculated without any preset parameters, and the 2D features of the original data can be explored so as to reduce the experiences of experts as appropriately.

3.3 Missing Data Reconstruction Algorithm

Countless data used to impute missing fragments can be generated through complete training set data in WGAN frame. Therefore, we need to select the data closest to real data and impute the data generated into the

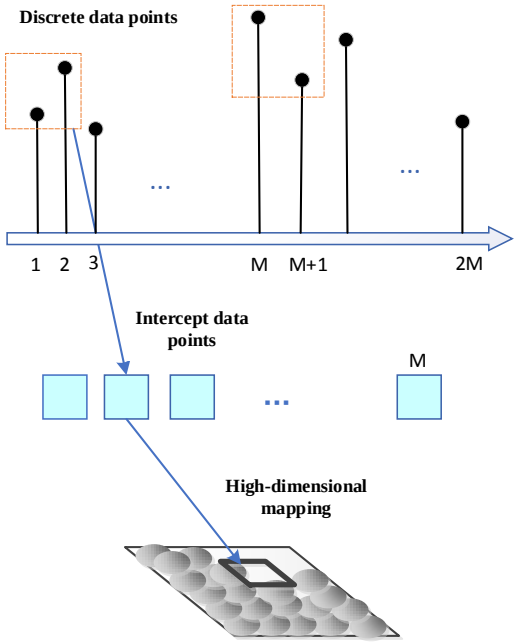


Fig. 3 Conversion of 1-D Data to 2-D Network Matrix

data not missing. The part generated by the generator and the part not missing meets the consistency of the context by comparing the issue of measurement data missing to image masking data reconstruction.

It is hard to describe for incomplete 1-D missing data. 1-D data are converted into 2-D tabular form in the paper in the way of data preprocessing, and a 2-D mask matrix of the same system measurement dimension is established for the measurement data missing. Data missing part is represented using 0 and the complete part is represented with 1 in the matrix.

To ensure that the complete data reconstructed are as close as possible to the real data and the data determined by the discriminator is real, the loss of the data imputed is defined as L_r.

$$L_r = D\left(G\left(z;\theta_G\right);\theta_D\right) \quad (4)$$

Where, $G(z;\theta_G)$ refers to the output data generated by the generator network; $D(\sim;\theta_D)$ refers to the output of Wasserstein distance between real data and reconstructed data.

To maintain the suitability of unmasked part and imputed part, similarity loss (L_s) is defined so that the generator will search continuously in the data generated for the most similar sample to the unmasked part in masked grid data.

$$L_s = \left\| G\left(z;\theta_G\right) \odot M_a, I \odot M_a \right\|_2 \quad (5)$$

In above equation, L2 norm is used to convert missing measurement data I to masked grid data. M_a is a binary mask matrix and $\odot$ refers to the operations between matrices.

In conclusion, the optimization goal for measurement data imputation is as below:

$$\min_{z\sim N(0,1)} \left(L_s + L_r\right) \quad (6)$$

Taking (6)as the final optimization goal using missing data, make the missing data generated close to the measurement data as far as possible using admin optimization function. The final imp-utation data consists of the unmasked part of the original data and the masked part generated [21] , i.e.

$$\hat{I} = I \odot M_a + G\left(\hat{z};\theta_G\right) \odot \left(1 - M_a\right) \quad (7)$$

4.EXPERIMENTAL RESULTS AND ANALYSIS

The program experiment of the paper is realized using python code. The program hardware environment includes CPU processor AMD Ryzen 5 2600X Six-Core Processor with the frequency of 3.60GHz; GPU: NVIDIA GeForce GTX 1660. The platform versions of Python 3.7.7 and torch 1.4.0 are used.

4.1 Data Set and Parameter Setting

GAN can learn well relevant features of image by the application of generative adversarial network in image field, and it is also relevant to the high correlation of image features. The learning of high-speed train measurement data must make sure high data correlation and large-scale data distribution.

The data set in this article uses the real data of a high-speed train running equipment for 32 consecutive days, and uses five types of characteristic values of each device (the maximum value and minimum value of AC positioning voltage, the maximum value and minimum value of DC positioning voltage and the average value of DC voltage), a total of 1280 sets of data. In order to maintain the learning ability and generalization ability of the model, the data is divided into the training set and the test set at a ratio of 4:1(1024 sets of training set and 256 sets of test set), and the number of model training is set to 100.

It should be noted that the training set used for the entire generative model must be 1024 sets of complete data sets. After the model training is completed, 256 sets of data with missing values are input into it.

To ensure rapid convergence and stability in network training, zero mean normalization must be carried out for the coordinates of complete data of the training set and the test set data with missing data, i.e. normalize the data point coordinates into the range of [-1，1]. The initial learning rate of the generator and discriminator model is set to 0.001. BATCH_SIZE is set to 1,290 for 100 Epoch cycles to improve the rate of the model for processing the data. Good learning monitoring effect can be obtained using the model structure in the paper.

As shown in Fig. 4, the loss functions of the generator and discriminator of the generative adversarial network in the paper are in 0-100 iterations. With the increase of the times of iterations, the images of the two loss functions tend to confrontation quickly, and the adversarial process of the whole model is very stable.

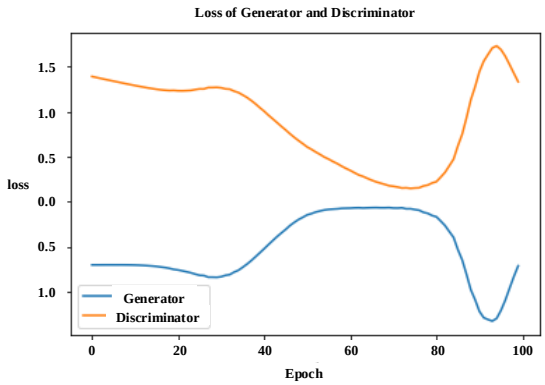


Fig. 4. Diagram for Loss of Generative Adversarial Network

4.2 WGAN Oriented Data Reconstruction Effect

The sample data is divided into training set (X_{train}) and test set (X_{test}) in the paper with the training set dimension of (1024, 40, 5) and test set dimension of (256, 40, 5). Missing at random is carried out for the test set based on the missing rate of 10%-50% respectively. Then binary mask matrix $Ms^{(40\times1)}$ is added for the test sets containing missing data. The part with missing data is 0 while the other part is 1. The number of zero in the mask matrix is

changed to realize the change in the number of missing data.

The test set (X_{test}) containing missing data is input into the model trained. If the data of a characteristic value is missing for a device, the characteristic value of other devices on the day or the passed day can be learned, and the missing part can be generated and reconstructed in combination with the law of other non-missing characteristic values of the device. The Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) obtained based on the reconstruction results are shown in Table 2 and 3.

As shown in Table 3 and 4, the data reconstruction effect is very good and relatively high accuracy is maintained when the data missing rate is within 10%-50%. With continuous increase of the system missing rate, both the MAE and MAPE of the five characteristic values change to different extents. This may be caused by the missing of critical values of measurement data to different extents due to random missing data. Therefore, when data reconstruction of high accuracy is maintained in case of a large number of missing of critical measurement data, it is demonstrated that the model trained is very superior in the capability of learning data context relationship and the reconstruction value of missing data is also very reasonable. Good reconstruction effect can still be maintained when the missing rate is at 50% in the paper. This demonstrates that the model can process a large amount of missing data for high-speed trains.

As shown in Table 2, when the missing rate is 50%, the method in this paper is compared with the traditional method. It can be found that the interpolation effect of the deep learning method in this paper is significantly better than the traditional method under the same missing rate.

Table 2 Comparison of Mean Absolute Percentage Error of different interpolation methods

Method	Max. value of DC positioning voltage
This article	1.173
Mean imputation	25.61
EM[4]	9.43

High-speed trains may cross through mountain area, tunnel and a variety of complex environments, and there may be network failure, harmonic interference and other special conditions causing random data missing. Long-term measurement data missing may also be caused by train skylight equipment maintenance. The part generated shall be imputed into the original data and the reconstruction effect of measurement data is observed considering if the measurement data generated comply with the contextual data distribution features. As shown in Fig. 5,the change law of the reconstruction results of the five groups of characteristic variables (i.e. maximum value and minimum value of AC positioning voltage, the maximum value and minimum value of DC positioning voltage and the average value of DC voltage) of the model in the paper is similar to that of actual measurement results. This demonstrates that, the reconstruction results generated by the WGAN model in the paper can reflect well the change trend of voltage and current and the reconstruction accuracy is high. Due to extreme historical data at some points, these extreme points may cause some minor deterioration occurred in the reconstruction effect.

Table 3 Mean Absolute Error (MAE) of Reconstruction Results

Missing rate	Max. value of DC positioning voltage	Min. value of DC positioning voltage	Average value of DC positioning voltage	Max. value of AC positioning voltage	Min. value of AC positioning voltage
0.1	0.048002	0.274403	0.042278	0.005850	0.3364296
0.2	0.101603	0.339498	0.112054	0.024648	0.7465155
0.3	0.097708	0.383544	0.102786	0.023610	0.6204315
0.4	0.123408	0.482997	0.087946	0.012937	0.4020587
0.5	0.100611	0.437346	0.080377	0.018360	0.4442114

Table 4 Mean Absolute Percentage Error (MAPE) of Reconstruction Results

Missing rate	Max. value of DC positioning voltage	Min. value of DC positioning voltage	Average value of DC positioning voltage	Max. value of AC positioning voltage	Min. value of AC positioning voltage
0.1	0.7600207	2.1878389	0.7193886	0.0883874	0.6998789
0.2	1.2674115	2.2842678	1.3353079	0.1783063	1.0890092

0.3	1.2004649	2.5261751	1.1927532	0.1731615	0.9901325
0.4	1.3142797	2.9769532	1.1141013	0.1282863	0.7672538
0.5	1.1730746	2.8098599	1.0829699	0.1577947	0.758148

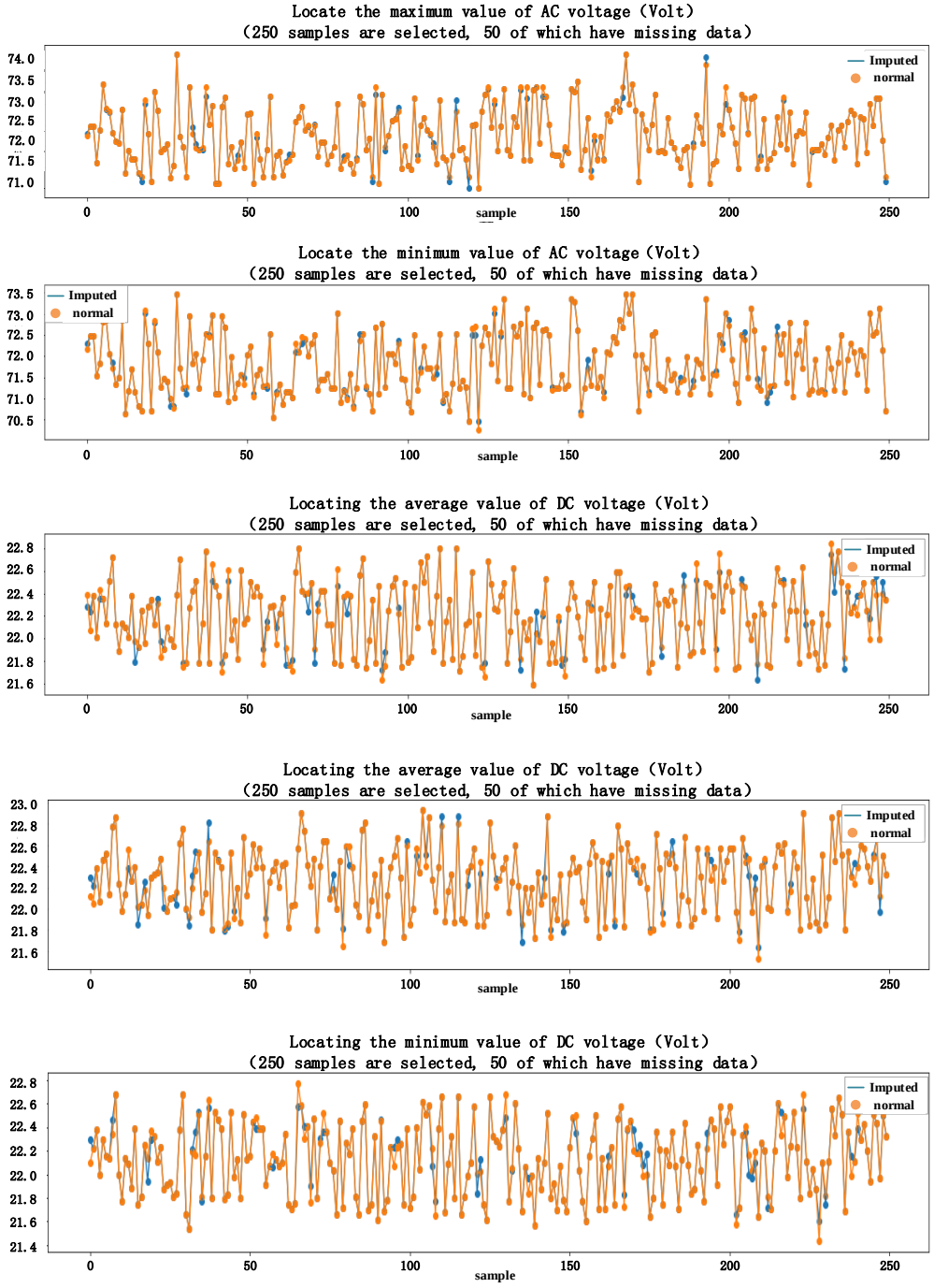


Fig. 5 Imputation of Reconstruction Data

5. CONCLUSIONS

For the missing of voltage and current measurement data of high-speed trains, a frame for WGAN oriented data reconstruction is proposed in the paper. Discrete 1-D data point is taken as the input in the frame to convert the discrete data point into 2-D grid form. The process can speed up the convergence of the model and maintain the stability of model training. The missing data can be imputed through the generative adversarial network frame. WGAN learns the correlation between characteristic values of each device in unsupervised pattern and generates countless data suitable for the distribution law for missing part based on the context of the missing part of measurement data. The discriminator guides continuously the optimization of the generator until the constraints are met. Then the reconstruction results can be output. Relatively

high reconstruction accuracy can still be maintained at different missing rates of measurement data for high-speed trains in the paper. Further study on reconstruction and imputation in case of multi-source heterogeneous data missing can be carried out, to understand how to impute the type of missing data with other types of data so that the learning of zero sample is realized.

Acknowledge

This work was supported by the Natural Science Foundation of China (U1934219, 61773159). Project of Hunan Provincial Department of Education (19A137), Hunan Provincial Natural Science Foundation of China (2020JJ6083).

References

[1] J. He, G.W. Liu, C.F. Zhang, et al. "Maximum likelihood identification method of adhesion performance parameters of Heavy-Duty locomotive", Journal of Electronic Measurement and Instrument,Vol. 31, No. 02, 2017, pp. 170-177.

[2] D. V. Mehrotra, F. Liu, T. Permutt. "Missing data in clinical trials: control-based mean imputation and sensitivity analysis". Pharma-ceutical Statistics,Vol. 16, No. 5 , 2017 , pp .378-392.

[3] X. Ma, F. Xu, B. Chen. "Interpolation of wind pressures using Gaussian process regression", Journal of Wind Engineering and Industrial Aerodynamics, Vol. 188 , 2019, pp. 30-42.

[4] S. Hua-Yan, L. Ye-Li, Z. Yun-Fei and H. Xu, "Accelerating EM Missing Data Filling Algorithm Based on the K-Means",2018 4th Annual International Conference on Network and Information Systems for Computers (ICNISC), 2018,pp. 401-406.

[5] V. Miranda, J. Krstulovic, H. Keko, et al. "Reconstructing Missing Data in State Estimation With Autoencoders", IEEE Transactions on Power Systems, Vol.27,No. 2, 2012, pp.604-611.

[6] Y. F. Li, X. Y. Cao, H. J. Chen, et al. "Method for detecting wear status of train external gear based on machine vision", Journal of the China Railway Society, Vol.40, No. 12, 2018, pp.33-41.

[7] B. Zhao, M. R. Dai, P. Li, et al. "Research on defect detection of railway key components based on deep learning", Journal of the China Railway Society, Vol.41, No.08, 2019, pp.67-73.

[8] W. Lao , L. Peng , Y. Li . "An image reconstruction method for improving resolution of capacitive wire mesh tomography", 2017 IEEE International Conference on Imaging Systems and Techniques (IST), 2017.

[9] J.S. Jiang, H.R. Ren, H.Y. Li. "Seismic data processing based on convolutional autoencoder", Journal of Zhejiang University(Engineering Science), Vol.54, No.5, 2020, pp.978-984.

[10] K. D. B. Mudavathu, M. V. P. C. S. Rao and K. V." Ramana, Auxiliary Conditional Generative Adversarial Networks for Image Data Set Augmentation", 2018 3rd International Conference on Inventive Computation Technologies (ICICT) ,2018, pp. 263-269.

[11] N.M. Yao, Q.P. Guo, F.C. Qiao ,et al. "Robust Facial Expression Recognition With Generative Adversarial Networks", Acta Automatica Sinica,Vol.44, No.05, 2018, pp, 865-877.

[12] X. Yao, H. Yang and Y. Li, "Modulation Identification of Underwater Acoustic Communications Signals Based on Generative Adversarial Networks", OCEANS 2019 - Marseille, Marseille, France, 2019, pp.1-6.

[13] M. Arjovsky，S. Chintala，L Bottou．Wasserstein GAN．arXiv preprint, 2017．

[14] S.X. Wang, H.W. Chen, Z.X. Pan, et al. "A Reconstruction Method for Missing Data in Power System Measurement Using an Improved Generative Adversarial Network", Proceedings of the CSEE, Vol.39, No.01, 2019, pp.56-64.

[15] T. Ince, S. Kiranyaz, L. Eren, et al. "Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks", IEEE Transactions on Industrial Electronics, Vol.63, No.11, 2016, pp. 7067-7075.

[16] B. Li, P. Zhang, D. Liu, et al. "Feature extraction for rolling element bearing fault diagnosis utilizing generalized S transform and two-dimensional non-negative matrix factorization", Journal of Sound and Vibration, Vol.40, No.330, 2011, pp.2388-2399.

[17] L. Wen , X. Li, L. Gao, et al. "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method", IEEE Transactions on Industrial Electronics, Vol.65, No.7, 2018, pp. 5990-5998.

[18] Y.W. Miao, J.Z. Liu, J.H. Chen, et al. "Structure-preserving shape completion of 3D point clouds with generative adversarial network" (in Chinese). Sci Sin Inform, Vol.50, No.05, 2020, pp. 675–691.

[19] C. Zhang, Y. Feng, B. Qiang and J. Shang, "Wasserstein Generative Recurrent Adversarial Networks for Image Generating", 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 242-247.

[20] H. Lou, Z. Qi and J. Li. "One-dimensional data augmentation using a wasserstein generative adversarial network with supervised signal", 2018 Chinese Control And Decision Conference (CCDC), 2018, pp. 1896-1901.

[21]]Luo, Yonghong, et al. "E²GAN: End-to-End Generative Adversarial Network for Multivariate Time Series Imputation", international joint conference on artificial intelligence ,2019,pp. 3094-3100.

Sparse Representation Based Googlenet for Indoor Scene Recognition

Wenhao Duan*,**, Luefeng Chen*,**,†, Min Li*,**, Min Wu*,**, Pingping Zhang*,**,
Kuanlin Wang*,**, and Witold Pedrycz***

*School of Automation, China University of Geosciences, Wuhan 430074, China
**Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
***Department of Electrical and Computer Engineering, University of Alberta, Edmonton T6R 2G7, Canada

† Correspinding author: Luefeng Chen (chenluefeng@cug.edu.cn)

Abstract

Indoor scene recognition is of great theoretical research significance and practical application value. Scene recognition achieve good results in the field of computer vision. How to identify the features of indoor complex scenes arises as the main question. In this paper, we propose a sparse representation algorithm to recognize indoor scene. It can not only solve the problem of over fitting and gradient disappearing, but also can effectively extract the characteristics of indoor scene while increasing the width and depth of neural network. The training of the method is so fast that it can meet the real-time demand of the recognition system to a certain extent. On this basis, a human-robot interaction system for indoor scene recognition can be built. This paper uses the GUI tool of MATLAB to build a graphical user interface, which can realize simple human-robot interaction. The experimental results show that this method comes with good recognition performance and the recognition accuracy reaches 92.8%.

Keywords: Indoor scene recognition, sparse representation algorithm, deep learning, auxiliary classifier, human-robot interaction.

1. INTRODUCTION

Scene recognition is the core research field of artificial intelligence [1], which focuses on the use of feature information of scene image to classify the scene of image. Scene recognition is widely used in human-robot interaction, so the machine can recognize the scene information in the image. Therefore, it plays an important role in image retrieval, intelligent robot, intelligent security and other fields [2].

Compared with the outdoor scene, the image content of indoor scene which includes multiple objects is more complex, and there is occlusion between objects [3], so it is difficult to extract scene features. Early indoor scene recognition mainly uses middle-level features and high-level semantic features, and the recognition effect depends on the selected features [4] which cannot effectively eliminate the interference of indoor objects and extract indoor scene features accurately. Deep learning algorithm has resulted in great achievements in many aspects [5], so more and more scholars begin to study deep learning [6] to solve the indoor scene classification problem.

In this paper, several sparse structure [7] convolution neural networks are used to extract and recognize the features of scene images, which has the characteristics of fast training speed, high recognition rate and significant classification effect.

2. INDOOR SCENE RECOGNITION

Aiming at solving the problem of extracting indoor scene features, this paper uses sparse representation algorithm to extract image information through multiple convolution kernels, which gets better representation of the image. The frame structure of the algorithm is shown in Fig.1.

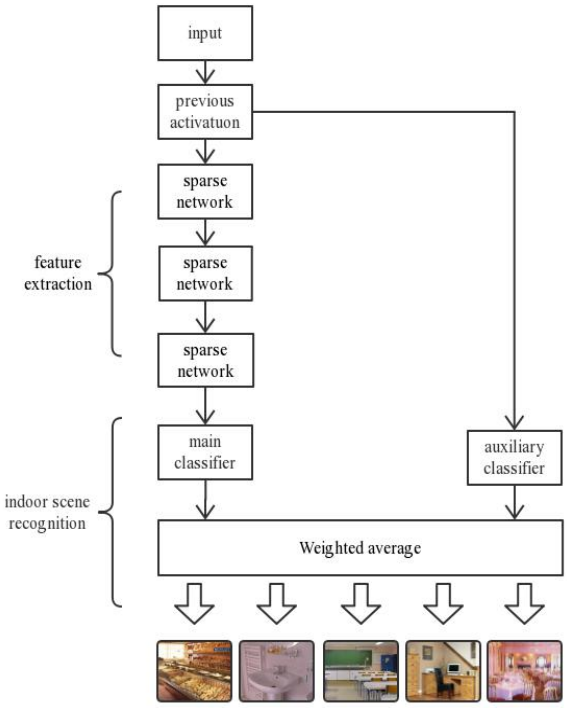


Fig.1. The framework of indoor scene recognition.

Multiple sparse convolution neural networks can also effectively solve the problem of over fitting and gradient disappearance [8], and extract the features of complex indoor scenes well and accelerate the training speed of neural networks. In addition, adding an auxiliary classifier can take advantage of the features of the middle layer. Combined with the main classifier, the recognition rate of indoor scene can be improved. Therefore, the indoor scene recognition algorithm based on convolution neural network has advantages of fast training speed, high recognition rate and significant classification effect. The frame structure of the algorithm is shown in Fig.1.

2.1 Indoor scene feature extraction

When the amount of data is small, the trained network is prone to over fitting [9], and it is easy to cause the phenomenon of gradient disappear when the network has a very deep depth. The image feature extraction method based on sparse representation algorithm can effectively solve these two side effects. In the same layer, there are 1×1, 3×3, 5×5 convolution and pooling layers. When using filter for convolution operation and pooling layer for pooling operation, zero padding [10] will be used to ensure that the output is of the same size. After these operations, the output results, namely feature maps, are all integrated together. The sparse network is shown in Fig.2.

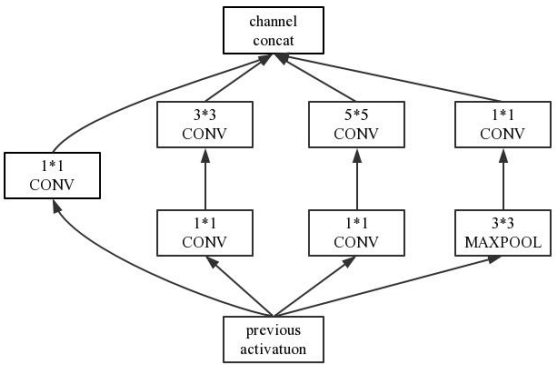


Fig.2. The sparse network.

The characteristic of sparse network is that in the same layer, different size features of input from the upper layer can be collected after passing through different size filters and then through pooling layer. This will increase the width of the network so that we can indirectly improve the performance of feature extraction. In the meantime, 1 × 1 convolution layer is added before 3×3 and 5×5 convolution layer, and after pooling layer. There are two advantages in this way: one is that it allows the number of cells in each step to be increased, and the computational complexity will not be out of control. The dimensionality reduction technique can make sure that a large number of input filters go from one level up to the next. Secondly, different scales of visual information are processed and fused, so that the features of different scales can be extracted simultaneously in the next step.

2.2 Indoor scene recognition

In the training stage, the main classifier is used in the top layer, and we add an auxiliary classifier in the middle one. The two classifiers are weighted according to different weights to get a classification model to minimize the loss value. In the training stage, the auxiliary classifier is removed and only one main classifier is used for indoor scene recognition and classification.

In image classification task, softmax function is often used as classifier because of its simple calculation and high efficiency. The cross entropy loss on account of softmax function is generally used in multi classification problems. The cross entropy loss based on softmax [11] activation is recorded as follows:

$$L = \frac{1}{N}\sum_i L_i = \frac{1}{N}\sum_i - log\left(\frac{e^{f_n}}{\sum_j e^{f_i}}\right) \quad (1)$$

where y_i is the tag of the ith input, $i \in [0, K)$, and K is the number of categories, f_j is the jth element of the class output vector f of the final full connection layer, $j \in [0, N)$, and N is the number of training samples. Since f is the output of the activation function of the fully connected layer, when the ith input of the fully connected layer is X_i and the ith weight is Wy_i, f_{y_i} can be expressed as follows:

$$f_{y_i} = W_{y_i}^T X_i = \|W_{y_i}\|\|X_i\| \cos(\theta_{y_i}) \quad (2)$$

The final loss function can be expressed as follows:

$$L_i = - log\left(\frac{e^{\|W_{y_i}\|\|X_i\| \cos(\theta_{y_i})}}{\sum_j e^{\|W_{y_j}\|\|X_j\| \cos(\theta_{y_j})}}\right) \quad (3)$$

where $0 \le \theta_i \le \pi$. Softmax is widely used in classification tasks based on deep convolution neural network. However, this form of learn has no difference for each category, so it may not effectively learn the features that are more com-pact within the class and more discrete between classes. Taking binary classification [12] as an example, the final output of the network is the probability of category 1 and category 2. If the category of input data belongs to category 1, the purpose of softmax activation function is to make the probability of output category 1 greater, even if $f_1 > f_2$, that is $W_1^T X > W_2^T X$, the above inequality is expressed as follows:

$$\|W_1\|\|X\| \cos(\theta_1) < \|W_2\|\|X\| \cos(\theta_2) \quad (4)$$

The correct classification result of X is obtained.

Therefore, we can get the categories of indoor scene images.

3. RESULT AND ANALYSIS ON INDOOR SCENE RECOGNITION

3.1 Data setting

The database of indoor scene used in this experiment is indoorcvpr_09 database, which includes 67 indoor categories, a total of 15620 images. And different kinds of scene categories have different number of images. Each category has more than 100 images, and all of the images are only in JPG format.

In the experiment, a 3:7 ratio of test samples is selected for the simulation test, that is, 70% of the images in the sample library are used to train, and the left 30% of the images are used to test.

3.2 Simulations and analysis

This section will analyze and discuss the influence of parameters on the performance of indoor scene recognition algorithm which is on account of deep learning. In addition, the recognition effect of this algorithm and Multi-resolution classical neural network on this database will be compared to verify the effectiveness of the algorithm.

A. The influence of learning rate

For deep learning, learning rate is an important parameter which will affect the speed of adjusting neural network weights on the basis of loss gradient. As shown in (5), the decline speed of loss gradient will be slower and the convergence time will be longer when the learning rate get smaller:

New_weight = Current_weight–Learning_rate*gradient (5)

However, if the learning rate is too small, the gradient descent will be too slow to work efficiently. And if the learning rate is too large, the gradient descent step may cross the optimal value.

In this section, we will discuss the influence of learning rate [13] on the recognition of indoor scene. Therefore, we design a control variable experiment. Each experiment only changes the size of learning rate while keeping other training parameters unchanged. We will get the indoor scene recognition rate and training time with learning rate changes as shown in Table 1 based on indoorcvpr_09 database original image samples.

The experimental results show that the modeling time increases as the learning rate increase, and the recognition rate reaches the highest when the learning rate is 0.0001. This shows that during the course of training, choosing the appropriate learning rate in the googlenet neural network can improve the recognition rate of the classifier and reduce the test time. In view of the influence of learning rate on test time and recognition rate, the optimal value of learning rate is 0.0001, which can effectively reduce the test time and obtain good recognition effect.

Table 1. Indoor scene recognition rate when changing the learning rate.

Learning rate	Recognition rate/%	Training time/min
0.1	63.4	11.3
0.01	72.3	13.2
0.001	81.1	17.6
0.0001	92.8	19.3
0.00001	79.4	23.1

B. Comparison and analysis

For purpose of verifying the accuracy of indoor scene recognition, we use the Multi-resolution CNNs [14] and Googlenet neural network for comparative experiments.

In the first group of experiments, Googlenet is used, and 100 images are selected as test pattern in the database, and the rest are used as training pattern; in the second group of experiments, Multi-resolution CNNs is used, using the same training pattern and test pattern. The indoor scene recognition results are shown in Table 2.

Table 2. Recognition effect of two methods with the same learning rate.

Model	Recognition rate/%	Training time/min
Googlenet	92.8	19.3
Multi-resolution CNNS	83.1	27.9

The results show that the accuracy rate of indoor scene recognition based on Googlenet is 92.8%. Compared with the Multi-resolution CNNs, the recognition rate is improved by 9.7% by using Googlenet.

The recognition rate and error classification of various scenes in the two groups are shown in the confusion matrix in Table 3 and 4.

Table 3. Confusion matrix of recognition results by using Googlenet.

	bakery	bathroom	classroom	office	restaurant
bakery	18	0	0	0	1
bathroom	0	19	1	1	0
classroom	0	0	14	0	0
office	0	0	3	13	0
restaurant	2	1	2	6	19

Table 4. Confusion matrix of recognition results by using Multi-resolution CNNS.

	bakery	bathroom	classroom	office	restaurant
bakery	18	1	0	0	1
bathroom	0	19	1	0	0
classroom	0	0	16	0	0
office	0	0	2	20	0
restaurant	2	0	1	0	19

The experimental results show that the proposed algorithm can effectively identify the above five indoor scenes, and has the characteristics of shorter training time, shorter test time and higher recognition rate compared with the traditional algorithm.

3.3 The visual design

Compared with the expression of language, graphics can be abstracted into concrete, which is easier to be accepted by the public, which is beneficial to the development of human-robot interaction.

Therefore, the visual design of indoor scene recognition system with Matlab GUI tools can make the indoor scene recognition system more intuitive, specific and convenient. The indoor scene recognition system has three simple functions: selecting pictures, testing pictures and displaying results. Its interface is shown in Fig.3.

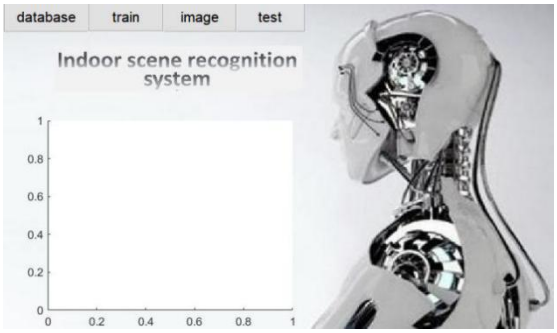


Fig.3. The interface of indoor scene recognition system.

Table 5. Confusion matrix of recognition results of indoor scene recognition system.

	bakery	bathroom	classroom	office	restaurant
bakery	10	0	0	0	1
bathroom	0	10	0	0	0
classroom	0	0	8	1	0
office	0	0	1	9	0
restaurant	0	0	1	0	9

The experimental confusion matrix shown in Table 5 indicates that the indoor scene recognition system obtains applicable performance.

4. CONCLUSION

The proposed method of feature extraction based on sparse representation algorithm put forward in this paper can effectively eliminate the interference of indoor objects and extract indoor scene features accurately. We establish a feature recognition model of Googlenet indoor scene using three sparse networks, and auxiliary classifier to extract middle layer features, and optimize the final classifier model. Therefore, the training and testing time of the algorithm are shorter and the recognition rate is higher. In addition, according to the proposed indoor scene recognition algorithm, the visual design of indoor scene recognition system is carried out by using Matlab GUI tool, which makes the indoor system too easy to understand and convenient.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants 61973286, 61603356, and 61733016, the 111 Project under Grant B17040, and the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) (No. 201839).

References

[1] D. Hubel, T. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 1962: 106-154.

[2] M. Pandey, S. Lazebnik. Scene recognition and weakly supervised object localization with deformable part-based models. In *Proceedings of International Conference on Computer Vision,* Barcelona, Spain, 2011: 6-10.

[3] F. Li. A bayesian hierarchical model for learning natural scene categories. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, California, 2005: 64-75.

[4] J. Wu, J. M. Rehg. Centrist: a visual descriptor for scene categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 33 (8): 489-501.

[5] D. Yoo, S. Park, J. Y. Lee, et al. Multi-scale pyramid pooling for deep convolutional representation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Boston, USA, 2015: 71-80.

[6] S. Khan, M. Hayat, M. Bennamoun, et al. A discriminative representation of convolutional features for indoor scene recognition. *IEEE Transactions on Image Processing*, 2016, 25 (7): 3372-3383.

[7] A. Krizhevsky, L. Sutskever, G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Proceedings of Advances in neural information processing systems*, Lake Tahoe, USA, 2012: 1097-1105.

[8] L. Feifei, P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, USA, 2005: 524-531.

[9] Y. Lin, P. Dollár, R. Girshick, et al. Feature pyramid networks for object detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, America, 2016: 1009-1013.

[10] A. Colombari, A. Fusiello. Patch-based background initialization in heavily cluttered video. *IEEE Transactions on Image Processing*, 2010, 19 (4): 926−933.

[11] A. Oliva, A. Tomalba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 2001, 42 (3): 145-175.

[12] M. Hofmann. Background segmentation with feedback: The pixel-based adaptive segmenter. In *Proceedings of IEEE Workshop on Change Detection*, Providence, RI, 2012: 38-43.

[13] C. Yang. C. Wang, Y. Zhao. LLR: Learning rates by LSTM for training neural networks. *Neurocomputing*, 2020: 12-19.

[14] L. Wang, G. Sheng, W. Huang, Y. Xiong, Y. Qiao. Knowledge guided disambiguation for large-scale scene classification with Multi-Resolution CNNs. *IEEE Transactions on Image Processing*, 2017, 17 (3): 39-43.

Numerical Simulation of Metal-Free Water Cannon

Tomomasa OHKUBO*, Ei-ichi MATSUNAGA*, Yuji SATO**
* Department of Mechanical Engineering, Tokyo University of Technology
Hachiouji, Tokyo 192-0982, Japan
** Joining and Welding Research Institute, Osaka University
Ibaraki, Osaka 567-0047, Japan

Abstract

Laser propulsion is expected as the next generation propulsion mechanism. Especially, metal-free water cannon realized propulsion without metallic target. In this paper, we developed numerical simulation code implementing C-CUP method to simulate laser induced bubble and metal-free water cannon. We succeeded in reproducing qualitative behavior of spouting water in 3-dimensional space when the metal-free water cannon is irradiated by laser. Furthermore, calculated results about time development of displacement of the metal-free water cannon are qualitatively agreed with experimental results. We simulated the behavior of the laser-induced bubble and realized simulating that the bubble inhales the water once spouted out and the target moves backward due to the pressure difference generated by the bubble expansion and collapsing and inhaling reaction. Furthermore, we realized simulating that the laser-induced bubble repeats this expansion and collapsing and the target moves forward while vibrating back and forth.

Keywords: Laser propulsion, laser induced bubble, C-CUP method, computational fluid dynamics, ns-pulse laser

1. INTRODUCTION

Laser propulsion is expected as the next generation of 20propulsion mechanism. The main feature of laser propulsion is that it does not require energy source on the target and energy can be supplied from outside to realize propulsion.

The basic concept of laser propulsion is proposed by Kantrowitz in 1972 [1]. Although the initial concept is proposed in 1972, authentic researches about it started in the end of 20th century. The first experiment of launching a target by CO_2 laser is realized in 1997 by the team of Myrabo [2]-[4].

In the 21st century, several methods of laser propulsion were proposed. Almost all of these researches realized propulsion getting reaction force from ablation of sine solid materials or breakdown in the air. However, considering the cannon ball theory which is studied for nuclear fusion[5]-[8], Phippse et al. realized getting several hundred of coupling coefficient using transparent materials to get larger reaction force [9].

Furthermore, we realized getting larger coupling coefficient of several thousand using water as transparent material and Nd:YAG laser as energy source [10]-[16].

Especially, we proposed the metal-free water cannon which does not require metallic target of laser ablation. The basic concept of the metal-free water cannon is shown in Fig. 1. The irradiated laser is focused into acryl pipe filled with water and the breakdown occurs inside the water. The breakdown generates a bubble of high density and high pressure and the bubble push the water out. The target is moved by the reaction of the spouting water. In the previous study, we realized several thousand of coupling coefficient using this metal-free water cannon [17].

Although the laser induced bubble inside water is studied these days [18][19], their target is thin water layer and they are experimental studies and the bubble behavior is not explained theoretically. It is very hard to perform theoretical study about these kinds of system because they have both compressible fluid of water and incompressible fluid of the laser induced buttle whose numerical solvers are usually completely different.

Furthermore, we observed the curious behavior of the metal-free water cannon using high-speed camera.

Therefore, in this study, we introduce the curious behavior of the metal-free water cannon at first and we discuss how the curious behavior was realized by considering results of our numerical simulation.

2. CURIOUS BEHAVIOR OF THE METAL-FREE WATER CANNON

We put the metal-free water cannon whose diameter is 6mm and length is 14mm shown in Fig. 1 on a pendulum. The target is an acryl pipe filled with water

with lens on the right side. We irradiate laser from lens side and the laser is focused in the water by the lens. A bubble is generated at the focal point and part of water is spouted out leftward. Therefore, as a reaction of spouted water, the target moves rightward.

We used the Q-switched Nd:YAG Laser of 5ns and 300mJ (Continum, Surelite SL-I 10). The laser power was measured by thermopile power meter (OPHIR Optris LTD's 30A) with display of AN/2.

The behavior of the target was measured by the high-speed camera (Phantom V4.0 by Vision Research). We took the pictures of the target with a frame rate of 7000 fps.

We changed the distance *d* which is distance between the water surface and the focal point inside the metal-free water cannon by changing the lens put on the target.

The time development of displacement of the targets observed by the high-speed camera are shown in Fig. 2. In cases of d > 0, the target shows a curious behavior. It moves forward at first, and then moves backward once, and then moves forward again. Furthermore, the backward displacement is larger when *d* is larger. The displacement of each target at time of 2.0 milliseconds are shown in Fig. 3.

Although we realized getting several thousand of coupling coefficient in the previous study, the curious behavior of the backward motion of the metal-free water cannon must be clarified. Therefore, we developed numerical simulation code to simulate the behavior of the metal-free water cannon.

3. CALCULATION MODEL AND CALCULATED RESULTS

3.1 Governing equation

Although the laser-induced bubble is compressible fluid, the water in the metal-free water cannon is incompressible fluid. It is difficult to solve problems of fluid dynamics of such system because the difference of sound speed of compressible fluid and incompressible fluid is huge.

In this study, we used CIP method [20] and C-CUP method [21] which is universal solver for compressible and incompressible fluid. The government equations are below.

$$\frac{\partial \rho}{\partial t} + (\boldsymbol{u} \cdot \nabla)\rho = -\rho \nabla \cdot \boldsymbol{u}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\boldsymbol{u} = -\frac{1}{\rho}\nabla p + \nabla \cdot (\nu \nabla \boldsymbol{u}) + \frac{\boldsymbol{F_s}}{\rho}$$

$$\frac{\partial p}{\partial t} + (\boldsymbol{u} \cdot \nabla)p = -\gamma p \nabla \cdot \mathbf{u}$$

$\rho, t, \boldsymbol{u}, p, \nu, \boldsymbol{F_s}, \gamma$ are density, time, velocity, pressure, dynamic viscosity, surface tension and specific heat ratio respectively.

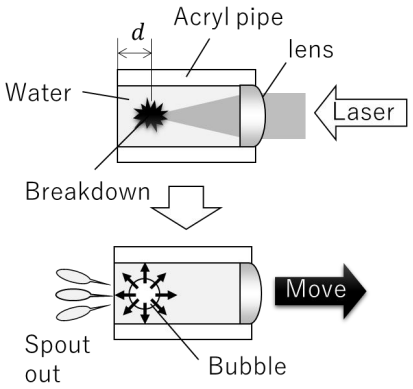


Fig. 1 Basic concept of the metal-free water cannon.

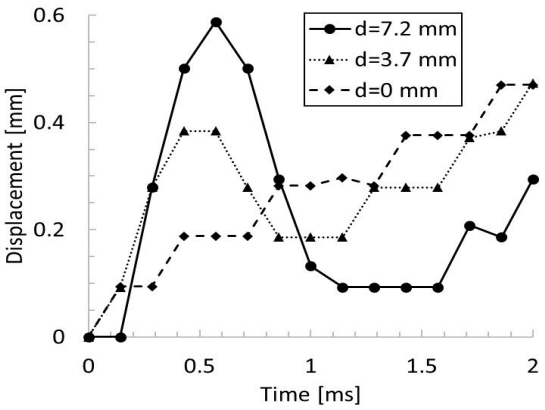


Fig. 2 Experimental results of displacement of the metal-free water cannon.

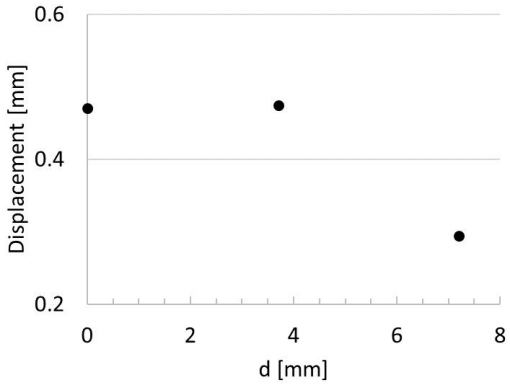


Fig. 3 Experimental results of displacement of the metal-free water cannon at time 2.0 ms.

3.2 Calculation geometry and parameters

We performed three-dimensional numerical simulation. The calculated geometry is shown in Fig. 4. We put an acryl pipe filled with water in the space of 30 x 10 x 10

mm. The acryl pipe of the outer diameter is 8mm and inner diameter is 6mm and length is 14mm. We put a bubble of diameter of 0.8mm inside water as initial state of the laser induced bubble. The distance from outlet of the acryl pipe to center of the bubble was parameter which is described as variable d d.

The coordinate of the calculation geometry is Cartesian coordinate and calculation mesh is 0.2 x 0.2 x 0.2 mm of structured grid.

The number of the grids is 150 x 50 x 50. We used the staggered mesh to stabilize solution of pressure p and velocity $\boldsymbol{u}$.

The boundary condition of the all boundaries was free boundary. The time step was controlled automatically considering stability of the advection term and the viscosity term.

The physical parameters used in our calculation are shown in Table 1.

Table 1 Physical properties used in the calculation

Pressure (water, air) [atm]	1.0
Pressure (bubble) [atm]	20.0
Density (water) [kg/m^3]	1.0×10^3
Density (air) [kg/m^3]	1.204
Density (bubble) [kg/m^3]	15.625
Dynamic viscosity (water) [m^2/s]	1.004×10^{-6}
Dynamic viscosity (air, bubble) [m^2/s]	15.12×10^{-6}
Specific heat ratio (air, bubble)	1.4
Specific heat ratio (water)	2.0×10^4

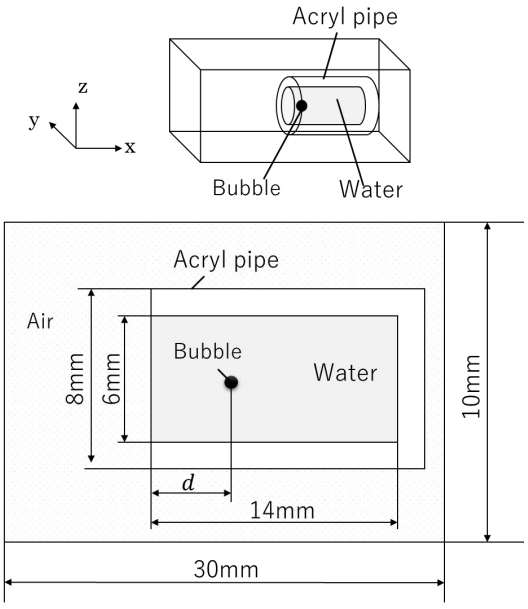


Fig. 4 Calculation geometry

3.3 Calculated results

Comparison between experimental results and simulation results is shown in Fig. 5 in three-dimensional space.

The state of the water ejection is qualitatively consistent with the calculation and the experimental results. Furthermore, the babble behavior of moving toward leftward is also consistent with the calculation and the experimental results. Therefore, this calculation model is reasonably valid at least qualitatively.

4. DISCUSSION

4.1 Displacement of the calculated results.

It is necessary to calculate the displacement of the target for comparison between experimental results shown in Fig. 2 and Fig. 3.

Because the position of the target is fixed in the calculation, it's necessary to consider acceleration to calculate the displacement of the target. The driving force of the target is pushing force on the lens of right side. Therefore, the acceleration of the target is calculated as below.

$$a = \frac{\int (p_A - p_0)\,\mathrm{d}S}{M}$$

$a, p_A, p_0, \mathrm{d}S, M$ are acceleration, pressure on the lens of right side, atmospheric pressure, area of each cell of y-z plane and total mass of the target respectively.

Using this acceleration a, we can calculate the displacement of the target D as below.

$$D = \iint a\ \mathrm{d}t\ \mathrm{d}t$$

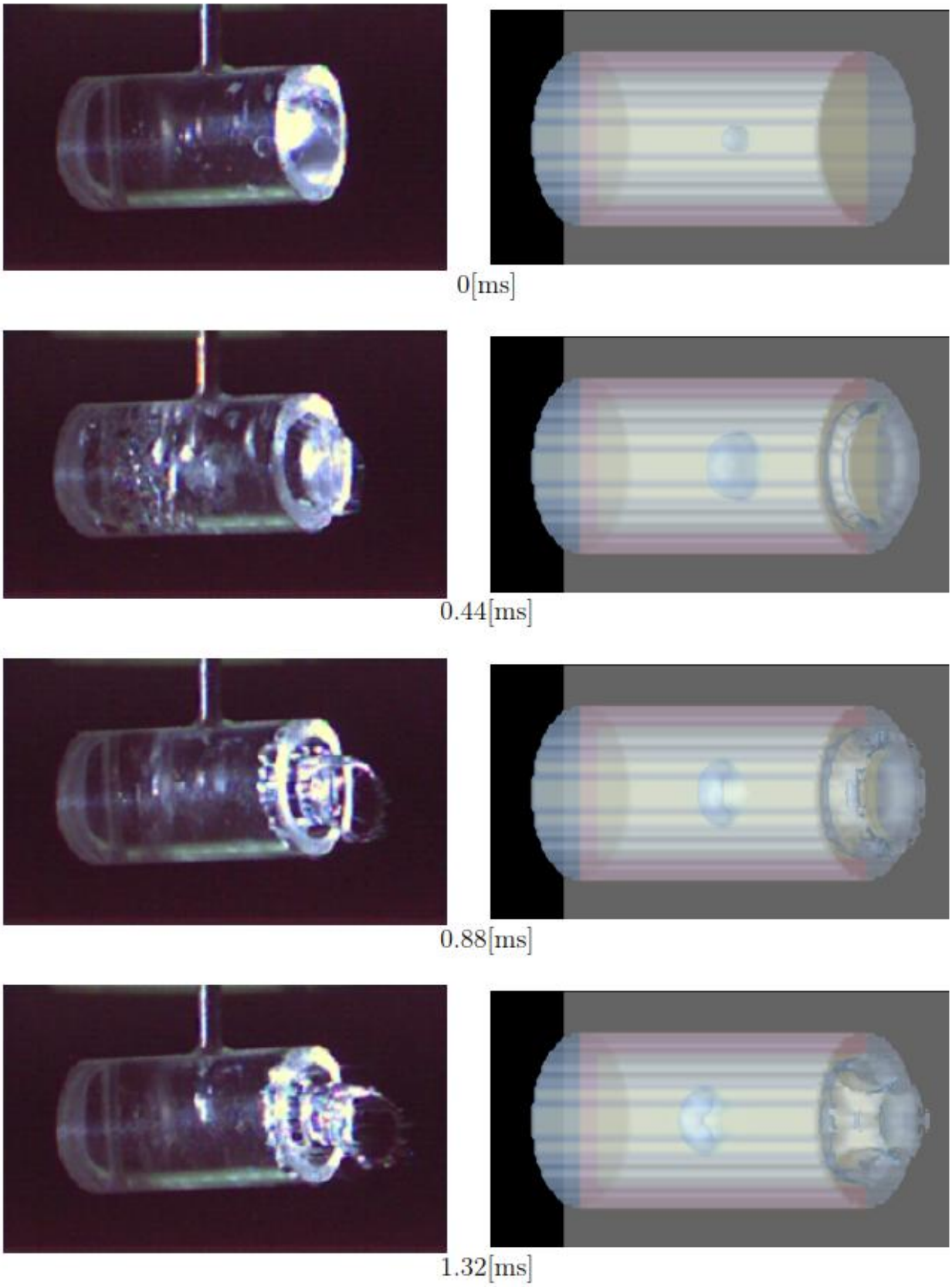
4.2 Time development of the displacement of the targets

The calculated time development of displacement of the metal-free water canon with each d are shown in Fig. 6. Because it was not possible to calculate stably in the case of d=0, the results of d=1 are shown instead.

Comparing the calculated results shown in Fig. 6 and the experimental results shown in Fig. 2, they are qualitatively comparable. When the bubble position d is small, although the displacement in the short time is small, it increases almost monotonically. On the other hand, when the bubble position d is large, it moves much larger distance in short time but it moves backward immediately and the final displacement is smaller than the case of smaller d.

4.3 The final displacement of the targets

The calculated results about the displacement of the target at time of 1.5ms when d is varied from 1 to 7 mm are shown in Fig. 7.



(A) **(B)**

Fig. 5 Comparison between the experimental result and the numerical simulation in case of d = 7.2 mm (A): Pictures taken by high-speed camera (B): Results of the numerical simulation. Iso surface shows the boundary of water and air. Cross section shows contour of material ID on y=0 plane.

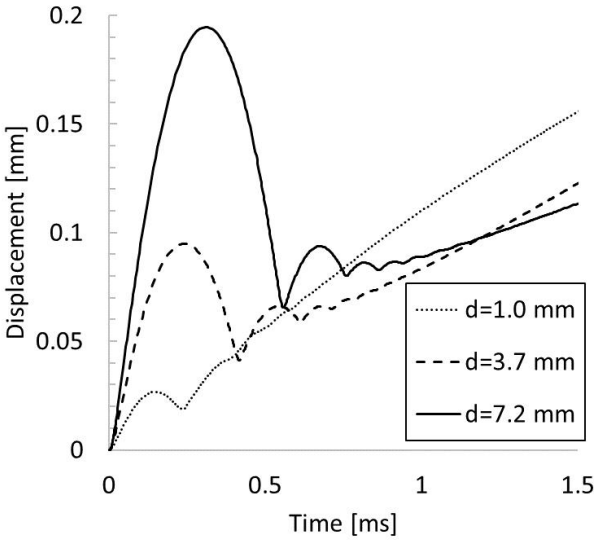


Fig. 6 The simulation results of the displacement of the target.

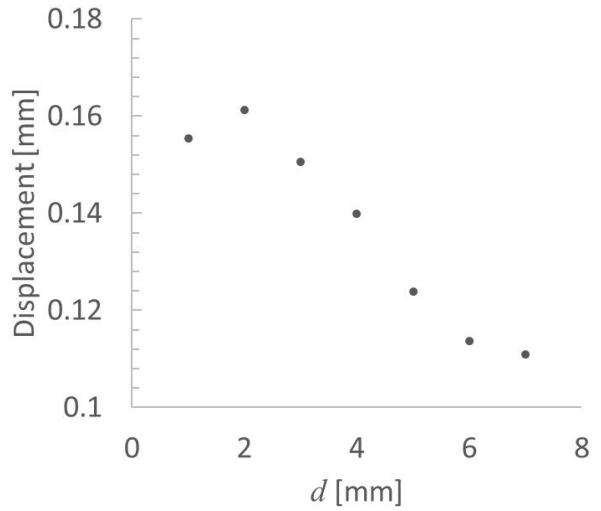


Fig. 7 The simulation results of the displacement at time 1.5 ms when d is varied from 1 to 7 mm.

Comparing the experimental results shown in Fig. 3 and the calculated results shown in Fig. 7, there is a qualitative agreement that the amount of the displacement decreases as d increases. Furthermore, it was found that there is a maximum displacement of d which is uncertain by the experiments.

4.4 The Bubble behavior and the reverse propulsion.

The contour of pressure and material ID on the cross section of $y = 0$ in case of $d = 7.2$ mm are shown in Fig. 8.

At first, due to the expansion of the bubble, the pressure on the right-side wall increases and it becomes higher than the atmospheric pressure (Fig. 8 (b)). Due to the pressure difference, the target moves forward (rightward).

And then, as the bubble expands, it loses its pressure and eventually begins to contract. Due to the collapsing of the bubble, the pressure on the bottom decreases and it becomes lower than the atmospheric pressure (Fig. 8(c)). Furthermore, the collapsing bubble inhales water which is once spouted out to the left. Due to the pressure difference and the reaction of the inhaling water, the target moves backward (leftward). Furthermore, due to the inhaling jet, the pressure on the right-side wall recovers and it becomes higher than the atmospheric pressure again (Fig. 8 (d)). Therefore, it moves forward (rightward) again. The laser-induced bubble repeats this expansion and collapsing and the target moves forward while vibrating back and forth.

As described above, we clarified the curious behavior of the metal-free water canon shown in Fig. 2.

5. CONCLUSION

We developed numerical simulation code to simulate the behavior of the laser-induced bubble and the metal free water cannon. We used CIP method and C-CUP method to realize universal solver for the system in which compressible fluid of the laser-induced bubble and incompressible fluid of water coexisted.

We succeeded in reproducing qualitative behavior of the spouting water and the movement of the laser-induced bubble when the metal-free water cannon is irradiated by laser by 3-dimensional numerical simulations. Furthermore, the calculated time development of the displacement of the metal-free water cannon are qualitatively agreed with the experimental results.

From our numerical simulations, we could explain the curious behavior of the metal-free water cannon which is observed by high-speed camera. If the laser-induced bubble is generated inside water, it expands and spouts out water and push the target forward once, and then the bubble collapse with inhaling water which is once spouted out and pull the target backward. The laser-induced bubble repeats this expansion and collapsing and the target moves forward while vibrating back and forth.

Acknowledgements

Part of this study is supported by SUZUKI FOUNDATION.

References

[1] Kantrowitz A., “Propulsion to Orbit by Ground-Based Laser”, Astronaut.Aeronaut.10, pp. 74. (1972)

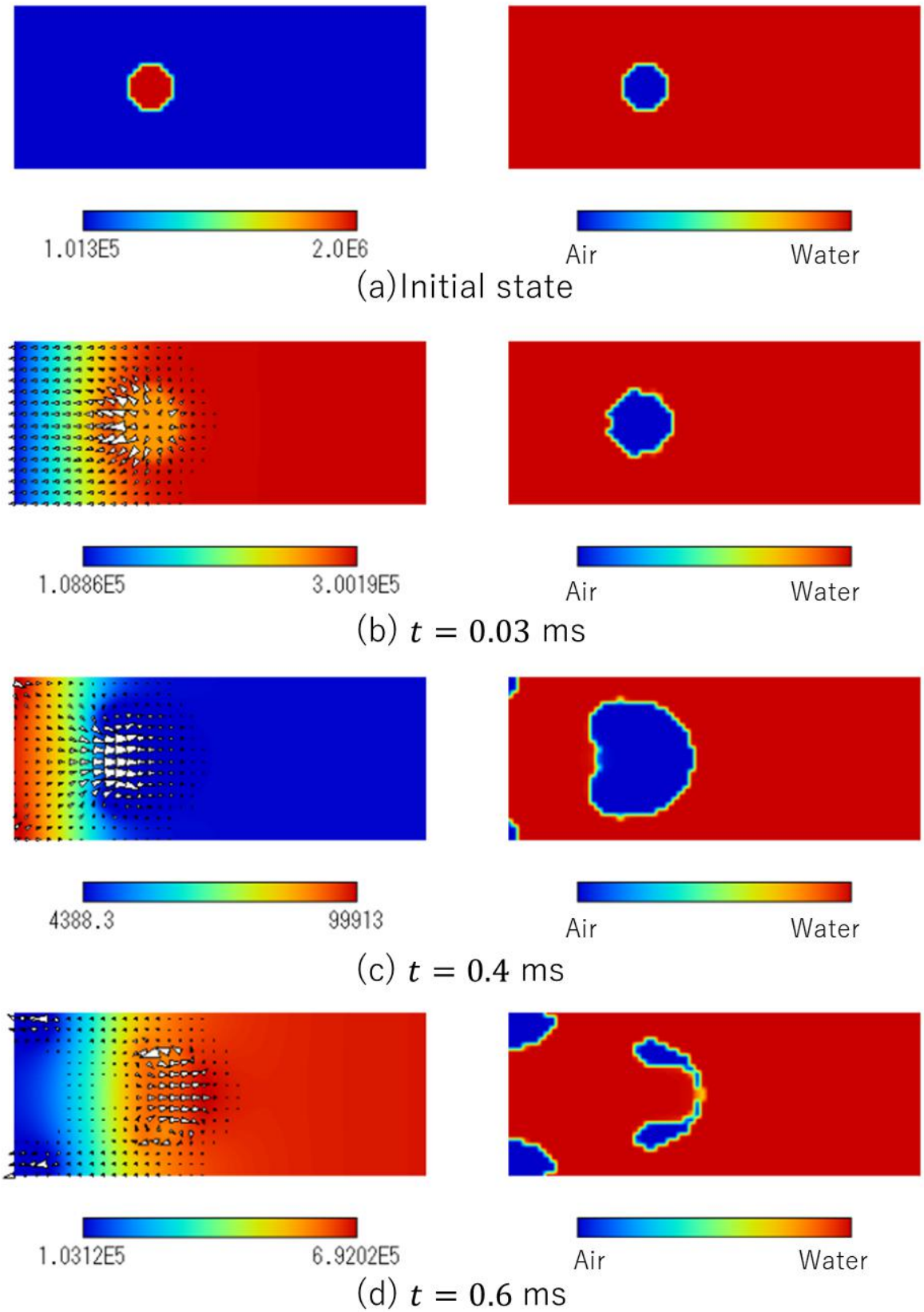


Fig. 8 The cross section of the numerical simulation results in case of d=7.2mm. The left side of pictures shows the distribution of pressure and velocity and the right side of pictures shows the distribution of water and air.

[2] Franklin B. Mead, Jr and Leik N. Myrabo, “FLIGHT EXPERIMENTS AND EVOLUTIONARY DEVEROPMENT OF A ASER PROPELLED, TRANSATOMOSPHERIC VEHICLE”, STAIF-98, Congres,Albuquerque(NM),USA, Jan.25-29,(1998)

[3] L.N.Myrabo and F.B.Mead, Jr., “Ground and Flight Tests of a Laser Propelled Vehicle”, AIAA98-1001,Aerospace Sciences Meeting & Exhibit,36th,Jan12-15,(1998)

[4] Franklin B. Mead, Jr and Leik N. Myrabo, “Flight and Ground Tests of a Laser-Boosted Vehicle”, AIAA 98-3735, AIAA/ASME/SAE/ASSEE Joint Propulsion Conference & Ex-hibit,36th July13-15, (1998)

[5] N.C.Anderholm, “Laser-Generated Stress Waves”, Appl.Phys.Lett. 16, pp.113-115, (1970)

[6] T.Yabe and K. Niu, “Numerical Analysis on Implosion of Laser-Driven Target Plasma”, J.Phys.Soc.Japan 40,pp.863-868,(1976).

[7] F. Winterberg, “Recoil Free Implosion of Large-Aspect Ratio Thermonuclear Microexplosion,” Lettere al Nuovo Cimento 16, pp.216-218, (1976).

[8] H.Azechi, N.Miyanaga, S.Sakabe, T.Yamanaka and C.Yamanaka, “Model for Cannonball-Like Accrleration of Laser-Irradiated Targets”, Jpn.J.Appl. Phys. 20, pp.868-868, (1981).

[9] C.R.Phipps, D. B. Seibert ,R. Royse, G. King and J. W. Campbell, “Very High Coupling Coe.cients at Low Laser Fluence with a Structured Target”, International Symposium on High Power Laser Ablation, Santa Fe, SPIE Vol. 4065 pp. 931-937. (2000)

[10] Masashi Yamaguchi, Ryou Nakagawa, Takashi Yabe, Choijil Baasandash, Keiichi Aoki, Tomomasa Ohkubo, Masashi Sakata, Youichi Ogata and Masamichi Nakagawa, “Laser-Driven Water-Powered Propulsion and Air Curtain for Vacuum Insulation”, ISBEP1, AIP 664, pp.557-568. (2002)

[11] Takashi Yabe, Hirokazu Ohzono, Tomomasa Ohkubo, CHoijil Baasandash, Masashi Yamaguchi, Takehiro Oku, Kazumoto Taniguchi, Sho Miyazaki, Ryosuke Akoh, Yoichi Ogata, Benjamin Rosenberg and Minoru Yoshida, “Proposal of Liquid Cannon Target Driven by Fiber Laser for Micro-Thruster in Satellite”, ISBEP2, AIP 702, pp.503-512. (2003)

[12] Takashi Yabe, “Prospect of Solar-Energy-Pumped-Laser-Driven Vehicles Powered by Water”, ISBEP3, AIP 766, pp.567-578. (2004)

[13] Takashi Yabe, Ryou Nakagawa, Masashi Yamaguchi, Tomomasa Ohkubo, Keiichi Aoki, Choijil Baasandash, Hirokazu Oozono, Takehiro Oku, Kazumoto Taniguchi, Masamichi Nakagawa, Masashi Sakata, Youichi Ogata and Gen Inoue, “Simulation and Experiments on Laser Propulsion by Water Cannon Target”, ISBEP1, AIP 664, pp.185-193, (2002)

[14] Tomomasa Ohkubo, Masashi Yamaguchi, Takashi Yabe, Keiichi Aoki, Hirokazu Oozono, Takehiro Oku, Kazumoto Taniguchi and Masamichi Nakagawa, “Laser-Driven Micro-Ship and Micro-Turbine by Water-Powered Propulsion”, ISBEP1, AIP 664, pp.535-544. (2002)

[15] YABE Takashi, OOZONO Hirokazu, TANIGUCHI Kazumoto, OHKUBO Tomomasa, MIYAZAKI Shou, BAASANDASH Choijil and UCHIDA Shigeaki, “Proposal for a Solar-Laser-Driven Vehicle”, J.Plasma Fusion Res. Vol.80, No.7, pp.547-548, (2004)

[16] Y. Ogata, T. Yabe, T. Ookubo, M. Yamaguchi, H. Oozono, and T. Oku, “Numerical and experimental investigation of laser propulsion,” Appl. Phys. A Mater. Sci. Process., vol. 79, no. 4–6, pp. 829–831, (2004).

[17] Tomomasa Ohkubo, Takashi Yabe, Sho Miyazaki, Choijil Baasandash, Kazumoto Taniguchi, Akito Mabuchi, Daisuke Tomita, Yoichi Ogata, Jun Hasegawa and Kazuhiro Horioka, “Laser Propulsion Using Metal-Free Water Cannon Target”, ISBEP3, AIP 766, pp.394-405. (2004)

[18] Thao Thi Phuong Nguyena, Rie Tanabe-Yamagishi, Yoshiro Ito, “Impact of liquid layer thickness on the dynamics of nano- to sub-microsecond phenomena of nanosecond pulsed laser ablation in liquid”, Appl. Surf. Sci. 470, 250 (2019)

[19] Senegačnik, M., Jezeršek, M. & Gregorčič, P. Propulsion effects after laser ablation in water, confined by different geometries. Appl. Phys. A 126, 136 (2020)

[20] H.Takewaki, A.Nishiguchi and T.Yabe, “The Cubic-Interpolated Pseudo-Particle (CIP) Method for Solving Hyperbolic-Type Equations”, J. Comput. Phys. 61, pp.261. (1985).

[21] T. Yabe, P.Y. Wang, “Unified numerical procedure for compressible and incompressible fluid”, J. Phys. Soc. Jpn. 60, pp.2105. (1991)

Computer Simulation of Pumping Cavity for Solar-Pumped Laser

Hayato KOSHIJI*, Tomomasa OHKUBO*,**, Takeru NAGAI**, Takumi SHIMOYAMA**,
Ei-ichi MATSUNAGA**, Yuji SATO***, Thanh-hung DINH****
* Sustainable Engineering Program, Tokyo University of Technology Graduate School
Hachioji-shi, Tokyo 192-0982, Japan
** Department of Mechanical Engineering, School of Engineering, Tokyo University of Technology
Hachioji-shi, Tokyo 192-0982, Japan
*** Joining and Welding Research Institute, Osaka University
Ibaraki-shi, Osaka 567-0047, Japan
**** National Institutes for Quantum and Radiological Science and Technology
Kizugawa-shi, Kyoto 619-0215, Japan

Abstract

Although the sunlight is promising renewable energy, it is incoherent light and difficult to use directly. Therefore, a solar-pumped laser, which directly converts sunlight into coherent light of laser, is promising technology. The solar-pumped laser collects sunlight into the laser medium to realize laser oscillation. In order to realize the efficient solar-pumped laser system, it is necessary to design a pumping cavity which absorbs more sunlight power into the laser medium with less thermal shock. In this research, the effective pumping cavity shape was studied using numerical simulation of ray tracing. As a result, it was found that the cone shaped pumping cavity can be expected to improve the absorption rate by about 30 % compared to the cylindrical shaped pumping cavity. Furthermore, the absorption power density distribution can be flattened by the vase shaped pumping cavity with same absorption efficiency. The vase shaped pumping cavity has almost half dispersion of the absorbed power density in the laser medium compared with cone shaped pumping cavity.

Keywords: Solar concentrator, Pumping cavity, Renewable energy, Sustainable engineering, Numerical simulation

1. INTRODUCTION

Development of renewable energy is essential for the realization of a sustainable society. Above all, the effective recycling of the energy of the sunlight that falls the earth inexhaustibly is very important. However, sunlight is incoherent light and difficult to use directly. Therefore, conversion system which converts sunlight into usable energy form is necessary.

The first solar-pumped laser was realized in 1965 that converts sunlight into laser light that is coherent and easy to use [1]. Solar-pumped lasers use solar energy as the energy source of laser pumping. Therefore, it is important to focus sunlight efficiently into a laser medium. However, since 1965, although various studies have been carried out, there has been no significant improvement in efficiency [2][3].

However, several researches of solar-pumped laser are performed these days because several applications of solar-pumped laser are proposed [4][5]. Solar-pumped laser system with fiber medium was proposed to reduce heat load of laser medium [6][7], and high efficiency of solar-pumped laser system of 31.5 W/m^2 was proposed [8]. However, their output power was less than 50W.

On the other hand, we developed several solar-pumped laser systems by using a Fresnel lens for the primary focusing system of sunlight, the output power and the conversion efficiency of the solar-pumped laser had been improved [9][10] and finally we realized output of 120 W and are efficiency of 20 W/m^2 [11]. Furthermore, we developed new pumping cavity with compound parabolic mirror in the previous study [12]. However, since a huge Fresnel lens of 4 m^2 was used for the system, it is difficult to manufacture the entire system. Furthermore, there was a problem that the sunlight transmittance of the Fresnel lens was as low as only 42 % [13].

We completely newly designed and manufactured a solar-pumped laser system using a small 850 mm x 850 mm Fresnel lens that is commercially available. 2.43W of laser output was realized by using this system [14]. In this system, the sunlight is focused into a pumping cavity by a Fresnel lens of primary concentrator. In the pumping cavity, the focused sunlight reflects inside the pumping cavity and it is absorbed into the laser medium which is put inside the pumping cavity.

In this research, we focused on the shape of the pumping cavity, and tried designing better pumping cavity considering absorption ratio of sunlight into the laser medium and its distribution.

2. SYSTEMS OF SOLAR-PUMPED LASER

The schematic figure of the solar-pumped laser system used in this study is shown in Fig. 1. The sunlight is focused using a Fresnel lens as the primary focusing system, and the focused light is irradiated into the pumping cavity through the glass window. The focal position of the Fresnel lens is designed to be inside the pumping cavity. A highly reflective film with visible light reflectance of 95% or more is attached inside the pumping cavity. In addition, the pumping cavity is filled with cooling water to cool a laser medium which is held inside it. The cooling water is circulated by a chiller. The irradiated solar light into the pumping cavity is re-focused into the laser medium by the reflective film inside the pumping cavity. A Nd:YAG crystal is used as the laser medium. The size is Φ 6.0 mm x 50 mm. It is difficult to install the high reflection mirror of laser resonator on the incident side of the focused sunlight due to its structure. Therefore, the end face of the laser medium on the incident side of the sunlight is HR (High Reflection) coated for the oscillation wavelength of 1064 nm. AR (Anti Reflection) coating is applied on the opposite side. Laser oscillation is realized by adjusting the output mirror provided outside the AR coat side.

Although the several laser medium for solar-pumped laser such as Ce/Cr/Nd:YAG, Cr/Nd:YAG ceramics are developed [15][16], we used Nd:YAG crystal as the laser medium because of the availability and reliability.

Furthermore, the entire system is developed on a sun tracking system to face the Fresnel lens to the sun.

In this study, we focused on the pumping cavity among these components and investigated the new shape using numerical simulation.

3. NUMERICAL SIMULATION OF PUMPING CAVITY

The sunlight condensed by the Fresnel lens is repeatedly reflected by the mirror inside the pumping cavity to be absorbed into the laser medium or emitted outside the pumping cavity. Therefore, we developed the designing simulation code based on ray tracing that simulates refraction of each materials and repeated reflection in the pumping cavity. Reflection and refraction of each ray and absorbed power inside the laser medium are calculated using following equations.

$$\boldsymbol{e_{reflect} = e_i - 2n(n \cdot e_i)}$$

$$\boldsymbol{e_{refract} = n_1/n_2\ (e_i - n(n \cdot e_i)) + \sqrt{1 - (n_1/n_2)^2(1 - (n \cdot e_i)^2)}}$$

$$\boldsymbol{P_{abosrob} = P_{in}(1 - \exp(\alpha L))}$$

$\boldsymbol{e_{reflect}, e_i, n, e_{refract}}, n_1, n_2, P_{abosrob}, P_{in}, \alpha, L$ are direction of the reflected ray, direction of the incident ray, normal vector of a surface of each material, direction of the refracted ray, refractive index of the input side, refractive index of the transmit side, absorbed power, absorption coefficient of the laser medium, ray transmission distance inside the laser medium. All the vectors are normalized.

We compared three shapes of pumping cavities and the schematic figures of each pumping cavity are shown in Fig. 2. Fig. 2 of A) shows the cylindrical shaped pumping cavity, B) shows the cone shaped pumping cavity and C) shows the vase shaped pumping cavity. All the cavities are rotational axis symmetrical shape. The left side of the figures are Fresnel lens side and the glass window and the laser medium are not shown in the figures. The surfaces indicated by the bold lines are the surface of high reflection films. The length of each pumping cavity was the same length of 50 mm because the target laser medium of length is 50 mm.

In this study, we varied diameter of the inner cylindrical volume of the cylindrical shaped pumping cavity from 10 to 100mm. Furthermore, we varied diameter of the input side of the cone shaped pumping cavity.
The divergence angle of each ray of the sunlight is randomly determined between 0 mrad and 4.6 mrad in radial direction. The spectrum of the incident sunlight is calculated at every 10 nm between 390 nm and 900 nm considering the solar spectrum. The rays from the Fresnel lens is calculated every 100 mm in the radial direction.
Under the calculation condition described above, we calculated absorbed power into the laser medium, absorption ratio and absorption distribution inside the laser medium.

4. SIMULATION RESULTS AND DISCUSSINOS

4.1 Improvement of Absorption Efficiency by Cone Shaped Pumping Cavity

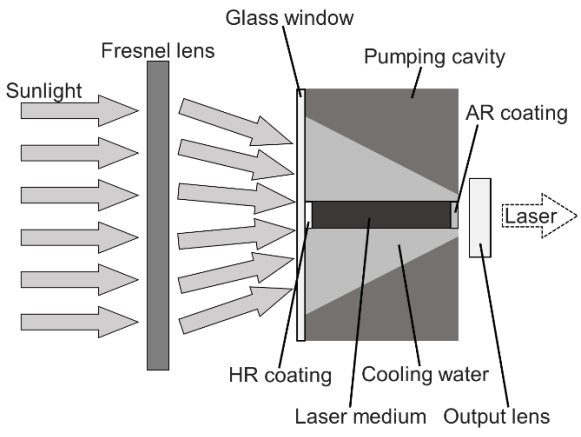


Fig. 1 Schematic figure of our solar-pumped laser system.

At first, we compared cylindrical shaped pumping cavity and cone shaped pumping cavity which are shown in Fig. 2 of A and B respectively. We changed the entrance aperture diameters of the cylindrical shaped pumping cavity and the cone shaped pumping cavity in the range of 10 to 100 mm. We calculated the absorption power and the absorption ratio. The absorption ratio is defined as absorbed power divided by the input power which passed through its entrance. The results of the cylindrical shaped pumping cavity are shown in Fig. 3 and those of the cone shaped pumping cavity are shown in Fig. 4.

Comparing Fig. 3 and Fig. 4., the cone shaped pumping cavity has higher absorption ratio and absorption power than the cylindrical shaped pumping cavity. Furthermore, in the cone shaped pumping cavity, both the absorption ratio and the absorption power were maximum when the entrance diameter was 50 mm. Therefore, the cone shaped pumping cavity is better than the cylindrical shaped pumping cavity.

However, the absorption ratio of the cone shaped pumping cavity decreases as the entrance aperture size increases. This is because the sunlight that once entered into the pumping cavity goes out through the enlarged entrance.

4.2 Confining Pumping Light and Flattening Absorption Power Density Distribution by Vase Shaped Pumping Cavity

Based on the discussion in 4.1, we designed vase shaped pumping cavity which is shown in Fig. 2 of C, to confine the sunlight that once entered into the pumping cavity.

Table 1 shows a comparison of absorption power and absorption ratio of the cone shaped pumping cavity whose entrance diameter is 50 mm and vase shaped pumping cavity. There is no great difference in absorption power and absorption ratio between cone shaped pumping cavity and the vase shaped pumping cavity.

Furthermore, the calculated absorbed power distribution inside the laser medium of each pumping cavity is shown in Fig. 5. The vase shaped pumping cavity has a smaller peak of the absorption power density distribution as the cone shaped pumping cavity. However, the peak height of the vase shaped pumping cavity is smaller than that of cone shaped pumping cavity. This means that the vase shaped pumping cavity has less thermal shock of laser medium than the cone shaped pumping cavity.

To evaluate flatness of the absorption density distribution, comparison of each dispersion is shown in Table 2. The vase shaped pumping cavity has only half of the dispersion of the absorbed power density distribution in the laser medium compared with the cone shaped pumping cavity. From these results, it was found that the vase shaped pumping cavity has less damage to the laser medium and is suitable as a pumping cavity for the solar-pumped laser system.

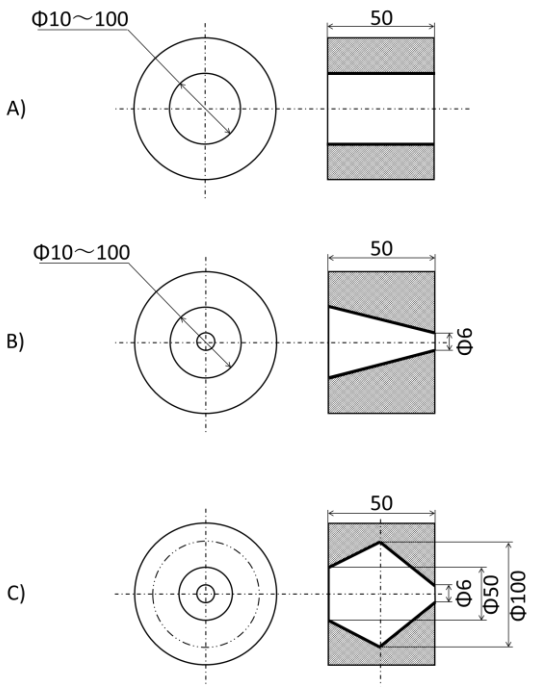


Fig. 2 The schematic figure of each pumping cavity. A): Cylindrical shaped pumping cavity. B): Cone shaped pumping cavity. C): Vase shaped pumping cavity.

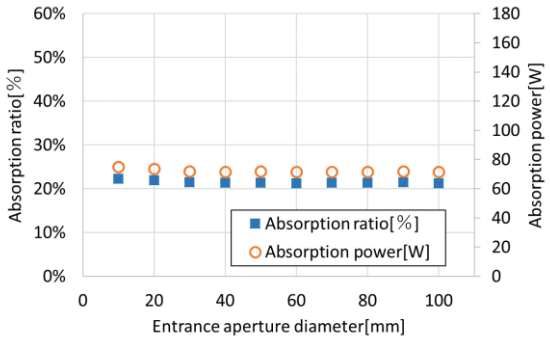


Fig. 3 Calculated absorption power and absorption ratio of cylindrical shaped pumping cavity when changing the entrance size.

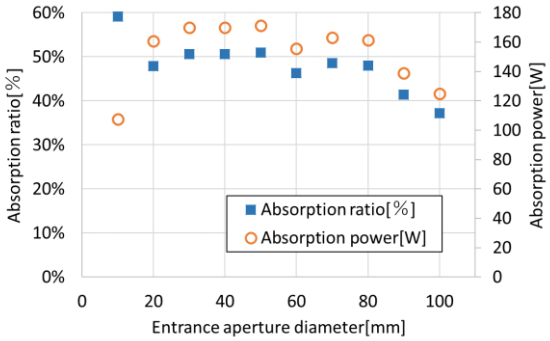


Fig. 4 Calculated absorption power and absorption ratio of cone shaped pumping cavity when changing the entrance size.

Table 1 Comparison of absorption power and absorption ratio of cone shaped pumping cavity and vase shaped pumping cavity.

Shape of pumping cavity	absorption power [W]	absorption ratio [%]
Cone	171.5	51 %
Vase	174.2	52 %

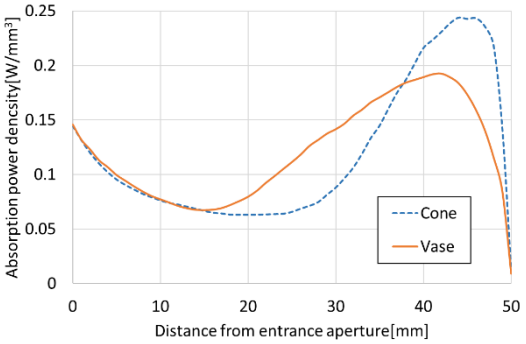


Fig. 5 Calculated absorbed power distribution in the laser medium.

Table 2 Dispersion of the absorbed power density of each pumping cavity.

Shape of pumping cavity	Dispersion
Cone	0.00405
Vase	0.00198

5. CONCLUSION

The numerical simulations for comparing shapes of the pumping cavity for solar-pumped laser system in this study have revealed the following.

- Comparing the cylindrical shaped pumping cavity and the cone shaped pumping cavity, it was found that the cone shaped pumping cavity is superior in absorption power and absorption ratio into the laser medium.

- Comparing the cone shaped pumping cavity and vase shaped pumping cavity, although the absorption ratio is almost the same, the vase shaped pumping cavity has more flat absorption power density distribution and less thermal damage to the laser medium. The vase shaped pumping cavity has almost half dispersion of the absorbed power density in the laser medium compared with cone shaped pumping cavity.

In the future, we will actually create a pumping cavity and perform experiment to verify these calculated results.

References

[1] C. G. Young, “A Sun-Pumped cw One-Watt Laser”, Appl. Opt. **5**, 6, pp.993-997 (1966).

[2] M. Lando, J. Kagan, B. Linyekin, and V. Dobrusin, “A solar-pumped Nd:YAG laser in the high collection efficiency regime”, Opt. Commun. 222, pp. 371-381 (2003)

[3] V. Krupkin, Y. Kagan, and A. Yogev, “ “Nonimaging optics and solar laser pumping at the Weizmann Institute”, Proc. SPIE, 2016, pp. 50-60, 1993.

[4] T. Yabe, S. Uchida, K. Ikuta, K. Yoshida, Choijil Baasandash, M. S. Mohamed, Y. Sakurai, Y. Ogata, M. Tuji, Y. Mori, Y. Satoh, Tomomasa Ohkubo, M. Murahara, A. Ikesue, M. Nakatsuka, T. Saiki, S. Motokoshi, C. Yamanaka, “Demonstrated fossil-fuel-free energy cycle using magnesium and laser”, Appl. Phys. Lett. 89, 261107 (2006)

[5] T. Motohiro, Y. Takeda, H. Ito K. Hasegawa, A. Ikesue, T. Ichikawa, K. Higuchi, A. Ichiki, S. Mizuno, T. Ito, N. Yamada, H. N. Luitel, T. Kajino, H. Terazawa, S. Kakimoto, K. Watanabe, “Concept of the solar-pumped laser-photovoltaics combined system and its application to laser beam power feeding to electric vehicles”, Jpn. J. App. Phys., 56, 852 (2017)

[6] T. Masuda, M. Iyoda, Y. Yasumatsu, M. Endo, “Low-concentrated solar-pumped laser via transverse excitation fiber-laser geometry”, Opt. Lett. 24, 17, pp.3427-3430 (2017)

[7] T. Masuda, M. Iyoda, Y. Yasumatsu, S. Dottermusch, I. A. Howard, B. S. Richards, J. F. Bisson, M. Endo, “A fully planar solar pumped laser based on a luminescent solar collector”, Comm. Phys. 3, 60 (2020)

[8] D. Liang, J. Almeida, C. R. Vistas, E. Guillot, “Solar-pumped Nd:YAG laser with 31.5 W/m2 multimode and 7.9 W/m2 TEM00-mode collection efficiency”, Sol. Energ. Mat. Sol. C., 159, pp.435-439 (2017)

[9] T. Yabe, T. Ohkubo, S. Uchida, K. Yoshida, M. Nakatsuka, T. Funatsu, A. Mabuti, A. Oyama, K. Nakagawa, T. Oishi, K. Daito, B. Bagheri, Y. Nakayama, M. Yoshida, S. Motokoshi, Y. Sato, and C. Baasandash, “High-efficiency and economical solar-energy-pumped laser with Fresnel lens and chromium codoped laser medium” Appl. Phys. Lett., 90, 261120, (2007)

[10] T. Ohkubo, T. Yabe, K. Yoshida, S. Uchida, T. Funatsu, B. Bagheri, T. Oishi, K. Daito, M. Ishioka, Y. Nakayama, N. Yasunaga, K. Kido, Y. Sato, C. Baasandash, K. Kato, K. Yanagitani, and Y. Okamoto, “Solar-pumped 80 W laser irradiated by a Fresnel lens”, Opt. Lett. 34, 175 (2009).

[11] T. H. Dinh, T. Ohkubo, T. Yabe, and H. Kuboyama, “120 watt continuous wave solar-pumped laser with a liquid light-guide lens and an Nd:YAG rod”, Opt. Lett. 37, 13 (2012)

[12] T. Ohkubo, “Design of New Pumping Cavity with Compound Parabolic Concentrator for Solar-Pumped Laser”, J. Adv. Comput. Intell. Intell. Inform., 20, 7, pp.1065-1069 (2016)

[13] T. OHKUBO, E. Matsunaga, “Design of solar collector for efficient solar-pumped laser”, Reports on the Topical meeting of the Laser Society of Japan, 476, (2015) (Japanese)

[14] T. Ohkubo, T. Nagai, S. Hisano, K. Mori, S. Kojima, K. Azato, H. Koshiji, E. Matsunaga, Y. Sato, T.H. Dinh, J. Yokota, “Demonstration of Solar-pumped laser with collection area of 1m^2”, Reports on the Topical meeting of the Laser Society of Japan, 533 (2019) (Japanese)

[15] K. Fujioka, M. Nakatsuka, T. Saiki, S. Motokoshi, K. Imasaki, Y. Fujimoto, H. Fujita, “Luminescence Properties of Ce/Cr/Nd:YAG Materials for Solar-Pumped Lasers”, 38, 3, pp.207-212 (2010) (Japanese)

[16] D. Liang, C. R. Vistas, B. D. Tiburcio, J. Almeida, “Solar-pumped Cr:Nd:YAG ceramic laser with 6.7% slope efficiency”, Sol. Energ. Mat. Sol. C., 185, pp.75-79 (2018)

Multi-robot Mobile Platform Design Based on Optimized Depth Q Network

Feng Liu*,**,†, Chang Chen*, Zhihua Li*,**, Zhi-Hong Guan***
* School of Automation, China University of Geosciences, Wuhan 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
*** College of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
† corresponding author's Email: fliu@cug.edu.cn

Abstract

Based on mobile robot path planning and depth of intensive study, designed the dynamic obstacle avoidance algorithm under complex environment, mobile robot through callback information for robot shooting environment information, network through pretreatment of image feature dimension reduction, improve the efficiency, the processed image by D3QN network parameter training information, will change the behavior of the output information to mobile robot of angular velocity, linear velocity. Through the reward and punishment mechanism and the maximization algorithm, the mobile robot can follow the optimal trajectory. Aiming at the multi-robot motion model, a multi-robot cooperative operation cloud network is designed, and a local depth Q network and a global network are designed. A single robot obtains the action information through its local depth Q network training parameters. Robots communicate with each other through the global network to realize information interaction and obstacle avoidance with other robots..

Keywords: Multi-robot network; intelligent robot; deep reinforcement learning; optimization algorithm

1. INTRODUCTION

Now, mobile robots are playing an increasingly important role in our lives. In the design module of mobile robots, path planning technology is one of the core contents of intelligent mobile robots. Path planning of mobile robot refers to giving a reasonable objective function in the search area and planning the optimal solution of the path from beginning to end, so that the robot can quickly and flexibly avoid all obstacles in the search environment without colliding with them. Reasonable path planning algorithm can make the intelligent robot perform better. For global path planning, currently mature algorithms include dynamic A* algorithm, Petri net algorithm, fuzzy rules based on data fusion, neural network algorithm, artificial potential field method, computational geometry method and genetic algorithm. In the aspect of local path planning, the current mature algorithms include artificial potential field, fuzzy logic, simulated annealing and other intelligent optimization algorithms, as well as genetic algorithm, ant colony algorithm, artificial neural network and other bionic optimization algorithms. Although these algorithms achieve good results in the past, but there can't solve the dynamic obstacle avoidance, easily plunged into local optimum, and synergy model for multiple robots, lack of good effect, and in the unmanned s is getting closer and closer to us, only able to handle multiple intelligence algorithm can meet the needs of today.

Reinforcement learning is a kind of learning algorithm between supervised learning and unsupervised learning. In the environment, robots that move autonomously do not know the state they are in, and can only learn through exploration and evaluation. The higher the evaluation value, the better the action in the current situation. Otherwise, this action is not what people expect.

Deep reinforcement learning is the product of the combination of deep neural network and reinforcement learning. Solve the problem that reinforcement learning is difficult to deal with complex states. ching-chi Tsai [1] studied how to input depth image into DQN algorithm and transfer the learned navigation strategy to unknown environment based on inheritance features. Qi [2] propose to join the storage pool mechanism, adding the trial and error samples to the storage pool to give priority to training.

In this paper, I designed an optimized depth Q network based on random sampling to guide robots to take actions in complex environments. The global communication network is designed to realize the cooperative motion and information interaction before the robot. The results show that the computational efficiency can be effectively improved and the problem of dynamic obstacle avoidance can be solved.

2. INDOOR POSITIONING BASED ON GLS AND ELS

Reinforcement learning representation response is a

mapping relationship between the state and behavior of the agent. The agent first perceiving the external environment, then making decisions and selecting actions, and continuously updating the decisions according to the new states and rewards returned by the actions to optimize the decisions, as shown in Fig. 1.

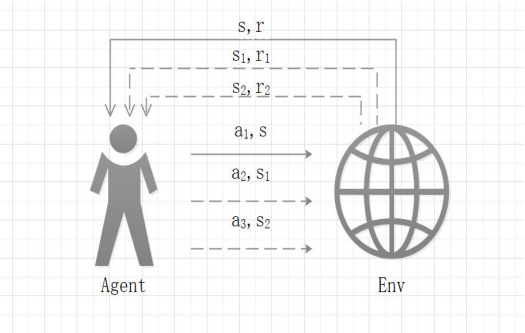


Fig.1. Agent-Environment interaction

The agent first selects the behavior A1 under the current states, and then the environment will return a return value R and a new state S1. The agent will select behavior A2 in the new state S_1, and then the agent will reach a new state S_2. Agents learn from the behavior of interacting with the environment and the mechanism of maximizing the return value, and finally get a model of intelligent decision-making.

The navigation of the robot is actually the task of interacting with the environment, in which we can design the controller according to a series of actions, observations and reward values. So it is natural to integrate reinforcement learning into robot path planning， In each step of time, the robot selects an action $A=\{1... ,K\}$, the action command is passed to the robot and interacts with the environment. Generally speaking, the environment may change at any time, so the navigation based on reinforcement learning can be used for obstacle avoidance in dynamic environment. The controller does not need to understand information about the environment; Instead, it observes an image x from the environment, which is a vector representing the original pixels of the current camera. In addition, it receives a reward r, which represents the change in the evaluation score. In addition, generally speaking, the score of the evaluation may depend on the whole previous action and observation; Feedback on an action may only be received after thousands of circular steps.

Since the robot only observes the image of the current camera, it can only observe the state and tasks of some robots, that is to say, it is impossible to fully understand the current state from the current camera x. Therefore, we consider a series of actions and observations of $s_t=x_1$, a_1 $x_2...$, a_t, x_t, and learn evaluation strategies that depend on these sequences. All of these sequences in the controller are assumed to terminate in a finite loop step, which is a large-scale but finite markov decision process (MDP)[3], where each sequence is a different state. Suppose we have a robot in the state of, and it has a variety of action choices to reach the termination state of,, but the benefits of executing each action are not the same. At this point, we need to make an algorithm to help the robot choose the action sequence, so as to ensure the highest profit when reaching the terminating state st, which requires the markov decision process.

Markov property is a concept in probability theory. When a random process is given the present state and all the past states, the conditional probability distribution of its future state only depends on the current state. In other words, given the present state, it is conditionally independent of the past state (that is, the historical path of the process), so the random process has markov properties. A process with markov properties is often called a markov process. State following markov means that the future state has nothing to do with the past state but only with the present state, i.e.

$$P[S_{t+1} \mid S_t = P[S_{t+1} \mid S_t,...,S_1] \quad (1)$$

where P - one state transition matrix. Markov decision process is composed of five key elements {S, A, P, R, γ}, where S represents the finite state set, A represents the finite action set, P is the state transition probability matrix, R is the reward function, and y represents the conversion factor γ in [0,1] .Among them[3].

$$P^{a}_{ss'} = P[S_{t+1} = s' \mid S_t = s, A_t = a] \quad (2)$$

Robot's goal is in the process of interaction with the environment, through the choice of action to maximize future reward value. We make a standard assumption that future rewards will be discounted by the factor γ of each time step, and define future discounted returns at time T, which is the time step of game termination. We make a standard assumption that future rewards will be discounted by the factor γ of each time step, and define future discounted returns at time T, which is the time step of game termination. The key problem of markov decision process is to find a strategy $\pi(s)$: select actions in states to form a function of actions. Our goal is to choose an action function two that maximizes the cumulative benefit R(T).

$$R(T) = \sum_{t+1}^{T} \gamma^{t} R_{a_t}(s_t, s_{t+1}) \quad (3)$$

Meanwhile, in order to evaluate the advantages and disadvantages of taking actions, the behavior value function $Q_\pi(s, a)$ is defined, which refers to the expected return obtained by taking actions according to strategy π after taking actions starting from state s and taking actions a.

$$Q_{\pi}(s, a) = E_{\pi}[R(T) \mid S_t = s, A_t = a] \quad (4)$$

Therefore, the value function from time t to t + 1 can be defined in this way.

$$Q_*(s, a) = \max_a Q_{\pi}(s, a) \quad (5)$$

By collating the above formula, an optimal value function will be defined as follows

$$Q_*(s, a) = R_s^a + \gamma \sum P_{ss'}^a v_{\pi}(s') \quad (6)$$

where v π (s') is an optimal value function.

Thus, the relationship between the optimal behavior value function at time t and time t+1 can be obtained

$$Q_*(s, a) = R_s^a + \gamma \sum P_{ss'}^a \max Q_{\pi}(s', a') \quad (7)$$

3. PATH PLANNING MODELING

This article is based on the Q - learning algorithm, the depth camera sensors to obtain the depth of the image as a state of the robot, this algorithm determines the behavior of the robot by controlling the angular velocity and the linear velocity of the robot. Seven behavior pools of the robot are formed by setting different angular velocity and the linear velocity.

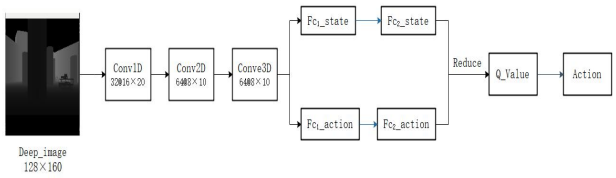


Fig. 2. The network structures.

Between the state and behavior of robot to interact through deep neural network, the network structure as shown in Fig.2, the parameters of the network as shown in table 1, the depth of the robot camera image after image preprocessing form 128 x 169 gray image, after the first layer of convolution kernel feature extraction into 32 pieces of 16 x 20 images, after the second and third layer output for 64 pieces of 8 x 10 images, smooth processing after a form value function connection layer, an act for the value function all connection layer, output behavior after through calculation of the mean square error.

Table 1 Convolution layer parameters

name	Size	Input	Output	Stride	Padding
conv1D	10*14	4	32	8	Same
conv2D	4*4	32	64	2	Same
conv3D	3*3	64	64	2	Same

For an environment with a finite number of states and a finite number of behaviors, it is possible to use the q-table method to find the optimal Q value one by one, but considering an environment with many states and multiple behaviors in each state, it would take a lot of time to traverse all the behaviors in each state. A better approach is to approximate the Q function, Q(s,a; θ) ≈ Q*(s,a), where a neural network weighted with theta is used to approximate the Q values for all possible behaviors in each state. Since the neural network is used to approximate Q function, it is called Q network. We adopt the new update rule:

$$Q(s, a) = Q(s, a) + a(r + \gamma \max Q(s', a') - Q(s, a)) \quad (8)$$

where, r+ γ maxQ(s',a') is the target value; Q(s,a) is the predictive value, the goal being to minimize Q(s,a) by learning a correct strategy.

$$Loss = (y_i - Q(s, a; \theta))^2 \quad (9)$$

$$y_i = r + \gamma \max_{a'} Q(s', a'; \theta) \quad (10)$$

The neural network model is shown in Fig. 3.

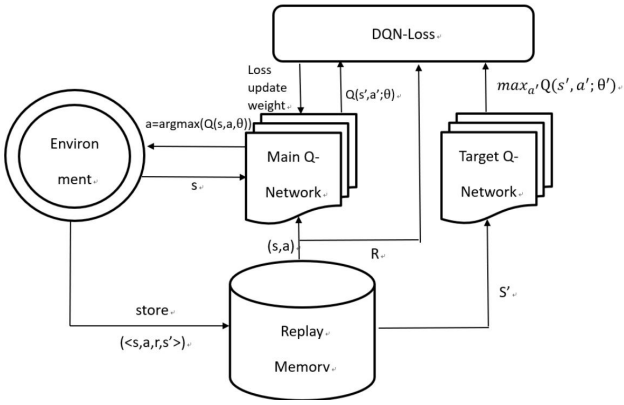


Fig. 3. The DQN structure diagram.

The neural network is used as function approximator to approximate the *Q* function, and the error is minimized by gradient descent.

It is clear to noticed that the parameters of the target *Q* are *θ*' rather than *θ*. The goal is to make the target network lag behind the actual network, and there is a difference of 500 step weight update times between the target network and the actual network.

In order to save the information obtained from interaction, this paper designs an experience playback pool, which stores key value pairs one by one. Each new state's' and the return *R* < *s*, *a*, *R*, *s*' > information obtained by action a under each state will be stored in the experience playback pool, and then train the deep neural network through the stored samples. In order to improve the reliability of experience playback pool samples and the efficiency of network training, in this algorithm, I adopt the priority playback mechanism based on random sampling. When

the experience playback pool reaches the storage capacity, first place the key value pairs with high return value in the priority position through sorting algorithm, and then set a random parameter ρ.

For the top of key value pair data, the data that can be divided by 3ρ is removed; for the middle of key value pair data, the data that can be divided by 2ρ is removed; for the last of key value pair data, the data that can be divided by ρ is removed, which can improve the data accuracy and reduce the possibility of local optimization.

4. ROBOT SIMULATION MODEL BUILDING

Gazebo and ROS are used to build the robot's simulation environment. The enhanced simulation environment of gazebo can well simulate the real environment, and various controller plug-ins can also be well combined with ROS. ROS has strong communication capability and coordinate transformation function, and controls the robot through node and communication. The main frame of the entire model is shown in Fig. 4.

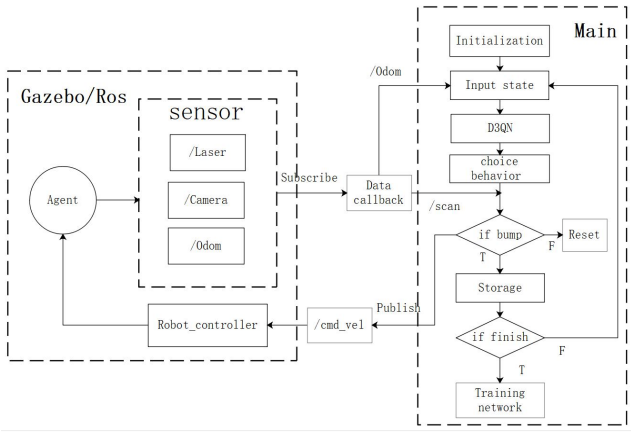


Fig. 4. Multi-robot model framework

The network first needs to initialize the parameters, and the list of parameters is shown in table 2.

The robot has laser, binocular camera and odometer, and when the program runs:

(1) The state information of the robot initialized by the Publish robot;

(2) Subscribe /camera node to get the current image taken by the robot binocular camera;

(3) Input the image into the depth neural network after preprocessing;

(4) The output behavior of neural network is processed;

(5) Release information to /cmd_vel node through the processed behavior to change the linear speed and angular speed of the robot;

(6) The obtain obstacle information near the robot by subscribing /scan node, and if there is a collision, the environment will be reset.

(7) Store the obtained in the playback pool.

(8) Run in a loop (2)-(7) until the specified number of times is reached;

(9) The data training network through playback pool.

(10) To get the model and complete the adaptive path planning.

Table 2 Parameters

Parameter	meaning	Size
Learning_rate	Update speed factor	0.1
γ	Attenuation factor	0.99
StackFrame	Stack frame number	4
Num_action	Behavior pool	7
Num_batch	Minibatch size	48
Num_replay_memory	Storage pool capacity	20000
Num_training	training steps	50000

A. Multi-robot environment

Based on multi robot platform environment as shown in Fig.5, three robots having their own DDQN network, learning from each other, as a result of the reinforcement learning decision-making mechanism. They can realize the robot mutual collision, accomplish their goal, through the main platform to realize communication between robots, back to their respective laser radar information by three robots, to get the distance between the three robots, by setting the barrier function, when the distance between the robots closer to get a negative return, so as to realize mutual coordination and mutual influence, robot action when image information is shown in Fig. 6.

B. Comparison and analysis

After modeling by gazebo and ROS, the operation results of the reinforcement learning network of deep confrontation are shown in the following figure. Figs. 7 to 10 show the three convolution layers of the target network and the main network in the deep neural network, and figure 10 shows the total return obtained by the robot network with the change of iteration times. It can be seen that from the overall analysis, the difference between the main network and the target network increases first and then gradually converges. This is because at the beginning of deep reinforcement learning, random trial and error scheme is used to put the samples into the priority error experience playback pool. The network will gradually approach the target network through the data of the experience playback pool, so that the robot can slowly choose the correct behavior.

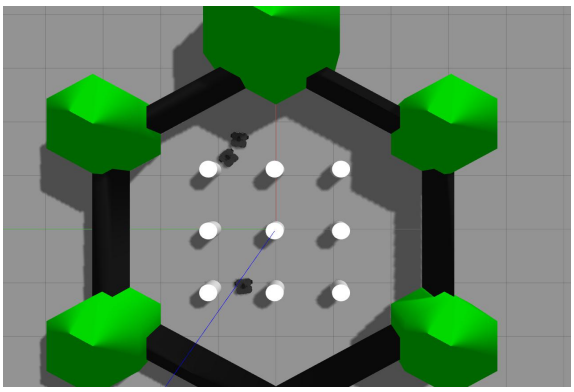


Fig.5. Gazebo environment

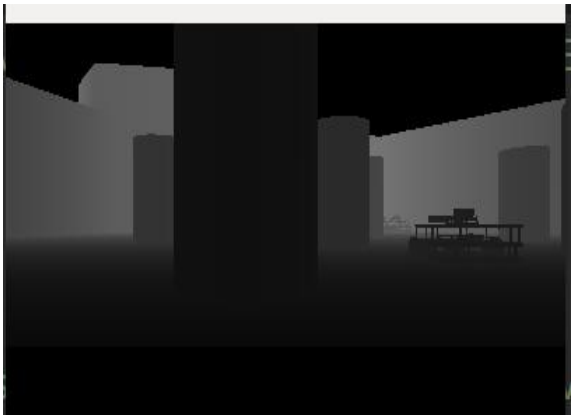


Fig. 6. Environmental information

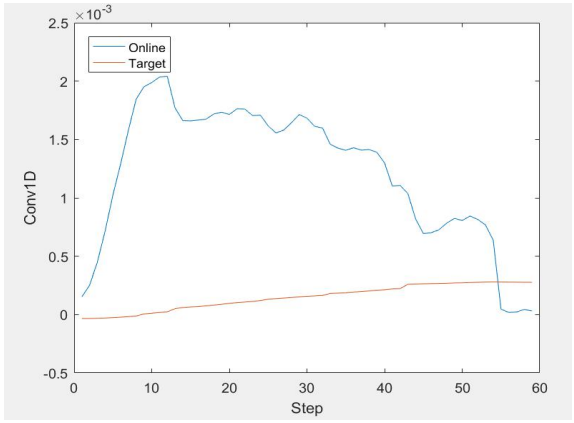


Fig. 7. Conv1D

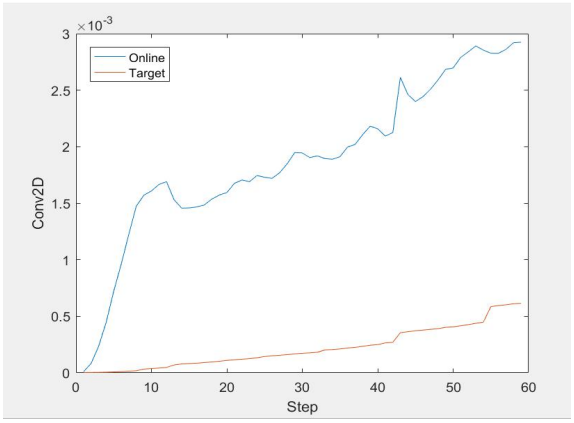


Fig. 8. Conv2D

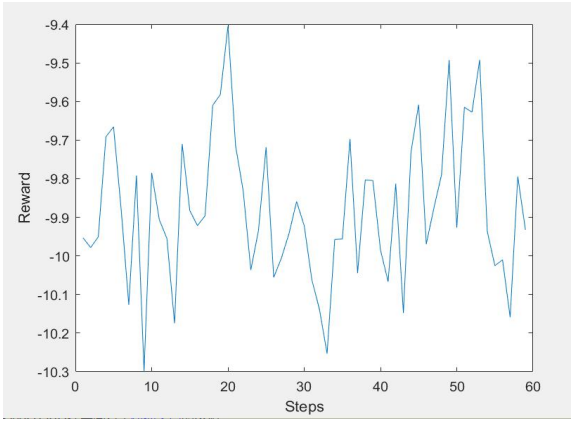


Fig.10. Reward

5. CONCLUSIONS

Mobile robot path planning, this paper first introduces several common algorithm, and then summarizes the development trends of path planning, Put forward the use of reinforcement learning method to solve the problem of mobile robot path planning, in the second chapter presents the source of the reinforcement learning, markov decision, State, and behavior, strategy, value function, returns five yuan group. In the third chapter, an algorithm model of reinforcement learning path planning is established, Combining reinforcement learning and deep neural network to solve the problem of excessive ergodic behavior in complex continuous environment, The simulation environment achieves the effect of the physical environment, and through the multi-robot network and the decision mechanism of reinforcement learning, the multi-robot network can realize mutual non-interference, which is in line with the unmanned multi-agent traffic environment in the physical environment.

Acknowledgements

This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61472374, 61503053, 61603358.

References

[1] Tsai, C.-C, Huang, H.-C, Chan, C.-K. Parallel Elite Genetic Algorithm and Its Application to Global Path Planning for Autonomous Robot Navigation. IEEE Transactions on Industrial Electronics, 2011, 58 (10): 271-275.

[2] Qi, X., et al. Deep reinforcement learning enabled self-learning control for energy efficient driving. Transportation Research Part C, 2019. 99: 67.

[3] Yuan, Y., Yu, Z. L., Gu, Z., Yeboah, Y., et al. A novel multi-step Q-learning method to improve data

efficiency for deep reinforcement learning. Knowledge-Based Systems, 2019,175: 107-117.

[4] Silver, D., et al. Mastering the game of Go with deep neural networks and tree search. Nature, 2016. 529(7587): 484-489.

[5] Liu, Z., Yao, C., Yu H. Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things. Future Generation Computer Systems, 2019, 97: 1-9.

[6] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. Human-level control through deep reinforcement learning. Nature, 2015, 518(7540):529.

[7] Wang, X., Sun, T., Yang, R. Quality-aware dual-modal saliency detection via deep reinforcement learning. Signal Processing Image Communication, 2019, 75: 158-167.

Analog Realization of Fractional-Order Capacitor and Inductor Defined by the Caputo-Fabrizio Derivative

Manjie Ran*, Xiaozhong Liao*, Da Lin*, Ruocen Yang*
* Automation Department, Beijing Institute of Technology
Beijing, 100081, China

Abstract

Capacitors and inductors have been proven to have fractional-order characteristics. Therefore, establishing fractional-order models for circuits containing such components are of great significance in practical circuit analysis. This work establishes the impedance models of the fractional-order capacitor and inductor based on Caputo-Fabrizio derivative, and further proposes the analog realization of fractional-order electronic components. The mathematical models of fractional RC, RL and RLC electrical circuits are deduced and verified by the comparison between the numerical simulation and the corresponding circuit simulation. The electrical characteristics of the fractional circuits are analyzed. This work not only enriches the models of fractional capacitors and inductors, but also can be applied to the description of circuit characteristics to obtain more accurate results.

Keywords: Fractional-order calculus, Caputo-Fabrizio derivative, Fractional circuits modeling, Analog realization.

1. INTRODUCTION

The concept of fractional calculus is generalized by integer-order calculus. At present, in the modeling and analysis of actual physical systems, differential equations are mostly considered as integer-order ones, so the research is usually based on the integer-order mathematical models. However, many actual physical systems exhibit fractional-order dynamic features due to their material and chemical properties. For example, the actual capacitors and inductors have irreversible dissipation effects due to the influence of electric and magnetic fields, such as nonlinearity, ohmic friction, internal friction, and thermal memory [1-4]. Therefore, their true electrical characteristics can be described more accurately by fractional models [5], making their mathematical models closer to actual components.

In the field of circuit analysis, the research on the fractional-order models of capacitors and inductors is getting deeper while the fractional calculus is introduced into the circuits to build a fractional model of the entire system.

In component modeling, I. Petras et al. established fractional-order mathematical models of capacitors, inductors, and memristors [6]; AA Raorane et al. proposed a fractional-order inductance model considering actual inductance losses [7]; Adhikary, A, et al. used GIC topology to study the unified modeling of fractional-order capacitors and inductors [8].

In terms of filter design and power circuit application, Radwan, A. G, et al. gave the general parameters of a fractional-order LC circuit as a band-pass filter [9]. Tripathy et al. designed a fractional KHN biquad filter, which can provide better performance compared with integer-order filters [10]. Martinez et al. studied the fractional model of DC-DC converters suitable for solar power generation systems [11].

The current modeling and realization of fractional-order components are mostly based on the Caputo definition, but it has singularities and cannot fully describe the memory effect of the system. Therefore, Caputo and Fabrizio proposed a new definition of fractional derivative without singular kernel in 2015 [12], and this definition uses the convolution of the first derivative of a given function and the exponential function to design. The novelty in this operator is that, the derivative has regular kernel, nevertheless, due to this property this operator has a better description of material structure characteristics [13]. So in this work, we use Caputo-Fabrizio derivative to model electrical components and investigate the behavior of the voltage and current of the fractional circuits, try to improve the accuracy of component models.

At present, many scholars applied this new definition in the field of circuit analysis. Alsaedi et al. compared the response waveforms of fractional-order circuits under the definitions of Riemann–Liouville, Caputo and Caputo-Fabrizio [14]. Abro K A et al. studied the fractional RLC network defined by the Caputo, Caputo-Fabrizio and Atangana-Baleanu definitions through precise numerical analysis [15]. Although the Caputo-Fabrizio definition has been used in circuit modeling, whether it can describe the properties of capacitors and inductors more accurately or not remains to be analyzed in detail.

Based on the Caputo-Fabrizio (CF) operator, this paper derives the impedance models of fractional-order capacitor and inductor. The mathematical models of the fractional circuits, including resistance-capacitance (RC) circuits, resistance-inductance (RL) circuits, and RLC circuits are also proposed in this work. In Section 2, the basic concepts of the CF operator are given, and the fractional mathematical models of RC, RL, and RLC circuits are established using Laplace transform together with its inverse transformation. In Section 3, through intuitive voltage and current response curves and the characteristic curves in frequency domain, the characteristics of the fractional-order circuits defined by CF derivative are analyzed. In Section 4, the conclusions are put forward.

2. MODELING OF FRACTIONAL COMPONENTS AND CIRCUITS

2.1 Caputo-Fabrizio Operator

The Caputo-Fabrizio definition of fractional derivative in time domain is defined as follows

$$ {}_{a}^{CF}D_{t}^{\alpha} f(t) = \frac{M(\alpha)}{1-\alpha}\int_{a}^{t} f'(\tau)\exp[-\frac{\alpha(t-\tau)}{1-\alpha}]d\tau, \quad (1) $$

where ${}_{a}^{CF}D_{t}^{a}$ is the Caputo-Fabrizio derivative with respect to t, $M(\alpha)$ is the normalized function such that $M(0)=M(1)=1$, and $\alpha\in[0,1]$, $a\in(-\infty,t)$.

Starting from time zero, that is, when $a=0$, the Laplace transform of the CF fractional derivative is

$$ L[{}^{CF}D_{t}^{\alpha+n} f(t)] = \frac{s^{n+1}L[f(t)]-\sum_{k=0}^{n} s^{n-k} f^{(k)}(0)}{s+\alpha(1-s)}, \quad (2) $$

where $n\geq 1$ and s is the Laplace transform operator. This article mainly studies the case where the fractional order is between 0 and 1, so we have n=0. Then the Laplace transform of CF fractional derivative can be simplified as

$$ L[{}^{CF}D_{t}^{\alpha} f(t)] = \frac{sL[f(t)]-f(0)}{s+\alpha(1-s)}, \quad (3) $$

where $f(0)$ is the initial value of $f(t)$. Under the zero-initial condition, that is, when $f(0)=0$, the Laplace transform of CF fractional derivative is further simplified as

$$ L[{}^{CF}D_{t}^{\alpha} f(t)] = \frac{s}{s+\alpha(1-s)}L[f(t)]. \quad (4) $$

According to similar analysis methods, we can obtain the Laplace transform of CF fractional integral as follows

$$ L[{}^{CF}D_{t}^{-\alpha} f(t)] = \frac{s+\alpha(1-s)}{s}L[f(t)]. \quad (5) $$

For complex representation in the time domain, Laplace transform is suitable for calculation.

2.2 Analog Realization Models of Fractional-order Capacitor and Inductor

In the integer-order network, the equivalent impedance of the capacitor and the inductor in the *s*-domain is

$$ Z_C = \frac{1}{sC}, \quad (6) $$

$$ Z_L = sL, \quad (7) $$

where Z_C represents capacitance impedance, and Z_L represents inductance impedance. Similar to the integer order, according to the Laplace transform (4), we obtain the fractional equivalent impedance of the capacitor under the definition of CF, namely

$$ \begin{aligned} Z_{C-\alpha} &= \frac{s+\alpha(1-s)}{s}\cdot\frac{1}{C} \\ &= \frac{(1-\alpha)s+\alpha}{sC} = \frac{1-\alpha}{C}+\frac{1}{s(C/\alpha)}, \end{aligned} \quad (8) $$

where α is the fractional order of capacitor. It can be seen from Eq. (8) that the fractional capacitor impedance model in the *s*-domain defined by CF derivative can be equivalent to integer-order resistors and capacitors. That is, a fractional-order capacitor with a capacitance of C and an order of α is equivalent to an integer-order resistor with a resistance of $(1-\alpha)/C$ in series with an integer-order capacitor with a capacitance of C/α. In the same way, the equivalent admittance of a fractional inductor via CF operator is expressed as

$$ \begin{aligned} G_{L-\beta} &= \frac{s+\beta(1-s)}{sL} \\ &= \frac{(1-\beta)s+\beta}{sL} = \frac{1-\beta}{L}+\frac{1}{s(L/\beta)}, \end{aligned} \quad (9) $$

where β represents the fractional order of inductor. Similarly, fractional-order inductor is composed of integer-order resistor and inductor.

Therefore, we obtain the analog realization models of the fractional-order capacitor and inductor defined by CF derivative shown in Figure 1.

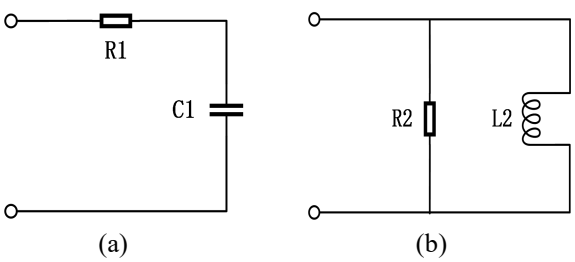


Fig.1 Analog realization models of fractional capacitor and inductor defined by CF derivative, (a) capacitor; (b) inductor.

The calculation method of the parameters in Fig. 1 is

$$\begin{cases} R1=\dfrac{1-\alpha}{C}, C1=\dfrac{C}{\alpha} \\ R2=\dfrac{L}{1-\beta}, L2=\dfrac{L}{\beta} \end{cases}, \tag{10}$$

where C represents the actual capacitance value, and L represents the actual inductance value. According to Eq. (10), the analog realization models of capacitors and inductors of any order in the range of 0 to 1 can be established.

2.3 Modeling of the Fractional RC Circuits

There are two types of simple RC circuits: series and parallel circuits, as shown in Figure 2.

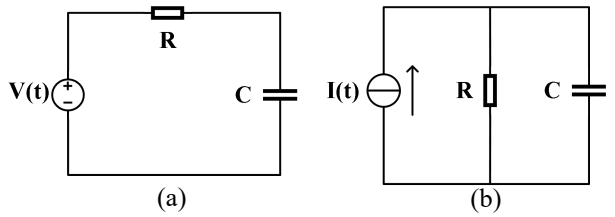


Fig.2 RC series circuit and RC parallel circuit, (a) series; (b) parallel.

In Figure 2 R is the resistance, C is the capacitance, $V(t)$ is the voltage source of the series circuit, $I(t)$ is the current source of the parallel circuit, and the source type is arbitrary. Based on the CF definition, the time-domain expressions of the capacitor voltages of the fractional-order RC circuits are derived under the indefinite input source.

According to Kirchhoff's Voltage Law (KVL) of the circuit, the RC series circuit can be modeled as

$$ {}_0D_t V_C(t)+\frac{1}{RC}V_C(t)=\frac{1}{RC}V(t), \tag{11}$$

where $V(t)$ is the voltage source, V_C is the voltage across the capacitor, and ${}_0D_t$ is the integer-order derivative with respect to t. Convert Eq. (11) into the fractional differential equation defined by CF derivative, and we can get

$$ {}^{CF}_{0}D_t^{\alpha} V_C(t)+\frac{1}{RC}V_C(t)=\frac{1}{RC}V(t), \tag{12}$$

where α represents the order of capacitor. Assuming that the capacitor voltage has an initial value of V_0, taking the Laplace transform to Eq. (12) by Eq. (3), we can obtain the transfer function of the capacitor voltage as follows

$$\begin{aligned} V_C(s)&=\frac{1-\alpha}{\Lambda}V(s)+\frac{(RC)V_0}{\Lambda s+\alpha} \\ &+\frac{1-\alpha}{\Lambda}\cdot\frac{\alpha/(1-\alpha)-(\alpha/\Lambda)}{s+(\alpha/\Lambda)}V(s), \end{aligned} \tag{13}$$

where $\Lambda = RC+1-\alpha$. Then we take the inverse Laplace transform to (13), and the time-domain expression of the capacitor voltage under the zero initial condition can be determined as follows

$$\begin{aligned} V_C(t)&=\frac{1-\alpha}{\Lambda}V(t) \\ &+\frac{1-\alpha}{\Lambda}\left(\frac{\alpha}{1-\alpha}-\frac{\alpha}{\Lambda}\right)\int_0^t V(t-\tau)\cdot\exp\left(-\frac{\alpha}{\Lambda}\tau\right)d\tau. \end{aligned} \tag{14}$$

In the same way, the capacitor voltage of the fractional RC parallel circuit under the zero initial condition can be expressed as

$$\begin{aligned} V_C(t)&=R\cdot\frac{1-\alpha}{\Lambda}I(t) \\ &+R\cdot\frac{1-\alpha}{\Lambda}\left(\frac{\alpha}{1-\alpha}-\frac{\alpha}{\Lambda}\right)\int_0^t I(t-\tau)\cdot\exp\left(-\frac{\alpha}{\Lambda}\tau\right)d\tau, \end{aligned} \tag{15}$$

where $I(t)$ is the current source.

2.4 Modeling of the Fractional RL Circuits

The RL series and parallel circuits are shown in Figure 3.

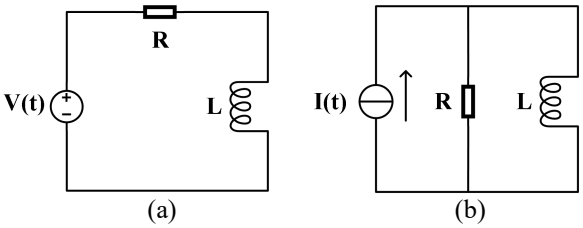


Fig.3 RL series circuit and RL parallel circuit, (a) series; (b) parallel.

R is resistance, L is inductance, $V(t)$ is any type of voltage source, and $I(t)$ is any type of current source. Similar to the numerical modeling of the RC circuit, the expressions of the inductor currents of the fractional-order RL circuits in time domain are derived.

If we apply the KVL, we can obtain the equation of the fractional RL series circuit as

$$ {}^{CF}_{0}D_t^{\alpha} I_L(t)+\frac{R}{L}I_L(t)=\frac{V(t)}{L}, \tag{16}$$

where $I_L(t)$ is the current of the fractional-order inductor. Applying Laplace transform on both sides of Eq. (16), the inductor current in the s-domain can be presented as

$$\begin{aligned} I_L(s)&=\frac{1-\alpha}{\Delta}V(s)+\frac{I_0 L}{\Delta}\frac{1}{s+(R\alpha/\Delta)} \\ &+\frac{1-\alpha}{\Delta}\cdot\frac{\alpha/(1-\alpha)-(R\alpha/\Delta)}{s+(R\alpha/\Delta)}V(s), \end{aligned} \tag{17}$$

where $\Delta = L\ +R\ -R\,\alpha$, I_0 is the initial value of the inductor current.

After taking the inverse Laplace transform to Eq. (17), the time-domain expression of the inductor current under the zero initial condition is given by

$$I_L(t)=\frac{1-\alpha}{\Delta}V(t)+\frac{1-\alpha}{\Delta}\left(\frac{\alpha}{1-\alpha}-\frac{R\alpha}{\Delta}\right)\int_0^t V(t-\tau)\cdot\exp\left(-\frac{R\alpha}{\Delta}\tau\right)d\tau. \quad (18)$$

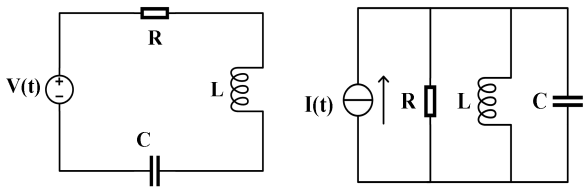
In the same way, we can derive the inductor current of the fractional-order RL parallel circuit under the zero initial condition as

$$I_L(t)=R\cdot\frac{1-\alpha}{\Delta}I(t)+R\cdot\frac{1-\alpha}{\Delta}\left(\frac{\alpha}{1-\alpha}-\frac{R\alpha}{\Delta}\right)\int_0^t I(t-\tau)\cdot\exp\left(-\frac{R\alpha}{\Delta}\tau\right)d\tau, \quad (19)$$

where $I(t)$ is the current source.

2.5 Modeling of the Fractional RLC Circuits

Simple RLC circuits include two forms of series and parallel, which are shown in Figure 4.



(a) (b)

Fig.4 RLC series circuit and RLC parallel circuit, (a) series; (b) parallel.

Where $V(t)$ is the voltage source, $I(t)$ is the current source. In this section, $Z_{C\text{-}\alpha}$ represents the impedance of a fractional capacitor, and $Z_{L\text{-}\beta}$ represents the impedance of a fractional inductor. According to Fig. 4, the impedance value of the fractional RLC series circuit can be determined as follows

$$Z_{sRL-\beta C-\alpha}=R+Z_{L-\beta}+Z_{C-\alpha}=R+\frac{sL}{(1-\beta)s+\beta}+\frac{(1-\alpha)s+\alpha}{sC}, \quad (20)$$

where β is the order of inductor, α is the order of capacitor. When the initial values of the capacitor voltage and inductor current in the circuit are 0, we derive the transfer function of the capacitor voltage according to Eq. (20) as

$$\frac{V_C(s)}{V(s)}=\frac{Z_{C-\alpha}}{Z_{sRL-\beta C-\alpha}}=\frac{A_1s^2+B_1s+\alpha\beta}{A_2s^2+B_2s+\alpha\beta}$$
$$\begin{cases}A_1=(1-\alpha)(1-\beta)\\B_1=\alpha+\beta-2\alpha\beta\\A_2=LC+(1-\beta)RC+(1-\alpha)(1-\beta)\\B_2=RC\beta+\alpha+\beta-2\alpha\beta\end{cases}, \quad (21)$$

where V_C is the voltage across the capacitor. It can be seen from Eq. (21) that the voltage across the capacitor is not only related to the capacitance and inductance value, but also to the orders of capacitor and inductor.

Similarly, the impedance model of the RLC parallel circuit can be derived. But because the form of impedance is more complex, so we express it in admittance, that is

$$G_{pRL-\beta C-\alpha}=\frac{1}{R}+\frac{1}{Z_{C-\alpha}}+\frac{1}{Z_{L-\beta}}=\frac{1}{R}+\frac{sC}{s(1-\alpha)+\alpha}+\frac{s(1-\beta)+\beta}{sL}. \quad (22)$$

When the current source I is input, the transfer function of the inductor current in the fractional-order parallel RLC circuit is derived according to Eq. (22) as

$$\frac{I_L(s)}{I(s)}=\frac{1}{G_{pRL-\beta C-\alpha}Z_{L-\beta}}=\frac{A_3s^2+B_3s+\alpha\beta}{A_4s^2+B_4s+\alpha\beta}$$
$$\begin{cases}A_3=(1-\alpha)(1-\beta)\\B_3=\alpha+\beta-2\alpha\beta\\A_4=LC+(1-\alpha)(L/R)+(1-\alpha)(1-\beta)\\B_4=(L/R)\alpha+\alpha+\beta-2\alpha\beta\end{cases}, \quad (23)$$

where I_L is the current across the inductor. So far, the mathematical models of the fractional RLC circuits in the complex frequency domain are established through the transfer function.

3. SIMULATION AND ANALYSIS OF FRACTIONAL-ORDER CIRCUITS

Since the parallel circuits are derived in the same way as the series circuits, only the series circuits are simulated in this section. According to the equations in Section 2, the numerical simulation programs by the MATLAB software are written to simulate the mathematical models of the series RC, RL and RLC circuits. Besides, the corresponding circuit simulations are performed by applying the analog implementation of fractional components in the Multisim software to verify the correctness of the deduced mathematical models.

3.1 RC Series Circuit Simulation

Fractional-order RC circuits are the basis for studying fractional-order circuits. On the basis of this research, it is convenient to expand more actual circuits into the field of fractional-order.

In order to analyze the effects of the fractional order α on the voltage across the capacitor, three different fractional orders 0.1, 0.5, 0.9, and the integer-order were selected to perform numerical simulations to obtain the voltage curves under different orders.

Given the values, V=15V, R=200Ω, C=1000μF, and V is the step voltage source. The numerical simulation

program is based on Eq. (14), and the result curves of different orders are plotted on the same graph, which is convenient for comparison and analysis.

Then we build the analog realization circuit of the fractional order circuit in the Multisim software to complete the RC circuit simulations, and get the simulation data to plot the result curves. The simulation model of the circuit is shown in Fig. 5.

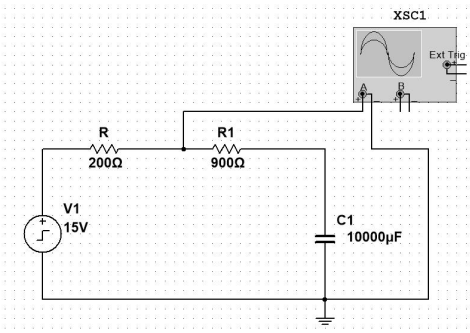
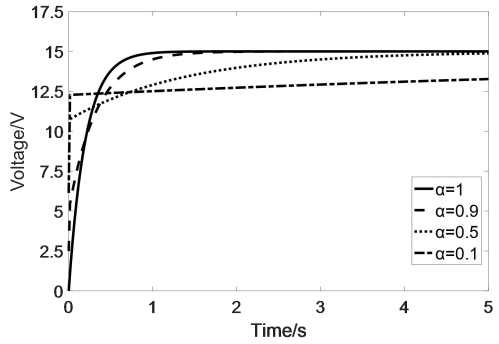
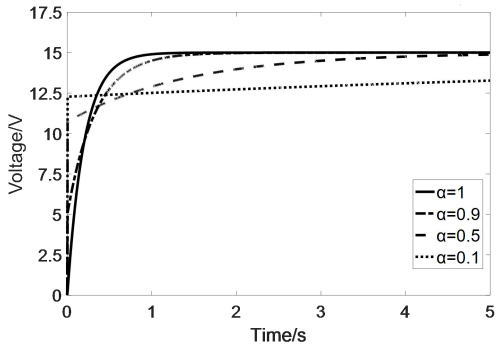


Fig.5 Simulation model of fractional RC series circuit.

The circuit parameters in Fig. 5 are the same as the numerical simulation parameters. *V1* is a 15V step voltage source, and *R1* and *C1* make up a 0.1-order fractional capacitor. According to Eq. (10) of the analog realization model of the fractional capacitor, the order changes with the values of *R1* and *C1*.



(a)



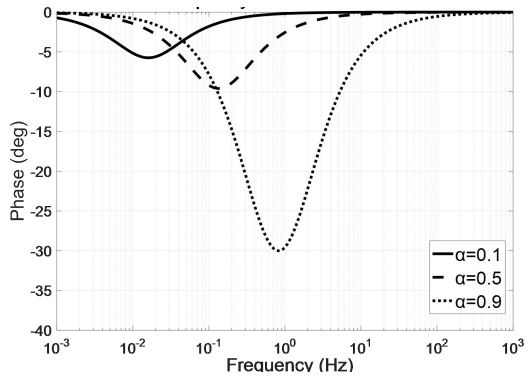
(b)

Fig.6 Voltage across the RC series circuit capacitor, in (a) numerical simulation and (b) circuit simulation, α is the fractional order of capacitor.

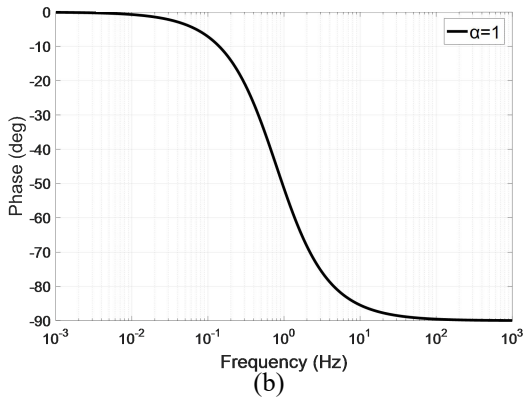
The results of the numerical simulation and the corresponding circuit simulation are shown in Figure 6.

The Figure 6 shows the behavior of the voltage across the capacitor, (a) and (b) are compared to verify the correctness of Eq. (14). It can be noted that when the circuit inputs a step voltage source, the integer-order capacitor voltage rises exponentially from 0, and the fractional-order capacitor voltage rises from a certain initial value. And the smaller the order of the capacitor, the higher the initial voltage value and the longer the charging time. This is because the analog realization model of the fractional capacitor contains the component of resistance. The smaller the order, the greater the influence of resistance, and therefore the time constant of the circuit is also affected. Compared with integer-order capacitors, this property of fractional capacitor voltage shows the heterogeneity of actual capacitor (resistance and capacitance), indicating that the fractional capacitor model defined by CF derivative can more accurately describe the actual physical properties of capacitors. The charging time of the fractional capacitor becomes longer, which reflects the memory characteristics of fractional calculus, and is suitable for describing components with memory effects such as capacitors.

Since the time-domain equations derived in this paper are based on any type of voltage source, we replaced the step voltage source with a sinusoidal voltage source and analyzed the response curves of the fractional circuit at different frequencies. Given a sinusoidal voltage source with an amplitude of 15V, the values of capacitance and resistance remain unchanged, and four different fractional orders are also used for numerical simulations. In order to analyze the behavior of the capacitor voltage at different frequencies more clearly, we obtain the characteristic curves of the voltage in frequency domain. The phase-frequency curve results are shown in Figure 7.



(a)



(b)

Fig.7 Phase frequency curve of RC series circuit, in (a) fractional order and (b) integer order, α is the fractional order of capacitor.

It can be noted from the results shown in Fig. 7 that at high frequencies, the nature of the components in which the fractional-order and integer-order capacitors play a leading role in the circuit is different. When the order is less than 1, the phase-frequency characteristic curve of fractional capacitor has a pole, and as the frequency increases, it eventually tends to 0°. Because the analog realization model of the fractional capacitor is a resistor and a capacitor in series, the capacitive reactance of the capacitor is very small at high frequencies, so the resistor plays a leading role. At this time, the impedance of the fractional capacitor becomes a pure resistance, and its value is $(1-\alpha)/C$. While the phase-frequency curve of integer-order capacitor eventually tends to -90° as the frequency increases, indicating that the capacitive reactance of the capacitor plays a leading role in the circuit. This phenomenon shows that the fractional order can better describe the characteristics of actual capacitor changing with frequency.

The amplitude-frequency curve results are shown in Figure 8.

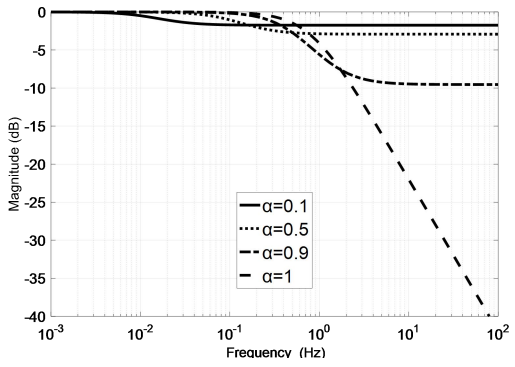


Fig.8 Amplitude-frequency curve of RC series circuit, where α is the order of capacitor.

The amplitude-frequency curves shown in Figure 8 show the variation of the capacitor voltage's amplitude and the shift of the cut-off frequency. In the frequency range of 0.001Hz-100Hz, when the frequency is increased from the cut-off frequency, the voltage amplitude decreases, and the amplitude below the cut-off frequency is almost constant. Above a certain frequency, the voltage amplitude of the fractional capacitor is greater than the integer order. And the smaller the order, the greater the voltage amplitude. This phenomenon can explain the irreversible dissipation effect of the actual capacitor, such as ohmic friction, which increases the voltage across the capacitor. Therefore, the voltage of the fractional order is higher than the integer order, indicating that using fractional calculus to describe the capacitor is more in line with the actual situation. Besides, as the order decreases, the cut-off frequency moves to the left, indicating that the frequency range of the attenuation of the fractional-order capacitor's voltage is greater than the integer order.

It can be noted from the above analysis results that the fractional order α makes the RC circuit more flexible and diverse.

3.2 RL Series Circuit Simulation

The fractional-order RL circuit is also the basis of the fractional-order circuit, and its simulation experiment is similar to the RC circuit. Given the values, V=15V, L=100mH, R=2Ω, and V is the step source. Select fractional orders 0.3, 0.7, 0.9 and the integer-order to complete the numerical simulations according to Eq. (18). The simulation circuit is shown in Fig. 9.

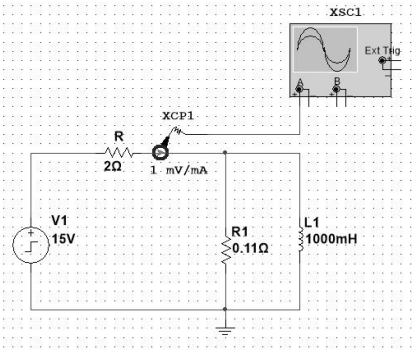
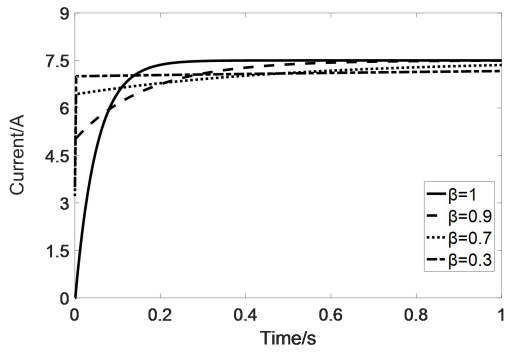


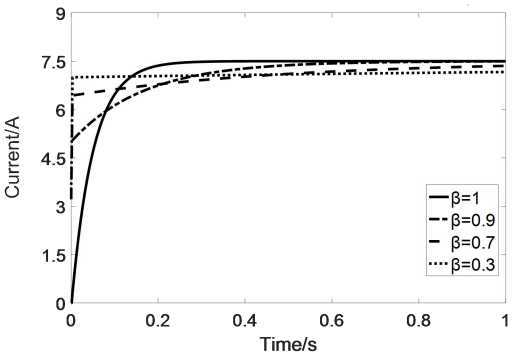
Fig.9 Simulation model of fractional RL series circuit.

The circuit parameters in Fig. 9 are the same as the numerical simulation parameters. Among them, *V1* is a 15V step voltage source, and *R1* and *L1* make up a 0.1-order fractional inductor. Changing the values of *R1* and *L1* can change the order of the inductor.

The results of the numerical simulation and the corresponding circuit simulation are shown in Figure 10.



(a)

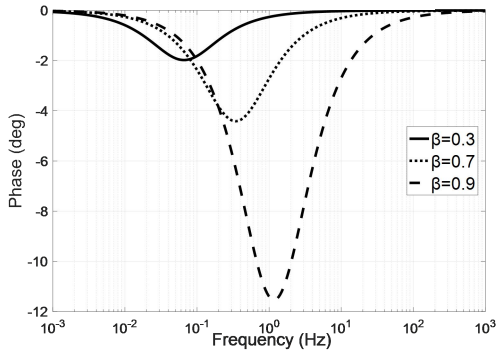


(b)

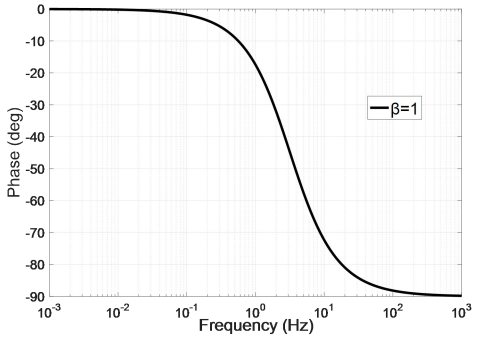
Fig.10 Current across the RL series circuit inductor, in (a) numerical simulation and (b) circuit simulation, β is the fractional order of inductor.

According to the comparison between (a) and (b) in Figure 10, the results of the numerical simulation are completely consistent with the circuit simulation, indicating that the Eq. (18) is correct. The behavior of the current across the inductor shows that when the circuit inputs a step voltage source, the integer-order inductor current rises from 0, but the fractional-order inductor current has an initial value. And the rise time and initial value of the current increase as the fractional order decreases. As the dual component of the capacitor, the changing trend of the fractional-order inductor's current is similar to the fractional-order capacitor's voltage, which also shows the heterogeneity (resistance and inductance) and memory characteristics of the actual inductor.

Then choose a sinusoidal voltage source with an amplitude of 15V, the values of inductance and resistance remain unchanged, and use four different orders of fractional inductor for numerical simulations to obtain the characteristic curves of the inductor current in frequency domain. The phase-frequency curves are shown in Fig. 11.



(a)



(b)

Fig.11 Phase-frequency curve of RL series circuit, in (a) fractional order and (b) integer order, β is the fractional order of inductor.

As the results shown in Fig. 11, it can be noted that the phase-frequency curve of the fractional-order inductor eventually tends to 0°. Because its analog realization model is a resistor and an inductor connected in parallel, and the inductive reactance is very large at high frequencies. Therefore the fractional-order inductor's impedance becomes a pure resistance at high frequencies, and its value is $L/(1-\beta)$. However, the phase-frequency curve of the integer-order inductor's current eventually tends to -90°. In reality, the actual inductors produce frequency dependent losses, so we find that classical integer-order inductors cannot consider these losses, and the fractional model of inductor can better describe the nature of actual inductors.

The amplitude-frequency curve of the inductor current is shown in Fig.12.

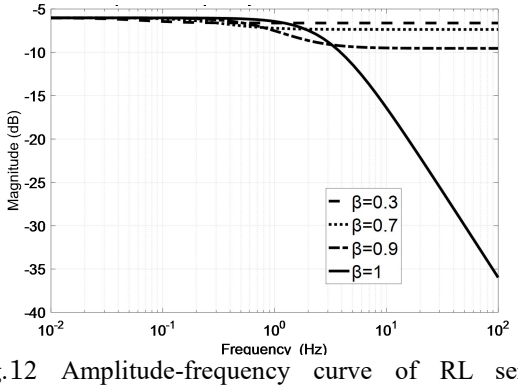


Fig.12 Amplitude-frequency curve of RL series circuit, where β is the order of inductor.

The amplitude variation of the inductor current shown in Fig. 12 is similar to that of the RC series circuit, so it will not be described in detail. In reality the actual inductor has nonlinearity. For example, eddy current or hysteresis loss causes the inductor with a ferromagnetic core to heat up, thereby increasing the current flowing through the actual inductor. The fractional-order current is higher than the integer-order, indicating that classical integer-order cannot consider these frequency dependent losses and using fractional calculus to describe inductor is more suitable for the actual situation.

3.3 RLC Series Circuit Simulation

The integer-order RLC circuit is a typical second-order circuit, while the fractional-order RLC circuit increases the orders of capacitor and inductor, and the circuit design is more complicated. For convenience, the orders of capacitor and inductor are the same in the simulation experiment in this section.

Given the values, V=15V, R=1Ω, L=100mH, C=0.022F, and V is the step source. Then select two orders of 0.99 and 0.995 for numerical simulations. According to Eq.(21) in Section 2, the transfer function model is established in the Simulink environment of MATLAB to complete the numerical simulation. The simulation model of fractional RLC series circuit is shown in Fig. 13.

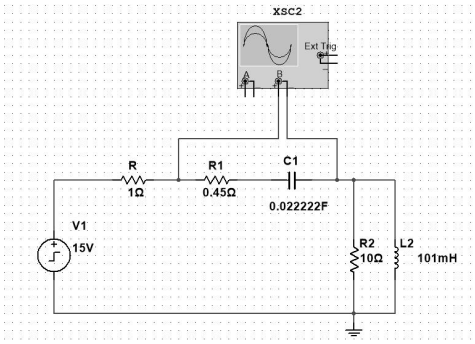
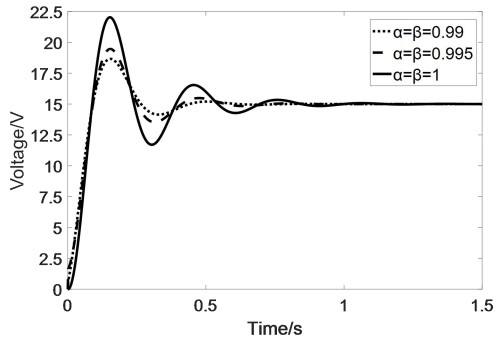


Fig.13 Simulation model of fractional RLC series circuit.

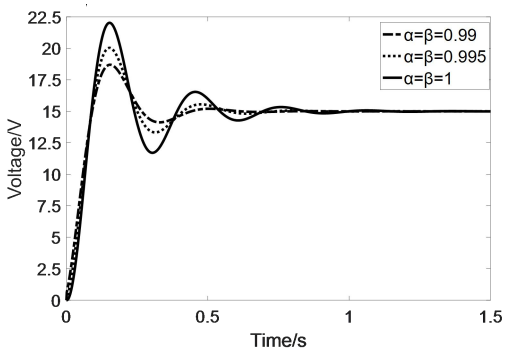
In Fig. 13, *R1* and *C1* constitute a fractional capacitor, *R2* and *L2* constitute a fractional inductor, and the remaining parameters are the same as the numerical simulation.

The results of the numerical simulation and the corresponding circuit simulation are shown in Fig. 14.

As shown in Fig. 14, the results of circuit simulation are almost identical to the numerical simulation, which proves that Eq. (21) is correct. Among them, the step response is the process of attenuated oscillation, the steady-state



(a)



(b)

Fig.14 Voltage across the RLC series circuit capacitor, in (a) numerical simulation and (b) circuit simulation, α is the fractional order of capacitor and β is the fractional order of inductor.

voltage is 15V, and there is no steady-state error. At this time, the circuit system is in an underdamped state. After comparing the step response curves at different orders in Fig. 14, we find that the smaller the order, the overshoot and the number of oscillations of the voltage across the capacitor decrease. This phenomenon shows that the fractional order increases the damping ratio of the second-order system.

4. CONCLUSIONS

In this work, we completed the modeling of capacitor and inductor via Caputo-Fabrizio operator, and established numerical models of the typical fractional-order circuits. The combination of numerical simulation and circuit simulation verifies the correctness of the numerical models. According to the simulation results, we found the following characteristics of the fractional circuit:

- When the RC and RL circuits input a step source, the fractional capacitor voltage has an initial value, and the smaller the order, the higher the initial value of the voltage, and the longer the charging time; the characteristic of the fractional inductor

current is the same as that of the capacitor. The reason is that the analog realization model of the fractional components under the CF definition contains resistance. The smaller the order, the greater the influence of resistance, which increases the time constant of the circuit and the initial value of the voltage or current of the fractional-order component.

- When RC and RL circuits input sinusoidal source, the nature of the fractional capacitor and inductor is related to frequency. As the frequency gradually increases, the fractional-order component finally exhibits resistance characteristics. Within a certain frequency range, the amplitude of the capacitor voltage and inductor current decreases as the fractional order increases.
- When the RLC series circuit inputs step source, the smaller the fractional order, the smaller the overshoot and the number of oscillations. It shows that the fractional order increases the damping ratio of the entire system compared to the integer order.

Through comparison with integer-order circuits, it is found that the fractional-order equivalent circuit is more in line with the actual physical properties of the circuit, and can describe the heterogeneity and memory characteristics of actual circuit components. Besides, the results show that the fractional order makes the circuit characteristics more flexible and the circuit design has more degrees of freedom.

Acknowledgements

This work was supported by the National Science Foundation of China under Grant 61873035.

References

[1] I. S. Jesus and J. A. Tenreiro Machado, “Development of fractional order capacitors based on electrolyte processes”, Nonlinear Dyn., Vol. 56, No. 1-2, 2009, pp. 45-55.

[2] A. Allagui, et al., “Review of fractional-order electrical characterization of supercapacitors”, J. Power Sources, Vol. 400, 2018, pp. 457-467.

[3] A. A. Raorane, M. D. Patil and V. A. Vyawahare, “Analysis of full-wave controlled rectifier with lossy inductive load using fractional-order models”, in 2015 IEEE International Conference on Industrial Instrumentation and Control, 2015, pp. 750-755.

[4] J. F. Gómez Aguilar, “Behavior characteristics of a cap-resistor, memcapacitor, and a memristor from the response obtained of RC and RL electrical circuits described by fractional differential equations”, Turk. J. Electr. Eng. Comput. Sci, Vol. 24, No.3, 2016, pp. 1421-1433.

[5] I. Petráš, “A note on the fractional-order Chua’s system”, Chaos Solitons Fractals, Vol. 38, No. 1, 2008, pp. 140-147.

[6] I. Petras and Y. Chen, "Fractional-order circuit elements with memory", in International Carpathian Control Conference, 2012, pp. 552-558.

[7] A. A. Raorane, M. D. Patil and V. A. Vyawahare, “Analysis of full-wave controlled rectifier with lossy inductive load using fractional-order models”, in 2015 IEEE International Conference on Industrial Instrumentation and Control, 2015, pp. 750-755.

[8] A. Adhikary, P. Sen, S. Sen and K. Biswas, “Design and Performance Study of Dynamic Fractors in Any of the Four Quadrants”, Circuits Syst. Signal Process., Vol. 35, No. 6, 2016, pp. 1909-1932.

[9] A. G. Radwan and K. N. Salama, “Passive and Active Elements Using Fractional $L_\beta C_\alpha$ Circuit”, IEEE Trans. Circuits Syst. I-Regul. Pap., Vol. 58, No. 10, 2011, pp. 2388-2397.

[10] M. C. Tripathy, K. Biswas and S. Sen, “A Design Example of a Fractional-Order Kerwin–Huelsman–Newcomb Biquad Filter with Two Fractional Capacitors of Different Order”, Circuits Syst. Signal Process., Vol. 32, No. 4, 2013, pp. 1523-1536.

[11] R. Martinez, Y. Bolea, A. Grau, and H. Martinez, “Fractional DC/DC converter in solar-powered electrical generation systems”, in IEEE International Conference on Emerging Technologies & Factory Automation, 2009, pp. 1475-1480.

[12] M. Caputo and M. Fabrizio, “A new definition of fractional derivative without singular kernel”, J. Progr. Fract. Differ. Appl., Vol. 1, No. 2, 2015, pp. 1–13..

[13] A. Atangana and R. T. Alqahtani, “Numerical approximation of the space-time Caputo-Fabrizio fractional derivative and application to groundwater pollution equation”, Adv. Differ. Equ., Vol. 2016, No. 1, 2016.

[14] A. Alsaedi, J. Nieto and V. Ventesh, “Fractional electrical circuits”, Advances in Mechanical Engineering, Vol. 7, No. 12, 2015, pp. 1-7.

[15] K. A. Abro and A. A. Memon, “Functionality of circuit via modern fractional differentiations”, Analog Integr. Circuits Process., Vol. 99, No. 1, 2019, pp. 11-21.

Design and Implementation of Multi-Function Servo Experiment System Based on High-Speed Bus

Yonghua Xiong*, Ke Li*, Zhentao Liu*, Jinhua She**, Min Wu*
* School of Automation, China University of Geosciences
Wuhan 430074, P.~R.~China
** School of Engineering, Tokyo University of Technology
Hachioji, Tokyo 192-0928, Japan

Abstract

In recent years, a number of breakthroughs have been made in the theoretical research of servo control algorithms, but most control algorithms only still stay in the simulation stage. They are difficult to be applied directly to practice platforms or complex industrial sites due to lack of experimental system suitable for verification of the effectiveness of different servo control algorithms. To address this problem, we design a multi-function servo control algorithm verification experiment system(MVES)that uses MATLAB/Simulink theoretical simulation model directly to communicate with TwinCAT3 PLC master program to carry out kinds of different servo control experiments. The MVES supports a variety of Simulink models and the operation process is simple and convenient, which greatly reduces the workload of algorithm test and has important practical value. In this work, two sets of comparative experiments are used to verify the versatility and superiority of MVES.

Keywords: Servo experiment system, Servo control, Multi-function experiment device, TwinCAT3.

1. INTRODUCTION

The Alternating Current servo system is one of the core components of industrial robots and an important tool in the field of automation and industrial production, which has been widely used in machinery manufacturing, metallurgy, transportation and other industries.

The simulation research of servo control theory based on MATLAB/Simulink has always been a research hotspot of universities and scientific research institutions[1-3].
There are many great breakthroughs have been made in theoretical research of servo control algorithms. However, in the process of applying theoretical models to practical industrial servo systems. There are three main problems in the following areas:

(1) The theoretical model cannot be directly applied to the industrial servo system. It needs to be converted into PLC industrial programming language or other assembly languages for system control[4]. The conversion workload is large and needs to be provided more equipment I/O interfaces, the operation process is inconvenient for researchers.
(2) Theoretical models cannot simulate complex industrial environments, and effective algorithm verification cannot be performed until the theoretical model is applied to actual industrial systems[5].
(3) Most servo algorithm verification devices are not very versatile[6]. A set of devices often only supports a few specific servo experiments, and it is impossible to carry out many different types of servo algorithm verification experiments.

In general, researchers apply servo control algorithms to practical industrial systems in two ways: The first way is to convert Simulink servo algorithms with good simulation effects into PLC industrial programming language or other assembly languages[7]. Then test the control effect through the test system. The disadvantages of this method are obvious: the algorithm conversion workload is large and requires a specific test system, and there are many restrictions, which are often only applicable to specific servo algorithm verification[8]. The second way is to use the simulation model to carry out the experiment, and then use the control parameters with good simulation effect to test on the specific test system of the same type. After achieving good results, it can be applied to practical industrial systems. This method is limited to a specific test system, and its test system and actual system are often the same type, which cannot meet the test requirements of different industrial systems. Usually, the control parameters obtained by simulation can not guide the complex test system to a large extent[9].

Aiming at the problems that most servo experimental devices are not universal and the experiment conversion steps are cumbersome, we designed a general-purpose servo control algorithm experimental system(MVES).

The MVES designed in this paper is specially used for the validity verification of many different types of servo control algorithms, without complicated conversion or configuration work. MATLAB/Simulink models can directly interact with the PLC master program in TwinCAT3 with our experimental system, and then we use EtherCAT communication bus to allow the PLC control the servo system to carry out experiments, thus eliminating the cumbersome steps of model conversion and making the model debugging convenient and feasible. Only the Simulink simulation model need to be adjusted during the debugging process to carry out the next experiment or collect the next set of data, which lets the algorithm testing process becomes easy to operate. Different Simulink models can be used to carry out different types of servo algorithm verification experiments. The system has strong versatility, it is suitable for Simulink models of various servo control algorithms. Finally, we used two sets of comparative experiments to verify the effectiveness of the MVES.

The rest of this paper is organized as follows. In the next section, the overall design of the MVES is discussed. Then section 3 presents physical system and testing. Section 4 describes the conclusions..

2. OVERALL SYSTEM DESIGN

This section begins with the whole system, introducing the composition and structure of the MVES and its overall working principle. And then we introduce the MVES from two aspects: servo motor experimental device design and communication framework design.

The experimental system designed in this paper has four main components, as shown in Fig.1, servo motor experimental device, industrial computer(IPC), driver and encoder.

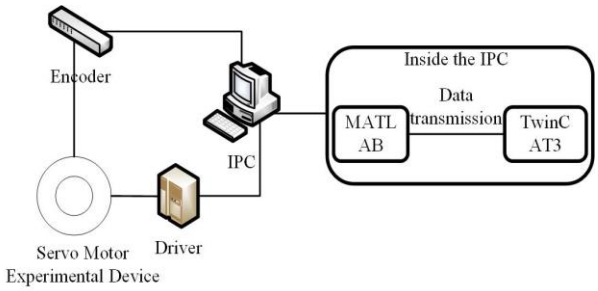


Fig.1 Experimental system overall structure.

The IPC runs the MATLAB/Simulink control algorithm in real time and sends the corresponding control command, and receives the real-time parameters returned by the encoder. Then the PLC sends the control command to the servo driver through the EtherCAT bus, and the servo driver control the motor to carry out relevant experiments. During the operation of the experimental system, we implemented the direct control of the servo motor by the MATLAB/Simulink control algorithm. It does not need to convert the control algorithm into other industrial programming languages to control the motor, which greatly reduces the workload of the theoretical simulation model to the actual industrial application. And due to the theoretical simulation model runs in real time in the MATLAB environment, which is independent of the PLC main control program. Therefore, this experimental system can support different types of servo control algorithms and has strong versatility.

2.1 Servo motor experimental device design

The servo motor experimental device is mainly composed of two 750 W servo motors of the IS620N series of Huichuan Company. Two motors were coaxially mounted to form the supporting experimental device, as shown in Fig.2.

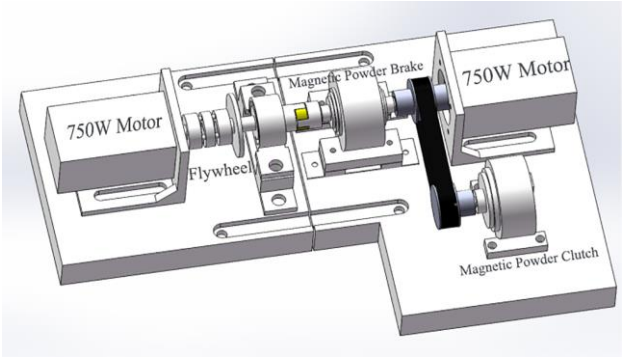


Fig.2 Experimental system overall structure.

The main body of the device consists of two 750 W motors, a magnetic powder brake, and a magnetic powder clutch. The right motor is connected to the magnetic powder clutch through an asynchronous transmission belt and directly connected to the magnetic powder brake. A magnetic slot on the right side of the magnetic clutch can be used to install the flywheel. Another motor is directly connected to the flywheel, and the two motors are mounted coaxially. On the one hand, experiments can be carried out independently, such as inertia identification, parameter self-tuning, etc. On the other hand, a pair of experimental devices can be formed, and a disturbance suppression experiment could be performed to verify the synchronization accuracy of the motion control of the servo system. The left motor is used for the inertia identification experiment by adding flywheels of different quality. The right motor can dynamically adjust the moment of inertia with the magnetic powder clutch to carry out the inertia identification experiment. At the same time, it could directly connect with the load torque continuously adjustable magnetic powder brake and

could carry out the disturbance suppression experiment of different forms of loading.

2.2 Communication framework design

Since the control algorithm and the servo motor device communicate through the main control program, under this premise, the communication feedback speed is guaranteed to be fast, which puts high requirements on the operating environment of the main control program. After research and comparison, we chose Beckhoff's TwinCAT3 platform, which is control software based on PC platform and Windows operating system. One of the biggest advantages of TwinCAT3 is its excellent scalability, supporting the control layer of the Matlab/Simulink module, which can extremely reduce the workload of developers. We achieve data communication through the Advanced Design System(ADS) protocol, which is the communication protocol within TwinCAT3.

The overall control framework of this experimental system is shown in Fig.3. We implement the real-time control effect of the data transmission between the control platform and the servo drive layer through the EtherCAT bus. TwinCAT3 is integrated into Visual Studio software, which is divided into the development layer and operation layer. Data transmission between the development layer and the runtime layer through the ADS protocol[10]. With the TwinCAT3 I/O port mapping, cyclic data can be acquired in the process mapping via different Fieldbuses. Different Fieldbuses can be run on different cycles with the same CPU. The configuration of the Fieldbus and process mapping is done in Visual Studio.

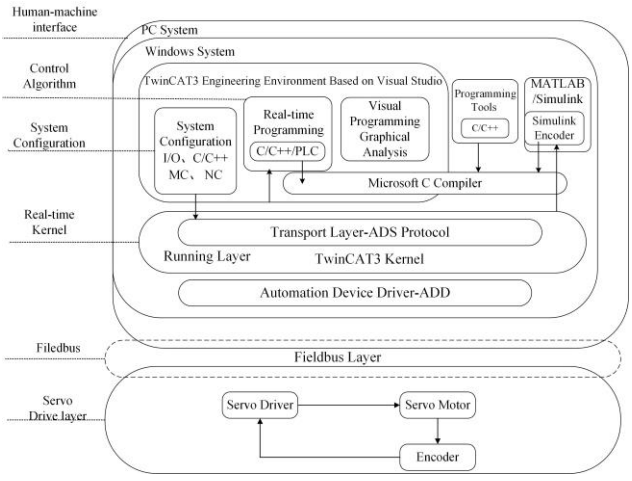


Fig.3 Control structure framework.

The biggest innovation of the MVES is that the theoretical simulation model does not need to be converted into a programming language in the TwinCAT3 environment.

The experimental data flow of the MVES is shown in Fig.4. The entire communication framework consists of three parts: the Simulink theoretical control algorithm model, the TwinCAT3 engineering environment, and the servo system. The Simulink theoretical control algorithm model is completed in MATLAB. We used the Beckhoff TE1410 plug-in to map Simulink variables to PLC variables in the TwinCAT3 engineering environment. The process variables under Simulink can be mapped to the PLC variables in TwinCAT3. The PLC project under the TwinCAT environment carries data transmission with the driver through the EtherCAT bus. The servo driver sends control commands to the motor. The motor then feeds back the real-time speed and other parameters to the driver, we implement two-way communication to carry out the experimental purpose of the verification control algorithm in this way.

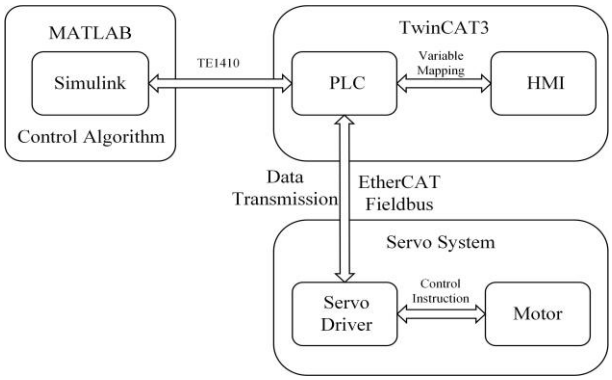


Fig.4 System control data flow.

3. PHYSICAL SYSTEM AND TESTING

This section will show the physical effects of the MVES designed in this work, including the overall hardware structure model and the actual experimental results.

3.1 Experimental device

The physical MVES is shown in Fig.¥label{fig:5}5. The industrial computer and the driver communicate with each other through the EtherCAT bus for motor control. The industrial computer is equipped with software such as Visual Studio , MATLAB and integrates the TwinCAT3 development environment. The servo driver is integrated into the control cabinet. The control cabinet can dynamically adjust the current flowing through the magnetic powder clutch to increase the inertia to simulate the continuous variable load environment. In this environment, the dynamic inertia identification experiment can be achieved easily.

3.2 Experimental effect

To test the practicability of the MVES, two representative experiments in the servo control process are selected for

verification in this work: inertia online identification experiment and external disturbance suppression experiment. All the experimental control algorithms can independently complete the simulation experiment in the Simulink environment, and can also carry out the physical experiment by using the MVES.

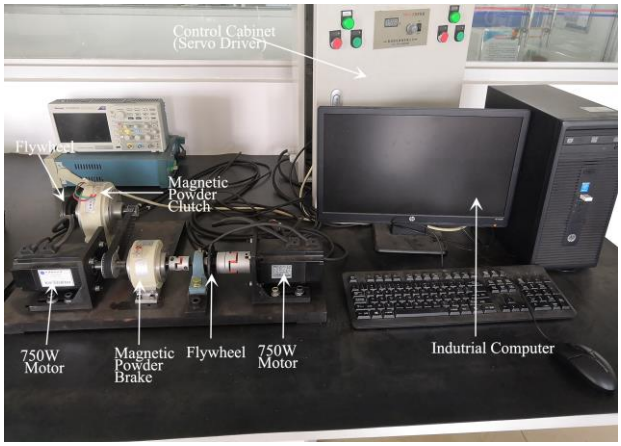


Fig.5 Whole system physical map.

(1) Inertia online identification experiment

The control performance of the industrial servo system is closely related to the accurate acquisition of the motor parameters. The research of high-precision inertia online identification has always been the most representative in the field of servo control. Therefore, this paper selects the inertia online identification experiment to verify the system effectiveness.

We use the MVES to carry out experiments, the inertia of the magnetic powder clutch current is adjusted through the control cabinet to simulate the variable load condition.The experimental result is shown in Fig.6:

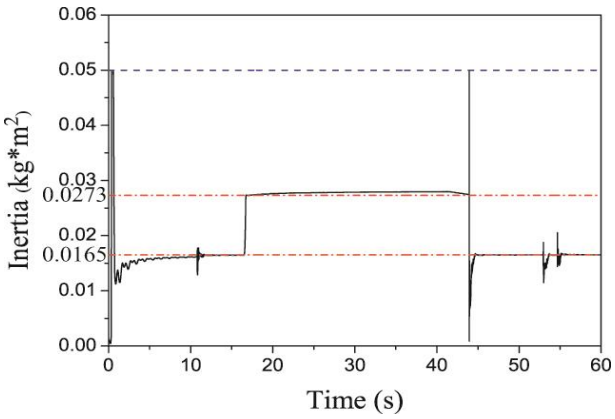


Fig.6 Inertia online identification(Physical experiment).

In the experimental process of Fig.6, after the motor starts for a period of time, the speed is stable and the moment of inertia is stable at around 0.0165kg*m^2. Then we increase the current flowing through the magnetic powder clutch to increase the inertia, the inertia increases to 0.0273kg*m^2, the current is canceled at the 44th second, and the inertia returns to the original state. We also added a slight disturbance at the 53th seconds in this experimental process.

From the above experimental result, the MVES designed in this work meets the high-precision identification requirements of inertia online, whose response speed is fast and the identification accuracy is high and it has high practical value in complex industrial sites.

(2) External disturbance suppression experiment

In the industrial production process, the servo system will inevitably be affected by disturbances from the external environment. External disturbances will seriously affect the control accuracy of the motor. Therefore, the research on the suppression of external disturbances has attracted widespread attention from many researchers. External disturbance suppression experiment is one of several basic types of experiments in the field of servo control. So we selected the external disturbance suppression as a verification experiment of the effectiveness and practicability of the system we designed.

The external disturbance suppression experiment was carried out by using the MVES. The left motor is the observation motor, the right motor forms the disturbance, and the observation control algorithm suppresses the disturbance. The experimental result which only used PI controller were shown as Fig.7:

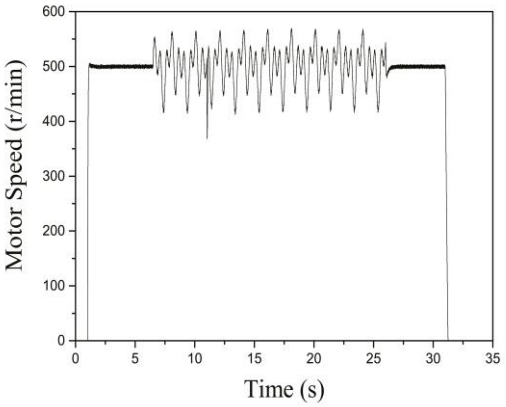


Fig.7 Disturbance suppression experiment(PI).

In the experiment which only used a PI controller, we added the disturbance at the 7th second, and the rotation began to fluctuate. At the 12th second, we dynamically adjusted the current of the magnetic clutch to increase the inertia, and the speed suddenly jumped and then immediately returned to the original target speed.

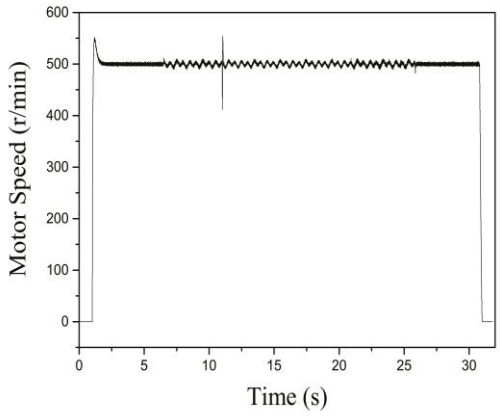


Fig.7 Disturbance suppression experiment(PI+EID).

When we use the sliding mode + EID controller, the disturbance suppression effect is significantly better than when only the PI controller is used, and the tracking speed fluctuation range is small. However, it can be seen from the figure that the speed of the inertia is still increasing instantaneously, but it also returns to normal, and the overall disturbance suppression effect is very obvious. Fig.8 show the experimental result of speed tracking.

It can be seen from the above comparison experiments that this experimental system can effectively verify the effectiveness of the servo system disturbance suppression algorithm, which can support the disturbance suppression algorithm verification using different controllers, and the experimental results are also consistent with the simulation results. The experimental system designed in this paper has good versatility, it can support the validity verification of a variety of servo control algorithms, and enables a variety of Simulink theoretical models to directly carry out motor control experiments.

4. CONCLUSION

In this work, we propose and implement the MVES that uses MATLAB/Simulink theoretical simulation model to communicate with TwinCAT3 PLC master program to carry out different types of servo control experiments, which effectively solves the problem of algorithm validation before the theoretical simulation model is applied to practical industry. The theoretical simulation model of MATLAB/Simulink can be directly applied to the system for verification of algorithm validity, without complicated conversion and configuration work. The MVES supports a variety of Simulink simulation models and the operation process is simple and convenient, which greatly reduces the workload of the theoretical researchers' test algorithms.

We have carried out two sets of experiments to verify the effectiveness and versatility of the system: online inertia identification and external disturbance suppression. The comparative analysis results show that the MVES supports multiple theoretical simulation models and have strong practicability and versatility. The MVES provides theoretical researchers with a convenient algorithm verification platform, which reduces their experiment workload and has good practical value.

References

[1] Wang J, Yang B J, and Xu Z X, “Development of a compact 750KVA three-phase NPC three-level universal inverter module with specifically designed busbar”, Twenty-Fifth Annual IEEE Applied Power Electronics Conference and Exposition (APEC), 2010.

[2] Leng C M, Chen C L and Lee C J, “A simple circuit to remove X-cap bleeder resistor for reducing standby power consumption”, IEICE Electronics Express, Vol. 13, No. 8, pp.20160174, February 2016.

[3] Cena G, Bertolotti I C, and Scanzio S, “Evaluation of EtherCAT distributed clock performance”, IEEE Transactions on Industrial Informatics, Vol. 8, No. 1, pp. 20-29, October 2011.

[4] Wu X, Zhou H X, and Yue H T, “The Optimization Design and Motion Characteristic Experiment of PMLSM Based on GA and TwinCAT”, International Journal of Simulation--Systems, Science and Technology, Vol. 17, No. 27, pp. 20-29, December 2016.

[5] Sartika E M, Maulidin I and Putra T A, “Perancangan dan Realisasi Alat Demonstrasi PC-based Control untuk Simulasi Keamanan Bangunan menggunakan Embedded PC, Setrum: Sistem”,Kendali-Tenaga-Elektronika-Telekomunikasi-Komputer, Vol. 7, No. 1, pp. 46-59.

[6] Sungmin Kim, “Moment of Inertia and Friction Torque Coefficient Identification in a Servo Drive System”, IEEE Transactions on Industrial Electronics, Vol. 66, No. 1, pp. 1-1, April 2018.

[7] Ren H P and Wang X, “PLC based variable structure control of hydraulic position servo system considering controller saturation”, 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), 2018.

[8] Ghaffari A and Ulsoy A G, “Experimental verification of dynamic contour error estimation for high-precision contouring of two-axis servo-systems”, ASME 2015 Dynamic Systems and Control Conference, January 2015.

[9] Chen Z W , Zhang K and Ding S X, “Improved canonical correlation analysis-based fault detection methods for industrial processes”, Journal of

Process Control, Vol. 41, pp. 26-34, February 2016.
[10] Wang T, Pan B, Fu Y, “Minimally invasive surgical robot based on EtherCAT bus and TwinCAT motion controller”. Journal of Huazhong University of ence and Technology(Natural ence Edition), Vol. 46, No. 12, pp. 1-7, 2018.

An Improved Proxy Re-Encryption Based Identity Combined with AES Storage Scheme in Cloud

Zhenwu Xu*, Jinan Shen*, †, Fang Liang**,Yingjie Chen*
* School of Information Engineering, Hubei Minzu University
Enshi, 445000, China
** School of Mathematics and Statistics, Hubei Minzu University
Enshi, 445000, China
† Corresponding author *Email address*: shenjinan@163.com

Abstract

Due to the exponential growth of cloud computing technology, cloud storage technology is favored by a large number of users. Storing data on the cloud can save the resources of local Storage configuration and reduce the cost of local hardware investment. However, the data stored in the cloud is out of the user's physical control. Based on the service characteristics of the cloud environment and the security requirements of user privacy data in the cloud environment, this paper combines the AES algorithm to improve on the basis of the identity proxy re-encryption algorithm. The performance of the proxy re-encryption algorithm is optimized by reducing the number of bilinear mapping operations that takes the most time to calculate in the proxy re-encryption scheme. In the improvement of this paper, only two bilinear mapping operation is needed. In addition, the encrypted data is tested to different degrees, and the experimental results show that the test results meet the user's demand for encryption and decryption performance.

Keywords: Privacy Protection, AES, Proxy Re-encryption, Fine-grained, Cloud Storage.

1. INTRODUCTION

With the expansion of cloud services and the further improvement of users' own needs, data privacy protection has gradually brought challenges that cannot be ignored to cloud applications [1]. Facing this form of distress, many scholars have proposed different schemes. The main schemes proposed are: Mandatory Access Control (MAC) [2,3], Discretionary Access Control (DAC), Identity-Based Encryption (IBE) [4,5], Proxy Re-Encryption (PRE) [6], Role Based Access Control (RBAC) [7,8], Role Based Encryption (Role Based Encryption, RBE) [9], Attribute-Based Encryption (ABE) [10], Fully Homomorphic Encryption (FHE) [11,12] and ciphertext search algorithm, etc. However, because the data is encrypted, users need to decrypt it when using it, which requires very powerful network resources and computing resources [13, 14]. In order to save local storage configuration and reduce the cost of local hardware investment, users store their data on cloud servers [15].

Shen et al. presented a multi-security level cloud storage system (FH-PRE) [6], including support for fine-grained control and performance optimization. By reducing the number of operations of bilinear mapping, they have achieved the goal of optimizing performance. Although the proposed scheme has been greatly optimized compared with standard proxy re-encryption scheme, it still has certain deficiencies in optimization compared with the scheme proposed in this paper.

Although ciphertext sharing has been proposed in many existing schemes, there is still a lot of improvement work to be done for the versatility and efficiency of the scheme.

(1) The re-encryption key can be used maliciously by the untrusted cloud to transform the ciphertext data generated by the data owner. If the untrusted cloud colludes with the recipient, the data of the data owner will be used maliciously. If the data is leaked, all data cannot be safely sent to the recipient;

(2) In the past identity-based proxy re-encryption, the number of bilinear operations is relatively large. When the number of users is relatively large, the operation of bilinear operations can lead to a large waste of network resources and cloud computing resources;

Based on this, we propose an improved identity proxy re-encryption algorithm based on AES, combined with AES symmetric encryption to encapsulate the symmetric key into ciphertext. While protecting user privacy data, the user can also control the data in the hands of the user, and achieve ciphertext sharing. At the same time, in the scheme proposed in this paper, bilinear operations are reduced, and our scheme makes full use of Network resources and cloud computing.

2. RELEVANT THEORETICAL KNOWLEDGE

Definition 1（Bilinear mapping）: When the mapping function $e: G_1 \times G_2 \to G_T$ satisfies the following conditions, it is called a bilinear mapping[16]:

（1） G_1 , G_T are groups of order , where is prime.
（2）for any $a, b \in Z_q^*, G_1$, $e(g^a, g^b) = e(g, g)^{ab}$;
（3） $e(g, g)$ is the generator of G_T ;
（4）for all $p, q \in G_1$, e can be computed.

Definition 2 （Advanced Encryption Standard algorithm）

Rijndael was designed by Joan Daemen and Vincent Rijmen in Belgium. Joan Daemen and Vincent Rijmen submitted the Rijndael algorithm in 2000, which was declared as Advanced Encryption Standard (AES) without modification [17]. It supports 128bit, 192bit, 256bit keys and at least 128bit video stream segmentation, and then use the key to convert each encoded plaintext block into the same 128-bit ciphertext block, and finally combine all the ciphertext blocks to generate the final ciphertext. In Figure 1, the AES algorithm uses a 128-bit key to encrypt and decrypt a 16-bit data block.

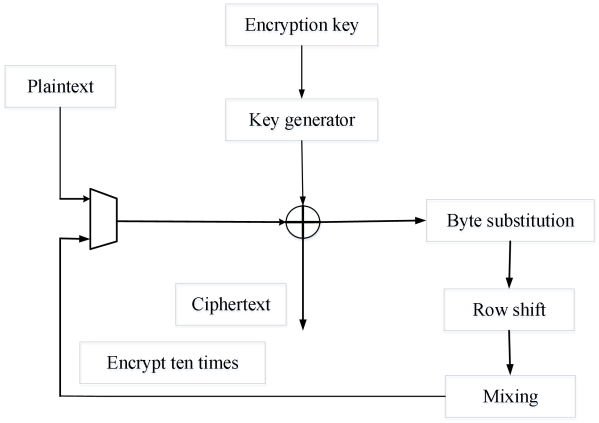


Fig.1 AES algorithm structure diagram.

3. AN IMPROVED RE-ENCRYPTION SCHEME BASED ON IDENTITY PROXY COMBINED WITH AES

3.1 The scheme design

The algorithm of this scheme contains seven algorithms mainly ($Setup_{PRE}$, $Extract_{PRE}$, Enc_{PRE} , $RKExtract_{PRE}$, $ReEnc_{PRE}$, $Dec1_{PRE}$, $Dec2_{PRE}$). In this paper, we describe the process of encrypted storage and ciphertext sharing in the cloud environment in our scheme. The process mainly includes three major blocks: local storage, re-encryption, and data sharing.

Program execution process:

（1） Initialization: The user enters the *ID* representing the identity. The *ID* is one of multiple information such as *mailbox*, *IP*, *ID* number, etc. The *KGC* (Key Generation Center) executes the initialization algorithm $Setup_{PRE}()$ and key generation algorithm $EXtract_{PRE}()$ to initialize the system.

（2） The data owner judges the user's request type for the data: When the data owner needs to store the data locally, the local secure storage plug-in is called, and the plaintext data is encrypted using the AES algorithm. The user can choose to store the ciphertext c_1 to the local host. You can also choose to store it in the cloud. When users need to share data, use the *AI-PRE* scheme to share the data with users who need to receive the data.

（3） Local storage or upload to the cloud: The data owner calls the secure storage plugin and first reads the stored data, and then executes the *AES()* encryption function in $Enc_{PRE}()$ to encrypt the data and generate ciphertext c_1 . After the ciphertext is generated, data owner can choose to store it locally. Data owner can also choose to upload to the cloud.

（4） Initial ciphertext download/decryption: the data owner obtains ciphertext c_1 , and executes function $Dec1_{PRE}()$, and obtains data m .

（5） Data sharing: After the cloud server judges the type of the data request, the user needs to call the client security sharing plug-in to share the data. Then the security plug-in judges whether it is the initial ciphertext, if it is the initial ciphertext, proceed to the next step, otherwise the corresponding operation information will appear.

（6） re-encryption: cloud server execute the $ReEnc_{PRE}$ function to generate the re-encrypted ciphertext c_2 , and use the ciphertext c_2 as the shared data.

（7） Re-encrypted ciphertext download/decryption: after obtaining the ciphertext, c_2 , the data requester uses his own public key and secret key to decrypt ciphertext c_2 by using the $Dec2_{PRE}()$ function. If it is a legal information recipient, obtain plaintext data *M*, otherwise, the corresponding operation message appears.

3.2 The scheme implementation

Based on the description of the basic structure in the previous section, this section gives a detailed introduction to the implementation of the algorithm with the above basic structure as the main idea:

(1) $Setup_{PRE}(s)$: it randomly generate a number $s \in G_T$ as the main safety parameter msk , which is generated by a random algorithm in the system described in this paper, and then select a prime number p to generate multiplicative groups G_1 and G_T of order q , and obtain the generator g of G_1 , Using the bilinear mapping function $\hat{e}: G_1 \times G_2 \to G_T$ and hash function $H_1: \{0,1\}^* \to G_1, H_2: \{0,1\}^* \to Z_P^*$ mentioned above to calculate the system parameter set,

$$params = (G_1, G_T, p, g, g^s, H_1, H_2, \hat{e}) \quad (1)$$

The parameters are managed by the key generation center.

(2) $KeyGen_{PRE}(msk, params, id)$: At this time, the user applies to the key generation center for a traditional secret key sk_{id} , and at the same time generates a symmetric key sk_{AES} for AES encryption. After generating two keys, the algorithm encapsulates the traditional secret key and symmetric key into sk_{id}^{I} . The key generation center generates a public and secret key pair (pk_{id}, sk_{id}) for the user according to the legal identity information id of the applying user, combining $params$ and the main security parameter msk , where

$$pk_{id} = H_1(id), sk_{id} = pk_{id}^{s}, sk_{id}^{I} = (sk_{id}, sk_{AES}) \quad (2)$$

Among them, $pk = H_1(ID)$, $sk = H_1(ID)^s$ and sk_{id}^{I} are used as public and secret keys for data encryption, decryption, and data sharing.

(3) $Enc_{PRE}(sk_{id}^{I}, params, m)$: After the user reads the information, when the data is encrypted, the data owner uses the legal identity information to generate the public key and secret key $(pk_{id}^{I}, sk_{id}^{I})$. Data plaintext m is encrypted as ciphertext $c_i = (c_{i1}, c_{i2})$ (ciphertext: when the ciphertext needs to be stored locally or uploaded to the cloud, only c_{i1} in c_i needs to be stored), where

$$\begin{aligned} c_{i1} &= AES\left(m, sk_{id}^{I}\right); \\ c_{i2} &= sk_{id}^{I} \oplus H_2\left(c_{i1}\right); \end{aligned} \quad (3)$$

(4) $Dec1_{PRE}(sk_{id}^{I}, params, c_i)$: During the first decryption, the user obtains the encrypted data file c_i . The user takes his secret key sk_{id}^{I} as input, uses the function $Dec1_{PRE}()$ to decrypt, and finally decrypts the plaintext *m*.

$$\begin{aligned} m &= AES.decryptAES(c_{i1}, sk_{id}^{I}); \\ c_{i2} &= sk_{id}^{I} \oplus H_2(c_{i1}) \end{aligned} \quad (4)$$

(5) $\mathrm{Re}KeyGen_{PRE}(sk_{id}^{I}, pk_{id}^{J}, params)$: after the data owner authorizes, *RKG* first randomly generates a positive integer $N \in N^*$, then calculate the set of K、N、pk_{id}^{I} 、pk_{id}^{J} , and then use the secret key sk_{id}^{I} generated based on the legal identity of the data owner and the data receiving user's public key pk_{id}^{J} , finally calculate the proxy re-encryption key from user i to user j based on the above parameters. The algorithm is executed in *RKG* (re-encryption key generator),

$$\begin{aligned} &N = Random(n); \\ &K = \hat{e}(sk_{id}^{I}, pk_{id}^{J}); \\ &\mathrm{concatArray} = K \| pk_{id}^{I} \| pk_{id}^{J} \| N; \\ &R = H_1(\mathrm{concatArray}) \bullet sk_{id}^{I}; \end{aligned} \quad (5)$$

Where, $rk_{id_i \to id_j} = (N, K, R)$.

(6) $\mathrm{Re}Enc_{PRE}(c_i, rk_{id_i \to id_j}, params)$: The re-encryption process is executed by a trusted third-party cloud server. The re-encryption function encrypts the original ciphertext into a re-encrypted ciphertext, and the re-encrypted ciphertext can be decrypted by user j using his own secret key. The system use the ciphertext $c_i = (c_{i1}, c_{i2})$ encrypted by user i and the proxy re-encryption key $rk_{id_i \to id_j}$ of users i to j and related system parameters as input. After the function calculation is performed at this stage, the re-encrypted ciphertext $c_j = (c_{j1}, c_{j2}, c_{j3})$ is output and stored in the cloud or sent to the data requester, where

$$\begin{aligned} r &= H_1(K, \mathrm{c}_{i2}) \\ c_{j1} &= c_{i1} \\ c_{j2} &= c_{i1} \oplus H_2(r) \\ c_{j3} &= \mathrm{c}_{i2} \end{aligned} \quad (6)$$

(7) $Dec_{PKE}(c_j, sk_{id}^{J}, pk_{id}^{I}, params)$: At this stage, the data receiver requests data sharing and executes the decryption function of the re-encrypted ciphertext. After receiving the ciphertext $c_j = (c_{j1}, c_{j2}, c_{j3})$ sent by the proxy server, it uses the secret key sk_{id}^{J} generated by the users own legal identity information and the data owner's public key pk_{id}^{I} generated by the identity information to participates in the calculation, and the data plaintext requested by the data receiver is calculated by the following algorithm

$$\begin{aligned} K &= \hat{e}(H_1(pk_{id}^{I}), sk_{id}^{J}) \\ r &= H_1(K, c_{j3}) \\ sk_{id}^{I} &= c_{j3} \oplus H_1(c_{j1}) \\ M &= AES(c_{j1}, sk_{id}^{I}) \end{aligned} \quad (7)$$

3.3 Algorithm and attack model

The attack model and algorithm are described in this section of the paper. First, we analyzed the attack model. Then we use code to test the algorithm, using pseudo-code to explain the first encryption, first decryption, re-encryption, and second decryption in the scheme.

Attack model:

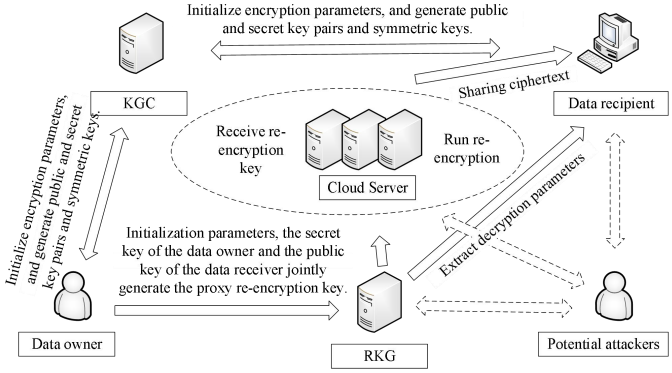


Fig.2 Attack model.

As shown in the figure 2, a potential attackers analyze the ciphertext in the system and attack functions such as *ReKeyGen*$_{PRE}$*()* and *ReEnc*$_{PRE}$*()*. In our scheme, although

the attacker can find the user's public key, he does not know the user's secret key. Because the first encrypted ciphertext is generated by the AES algorithm, it is impossible to decrypt the ciphertext when the attacker does not know the secret key. In the *ReEnc$_{PRE}$()* algorithm, the relevant information can only be known by knowing the user's secret key. At the same time, the attacker cannot obtain encryption parameters and decryption parameters. In summary, the attacker cannot easily obtain the plaintext.

In our paper, we used pseudo code to describe the algorithm, which contains *Enc$_{PRE}$()*, *Dec1$_{PRE}$()*, *ReEnc$_{PRE}$()* and *Dec$_{PRE}$()*. As follows:

Algorithm1 Encryption

```
Switch(Encryption){
  case Upload;
  Ci1 = AESUtil.encryptAES(msg), sk_I);
  break;
  case Share ciphertext;
  Ci1 = AESUtil.encryptAES(msg), sk_I);
  Ci2 = improve_mine.XOR(sk_P1.toBytes(), H2(Ci1));
break;
                    }
```

Algorithm2 Decryption

```
if(user = legitimate user){
    M = AESUtil.decryptAES(ci1, sk_I);
    }else if(user ≠ legitimate user){
    System.out.println("Error");
    }
```

Algorithm3 Re-encryption

```
Switch(Re-encryption){
    case Share ciphertext;
    r = H1(K, Ci2); cj1=ci1;
    Cj2=Ci1 XOR H2(r); cj3=ci2;
    break;
                    }
```

Algorithm4 Decryption

```
if(user = legitimate user){
    K = e(H1(ID_I), sk_J); r = H1(K, cj3);
    sk_I=cj3 XOR H1(cj1); M` = AES(cj1,sk_I) = m;
      }else if(user ≠ legitimate user){
      System.out.println("Error");
      }
```

4. SECURITY AND PROOF OF THE SCHEME

4.1 Security of the scheme

Definition4.1 Bilinear Diffie-Hellman problem:
Group (G_1,G_t) is a group supporting a pair of computabing Bilinear mapping $e:G_1\times G_2\to G_T$, and g is a random generator of G_1. DBDH problem is a $BGen(1^k)\{q,G_1,G_t,g,e\}$ problem. For all input tuple $(g,g^a,g^b,g^c,T)\in G_1\times G_t$, it is determined whether T is equal to $e(g,g)^{abc}$ or a random element of group G_T in a set of values (where $a,b,c\in R\ Z_q^*$) [6,18].

Assume k is a sufficiently large safety parameter, for (G_1,G_t) polynomial algorithm A, the following conditions are satisfied:

$$\left|pr\left[a,b,c\overset{\$}{\leftarrow}Z_q^*;1\leftarrow A(g,g^a,g^b,g^c,e(g,g)^{abc})\cdot\right]-\right.$$

$$\left.pr[a,b,c\overset{\$}{\leftarrow}Z_q^*;T\overset{\$}{\leftarrow}G_T;1\leftarrow A(g,g^a,g^b,g^c,T)\cdot]\right|\le v(k)$$

Where $v(\cdot)$ is a minimum value, for all function satisfied $p(\cdot)$, $v(k)$ $<1/p(k)$.

Definition4.2 Discrete logarithm difficult problem:
Eliptic Curve Discrete Logarithm Problem (ECDLP), Given a prime number p and an elliptic curve E, for $Q=kP$, when p and Q are known, find a positive integer k less than p. It can be proved that calculating Q from k and P is easier, but calculating k from Q and P is more difficult.

In addition, in any group G, a power of s can be defined for the generator g. It is easier to calculate $g^s=G_T$, but it is more difficult to find s by inverse operation.

Definition4.3 The security of AES encryption algorithm:

(1) Resistance to brute force attacks: packet length and key length are the main factors for the complexity of brute force attacks. The key length of the AES algorithm must meet 128 bits and be the minimum. During the attack, if 256 keys can be searched in one second, it would take at least 149 trillion years to crack the key. Therefore, the AES algorithm can be immune to brute force attacks.

(2) Resistance to differential analysis and linear cryptanalysis: differential analysis is mainly to find out whether the features that appear meets a certain

probability. If it can be found, then it is possible to analyze some of the keys in the last round of S-box analysis. To get all the keys, you need to use an exhaustive method to test all the keys. The linear analysis is mainly to find a linear formula that satisfies the maximum probability of occurrence. After reading the paper [17], it is known that the 4-round Rijndael algorithm is sufficient Resist differential cryptanalysis and linear cryptanalysis attacks.

(3) Ability to resist algebraic calculation attacks: algebraic calculation attacks are a block cipher attack method that uses the input or output of a password to construct a polynomial. In Formula (8), we use an algebraic formula to briefly describe the Rijndael algorithm.

$$C_{i,j} = k^* \oplus \sum^{W^*} \underset{e_{10}\in\varepsilon,\, d_{10}\in\phi}{} k^* \oplus \cdots\cdots \oplus \sum^{W^*} \underset{e_2\in\varepsilon,\, d_2\in\phi}{} k^* \oplus \sum^{W^*} \underset{e_1\in\varepsilon,\, d_1\in\phi}{} k^* \oplus P^* \quad (8)$$

The main idea of the attack is to use sufficient plaintext and ciphertext pairs, and then use the Lagrangian interpolation formula to analyze an approximate polynomial approximation and obtain a cryptographic algorithm. If the Rijndael algorithm is attacked by algebraic calculations, the analysis and research are carried out with polynomials, but if you want to eliminate a layer of substructure in formula (8), you must eliminate all operations contained in the S-box. At the same time, the expression of the S-box is not only very complicated but also difficult to attack, so the attack is not established [19].

(4) Resistance to XSL attacks: Nicolas Courtois and Josef Pieprayk proposed an attack method called algebraic calculation attack, and used redundant quadratic formulas to redefine the Rijndae algorithm in the paper. If the key is analyzed from a single plaintext in the 128-bit Rijndael algorithm, it can be converted to a problem of 8000 quadratic formulas, among which there are 1600 algebraic problems, and it is still unknown. The security of the Rijndael algorithm is mainly because there is no effective method to solve the above formulas. Shamir et al. published the XL (ORFXL) algorithm, which pointed out that the problem of the above algorithm can be solved in subexponential time. But in actual application, the Rijndael algorithm is not broken by the XL algorithm. In the paper published by Nicolas Courtois and Josef Pieprayk, the idea of improving the XL algorithm is described, and a new type of XSL attack is designed in the article. Although in theory, XSL attacks can be used for any method belonging to block encryption, attacks on AES in actual systems have no ideal effect [19].

5. EXPERIMENT AND PERFORNANCE ANALYSIS

The computer used in the test environment of this program is Lenovo, a computer with Intel i3 2.3GHz processor and 12GB (memory bar added later) memory. Using the 2014 version of the Myeclipse console on this machine, it was implemented with code to test the performance of the cloud storage system combined with the AES improved proxy re-encryption module.

(1) $Enc_{PRE}(sk_{id}^{I}, params, m)$

The data owner uses the $Enc_{PRE}(sk_{id}^{I}, params, m)$ encryption algorithm to be executed on the client. In our scheme, the algorithm at this stage only uses the AES encryption algorithm, and the number of times the bilinear mapping participates in the operation is 0, which greatly reduces the consumption of the system. Therefore, compared with the traditional proxy re-encryption algorithm (the standard proxy re-encryption scheme) and the FH-PRE [6] in this paper, the time consumption in the encryption process has been greatly optimized.

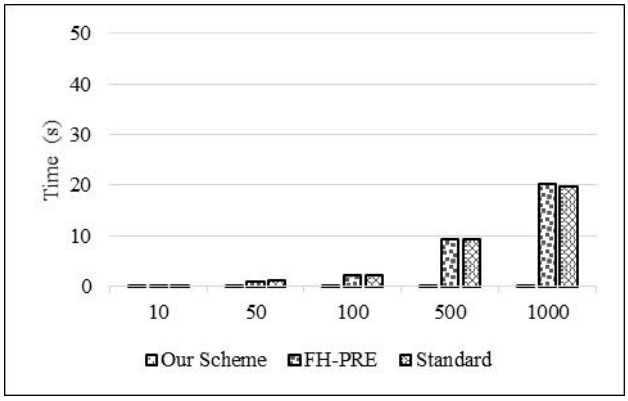


Fig.3 Computational overhead of $Enc_{PRE}()$ function.

As shown in figure 3, the time consumed by this part of the algorithm in the encryption process has obvious advantages over the traditional proxy re-encryption algorithm and FH-PRE. The average time spent on each encryption is less than 5ms. According to experimental data, the more repetitions, the less time it takes. In this scheme, a 128-bit symmetric encryption key is used, and the length of the key is acceptable in the system. In the encryption process, the algorithm in this scheme has obvious advantages and can meet the performance requirements of small clients.

(2) $Dec1_{PRE}(sk_{id}^{I}, params, c_1)$

In our scheme, the data owner executes a decryption function on the client to decrypt the original ciphertext. After executing this decryption function, the data owner can find the data he uploaded. In this part of the algorithm, only AES is involved in the calculation, and the number of times that bilinear calculations are involved in the calculation is 0, so this function has huge

advantages in time consumption and system consumption, and it can be smooth on small clients. Decrypt the information.

As can be seen from figure 4, compared with FH-PRE and traditional proxy re-encryption algorithm, the algorithm at this stage has a very obvious advantage in terms of time cost. The average encryption operation takes less than 5ms. Although the decryption object in the scheme is only the ciphertext of the 128-bit symmetric key, it will not have a great impact on the overall performance.

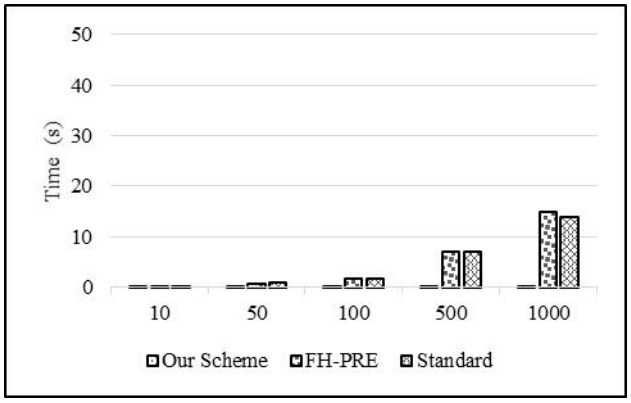


Fig.4 Computational overhead of $Dec1_{PRE}()$ function.

(3) $\mathrm{Re}\,KeyGen_{PRE}(sk_{id}^{I}, pk_{id}^{J}, params)$

$\mathrm{Re}\,KeyGen_{PRE}(sk_{id}^{I}, pk_{id}^{J}, params)$ is a function that uses the secret key of the proxy authorizer and the public key of the proxy acceptor to generate the proxy re-encryption key. Different from the previous encryption and decryption functions, users only need to perform a calculation for each file when uploading and downloading. The authorization operation is that the data owner authorizes a user to execute it once, so in the case of multiple users, this function will be executed multiple times, so the performance of this function has a relatively large impact on the performance of the overall solution. In this system, the number of bilinear mapping used in this step is only once.

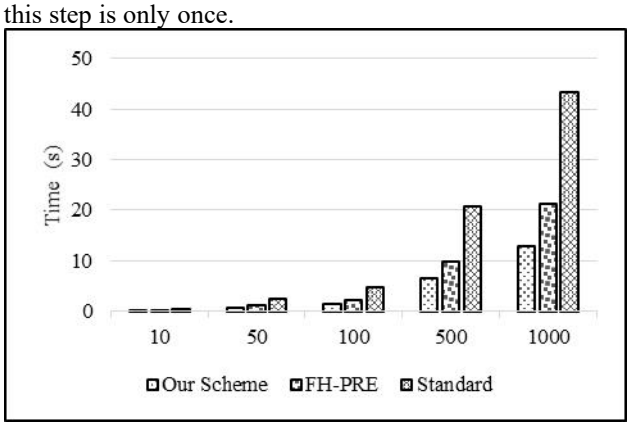


Fig.5 Computational overhead of $\mathrm{Re}\,KeyGen_{PRE}()$ function.

It can be seen from figure 5 that the average time for an authorization proxy re-encryption in this scheme is between 18ms and 20ms. An authorization in FH-PRE takes 20 to 24ms, while this process in the standard scheme takes 43.5 to 48ms, it can be seen from the experimental data that the re-encryption key generation function in this scheme requires lower system consumption and shorter time consumption, and has certain advantages in performance.

(4) $\mathrm{Re}\,Enc_{PRE}(C_i, rk_{id_i \to id_j}, params)$

It can be seen from figure 6 that the proxy re-encryption cloud storage scheme improved by AES takes less than 5ms each time during the re-encryption process, and the time cost of FH-PRE in this process is the same as the traditional proxy re-encryption algorithm. The cost is similar, and each encryption operation takes about 11ms on average. The improvement plan has an obvious effect on the calculation cost. The more cycles, the less time it takes each time. At the same time, the object of re-encryption is also the ciphertext of the 128-bit symmetric key, which does not affect the overall usability.

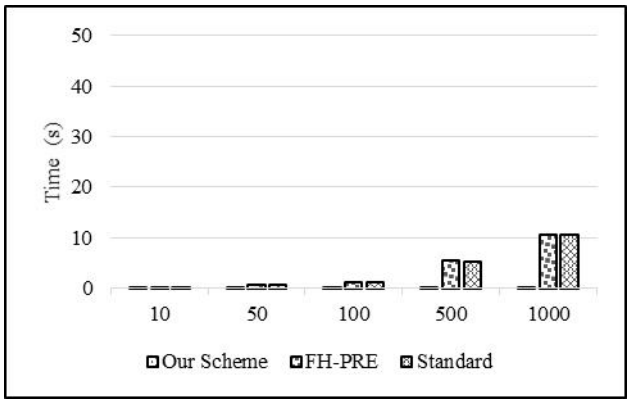


Fig.6 Computational overhead of $\mathrm{Re}\,Enc_{PRE}()$ function.

(5) $Dec_{PKE}(c_j, sk_{id}^{J}, pk_{id}^{I}, params)$

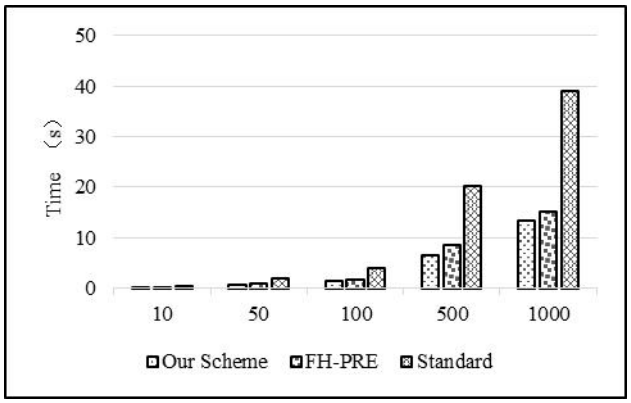


Fig.7 Computational overhead of $Dec_{PKE}()$ function.

The decryption re-encrypted ciphertext function is used by the user to decrypt the re-encrypted ciphertext when using the data shared by the data owner, and is executed on the client. In this system, the number of times of bilinear mapping used in this step is only once, so the performance is also greatly improved. It can be seen from

figure 7 that that the decryption and re-encryption ciphertext process time of this scheme and FH-PRE scheme is similar. This scheme and FH-PRE takes 15 to 17 ms to decrypt a re-encryption ciphertext, while this process in the standard scheme It takes 40 milliseconds,. In comparison, this scheme is also greatly improved.

6. CONCLUSION

Combined with the characteristics and requirements of the cloud environment, this scheme proposes an improved identity-based proxy re-encryption cloud storage scheme combined with AES. The improved proxy re-encryption scheme has the following advantages:

(1) In the proposed improved proxy re-encryption scheme, the data owner decides whether to share the data with other users. At the same time, we use the Java programming language to implement the improved algorithm;

(2) Usually in the proxy re-encryption scheme, the client relies on the proxy in the cloud for computing overhead, leading to the PRE scheme Brings huge system consumption to the agent in the cloud. In the improvement of this scheme, the time-consuming re-encryption of some functions has obvious advantages, which reduces the system consumption of cloud service providers;

(3) The idea of this scheme through key encapsulation, use symmetric encryption to encrypt user data to reduce the amount of calculations operated by the public key cryptographic algorithm; then, an improved identity-based re-encryption algorithm is used to encrypt the symmetric key, which reduces the number of bilinear mappings and optimizes performance without affecting security.

Acknowledgment

The work was supported by the National Natural Science Foundation of China under grant No.61662022 and Incubation Project for High-Level Scientific Research Achievements of Hubei Minzu University under grant No. PY20008.

References

[1] Y. Liu, “Image Encryption Algorithm Based on a Hyperchaotic System and Fractional Fourier Transform,” Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 23, no. 5, pp. 805-809, 2019.

[2] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. D. Konwinski et al. “A view of cloud computing,” Commun. ACM, 2010.

[3] D. Elliott Bell, et al. “Secure computer systems: mathematical foundations and model,” Technical Report M74–244, MITRE Corporation, Bedford, MA, 1973.

[4] A. Shamir, “Identity-based cryptosystems and signature schemes," In: CRYPTO, Lecture notes in computer science, vol. 196. Springer, Berlin, pp. 47–53, 1985.

[5] C. Delerablée, P. Paillier, D. Pointcheval, “Fully collusion secure dynamic broadcast encryption with constant-size ciphertexts or decryption keys,” In: Pairing, Lecture notes in computer science, vol. 4575. Springer, Berlin, pp. 39–59, 2007.

[6] J. Shen, X. Deng, Z. Xu, et al. “Multi-security-level cloud storage system based on improved proxy re-encryption,” Eurasip Journal on Wireless Communications and Networking, vol. 2019, no. 1, pp. 1–12, 2019.

[7] J. Crampton, G. Loizou, “Administrative scope: a foundation for role-based administrative models,” ACM Transactions on Information and System Security, vol. 6, no. 2, pp. 201–231, 2003.

[8] J. Crampton, “Understanding and developing role-based administrative models,” In: ACM conference on computer and communications security, pp. 158–167. 7–11 Nov 2005.

[9] L. Zhou, V. Varadharajan, M. Hitchens, et al. “Enforcing Role-Based Access Control for Secure Data Storage in the Cloud [J],” The Computer Journal, vol. 54, no. 10, pp. 1675–1687, 2011.

[10] J. Bethencourt, A. Sahai, B. Waters, “Ciphertext-policy attribute-based encryption. IEEE symposium on security and privacy,” IEEE Computer Society, pp. 321–334, 2007.

[11] C. Gentry, S. Halevi, “Implementing Gentry’s fully-homomorphic encryption scheme,” in: Proceeding of 30th Annual International Conference on Theory and Applications of Cryptographic Techniques. Paterson, Kenneth: Springer press, pp. 129–148, 2019.

[12] Z. Brakerski, V. Vaikuntanathan, “Efficient fully homomorphic encryption from (standard) LWE,” in: Proceeding of IEEE Symposium on Foundations of Computer Science. California, USA: IEEE press, pp. 97-106, 2011.

[13] B. Liu, et al. “HEVC Video Encryption Algorithm Based on Integer Dynamic Coupling Tent Mapping,” Journal of Advanced Computational Intelligence and Intelligent Informatics, vol. 24, pp. 335-345, 2020.

[14] A. Haggag, M. Ghoneim, J. Lu, et al. “Access Control and Scalable Encryption Using a Stream Cipher for JPEG 2000 Encoded Images,” Journal of Advanced Computational Intelligence and

Intelligent Informatics, vol. 11, no. 7, pp. 718-734, 2007.
[15] J. Hu, J. Shen, F. Liang, "An improved identity-based broadcast proxy re-encryption storage scheme in cloud [J]," Application Research of Computers, vol. 34, no. 5, pp. 1520–1524, 2017.
[16] CHEN Xiao-Feng, FENG Deng-Guo, "Direct Anonymous Attestation Based on Bilinear Maps [J]," Journal of Software, vol. 21, no. 8, pp. 2070-2078, 2010.
[17] J. Daemen, V. Rijmen, "The Design of Rijndael: AES - The Advanced Encryption Standard [M]," DBLP, 2002.
[18] Xu Peng, Cui Guohua, and Lei Fengyu, "An Efficient and Provably Secure IBE Scheme Without Bilinear Map [J]," Journal of Computer Research and Development, vol. 45, no. 10, pp. 1687-1695, 2008.
[19] W. Han, "Analysis of AES encryption algorithm and its security research," Computer Applications of Petroleum, vol. 16, no. 2, pp. 46-48, 2008.

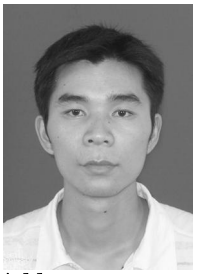
Name:
ZhenWu Xu

Affiliation:
School of Information Engineering, Hubei Minzu University, Enshi, China

Address:
No.39 Xueyuan Road, Enshi, China.

Brief Biographical History:
2018, he graduated from school of information engineering and technology,Henan Institute of Science and Technology (HIST) with a BS degree.
Currently, he is a MSD candidate at Hubei Minzu University, Enshi, China.

.Main Works:
- software development, security in the area of cloud computing security and network security testing



Name:
Jinan Shen

Affiliation:
School of Information Engineering, Hubei Minzu University, Enshi, China.

Address:
No.39 Xueyuan Road, Enshi, China.

Brief Biographical History:
He is an Associate Professor of computer science at Hubei Minzu University, Enshi, China.
2020, He received his PhD at Huazhong University of Science and Technology (HUST), Wuhan, China.

Main Works:
- in the area of security in cloud computing, focusing on privacy preserving in the cloud

Name:
Fang Liang

Affiliation:
School of Mathematics and Statistics, Hubei Minzu University, Enshi, China.

Address:
No.39 Xueyuan Road, Enshi, China.

Brief Biographical History:
She is an Associate Professor of computer science at Hubei Minzu University, Enshi, China.
2012, She received her MS degree in computer school at Wuhan University, China.

Main Works:
- in the area of security in cloud computing

Name:
Yingjie Chen

Affiliation:
School of Information Engineering, Hubei Minzu University, Enshi, China

Address:
No.39 Xueyuan Road, Enshi, China.

Brief Biographical History:
2015, he graduated from South-Center University for Nationalities.
Now. he is study for a master's degree in Hubei Minzu University.

.Main Works:
- software development, security in the area of cloud computing security and network security testing.

An Intelligent Compensating Method for MPC-Based Deviation Correction with Stratum Uncertainty in Vertical Drilling Process

Dian Zhang[*, **], Min Wu[*, **, †], Chengda Lu[*, **], Luefeng Chen[*, **], Weihua Cao[*, **], Jie Hu[*, **]
* School of Automation, China University of Geosciences, Wuhan 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent
Automation for Complex Systems, Wuhan 430074, China

Abstract

Model predictive control (MPC) is widely accepted and has been chosen as a new way for deviation correction in vertical drilling. However, the accuracy of the prediction model is affected due to the uncertainty of the stratum, and lead to model mismatch and reduction of the control performance. In this paper, an intelligent compensating method is proposed for MPC-based deviation correction with stratum uncertainty in vertical drilling process to increase the control accuracy. Firstly, the trajectory extension model is introduced as the predictive model for MPC and the uncertainty of the stratum is discussed. Then, the compensation for MPC is acquired based on the hybrid bat algorithm. Finally, based on the raw data collected from a vertical drilling site, a simulation is carried out to demonstrate the effectiveness of the proposed method.

Keywords: Model predictive control, deviation correction control, vertical drilling, intelligent compensating method.

1. INTRODUCTION

Deviation correction control is one of the key technologies in vertical drilling process. The purpose of the deviation correction control is to correct inclination angle and position deviation to be zero, and maintain the drilling trajectory to be straight along with the plumb line of the wellhead. The quality of drilling trajectory is determined in large part by the performance of the correction control.

The early deviation correction control methods are mainly manual control methods based on manual experience [1]. However they have gradually been replaced by many modern control methods, such as deviation vector control theory, attitude-hold control method and model-based robust Control et. al [2-4]. In recent years, model predictive control (MPC) is widely accepted and has been chosen as a new way for the deviation correction control in vertical drilling, because of its strong ability to cope with constraints and higher control performance. Martin et al. provided a MPC scheme for drilling to deal with system delay [5]. Demirer et al. considered a long-range behavior of bottom hole assembly (BHA) and built a MPC tracking controller considering curvature constraints for directional drilling, and had applied successfully in field tests [6]. Zhang et al. established a MPC based on a trajectory extension model considering angle limit and build up rate constraints in vertical drilling [7].

However, due to the uncertainty of the stratum, it is hard to establish accurately drilling model. Poor modeling accuracy will lead to model mismatch in MPC, as well as the bad correcting performance. The most common way for dealing with this problem is compensation. Farina et al. provided an output feedback MPC with the Chebyshev–Cantelli inequality to deal with a possibly unbounded additive noise [8]. Tang et al. established an observer-based output feedback MPC for T-S fuzzy system with bounded disturbance [9]. Shi et al. used a Kalman filter to compensate the MPC for attenuating the effect of the disturbance on yaw performance [10]. Buddhadeva et al. provided an adaptive Lyapunov-based model to deal with the model mismatch of MPC [11]. Although there are many works about feedback adjustment method in many other industry applications, it is still no effective way to deal with the model mismatch for MPC-based deviation correction in vertical drilling.

In this paper, an intelligent compensating method for MPC-based deviation correction with stratum uncertainty in vertical drilling process is proposed for MPC to increase the control accuracy. Firstly, the trajectory extension model is introduced as the predictive model for MPC and the uncertainty of the stratum is discussed. Then, the compensation for the MPC is acquired based on the predictive model by the hybrid bat algorithm in [12]. Finally, based on the raw data collected from a practical vertical drilling site, a simulation is carried out to

† Corresponding author: Min Wu (wumin@cug.edu.cn)

demonstrate the effectiveness of the proposed method. The contribution of this paper is to find a way to compensate the control output, in order to deal with the model mismatch.

2. TRAJECTORY EXTENSION MODEL AND PROBLEM DESCRIPTION

This section describes the trajectory extension model given by [7]. Then, control problems of this paper are discussed based on this model.

2.1 Trajectory extension model

The deviation correction process discussed here can be described as Fig. 1. The schematic shows the movement of the BHA and the formation of drilling trajectory in comprehensive perspective. In order to quantitative analyze the drilling trajectory, an orthonormal Cartesian coordinate system under ground is established as shown in Fig. 1, where XOZ is parallel to the Earth plane and the Y-axis goes along the north. The curve is defined as the drilling trajectory, and BHA's states of point P are shown by the curve. According to [7], the trajectory extension model can be established to describe the deviation correction process as (1):

$$\begin{cases} \tan\alpha_x = \tan\alpha\sin\beta \\ \tan\alpha_y = \tan\alpha\cos\beta \\ \dot{S}_z = \dot{S}\cos\alpha \\ \dot{S}_x = \dot{S}\tan\alpha_x \\ \dot{S}_y = \dot{S}\tan\alpha_y \\ \dot{\alpha}_x = \omega_x + \varepsilon_x = r\omega_{SR}\sin\tilde{\theta}_{tf} + \varepsilon_x \\ \dot{\alpha}_y = \omega_y + \varepsilon_y = r\omega_{SR}\cos\tilde{\theta}_{tf} + \varepsilon_y \end{cases} \tag{1}$$

where S_x and S_y are components of the position deviation, as the α_x and α_y are components of the inclination angle. $\tilde{\theta}_{tf}$ and ω_{SR} are magnetic tool face angle and steering ratio respectively. Build up rate r is the maximum deflection capability of BHA, and $r\omega_{SR} \sim [0, r]$ denotes the real deflection capability that BHA provided. As the existence of the stratum uncertainty, there are some uncertainty parameters in this model. The stratum uncertainty is independent with BHA, the additive uncertainties ε_x and ε_y are utilized in this paper to quantitative describe the uncertainty.

Trajectory extension model is used to describe the vertical drilling process in the form of mathematic, and it is also an important basis for establishing the predictive model of the MPC controller in this paper.

2.2 Problem description

Deviation correction control is aiming to decrease

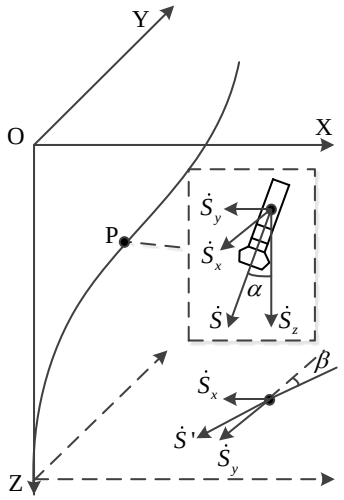


Fig. 1 Deviation correction process

the deviation of drilling trajectory include position deviations S_x, S_y, and inclination angles α_x and α_y by adjusting the magnetic tool face angle $\tilde{\theta}_{tf}$ and the steering ratio ω_{SR}.

Besides, as the limited drilling condition, there are some constraints in practical drilling. The measurement interval is the time of drilling for a certain distance according to process requirement. In order to ensure the quality of vertical trajectory, the inclination angle is required to be less than around $\alpha_{\max}$. Moreover, the build up rate r which is provided by BHA should be less than its deflecting limit, this limit is defined by the parameters of BHA [13].

The control problem shown above had been well addressed with a MPC controller in [7]. The control construction is shown in Fig. 2.

where r_{in} is the reference values of $[S_x, S_y, \alpha_x, \alpha_y]$, O_{out} is trajectory parameters $[S_x, S_y, \alpha_x, \alpha_y]$, and O'_{out} represents the trajectory parameters $[S_x, S_y, \alpha_x, \alpha_y]$ calculated by the trajectory calculation model. Vertical drilling process can be defined as trajectory extension model, and the minimum curvature method is selected to be the trajectory calculation model [14].

However, as the uncertainty ε_x and ε_y are not zero, it will make the accuracy of predictive model be decreased, and finally lead to model mismatch in MPC control. That is a serious problem for deviation correction control, which will make it difficult to adjust drilling trajectory. For this reason, a compensation generator is going to be established to find the proper compensation $\tilde{\theta}'_{tf}$, ω'_{SR} is going in this paper for the MPC controller to counteract the effects of model mismatch cause by the uncertainty ε_x and ε_y, as shown in Fig. 2.

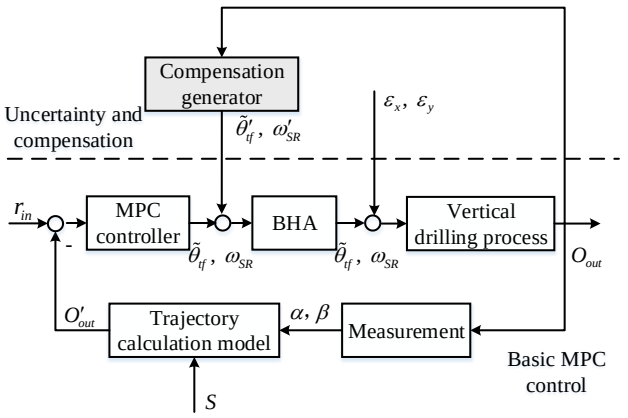


Fig. 2 MPC control and the compensation

3. INTELLIGENT COMPENSATION METHOD

In this section, Gaussian fitting method is explored to analyze the probability density distribution of the uncertainty, and the hybrid bat algorithm is utilized to acquire the compensation for MPC controller.

3.1 Structure of the compensation generator

In order to make the compensation generator be able to deal with various uncertainties, it is mainly established with an intelligent optimization algorithm. Structure of the compensation generator is shown in Fig. 3. Two stages are included in the compensation generator.

The main purpose of stage 1 is to get the probability density distribution of the uncertainty from the trajectory parameters, thereby determining parameters of the trajectory extension model (1). Gaussian fitting method is selected to acquire this probability density distribution.

The main purpose of stage 2 is to get the most proper compensation by a hybrid bat algorithm based on the trajectory extension model acquired from stage 1. The hybrid bat algorithm ensures the compensation generator be able to deal with various uncertainties.

3.2 Probability density distribution of the uncertainty

We consider that the distribution of the uncertainty in one well section is constant or changes slowly，as one drilling trajectory is consist of several well sections. Probability density distribution of the uncertainty can be acquired from the trajectory parameters.

Fig. 4 shows the build up rates provided by BHA in one well section from a practical drilling site. It can be seen after performing statistical analysis on this data that the build up rate is mainly within 4.5~7.5 °/30 m, the probability that build up rate is within 0~1.5 °/30 m or

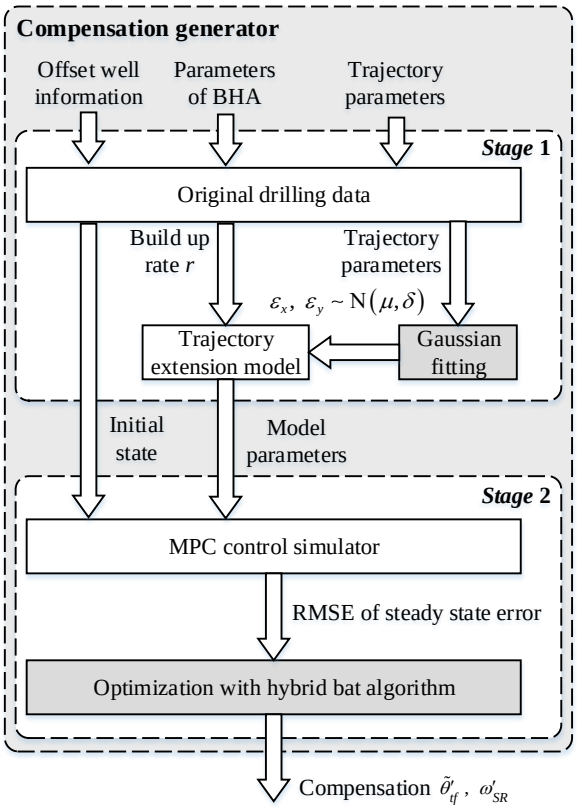


Fig. 3 Structure of the compensation generator

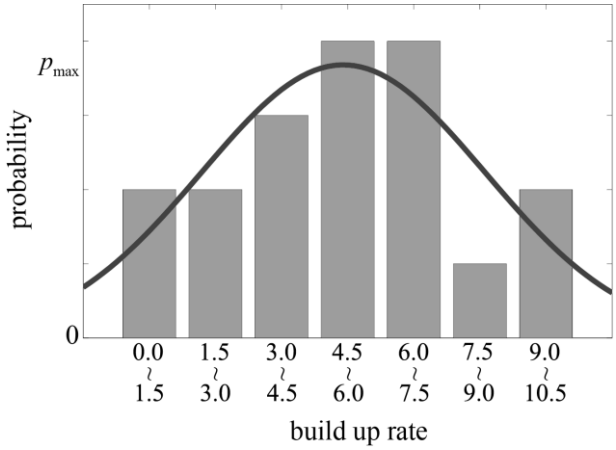


Fig. 4 Distribution of the uncertainty

9.0~10.5 °/30 m is much smaller. That means the probability density distribution of build up rate provided by BHA is unimodal. So Gaussian fitting method can be used to acquire the probability density distribution of the uncertainty.

3.3 Optimization of the compensation

Based on the probability density distribution of the uncertainty and the trajectory extension model, the optimization problem is organized as shown in Fig. 5.

In order to ensure that the compensation can counteract the effect of the uncertainty under the probability density

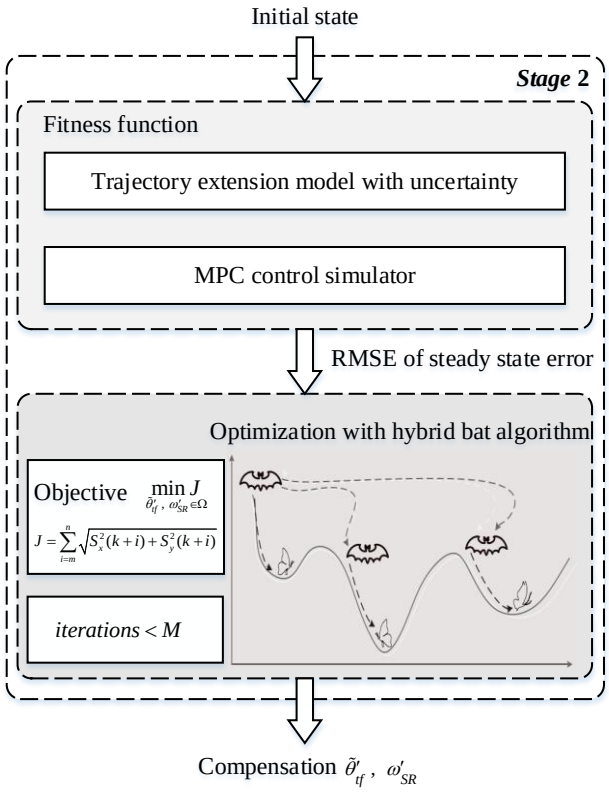


Fig. 5 Optimization of the compensation

distribution, a optimization method is selected to find the most proper compensation during vertical drilling. The results of basic MPC in [7] is selected to be the fitness function. What the difference is that the trajectory extension model suffers uncertainties with the probability density distribution shown in section 3.2. The output of the fitness function is root mean square error (RMSE) of steady state error. RMSE will be different by adjusting the compensations $\tilde{\theta}'_{tf}$ and ω'_{SR}.

Based on the fitness function, the optimization objective can be written as (2):

$$\min_{\tilde{\theta}'_{tf},\,\omega'_{SR}\in\Omega} J = \sum_{i=m}^{n}\sqrt{S_x^2(k+i)+S_y^2(k+i)} \qquad (2)$$

where $S_x(k+i)$, $S_y(k+i)$ is the position deviation at time $(k+i)$, m is the time when the system enters steady state phase, n is the duration of the simulation. J defines the performance of the compensation. Then a global optimization algorithm named hybrid bat algorithm is used to find the most proper compensation of this deviation correction process. In addition, the end condition of the optimization is $iterations < M$, where M is a positive integer.

As the density distribution reflects the uncertainty of the whole well section, the compensation which is found in stage 2 is unnecessary to be updated until the drill bit reaches the next well section. The MPC controller can use the same compensation in one well section.

4. SIMULATION AND RESULT ANALYSIS

Based on the raw data collected from a vertical drilling site, a simulation is carried out to demonstrate the effectiveness of the proposed method. Using the improved trajectory extension model to describe the vertical drilling process, simulations are carried out to test the deviation correction capacity for the case of vertical drilling. According to [7], the parameters of simulation are selected as follow. Rate of penetration ROP $\dot{S}$ is 30 m/hr, the control cycle T is 0.3 hr. For constraints the maximum deflection capability of BHA r is 6 ?30 m, and $\alpha_{\max}$ is 3 °. Parameters of MPC are that: p and c is 5, R is diag(50000,50000), Q is diag(0.1,10,0.1,10).

According to data from an actual drilling site, the horizontal deviation between actual trajectory and the reference is 8.82 m in XOZ plane at 600 m measured depth (MD), meanwhile the horizontal deviation is 1.51 m in the YOZ plane, the inclination angle is 1.5 °, the azimuth angle is 35.9 °. The uncertainty of the hole well section is assumed to be $\varepsilon_x, \varepsilon_y \sim$ N(0.54, 0.36). In order to validate the validity of the proposed method, a comparison with MPC in [7] has been conducted. Fig. 6 shows simulation results of S_x and S_y, Fig. 7 shows simulation results of S_{xy} and α, Fig. 8 shows simulation results of $\tilde{\theta}_{tf}$ and ω_{SR}. Table 1 is the results of 20 times Monte Carlo simulations, it assumes that the deviation adjustment of these two methods enters steady state phase at 780 m.

Here, both of these two methods can correct the deviation of drilling trajectory, but the results are different. The position deviation S_x and S_y of the proposed method are eliminated to zero at nearly 780 m MD, as it can deal with constrains well. The position deviation S_x and S_y of the MPC is corrected to be small at nearly 870 m. The RMSE of the proposed method is 0.0183, and the RMSE of the MPC is 4.6136. So, as the results show, the proposed method has better converge and smaller steady state error than MPC.

In conclusion, the proposed method can efficiently address the problem of model mismatch for deviation correction control of vertical drilling, and increases the accuracy of MPC.

5. CONCLUSION

In this paper, an intelligent compensating method for MPC-based deviation correction with the stratum uncertainty in vertical drilling process has been proposed for model predictive control to increase the control

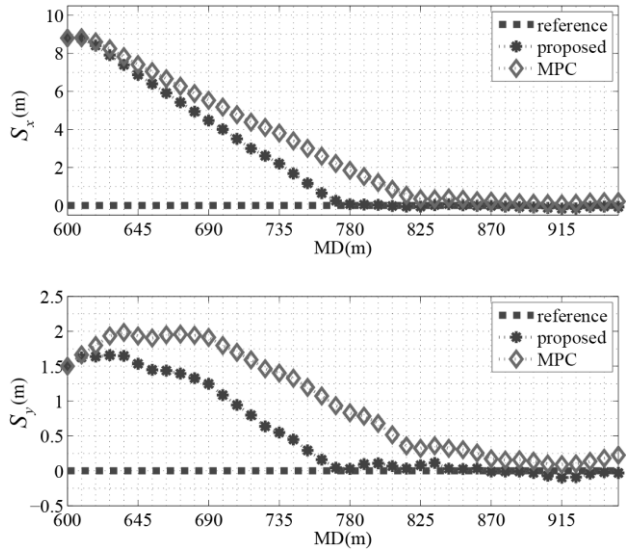


Fig. 6 Simulation results of S_x and S_y

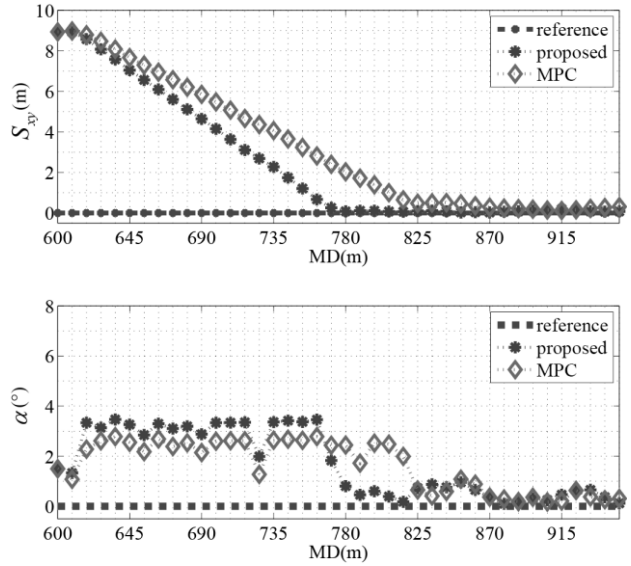


Fig. 7 Simulation results of S_{xy} and α

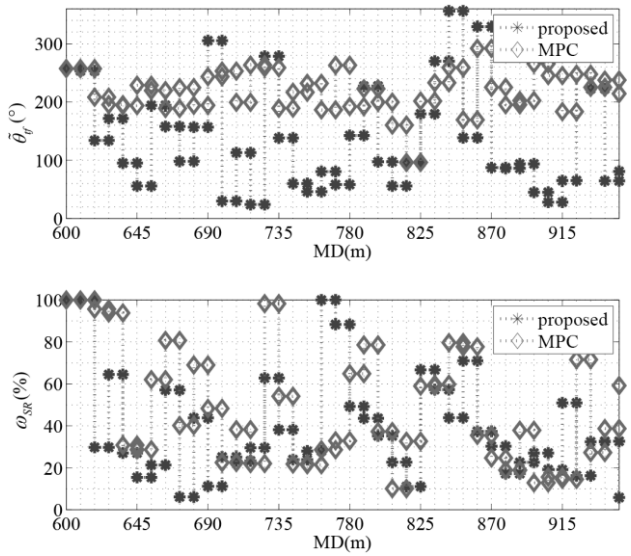


Fig. 8 Simulation results of $\tilde{\theta}_{tf}$ and ω_{SR}

accuracy. Two steps compensation generator has been provided, and the probability density distribution of the uncertainty from trajectory parameters is acquired through Gaussian fitting method and parameters of the trajectory extension model are determined at step 1. The most proper compensation is acquired by the hybrid bat algorithm to resolve the effect of the uncertainty of stratum at step 2.

According to the simulation results, the proposed method has better converge and smaller steady state error than MPC. It can efficiently address the problem of model mismatch for deviation correction control of vertical drilling, and increases the accuracy of MPC.

Acknowledgement

This work was supported in part by the National Natural Science Foundation of China under Grants 61733016; in part by the National Key R&D Program of China under Grant 2018YFC0603405; in part by the Hubei Provincial Technical Innovation Major Project under Grant 2018AAA035; in part by the 111 Project under Grant B17040; and in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan), under Grant CUGCJ1812.

References

[1] Y. D. Lu, L. C. Zeng, F. L. Zeng, and G. S. Kai, “Dynamic Simulation and Research on Hydraulic Guide System of Automatic Vertical Drilling Tool”, *Materials Research Innovations*, Vol. 18 No. s6, 2015, pp. 170-174.

[2] N. Panchal, M. T. Bayliss, and J. F. Whidborne, “Attitude Control System for Directional Drilling Bottom Hole Assemblies”, *IET Control Theory and Applications*, Vol. 6, No. 7, 2012, pp. 884-892.

[3] N. A. H. Kremers, E. M. Detournay, and V. D.Wouw, “Model-Based Robust Control of Directional Drilling Systems”, *IEEE Transactions on Control Systems Technology*, Vol. 24, No. 1, 2016, pp. 226-239.

[4] Z. Cai, X. Z. Lai, M. Wu, L. F. Chen, and C. D. Lu, “Compensation Control for Tool Attitude in Directional Drilling Systems”, *Proceedings of the 12th Asian Control Conference (ASCC 2019)*, Kitakyushu, Fukuoka, Japan, Jun. 9-12, 2019, pp. 376-380.

[5] B. Martin, B. Chris, and W. James, “MPC-based Feedback Delay Compensation Scheme for Directional Drilling Attitude Control”, *SPE/IADC Drilling Conference and Exhibition*, London, UK, 2015, SPE-173009-MS, 13 pages.

[6] N. Demire, U. Zalluhoglu, J. Marck, R. Darbe, and

M. Morari, “Autonomous Directional Drilling with Rotary Steerable Systems”, *2019 American Control Conference (ACC)*, PA, USA, July 10-12, 2019, pp. 5203-5208.

[7] D. Zhang, M. Wu, L. F. Chen, C. D. Lu, and W. H. Cao, “Model Predictive Control Strategy Based on Improved Trajectory Extension Model for Deviation Correction in Vertical Drilling Process”, *21st IFAC World Congress*, Berlin, Germany, July 12-17, 2020.

[8] Marcello Farina, Luca Giulioni, Lalo Magni, Riccardo Scattolini, “An approach to output-feedback MPC of stochastic linear discrete-time systems”, *Automatica*, Vol. 55, 2015, pp. 140-149.

[9] X. Tang, L. Deng, N. Liu, S. Yang and J. Yu, “Observer-Based Output Feedback MPC for T–S Fuzzy System With Data Loss and Bounded Disturbance”, *IEEE Transactions on Cybernetics*, vol. 49, no. 6, 2018, pp. 2119-2132.

[10] K Shi, X. F. Yuan, G. M. Huang, and Q. He, “MPC-based compensation control system for the yaw stability of distributed drive electric vehicle”, *International Journal of Systems Science*, Vol. 49, No. 7, 2018, pp. 1795-1808.

[11] D. Buddhadeva and M. Prashant, “Adaptive output-feedback Lyapunov-based model predictive control of nonlinear process systems”. International Journal of Robust and Nonlinear Control, Vol. 28, No. 5, 2018, pp. 1597-1609.

[12] G. Gan, W. H. Cao, K. Z. Liu, M. Wu, F. W. Wang, S. B. Zhang, “A new hybrid bat algorithm and its application to the ROP optimization in drilling processes”, *IEEE Transactions on Industrial Informatics*, https://doi.org/10.1109/TII.2019.2943165.

[13] A. Wilson, “Bending Rules with High-Build-Rate Rotary-Steerable Systems”, *Journal of Petroleum Technology*, Vol. 68, No. 12, 2016, pp. 62-63.

[14] “Bulletin on Directional Drilling Survey Calculation Methods and Terminology”, *API Bull D20*, Washington, 1985, pp. 1862-1867.

Research on the Single-Phase Photovoltaic Grid-Connected Inverter Based on Fuzzy Neural Network

Shenping XIAO, Zhouquan OU, Junming PENG,Yang ZHANG
College of Electrical and Information Engineering, Hunan University of Technology
Key Laboratory for Electric Drive Control and Intelligent Equipment of Hunan Province
NO.88 Taishan Xi Road,Tianyuan District,Zhuzhou City,Hunan 412007, China

Abstract

Based on the single-phase photovoltaic grid-connected inverter, a control strategy combining traditional PID and dynamic optimal control algorithm with fuzzy neural network was proposed to effectively improve the dynamic characteristics of grid-connected inverter system. The fuzzy inference rule was established after an analysis of the PID controller for its proportional coefficient, integral coefficient and differential coefficient. A fuzzy neural network was applied to make automatic adjustment to the parameters of PID controller, based on which the proposed dynamic optimization algorithm was deduced in theory. According to the simulation and experimental results, the method is effective in making the system more robust to external disruption due to its excellent steady-state adaptivity and self-learning ability.

Keywords: Single-phase inverter; Fuzzy neural network; PID controller.

1. INTRODUCTION

PWM inverter is a device that converts DC power into AC power by exercising control over the switch of semiconductor. Due to its simple principle, excellent universality, and stable output voltage, it has been widely applied in such fields as photovoltaic power generation, wind power generation, industrial control and so on[1]. At present, the control strategies commonly adopted by PWM inverter system include conventional PI control, double closed-loop control, repetitive control, frequency control, fuzzy control and neural network control, etc. Conventional PI control is characterized by the simplicity in structure, high reliability and excellent stability. However, the system is incapable of quickly restoring the equilibrium state in case of overset induced by abrupt disruption. Despite the remarkable dynamic and static characteristics shown by double closed loop control, there will be corresponding control effect produced to eliminate the deviation if the controlled variable deviates from the specified value[2,3]. Consequently, the interference difference would be reduced by control accuracy. Though repetitive control is effective in making the system capable of static error-free tracking, but it is inevitable to cause a significant deterioration of dynamic performance[4]. Frequency control shows such advantages as instant response, the direct adjustment to oscillation frequency and low distortion. Nevertheless, it lacks robustness[5]. Fuzzy control demonstrates an impressive robustness and adaptability. However, it lacks systematicness and the ability to define control objectives[6-8]. Neural network control is capable of self-learning and highly efficient in finding the optimal solution, which makes it suitable for both linear and nonlinear systems[9,10].

Furthermore, fuzzy neural network control with PID control were combined in this paper to develop a PID controller parameter optimization technology based on fuzzy neural network. Based on the uncertainty information processing ability of fuzzy theory and the self-learning ability of neural network, the fuzzy neural network was applied to adjust the PID controller parameters, thus achieving the effective control over the output voltage of the inverter

2. MATHEMATICAL MODEL OF PHOTOVOLTAIC GRID-CONNECTED INVERTER

Photovoltaic grid-connected inverter is diversified in terms of topological structure. In this paper, single-phase grid-connected photovoltaic inverter was taken as the research object. As shown in Figure 1, the first half is the boost chopper circuit, which is capable to obtain the power point of the system. The second part is comprised of the inverter and filter circuit in the photovoltaic inverter system, for converting electric energy from direct current to alternating current without changing the frequency and phase of the grid voltage. It not only plays a crucial role in making the whole system interconnected, but also determines the quality of power from the photovoltaic system to the grid. The control of grid-connected inverter is a closed-loop double loop that consists of a grid-connected current inner loop and a DC voltage outer loop. The inner loop of grid-connected current is a PID controller based on fuzzy neural network.

In this paper, LC filter was adopted to improve the quality

of current for the power grid, thus preventing the sensitive equipment in the power grid from being disrupted by the harmonics in the alternating current outputted by the inverter[11,12]. The switching tubes T_1 、 T_2 、 T_3 and T_4 comprise two pairs of bridge arms. In order to avoid the short circuit of the power supply, the upper and lower switching tubes of the same bridge arm work in a complementary way, with the inductor on the inverter side L_1 and the filter capacitor C_1 comprising the LC filter.

In Figure 1, current i_1 was taken as the state variable, which came from the output of the inverter and flew through the filter inductor L. According to KVL law:

$$U_1 = U_2 + \frac{L_1 di_1}{dt^3} + i_1 R_L \tag{1}$$

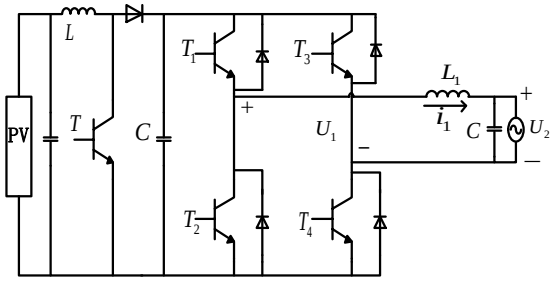


Fig.1 Schematic diagram of two-stage grid-connected inverter.

According to Formula (1), the transfer function of i_1 after Laplace transform was:

$$I_1(\mathrm{s}) = \frac{1}{L\mathrm{s}+\mathrm{R}_L}[U_1(\mathrm{s}) - U_2(\mathrm{s})] \tag{2}$$

In the aforementioned formula, U_1 represents the output voltage of the inverter bridge, U_2 indicates the output voltage of the inverter, i_1 denotes the current flowing through the inductance at the side of the network, and "s" refers to the Laplace operator. R_L is defined as the equivalent resistance of the inductor coil and AC feedback inductor coil. After passing through the LC filter and being incorporated into the power grid, U_1 is treated as the standard sinusoidal wave.

Herein, since the unstable load system is connected to the grid, which means the load changes on a frequent basis, the fluctuation of net voltage results. Therefore, in order to suppress the oscillation caused by the fluctuations of power grid voltage, the control system is supposed to be added with the power grid voltage feed-forward link. I_{ref} indicates the current reference, I_g denotes the current feedback value of the grid side, and $\frac{K}{Ts+1}$ refers to the inverter unit. Figure 2 illustrates how the grid current control works.

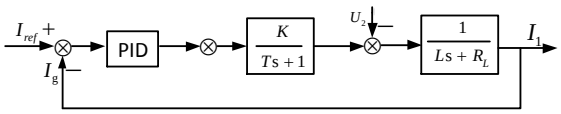


Fig.2 Structure chart of grid-connected current control.

3. ESTABLISHMENT OF FUZZY NEURAL NETWORK

3.1 Principle Of Fuzzy Neural Network PID Control

Referred to as a technology that integrates the powerful structural knowledge expression enabled by fuzzy logic reasoning and the excellent self-learning ability of the neural network, fuzzy neural network (FNN) is a combination of fuzzy logic reasoning and neural network. In this paper, the fuzzy neural network was applied to construct three fuzzy neural networks for the three parameters of PID, based on which the changes in parameters can be predicted. Figure 3 shows the schematic diagram of the fuzzy neural network PID closed-loop control system.

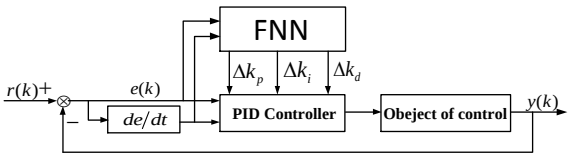


Fig.3 Structure chart of fuzzy neural network PID closed-loop control system.

In this network, the input error e and the error change rate e_c are taken as the changes to PID control parameters according to the rules applied to the fuzzy neural network.

$$e(k) = r(k) - y(k) \tag{3}$$

When the input signals are propagated forward, the input signals e and e_c would enter the neural network from the input layer. After the sum of each layer is weighted and various functions of the activation function are fulfilled, the input signals e and e_c would be inputted by the output layer.If the expected error value is set to "e<0.01",In case that the deviation between the output results obtained by network Δk_p、Δk_i 、Δk_d and expectations fails to achieve the goal, it would proceed to the stage of error back propagation, while the error signal would be propagated back from the output layer along the direction of the network to the input layer in steps. In the meantime, the network weights and thresholds at this stage would be adjusted according to the error gradient value, with the distance between the network output and the desired results reduced layer by layer, thus approaching the

expected value. The loop iteration is conducted on a continued basis in these two stages until the desired output of the network is achieved, thus improving the dynamic and static performance for the controlled object.

3.2 Forward Propagation

The input error signal e and error rate signal e_c collected by the fuzzy neural network PID controller are applied to fuzz the input signal and convert it into a language variable value with a specific meaning. There are two input variables required as input for the fuzzy controller, including deviation e and deviation change rate e_c. The language variable value is selected as "negative (N), zero (Z) and positive (P)". The Gaussian membership function is treated as the membership function.

The fuzzy rule base is obtained by means of induction. Based on the fuzzed input variables, the language variable values of output variables Δk_p、Δk_i and Δk_d are obtained by looking up the table. As the input variables e and e_c contain 3 language variable values, respectively, the number of fuzzy rules is set to 9. After fuzzing and fuzzy reasoning, the solution is fuzzily transmitted to the output layer, based on which three adjustable parameters Δk_p、Δk_i and Δk_d controlled by PID are obtained. Figure 4 shows the schematic diagram of the fuzzy neural network of Δk_p that is, 2x3x9x9x1. It consists of five layers, which are input layer, membership layer, rule layer, back layer and output layer.

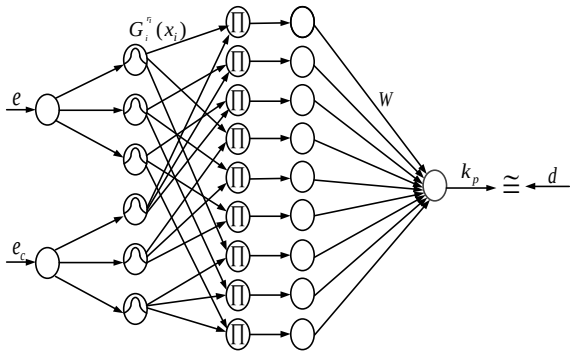


Fig.4 Topology diagram of fuzzy neural network.

The processing of each layer of fuzzy neural network controller is detailed as follows:

The first layer: input layer, error e and error change e_c are treated as the control input.

$$x_1 = e, x_2 = e_c, X(m,n) = x_m, m = 1,2; n = 1,2,3 \quad (4)$$

The second layer: fuzzy layer. The error e and error change e_c are fuzzily processed by Gaussian membership function. ξ_{mn} represents the average value of the n th Gaussian function mapping of the m th fuzzy variable, and σ_{mn} refers to the variance value of the n Gaussian function mapping of the m th fuzzy variable:

$$G_m^n(x) = \exp[\frac{-(x-\xi_{mn}^2)}{2\sigma_{mn}^2}] \quad (5)$$

The third layer is also known as the "sum" layer, where the number of nodes in this layer is the product of the number of fuzzy sets (that is, the number of rules) of each input variable, with each node of the layer indicating the first part of the rule.

$$\begin{aligned} &\mu(1) = \mathrm{G}_1^1 * \mathrm{G}_2^1 \ \ \mu(2) = \mathrm{G}_1^2 * \mathrm{G}_2^1 \ \ \mu(3) = \mathrm{G}_1^3 * \mathrm{G}_2^1 \\ &\mu(4) = \mathrm{G}_1^1 * \mathrm{G}_2^2 \ \ \mu(5) = \mathrm{G}_1^2 * \mathrm{G}_2^2 \ \ \mu(6) = \mathrm{G}_1^3 * \mathrm{G}_2^2 \\ &\mu(7) = \mathrm{G}_1^1 * \mathrm{G}_2^3 \ \ \mu(8) = \mathrm{G}_1^2 * \mathrm{G}_2^3 \ \ \mu(9) = \mathrm{G}_1^3 * \mathrm{G}_2^3 \end{aligned} \quad (6)$$

The fourth layer is the output layer, the purpose of which is to perform defuzzification. The neurons in this layer represent the output variables, and the output can be obtained by means of area-center defuzzification.

$$\begin{aligned} &y_k^p = \sum_{n=1}^{L} W_{nk}\mu_n^p \Big/ \sum_{n=1}^{L} \mu_n^p \\ &(\mathrm{k} = 1,...,\mathrm{M}; P = 1,...,P) \end{aligned} \quad (7)$$

P indicates the number of training modes, L denotes the total number of the fuzzy rules of the third layer, while W_{nk} represents the weight coefficients of, k_p, k_i, and k_d, respectively.

3.3 Error Back Propagation

When the outputs Δk_p, Δk_i, and Δk_d as obtained by the network fail to approach the expected values, it would enter the stage of error back propagation. To be specific, the error signal would undergo a back propagation from the output layer along the direction of the network to the input layer in steps. At this stage, the weights and thresholds of the network would be adjusted according to error gradient values, while the deviation between the output results of the network and expected results would be progressively reduced layer by layer, thus approaching the expected value. The loop iteration is continuously conducted in these two stages, until the desired output of the network is obtained.

The actual output is defined as:

$$y_z = \sum_{l=1}^{L} r_l w_z^l = r^T w_z \quad (8)$$

$z = \{1,2,3\}$, L indicates the number of fuzzy rules, r denotes the output number of Gaussian function mapping, and d stands for the expected value. Given all the p training vectors as $\{r_i = [r_i^1 \ r_i^2 ... r_i^L]^T \mid i = 1,...,p\}$.

In the matrix symbol, it is specified that:

$$R=[r_1,r_2...,r_p]\in L\times P$$
$$Y=[y_1,y_2...,y_p]\in P\times Z \quad (9)$$
$$D=[d_1,d_2...,d_p]\in P\times Z$$

The effective output matrix Y can be expressed as:

$$Y=R^TW \quad (10)$$

The aforementioned J can also be reorganized using matrix notation. Thus, the error function E is defined as:

$$E=Y-D=R^TW-D \quad (11)$$

As the weighted matrix W is required to be updated (or trained), the effective output y_z is capable of convergence to the required output d_z. Thus, the objective function of total squared error J is expressed as follows:

$$J=\frac{1}{2PZ}\sum_{P=1}^{P}\sum_{Z=1}^{Z}(y_z^p-d_z^p)^2 \quad (12)$$

In order to update W, the error back propagation method is applied to the weight, based on which the chain rule is applied to obtain:

$$W_{\tau+1}=W_\tau-\beta_\tau\frac{\partial J}{\partial W}|_\tau=W_\tau-\beta_\tau\frac{1}{PZ}RE \quad (13)$$

The update of fuzzy neural network (FNN) premise part:

$$\left\{\begin{array}{l}\xi_{mn}(\tau+1)=\xi_{mn}(\tau)-\alpha\dfrac{\partial J}{\partial\xi_{mn}}\\ \sigma_{mn}(\tau+1)=\sigma_{mn}(\tau)-\alpha\dfrac{\partial J}{\partial\sigma_{mn}}\end{array}\right\} \quad (14)$$

ξ_{mn} is the mean of n^{th} membership function of m^{th} input. σ_{mn} is the variance of n^{th} membership function of m^{th} input. J is the mean square error. α is the fixed learning rate for the BP algorithm in the premise part.

After training, it is assumed that the error is zero to obtain the matrix form $D=R^TW$. It is worth noting that the learning rate for each iteration in the back propagation process is assumed to vary, which means the learning rate is not fixed. The theorem of optimal learning rate β_τ is expressed as follows.

Theorem: The minimum value of optimal learning rate β_τ defined in Formula (13) is obtained from quadratic polynomial $a\beta_\tau^2+b\beta_\tau$, where $a>0$ and $b<0$ are determined by training vector r, expected output vector d and weighted matrix W.

Proof: First of all, the stable range of β_τ must be found in the first place. To achieve this purpose, the Lyapunov function is defined as $U=J^2$, and the change of the Lyapunov function is $\Delta U=J_{\tau+1}^2-J_\tau^2$. If system response can be stabilized if $\Delta U<0$, then $J_{\tau+1}-J_\tau<0$ can be obtained through detailed deduction:

$$\begin{aligned}J_{\tau+1}-J_\tau&=\beta_\tau\frac{-1}{PZ}Tr[(PZ)^{-1}E_\tau^TR^TRE_\tau]\\&\quad+\beta_\tau^2\frac{1}{2PZ}Tr[(PZ)^{-2}R^TRE_\tau E_\tau^TR^TR] \quad (15)\\&=a\beta_\tau^2+b\beta_\tau<0\end{aligned}$$

To satisfy Formula (15), $a\beta_\tau^2+b\beta_\tau<0$. As $a>0$, there is a clear range of stability. Then a and b are expressed as:

$$\begin{aligned}a&=\frac{1}{2}(PZ)^{-3}Tr[R^TRE_\tau E_\tau^TR^TR]\\&=2(PZ)^{-3}\sum_{P=1}^{P}\sum_{Z=1}^{Z}\left(\sum_{l=1}^{L}r_l^p(\tau)\sum_{i=1}^{p}r_l^i(\tau)(y_z^i-d_z^i)\right)^2>0\end{aligned} \quad (16)$$

$$\begin{aligned}b&=-(PZ)^{-2}Tr[E_\tau^TR^TRE_\tau]\\&=\frac{-1}{(PZ)^2}\sum_{P=1}^{P}\sum_{Z=1}^{Z}\left((y_z^p-d_z^p)\sum_{l=1}^{L}r_l^p(\tau)\sum_{i=1}^{p}r_l^i(\tau)(y_z^i-d_z^i)\right)<0\end{aligned} \quad (17)$$

Apparently, Formulas (16) and (17) contain a quadratic matrix, for which a is supposed to be greater than zero and b ought to be lower than zero. Thus,

$$J_{\tau+1}-J_\tau=a\beta^2+b\beta<0 \quad (18)$$

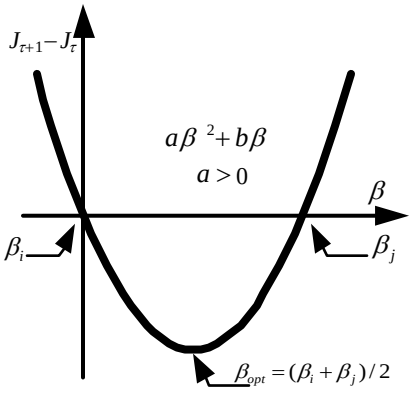


Fig.5 Parabolic path.

In order to satisfy Formula (15), there must be $a\beta^2+b\beta<0$. As $a>0$, it is evident that the range of stability for β is $(\beta_i;\beta_j)$, where β_i and β_j are the two roots of $a\beta^2+b\beta=0$. From Figure 5, it can be concluded that the optimal is β_{opt} the median of β_i and β_j

$$\beta_{opt} = (\beta_i + \beta_j)/2 \tag{19}$$

This is because the symmetry of the parabola shown in Figure 5 not only ensures the stability of the training process, but also achieves convergence at the fastest pace.

4. SIMULATION EXPERIMENT

For a comparison to be performed between the fuzzy neural network PID control and BP neural network, the model of single-phase photovoltaic (PV) grid inverter was constructed in Matlab using the control strategy that combined the proposed fuzzy neural network and PID control, while the fuzzy neural network algorithm was applied to set the three parameters related to PID controller, including k_p, k_i, and k_d, thus making the output error satisfy the requirements for synchronization. The main circuit parameters of the simulation include: ambient temperature:25℃, light intensity: $1000W/m^2$, switching frequency:$10k$ Hz, filter inductance: $L = 4mH$, capacitance: $C = 0.6\mu F$, AC output voltage of inverter: 220V/50Hz.

Figure 6 shows the voltage and current waveform of single-phase grid-connected photovoltaic inverter under the control of fuzzy neural network PID. The optimization strategy proposed in this paper is proven to be effective in enabling the grid-connected current to track the grid voltage in a more efficiently and accurate way, thus achieving the identical frequency and phase for the two.

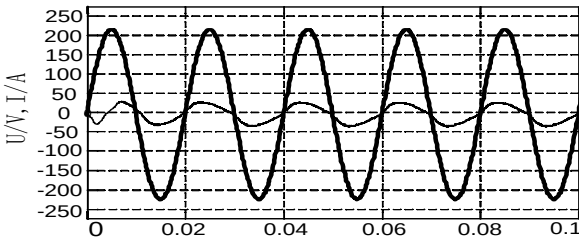


Fig.6 Load output voltage and current signal of fuzzy neural network PID control.

As shown in Figures 7 and 8, in comparison with BP neural network control, the load voltage waveform of the inverter controlled by fuzzy neural network PID is smoother and closer to the reference voltage. Therefore, the output voltage of the inverter controlled by fuzzy neural network PID control achieves a higher steady state precision and a greater response speed. According to Figures 9 and 10,through comparison and analysis of the load voltage THD controlled by the fuzzy neural network PID, it can be known that the harmonic count of output voltage for the inverter using the fuzzy neural network PID control is lower. Under the context of fuzzy neural network PID control, the grid-connected current error is 0.02s. In the meantime, the current tracking error enters its steady state.

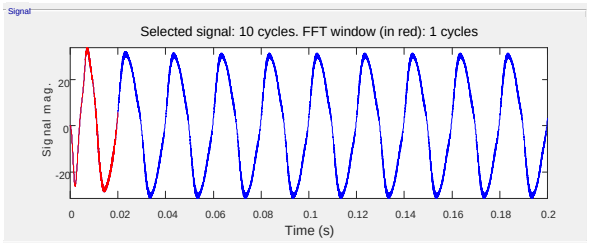


Fig.7 Load output voltage signal of BP neural network PID control.

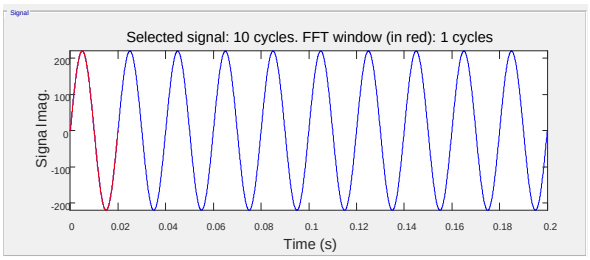


Fig.8 Load output voltage signal of fuzzy neural network PID control

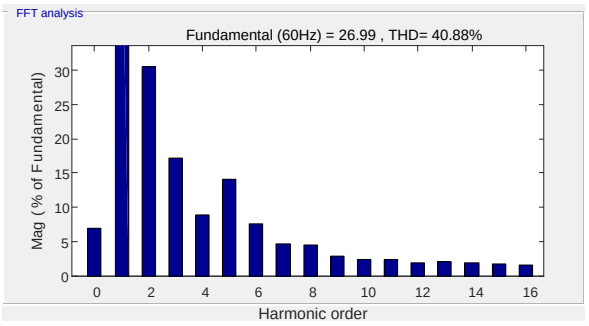


Fig.9 Load voltage THD analysis of BP neural network PID control.

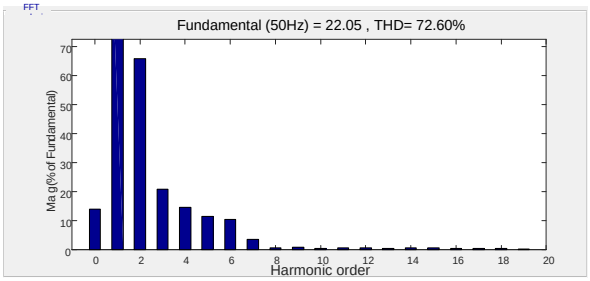


Fig.10 Load voltage THD analysis of fuzzy neural network PID control.

5. CONCLUSION

In this paper, the fuzzy neural network PID controller was constructed based on the uncertainty information processing ability of fuzzy theory as well as the self-learning ability of neural network. Then, the real-time parameter update to a new fuzzy neural network was carried out. As indicated by the simulation results, the PID parameters based on fuzzy neural network demonstrate an outstanding

self-learning ability and adaptive ability. After training, the parameters could be effectively controlled in the grid-connected current, with the load voltage showing not only a lower THD value but also a superior steady and dynamic performance.

References

[1] Yu W , Department E E . “A Research and Design of Inverter Based on Bipolar SPWM Modulation”, Journal of Electric Power, 2014.

[2] C. T. Wu and L. Y. Wang, “Double-loop SVPWM control strategy for multi-level inverter based on LC filter”, 2015 IEEE International Conference on Applied Superconductivity and Electromagnetic Devices (ASEMD),2015, pp. 395-396.

[3] W. D. P. Vallejos,“Standalone photovoltaic system, using a single stage boost DC/AC power inverter controlled by a double loop control”, 2017 IEEE PES Innovative Smart Gri Technologies Conference - Latin America (ISGT Latin America),2017, pp. 1-6.

[4] Wang Fang, Zhang Yi and Wang Zhen, “Research on a kind of multiple control strategy for parallel connected inverters”, IEEE 2011 10th International Conference on Electronic Measurement & Instruments,Vol. 4, 2011, pp. 267-269.

[5] M. Jahanbakhshi, B. Asaei and B. Farhangi, “A novel deadbeat controller for single phase PV grid connected inverters”, 2015 23rd Iranian Conference on Electrical Engineering,2015, pp. 1613-1617.

[6] P. Mitra, C. Dey and R. K. Mudi, “An improved fuzzy PID controller with fuzzy rule based set-point weighting technique”,2016 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC),2016, pp. 40-44.

[7] HE Yun-sheng,“Grid-Connected Fuzzy-PID Control of PV Power Generation System”,Power System and Clean Energy, Vol. 29, No.2, 2013, pp. 85-89.

[8] P. Mohammadi, B. Azimian and A. Shahirinia, “A Novel Double-Loop Control Structure Based on Fuzzy-PI and Fuzzy-PR Strategies for Single-Phase Inverter in Photovoltaic Application”, 2018 North American Power Symposium (NAPS),2018,pp. 1-6.

[9] Z. Yunbo, L. Zhiguo, L. Shengzhu and Z. Hong, “Research on BP neutral network based grid-connected photovoltaic inverter”,2016 Chinese Control and Decision Conference (CCDC), 2016, pp. 506-509.

[10] G. Han, Y. Xia and W. Min, “A grid-connected current control technique of single-phase voltage source inverter based on BP neural network”, 2012 IEEE International Conference on Computer Science and Automation Engineering (CSAE), Vol. 1, 2012, pp. 547-551.

[11] Liu J, “MATLAB simulation of advanced PID control”, Publishing house of electronics industry, 2004.

[12] Dehong X,“Modeling and control of power electronic system”Machinery Industry Pressu1,2006.

Repetitive Control Based on Multi-Stage PSO Algorithm with Variable Interval for T-S Fuzzy Systems

Yibing Wang*,**, Manli Zhang*,**, Min Wu *,**, Luefeng Chen*,**
* School of Automation, China University of Geosciences, Wuhan 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China

Abstract

This paper presents a repetitive-control method based on particle swarm optimization (PSO) with variable interval to enhance tracking performance for T-S fuzzy systems. First, T-S fuzzy model is used to describe a nonlinear system. A modified repetitive control structure with two repetitive loops guarantees the tracking accuracy of periodic signals. Taking advantage of the two-dimensional (2D) property with continuous control and discrete learning, a continuous-discrete 2D model is presented to describe the nonlinear repetitive-control system. Next, a multi-stage PSO algorithm with variable interval searches for the best parameter combination in the linear matrix inequality-based stability condition to regulate the control and learning actions, which avoids falling into a sub-optimal solution and guarantees high control accuracy. Finally, an application to speed control of a permanent magnet synchronous motor demonstrates the validity of the method and comparisons with related methods show the superiority.

Keywords: Two-dimensional repetitive control, Multi-stage particle swarm optimization algorithm with variable interval, Takagi-Sugeno fuzzy model, Permanent magnet synchronous motor.

1. INTRODUCTION

In industrial applications, many servo systems need to deal with periodic control tasks. The repetitive control method [1] based on the internal model theory is an alternative to solve this problem. But it's difficult to stabilize the system due to a pure-delay positive-feedback loop in the repetitive controller. The modified repetitive controller (MRC) [2] was developed by inserting a low-pass filter into the delay line. Moreover, the repetitive controller guaranteed the system stable by sacrificing high-frequency performance.

There are two kinds of actions in a repetitive control process: continuous control and discrete learning, which means that repetitive control has two-dimensional (2D) properties. A continuous-discrete 2D hybrid model was proposed [3, 4], and it converted the control problem of a modified repetitive control system (MRCS) to the stability of the 2D system. And two parameters were used during the Lyapunov stability analysis to adjust control and learning actions. It's noticed that those researches were considered mainly in the linear system. However, most physical systems have various nonlinearities, which will deteriorate the control performance of the systems.

On the other hand, a Takagi-Sugeno (T-S) fuzzy model [5] is an alternative way to deal with nonlinear systems stabilization, regulation and tracking control problems [6, 7]. It uses local linear models combined with fuzzy membership functions to describe a nonlinear system. Therefore, the linear system theory can be directly used to the nonlinear system, which makes it receive a great deal of attention. A T-S fuzzy model has been widely used to complex nonlinear systems, such as a nonlinear system with parameter some uncertainties and disturbances [8] and a discrete nonlinear system [9], etc.

An MRC with two repetitive loops was used to improve the tracking performance of the nonlinear system. The parameter in one of the repetitive loop can balance control and learning action [10]. However, those two actions were blent up by a low-pass filter in the MRCS, which made the parameter selection more difficult. Therefore, it's necessary to find an approach to select parameters efficiently. Intelligent optimization methods, such as genetic algorithm (GA) and particle swarm optimization (PSO) algorithm, etc, have the advantages of fast, high efficiency and high precision. Among them, the PSO algorithm features a simple structure and easy implementation. It could be used to search for the parameter combination [11]. However, the complex relationships among parameters made the classical PSO algorithm easy to fall into a sub-optimal solution, which was a difficulty to be solved urgently. Some researchers improved the ability of the PSO algorithm to jump out of a sub-optimal solution by adjusting inertia weight [12] or initializing particle population [13], etc.

In this paper, a PSO-based MRCS is designed to deal with the problem of periodic tracking for the nonlinear system. A T-S fuzzy model is designed to describe the nonlinear system. An MRC, which has an additional rep-

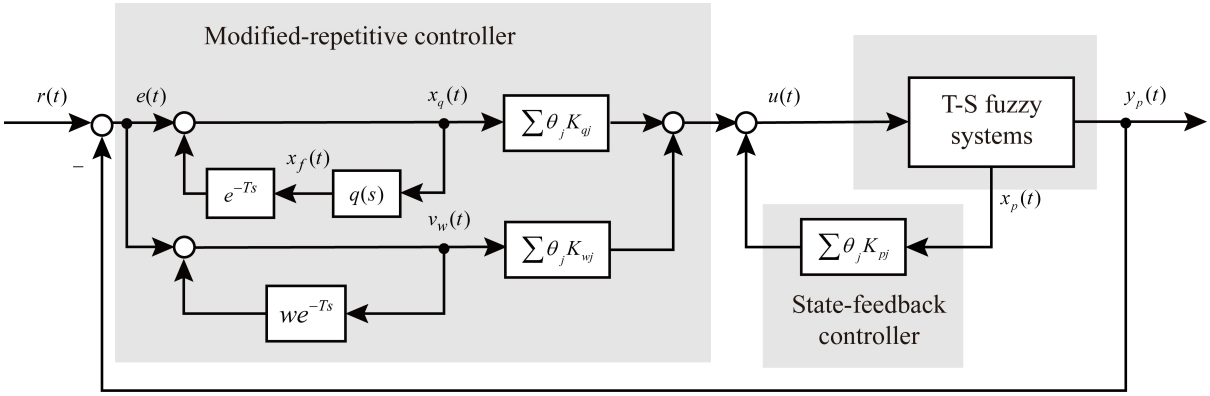


Fig. 1. Configuration of MRCS

etitive loop, is used to obtain the satisfying tracking performance. A 2D hybrid model is established and an LMI condition guarantees the asymptotical stability of the system. And a multi-stage PSO algorithm with variable interval is given to search for the optimal parameter combination so that both transient and steady-state performance can be improved. Finally, the effectiveness is verified by numerical simulation and comparative experiments.

2. PROBLEM FORMULATION

In this paper, the periodic tracking control problem is considered for the nonlinear system based on a T-S fuzzy model, which is capable of approximating any smooth nonlinear system using IF-THEN fuzzy rules. The configuration of MRCS (Fig. 1), has three parts: T-S fuzzy systems, an MRC with two repetitive loops and a state-feedback controller.

The nonlinear system can be described as:
Rule *i*: If $z_1(t)$ is θ_{1i} and $\cdots$ $z_p(t)$ is θ_{pi}, **Then**

$$\begin{cases} \dot{x}_p(t) = A_i x_p(t) + B_i u(t), \\ y_p(t) = C_i x_p(t), \ 0 < i \le r. \end{cases} \tag{1}$$

where $z(t)=[z_1(t),\ z_2(t),\ \cdots,\ z_p(t)]$ is known premise variables, $[\theta_{1i},\ \theta_{2i},\ \cdots,\ \theta_{pi}]$ are fuzzy sets, r is the number of fuzzy rules; $x_p(t)$, $u(t)$ and $y_p(t)$ are the system state, input and output, respectively; A_i, B_i and C_i are known real constant matrices with appropriate dimension. It assumes that the output matrices are all the same [8].

By fuzzy blending, the fuzzy system can be represe- nted as follows:

$$\begin{cases} \dot{x}_p(t) = \sum_{i=1}^{r} \theta_i(z(t))(A_i x_p(t) + B_i u(t)), \\ y_p(t) = \sum_{i=1}^{r} \theta_i(z(t)) C x_p(t). \end{cases} \tag{2}$$

where $\theta_i(z(t)) = \sigma_i(z(t)) \Big/ \sum_{i=1}^{r} \sigma_i(z(t))$, with $\sigma_i(z(t))$ $\Pi_{m=1}^{p} F_{mi}(z(t))$, $F_{mi}(z(t))$ is the grade of membership value of $z_m(t)$ in F_{mi} , $\sum_{i=1}^{r} \theta_i(z(t))=1$, $0 < \theta_i(z(t)) < 1$.

The state-space expression of the MRC is

$$\begin{cases} \dot{x}_q(t) = -\omega_c x_q(t) + \omega_c x_q(t-T) + \omega_c e(t) + \dot{e}(t), \\ v_w(t) = w v_w(t-T) + e(t). \end{cases} \tag{3}$$

where $x_a(t)$ and $v_w(t)$ stand for the state vectors of MRC; T is the period of the reference input; ω_c is the cut-off frequency of the low pass filter; w is a constant value and $e(t)$ is the tracking error.

The state-feedback controller is designed as:
Rule *i*: If $z_1(t)$ is θ_{1i} and $\cdots$ $z_p(t)$ is θ_{pi}, **Then**

$$u(t) = K_{pj} x_p(t) + K_{qj} x_q(t) + K_{wj} v_w(t). \tag{4}$$

where K_{pj} , K_{qj} and K_{wj} are controller gains.The overall output of the fuzzy controller is

$$u(t) = \sum_{j=1}^{r} \theta_j(z(t))[K_{pj} x_p(t) + K_{qj} x_q(t) + K_{wj} v_w(t)]. \tag{5}$$

By analyzing the MRC, the state-feedback controller and using the lifting technology [11], the 2D hybrid model of the closed-loop augmented system is as follows, where k is the number of learning discrete variables, τ is the time continuous variables in a period:

$$\begin{cases} \dot{\phi}(k,\tau) = \sum_{j=1}^{r} \sum_{i=1}^{r} \theta_i(z(k,\tau)) \theta_j(z(k,\tau)) [\tilde{A}\phi(k,\tau) \\ \qquad + \tilde{A}_d \phi(k-1,\tau) + \tilde{B} v_w(k-1,\tau)], \\ v_w(k,\tau) = \sum_{j=1}^{r} \sum_{i=1}^{r} \theta_i(z(k,\tau)) \theta_j(z(k,\tau)) [\tilde{C}\phi(k,\tau) n \\ \qquad + w v_w(k-1,\tau)]. \end{cases} \tag{6}$$

where

$$\phi(k,\tau) = \begin{bmatrix} x_p^{\mathrm{T}}(k,\tau) & x_q^{\mathrm{T}}(k,\tau) \end{bmatrix}^{\mathrm{T}},$$

$$\tilde{A} = \begin{bmatrix} A_i + B_i \tilde{K}_{pj} & B_i K_{qj} \\ -\omega_c C - C A_i - C B_i \tilde{K}_{pj} & -\omega_c I - C B_i K_{qj} \end{bmatrix},$$

$$\tilde{A}_d = \begin{bmatrix} 0 & 0 \\ 0 & -\omega_c I \end{bmatrix}, \tilde{B} = \begin{bmatrix} w B_i K_{wj} \\ -w C B_i K_{wj} \end{bmatrix}, \tilde{C} = \begin{bmatrix} -C & 0 \end{bmatrix}.$$

and the fuzzy control law is

$$
\begin{aligned}
u(k,\tau) &= \sum_{j=1}^{r}\theta_j(z(k,\tau))[K_{pj}x_p(k,\tau)+K_{qj}x_q(k,\tau) \\
&\quad +K_{wj}v_w(k,\tau)] \\
&= \sum_{j=1}^{r}\theta_j(z(k,\tau))(\begin{bmatrix}\tilde{K}_{pj} & \tilde{K}_{qj}\end{bmatrix}\phi(k,\tau) \\
&\quad +\tilde{K}_{wj}wv_w(k-1,\tau)).
\end{aligned} \tag{7}
$$

with $\tilde{K}_{pj}=K_{pj}-K_{wj}C$, $\tilde{K}_{qj}=K_{qj}$ and $\tilde{K}_{wj}=K_{wj}$.

Remark 1. The control input $u(k, \tau)$ in (7) stands for the direct sum of the effects of the control and learning actions. Thus, by adjusting K_{pj}, K_{qj} and K_{wj}, we can regulate those two actions.

3. STABILITY ANALYSIS AND CONTROL SYSTEM DESIGN

This section first presents a stable condition of the closed-loop system (6), then the main idea of multi-stage PSO algorithm with variable interval is described.

3.1 Stability Analysis

The stability of the MRCS is guaranteed by **Lemma 1**.

Lemma 1. [11] For the given constants α, β, w and a cut-off frequency ω_c, if there exist the symmetric and positive-definite matrices X_1, X_2, Y_1 and Y_2; arbitrary matrices W_{1i}, W_{2i} and W_{3i} such that the LMIs hold for $1\le i\le j\le r$:

$$
\begin{cases}
\Lambda_{ii}<0, \\
\Lambda_{ij}+\Lambda_{ji}<0.
\end{cases} \tag{8}
$$

where

$$
\Lambda_{ij}=\begin{bmatrix}
\Lambda_{1,1}^{ij} & \Lambda_{1,2}^{ij} & 0 & 0 & \Lambda_{1,5}^{ij} & \alpha X_1 & 0 & \Lambda_{1,8}^{ij} \\
* & \Lambda_{2,2}^{ij} & 0 & \omega_c Y_2 & \Lambda_{2,5}^{ij} & 0 & \beta X_2 & 0 \\
* & * & -Y_1 & 0 & 0 & 0 & 0 & 0 \\
* & * & * & -Y_2 & 0 & 0 & 0 & 0 \\
* & * & * & * & \Lambda_{5,5}^{ij} & 0 & 0 & 0 \\
* & * & * & * & * & -Y_1 & 0 & 0 \\
* & * & * & * & * & * & -Y_2 & 0 \\
* & * & * & * & * & * & * & -Y_2
\end{bmatrix},
$$

$$
\begin{aligned}
&\Lambda_{1,1}^{ij}=\alpha(A_iX_1+B_iW_{1j})+\alpha(A_iX_1+B_iW_{1j})^{\mathrm{T}}, \\
&\Lambda_{1,2}^{ij}=-\alpha X_1(\omega_cC+CA_i)^{\mathrm{T}}-\alpha(CB_iW_{1j})^{\mathrm{T}}+\beta B_iW_{2j}, \\
&\Lambda_{1,5}^{ij}=wB_iW_{3j}-\alpha wX_1C^{\mathrm{T}}, \Lambda_{1,8}^{ij}=-\alpha X_1C^{\mathrm{T}}, \\
&\Lambda_{2,2}^{ij}=-\beta(2\omega_cX_2+CB_iW_{2J}+(CB_iW_{2j})^{\mathrm{T}}), \\
&\Lambda_{2,5}^{ij}=-wCB_iW_{3j}, \Lambda_{5,5}^{ij}=(w^2-1)Y_2.
\end{aligned}
$$

Then the closed-loop system is asymptotically stable.Moreover, the 2D feedback gains are given by

$$
\tilde{K}_{pj}=W_{1j}X_1^{-1}, \tilde{K}_{qj}=W_{2j}X_2^{-1}, \tilde{K}_{wj}=W_{3j}Y_2^{-1}. \tag{9}
$$

Remark 2. The 2D feedback gains in (9) can be adjusted by the constants α, β and w, which can also adjust the performance of transient and steady-state.

3.2 Repetitive Control Design Based on Multi-Stage PSO with Variable Interval

Repetitive control features self-learning, and studying it from a 2D perspective can enhance the tracking accuracy of the MRCS essentially. An effective method is to choose a suitable parameter combination in the Lyapunov functional to adjust control and learning actions in the MRCS. But a low-pass filter mixes up those two actions, which makes parameters correlated to each other. So it is difficult to choose a suitable parameter combination through trial and error. Intelligent optimization methods with fast, efficient and high accuracy, can be used to solve this problem. Among them, the PSO algorithm is highly potential due to its simple structure, easy implementation and fast computation capability.

PSO algorithm [11] is used to search for the optimal combination of repetitive-controller gains and state-feedback controller gain. The brief idea is as below:

- **Parameter Initialization:** Select an optimization precision J_{set}; time constant T; cut off frequency ω_c; inertia weight b; learning coefficients c_1, c_2, the number of particles n and suitable search ranges;
- **Particle Initialization:** Generate n particles satisfying (8) in search space, and calculate J_i^{ι}, update J^{ι}_{pest} and J^{ι}_{gest}, where ι is the times of iterations;
- **Iteration Optimization:** According to the updating laws for the velocity and position in PSO, update n particles under the limitation of (8). Calculate J_i^{ι}, change J^{ι}_{pest} and J^{ι}_{gest} simultaneously. According to the global performance value, determine whether to stop iterative optimization.

The performance value J_i^{ι} is calculated by

$$
J_i^{\iota}=\frac{1}{2}\sum_{k=1}^{\rho}\int_{(k-1)T}^{kT}e_i^{\iota\,2}(t)dt
$$

with ρ is the number of repetitive periods. $e_i^{\iota}(t)$ is the tracking error of the $i-$th particle.

However, the parameters are not independent to each other. Also, the relationship between parameters and performance value is complex. Both make the classical PSO algorithm easy to fall into a sub-optimal solution. To solve this problem, a multi-stage PSO algorithm with variable interval is designed as follows:

step 1: Set $m=0$, , $\sigma=1$, where σ stands for the $\sigma-$th optimization and m is used to record the number of times that iteration optimization doesn't work much;

step 2: Start the $\sigma-$th PSO, record the optimal global value for the $\iota-$th iteration in the $\sigma-$th opti-

mization as $J_{gbest}^{\iota,\sigma}$.

step 3: If $\iota < 1$, then go to ***step 4***; else if

$$J_{gbest}^{\iota-1,\sigma} - J_{gbest}^{\iota,\sigma} \le k * J_{gbest}^{\iota,\sigma},$$

then $m = m+1$, otherwise $m = 0$;

step 4: If $m > M$, adjust the parameter ranges as $\alpha \in (\alpha_\sigma - \Delta\alpha, \alpha_\sigma + \Delta\alpha)$, $\beta \in (0.1\beta_\sigma, 10\beta_\sigma)$ and $w \in \lceil w_\sigma - \Delta w, w_\sigma + \Delta w)$, where $(\alpha_\sigma, \beta_\sigma, w_\sigma)$ is the result of $\sigma - th$ optimization, $\sigma = \sigma + 1$, $m = 0$, and return ***step 2***.

Remark 3. m indicates the potential of the optimization intervals. When $m > M$, we think that the optimization falls into a sub-optimal solution, and the optimization intervals need to be readjusted. The value of k, M, $\Delta\alpha$ and Δw depends on the actual problem.

Remark 4. Compared to other optimization methods like GA, the multi-stage PSO algorithm with variable interval does not possess the crossover and mutation processes. It has much more profound intelligent background and could be performed more easily.

4. NUMERICAL SIMULATION

In this section, we apply our method to the nonlinear system of speed control of PMSM (Fig. 2). The system has two motors: PMSM1 is used for speed control, and PMSM2 is used to generate a nonlinear torque.

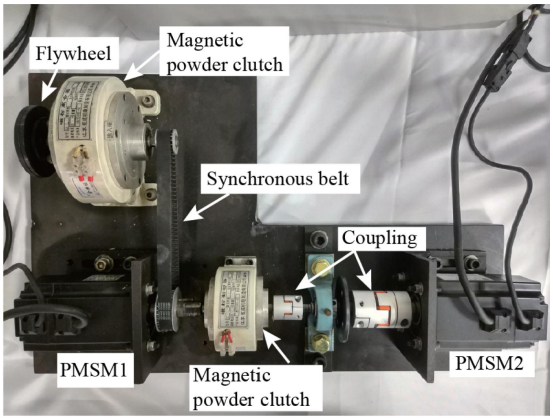


Fig. 2. Nonlinear system of PMSM.

By analyzing the control process, the mechanical motion equation of the controlled motor is

$$J\frac{d\omega(t)}{dt} = -B_v\omega(t) + \tau_e(t) - f(t). \qquad (10)$$

where ω is angular velocity $[\text{rad/s}]$; τ_e is electromagnetic torque [Nm]; J is moment of inertia of PMSM $[\text{Kg}\cdot\text{m}^2]$; B_v is the coefficient of friction $[\text{Nm}\cdot\text{s/rad}]$; $f(t)$ is the nonlinearity, and $f(t) = 0.5J\omega\sin\omega$.

Choose the input, output and state variables as $u(t)=\tau_e(t)$, $y_p(t)=\omega(t)$ and $x_p(t)=\omega(t)$. The nonlinear system (7) can be represented by the fuzzy model with two fuzzy rules [14], and the system matrices are given as follows:

$$A_1 = -8.738,\ B_1 = 2272,\ C = 1,$$
$$A_2 = -7.738,\ B_2 = 2272.$$

Let the periodic reference input be $r(t) = 13\sin\pi t + 12\sin 2\pi t$. So the period is $T = 2$ s. And carry out the optimization algorithm in Section 3, we first choose

$$\begin{cases} J_{set} = 0.1,\ \omega_c = 100,\ k = 0.01,\ M = 20, \\ \alpha \in (1, 100),\ \beta \in (0, 10),\ w \in [0, 1),\ \rho = 10, \\ n = 20,\ b = 1,\ c_1 = c_2 = 1,\ \Delta\alpha = 10,\ \Delta w = 0.05. \end{cases}$$

The whole process of the multi-stage PSO with variable interval can be represented as Table 1. The parameters are finally selected as below:

$$\alpha = 39.9731,\ \beta = 0.0001,\ w = 0.14. \qquad (11)$$

Solving LMIs in **Lemma 1** with (11), the gains are

$$K_{p1} = -0.0022,\ K_{p2} = -0.0026,$$
$$K_{q1} = 2.3717,\ K_{q2} = 2.3717,$$
$$K_{w1} = 0.0204,\ K_{w2} = 0.0204$$

and simulation results are shown in Fig. 3.

Table 1. Whole Process of optimization.

Parameter	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$
α	39.82	31.0046	39.9731
β	0.01	0.001	0.0001
w	0.22	0.19	0.14
J	448.5550	17.2847	0.0567

From Fig. 3, we can know after approximately one period, the system output $y_p(t)$ can track the given periodic signal $r(t)$. The maximum tracking error is only 5.792×10^{-2} rad/s and the steady-state error is less than 3×10^{-3} rad/s. The control input $u(t)$ in transient response is similar to that in the steady state. This shows the system has satisfactory tracking performance.

Remark 5. The model of PMSM used for numerical simulation was obtained through the identification experiment, which indicated our method could be applied to an experimental system to a certain extent. And the related experiment will be done in the future to verify the practicability.

To verify the effectiveness of our method, we use the methods in [10] and [11] to carry out a comparison.

Method 1 [10]: The parameters are chosen as:

$$\alpha = 40,\ \beta = 0.001,\ w = 0.2,\ \gamma = 1.$$

and the control gains are

$$K_{p1} = -0.0039,\ K_{p2} = -0.0044,$$
$$K_{q1} = 0.1552,\ K_{q2} = 0.1552,$$
$$K_{w1} = 0.0131,\ K_{w2} = 0.0131.$$

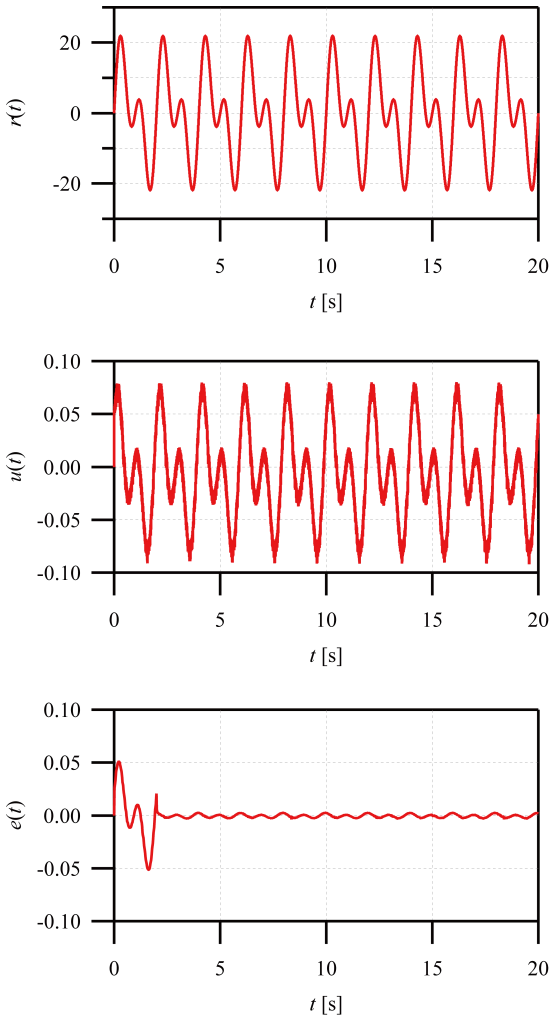


Fig. 3. Simulation results

The simulation results (Fig. 4) show the maximum transient tracking error is 0.9062 rad/s and the state-stable tracking error is 0.05271 rad/s . The system enters the steady-state in the third period and $J = 13.8652$.

Method 2 [11]: The parameters are chosen as

$$\alpha = 49.94,\ \beta = 1.06\times10^{-4},\ w = 0.2302.$$

They provide the gains

$$K_{p1} = -0.0032,\ K_{p2} = -0.0036,$$
$$K_{q1} = 0.6364,\ K_{q2} = 0.6364,$$
$$K_{w1} = 0.0194,\ K_{w2} = 0.0194.$$

Fig. 5 shows the simulation results. The maximum transient tracking error is 0.2157 rad/s , the steady-state tracking error is 1.133×10^{-2} rad/s and $J = 0.7761$.

The tracking performance comparison of three method has been shown in Table 2.

When we use the PSO algorithm, J is 0.7761 (0.056 of that in ***Method 1***), the maximum transient tracking error steady-state tracking error is 1.133×10^{-2} rad/s (0.215

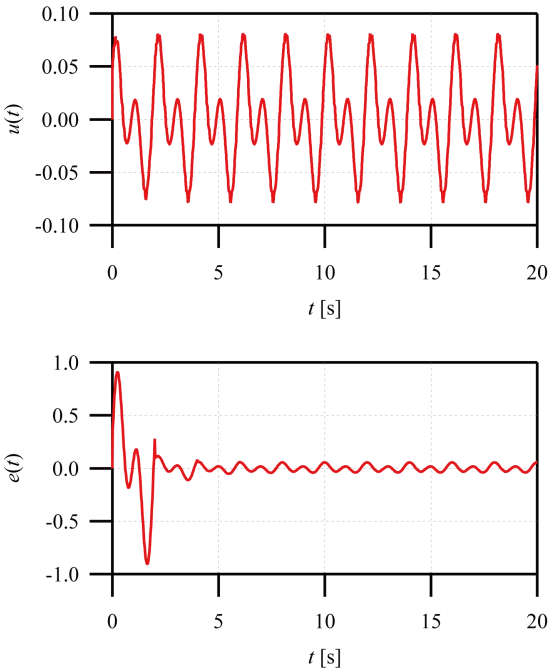


Fig. 4. Simulation results of ***Method 1***

of that in ***Method 1***). This indicts that the parameter combination searching by the PSO algorithm can provide better tracking performance.

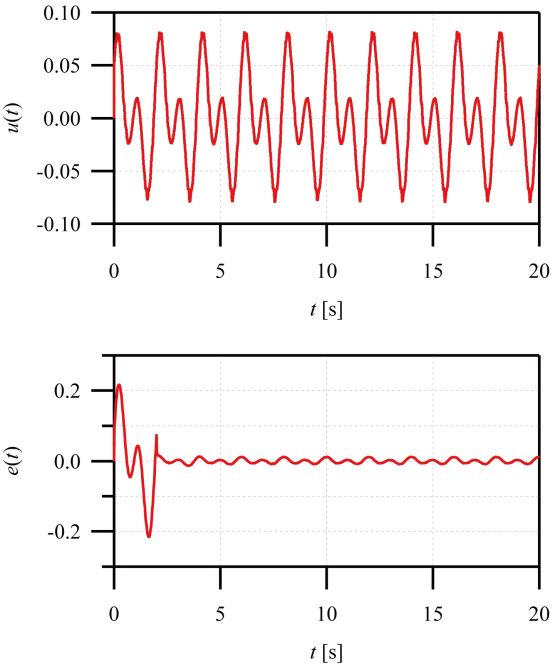


Fig. 5. Simulation results of ***Method 2***

By using the multi-stage PSO with variable interval, J is only 0.0567 (0.0041 times of that in ***Method 1*** and 0.073 times of that in ***Method 2***), the maximum transient tracking error is 5.792×10^{-2} rad/s (0.064 times of that in ***Method 1*** and 0.27 times of that in ***Method 2***), and the steady-state error is just 2.895×10^{-3} rad/s (0.055 times of that in ***Method 1*** and 0.255 times of that in ***Method 2***).

It means that our method has the best tracking performance.

Table 2. Tracking performance comparison.

Method	Maximum transient tracking error	Steady-state tracking error	Steady-state tracking error ratio
Our method	0.05792	0.002895	0.0132%
[10]	0.9062	0.05271	0.229%
[11]	0.2157	0.01133	0.0493%

5. CONCLUSIONS

This paper has presented a multi-stage PSO algorithm with variable interval to improve the control performance. First, T-S fuzzy model is used to describe the nonlinear system. Then, a continuous-discrete 2D model is established to adjust two actions. By using the Lyapunov stability theory, a sufficient condition is given in terms of LMIs. The parameters in the Lyapunov functional are chosen by a multi-stage PSO optimization with variable interval to produce the optimal gains of controllers. Simulations and comparisons show that our method can provide the best tracking performance.

The related experiment based on the experimental system of PMSM will be done to verify the practicability of our method in the future. Also, The method of choosing the value of constant k, M, $\Delta\alpha$ and Δw will be discussed.

Acknowledgements

This work was supported by the National Key R&D Program of China [2018YFC0603405] and the National Natural Science Foundation of China [61733016] and the Hubei Provincial Technical Innovation Major Project [2018AA035] and the Fundamental Research Funds for the Central Universities [CUG160705] and the 111 project [B17040].

References

[1] T. Inoue, M. Nakano, and S. Iwai, "High accuracy control of a proton synchrotron magnet power supply", Proceedings of the 8th International Federation of Automatic Control World Congress, 1981.

[2] S. Hara, Y. Yamamoto and T. Omata, "Repetitive control system: a new type servo system for periodic exogenous signals", IEEE Transactions on Automatic Control, Vol. 33, No. 7, 1988, pp. 659-668.

[3] M. Wu, L. Zhou, and J. H. She, "Design of observer-based H_∞ robust repetitive control system", IEEE Transactions on Automatic Control, Vol. 56, No. 6, 2011, pp. 1452-1457.

[4] L. Zhou, J. H. She, and M. Wu, "Design of robust observer-based modified repetitive control system", ISA Transactions, Vol. 52, No. 3, 2013, pp. 375-382.

[5] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 15, No. 1, 1985, pp. 116-132.

[6] S. C. Tong, B. Y. Huo, and Y. M. Li, "Observer-based adaptive decentralized fuzzy fault-tolerant control of nonlinear large-scale systems with actuator failures", IEEE Transactions on Fuzzy Systems, Vol. 22, No. 1, 2014, pp. 1-15.

[7] H. G. Zhang, J. Han, Y. C. Wang, and X. H. Liu, "Sensor fault estimation of switched fuzzy systems with unknown input", IEEE Transactions on Fuzzy Systems, Vol. 26, No. 3, 2018, pp. 1114-1124.

[8] Y. C. Wang, R. Wang and X. Xie, "Observer-based H_∞ fuzzy control for modified repetitive control systems", Neurocomputing, Vol. 286, 2018, pp. 141-149.

[9] X. P. Xie, D. Yue, H. Zhang, and Y. Xue, "Control synthesis of discrete-time T-S fuzzy systems via a multi-instant homogenous polynomial approach", IEEE Transactions on Cybernetics, Vol. 47, No. 9, 2017, pp. 2480-2491.

[10] M. L. Zhang, M. Wu, L. F. Chen, and P. Yu, "Design of modified repetitive controller for T-S fuzzy systems", Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 23, No. 3, 2019, pp. 602-610.

[11] M. L. Zhang, M. Wu, L. F. Chen, S. N. Tian and J. H. She, "Optimization of control and learning actions for repetitive-control system based on Takagi-Sugeno fuzzy model", International Journal of Systems Sciences, 2020, DOI:10.1080/00207721.2020.1807651.

[12] H. B. Dong, D. J. Li, and X. P. Zhang, "A particle swarm optimization algorithm for dynamically adjusting inertia weights", Computer Science, Vol. 45, No. 2, 2018, pp. 98-102.

[13] W. Li, and X. Q. Chao, "Improved particles swarm optimization method for feature selection", Journal of Frontiers of Computer Science and Technology, Vol. 13, No. 6, 2019, pp. 990-1004.

[14] K. Tanaka, and H. O. Wang, "Fuzzy control systems design and analysis: a linear matrix inequality approach", New York: Wiley, 2001, pp. 5-23.

Neural Network-Based Optimal Control for a Class of Unknown Nonlinear System via Output Information

Can DING*, Jing ZHANG*, Yingjie ZHANG **, Zhe ZHANG*, Feng LIU***
* College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China
** College of Computer Science and Electronic Engineering, Hunan University, Changsha, 410082, China ***
School of Automation, China University of Geosciences, Wuhan , 430074, China

Abstract

This paper develops an adaptive dynamic programming (ADP) based output feedback optimal approach for a class of unknown nonlinear system. First, considering the problems that the system states is unmeasurable and the exact system model is hard to obtain in practical applications, a neural network state observer is developed to estimate the unmeasurable system states and reconstruct the internal states of system via output information. After acquiring the knowledge of the system states, a critic network is proposed to approximate the solution of Hamilton-Jacobi-Bellman (HJB) equation. Then an output feedback optimal control was developed. Throughout the Lyapunov theory, the update laws of neural network and critic network was obtained and the stability of closed loop control system was proved. Finally, a simulation experiment was carried out to validate the effectiveness of the proposed method.

Keywords: ADP, Unknown Nonlinear Systems, Neural Network, Output Feedback.

1. INTRODUCTION

With the improvement of control performance requirements, the optimal control of nonlinear systems has received widespread attention in the field of control [1-3]. The key to optimal control lies in the solution of the system's HJB equation, and dynamic programming (DP) [4] is widely used in optimal control strategies as a solution. However, dynamic programming is performed backwards in time, so DP is an offline method. And because of the "dimensional curse [5]" problem in the high-dimensional optimization problem, DP is difficult to apply in practical control. In order to avoid the above problems, Werbos [6-7] proposed an adaptive dynamic programming (ADP) strategy based on reinforcement learning (RL). It combines reinforcement learning methods, Actor-Critic Structure structures and neural networks to solve optimal control problems. The critic network is used to estimate the cost function in dynamic programming, thereby solving the "dimensional curse" problem. In recent years, ADP and related fields have received extensive attention from scholars. Reference [8] designed a robust neural network controller based on ADP to solve the optimal control problem for uncertain nonlinear systems. Reference [9] proposed a new robust ADP control strategy for the optimal control of uncertain nonlinear systems with output constraints. Reference [10] applies the ADP method to multi-agent formation control with unknown dynamics.

In the practical application of optimal control, the problems of unknown system model and unmeasured internal states of the system also need to be considered [11]. It should be noted that in many practical applications, only the output and control input of the system are usually measurable, and only measuring the output can greatly reduce the need for measurement equipment in the control system. In order to estimate unmeasured state quantities, observers are widely used in output feedback control. Reference [12] estimated the internal states of the system by constructing an observer-evaluation structure, and combined with the ADP algorithm to solve the nonlinear system control problem. Reference [13] uses high-gain observers (HGOs) and event-driven technology to achieve multi-agent output feedback control goals. In this paper, the state quantity of the system is reconstructed by the neural network observer first, and then the output feedback optimal control strategy based on the ADP algorithm is designed from the observed state quantity. The weight update laws of the neural network observer and the weight update laws of the evaluation network are obtained through Lyapunov theory. The stability of the closed-loop system has also been strictly proved. Finally, simulation experiments verify the effectiveness of the control algorithm.

The rest of the paper is organized as follows. In Section 2, the unknown nonlinear system and control objective are described. In Section 3, an ADP based optimal controller with neural network observer was proposed and the stability of all close looped signal was proved through Lyapunov method. In Section 4, simulation experiment are represented to illustrate the validity and feasibility of the proposed approach. Finally, in Section 5 the main conclusions are outlined.

Notations: R represents the sets of real numbers. R^n and $R^{m\times n}$ denote the set of $n\times 1$ vectors and the set of $m\times n$ matrix respectively. $\|\bullet\|$ denotes the 2-norm of $\bullet$. T means the transposition symbol of matrix. $tr(\bullet)$ represents the trace of matrix $\bullet$.

2. SYSTEM DESCRIPTION

Considering the following unknown nonlinear system

$$\begin{cases}\dot{x}(t)=G(x(t),u(t))\\ y(t)=Cx(t)\end{cases} \tag{1}$$

where $x(t)\in R^n$ denote the system states, $y(t)\in R^m$ represents the system outputs which are measurable, $u(t)\in R^p$ are the control input, $C\in R^{m\times n}$ is the constant matrix, $U(x(t),u(t))$ denotes the unknown dynamic of system. It should be noted that in actual applications, the state quantity of the system is not necessarily completely measurable and the state equation of the system is not necessarily accurate, so the system (1) conforms to the actual application scenario.

In order to facilitate subsequent observer design, the system (1) is written as follows

$$\begin{cases}\dot{x}(t)=Ax+U(x(t),u(t))\\ y(t)=Cx(t)\end{cases} \tag{2}$$

where $U(x(t),u(t))=G(x(t),u(t))-Ax$.

Assumption 1. The system unknown dynamic $U(x(t),u(t))$ is smooth function and satisfies

$$\|\partial U(x(t),u(t))/\partial u(t)\|\le\delta \tag{3}$$

where δ is positive constant.

The control goal of this paper is to design an output feedback controller for the system (1) and ensure that all signals of the closed-loop control system are ultimately consistent and bounded (Uniformly Ultimate Boundedness , UUB).

3. SYSTEM DESCRIPTION

This section is divided into two parts. Firstly, a state estimator based on neural network is designed to estimate the state quantity of the system. Then the output feedback optimal control strategy based on adaptive evaluation is designed.

3.1 Neural Network based State Estimator

Combine the system (2) to design the following state observer

$$\begin{cases}\dot{\hat{x}}=A\hat{x}+\hat{U}+T(y-C\hat{x})\\ \hat{y}=C\hat{x}\end{cases} \tag{4}$$

where $\hat{x}$ and $\hat{y}$ are estimation of the system states and output respectively. $\hat{U}$ is the estimated value of system unknown dynamic U which approached by neural network. The selection of $T\in R^{n\times m}$ satisfied that $D=A-TC$ is Hurwitz matrix and possess the following property

$$D^{\mathrm{T}}P+PD=-M \tag{5}$$

where $P=P^{\mathrm{T}}$, $M=M^{\mathrm{T}}$.

The following neural network approximator is designed to approximate the unknown function U in the system (1)

$$U=\theta^{\mathrm{T}}s(\hat{x},u)+\alpha(x) \tag{6}$$

$$\hat{U}=\hat{\theta}^{\mathrm{T}}s(\hat{x},u) \tag{7}$$

where $\hat{U}$ is the estimated value of U, $\tilde{U}=U-\hat{U}$ is the estimation error of neural network and is bounded, which means $\alpha(x)$ satisfy $\|\alpha(x)\|\le\alpha_m$, α_m is positive constant. The activation function of the hidden layer selects the hyperbolic tangent function. The input of the neural network is the estimated state and control output of the system.

The update laws of network weights are selected as follow

$$\dot{\hat{\theta}}=-a(\tilde{y}^{\mathrm{T}}CD^{-1})^{\mathrm{T}}s^{\mathrm{T}}(\hat{x},u)-\lambda\|\tilde{y}\|\hat{\theta} \tag{8}$$

where $\tilde{y}=y-\hat{y}$, λ is positive constant.

Theorem 1. For the system (1), the state observer (4), the neural network (6), and the weight update laws (8), the estimated error of the state observer $\tilde{x}=x-\hat{x}$ and the estimated weight error $\tilde{\theta}$ of the neural network are UUB.

Proof: Define the following Lyapunov candidate function

$$V_1=\frac{1}{2}\tilde{x}^{\mathrm{T}}P\tilde{x}+tr(\tilde{\theta}^{\mathrm{T}}\tilde{\theta}) \tag{9}$$

The derivative of (9) is

$$\dot{V}_1=\frac{1}{2}\dot{\tilde{x}}^{\mathrm{T}}P\tilde{x}+\frac{1}{2}\tilde{x}^{\mathrm{T}}P\dot{\tilde{x}}+tr(-\tilde{\theta}^{\mathrm{T}}\dot{\hat{\theta}}) \tag{10}$$

Combining (2) (4) (6) (7) (8), formula (10) can be written as

$$\begin{aligned}\dot{V}_1=&-\frac{1}{2}\tilde{x}^{\mathrm{T}}M\tilde{x}+s(\hat{x},u)^{\mathrm{T}}\tilde{\theta}P\tilde{x}+\alpha^{\mathrm{T}}P\tilde{x}\\&+tr(\tilde{\theta}^{\mathrm{T}}a((C\tilde{x})^{\mathrm{T}}CD^{-1})^{\mathrm{T}}s(\hat{x},u)^{\mathrm{T}}+\lambda\tilde{\theta}^{\mathrm{T}}\|\tilde{y}\|\hat{\theta})\end{aligned} \tag{11}$$

Let $b=a(D^{-1})^{\mathrm{T}}C^{\mathrm{T}}C$, then (11) be simplified to

$$\begin{aligned}\dot{V}_1=&-\frac{1}{2}\tilde{x}^{\mathrm{T}}M\tilde{x}+s(\hat{x},u)^{\mathrm{T}}\tilde{\theta}P\tilde{x}+\alpha^{\mathrm{T}}P\tilde{x}\\&+tr(\tilde{\theta}^{\mathrm{T}}b\tilde{x}s(\hat{x},u)^{\mathrm{T}}+\lambda\|\tilde{y}\|\tilde{\theta}^{\mathrm{T}}\hat{\theta})\end{aligned} \tag{12}$$

Consider the following inequality

$$tr(\tilde{\theta}^{\mathrm{T}}(\theta-\tilde{\theta}))\le\theta_m\|\tilde{\theta}\|-\|\tilde{\theta}\|^2 \tag{13}$$

Bring inequality (13) into (12) , then we have

$$\begin{aligned}\dot{V}_1 \le & -\frac{1}{2}\lambda_{\min}(M)\|\tilde{x}\|^2 + \|\tilde{x}\|(s_m\|\tilde{\theta}\|\|P\| + \alpha_m\|P\| \\ & + s_m\|\tilde{\theta}\|\|b\|) + \lambda\|C\|\|\tilde{x}\|(\theta_m\|\tilde{\theta}\| - \|\tilde{\theta}\|^2)\end{aligned} \tag{14}$$

where $\lambda_{\min}(M)$ denotes the minimum eigenvalue of matrix M.

In order to eliminate the term containing $\|\tilde{\theta}\|$ in (14), let

$$H = \frac{s_m\|P\| + s_m\|b\| + \lambda\theta_m\|C\|}{2\lambda\|C\|} \tag{15}$$

Then (14) becomes

$$\begin{aligned}\dot{V}_1 \le & -\frac{1}{2}\lambda_{\min}(M)\|\tilde{x}\|^2 + \|\tilde{x}\|(\alpha_m\|P\| \\ & + \lambda\|C\|H^2 - \lambda\|C\|(H - \|\tilde{\theta}\|)^2) \\ \le & -\frac{1}{2}\lambda_{\min}(M)\|x\|^2 + \|\tilde{x}\|(\alpha_m\|P\| + \lambda\|C\|H^2)\end{aligned} \tag{16}$$

Therefore, in order to ensure that $\dot{V}_1 < 0$, the state errors meet the following condition

$$\|\tilde{x}\| \ge \frac{2(\alpha_m\|P\| + \lambda\|C\|H^2)}{\lambda_{\min}(M)} \tag{17}$$

According to Lyapunov theory, as long as the condition (17) is satisfied, the estimation error $\tilde{x}$ of the observer and the estimation weight error $\tilde{\theta}$ of the neural network satisfy UUB. Theorem 1 is proved.

3.2 ADP based Output Feedback Optimal Controller

For the system (1), design the following performance indicators

$$V(x(t)) = \int_0^\infty y^{\mathrm{T}}(s)Qy(s) + u^{\mathrm{T}}(s)Ru(s)ds \tag{18}$$

where $Q = Q^{\mathrm{T}} \in R^{m\times m}$ is positive definite constant matrix.

If the control input u is admissible on Ω [14], and the performance index function is first-order derivable, then

$$V_x^{\mathrm{T}}U(x,u) + y^{\mathrm{T}}Qy + u^{\mathrm{T}}Ru = 0 \tag{19}$$

Define the following Hamiltonian function

$$H(x,V_x,u) = V_x^{\mathrm{T}}U(x,u) + y^{\mathrm{T}}Qy + u^{\mathrm{T}}Ru \tag{20}$$

Then, the optimal performance index function can be obtained by solving the following HJB equation

$$\min_{u\in\Omega} H(x,V_x^*,u) = 0 \tag{21}$$

After solving the optimal performance index, the following optimal control strategy can be obtained

$$u^* = -\frac{1}{2}\left(\frac{\partial U(x,u^*)}{\partial u}\right)^{\mathrm{T}}V_x^* \tag{22}$$

The optimal control strategy (22) is the ideal optimal control solution. This paper uses the following evaluation neural network to approximate the optimal performance index function

$$V^* = W^{\mathrm{T}}s(x,u) + \sigma(x,u) \tag{23}$$

where W is the ideal weight of critic network, the activation function $s(x,u)$ of hidden layer is selected as hyperbolic tangent function. $\sigma(x,u)$ is the estimation error of network, then we have

$$\hat{V} = \hat{W}^{\mathrm{T}}s(\hat{x},u) \tag{24}$$

$$\tilde{V} = V^* - \hat{V} = \tilde{W}^{\mathrm{T}}s(\hat{x},u) + \varepsilon \tag{25}$$

where $\hat{V}$ is the real performance index, $\hat{W}$ is real weight of the critic network, $\varepsilon = W(s(x,u) - s(\hat{x},u)) + \delta(x,u)$ denotes the total estimation error, which meets $\|\varepsilon\| \le \varepsilon_m$ and ε_m is positive constant.

According to (24), we have

$$\hat{V}_x = \hat{W}^{\mathrm{T}}\frac{\partial s(\hat{x},u)}{\partial \hat{x}} \tag{26}$$

Combining (26) and (7), the actual control law is designed as follows

$$u = -\frac{1}{2}\left(\hat{\theta}^{\mathrm{T}}\frac{\partial s(\hat{x},u)}{\partial u}\right)^{\mathrm{T}}\hat{W}^{\mathrm{T}}\frac{\partial s(\hat{x},u)}{\partial \hat{x}} \tag{27}$$

Taking (26) (27) into (20), the estimation of Hamiltonian can be expressed as

$$\hat{H}(x,V_x,u) = \hat{V}_x^{\mathrm{T}}\hat{U}(\hat{x},u) + y^{\mathrm{T}}Qy + u^{\mathrm{T}}Ru \tag{28}$$

Since $H(x,V_x^*,u^*) = 0$, the Hamiltonian estimation error is expressed as follows

$$\begin{aligned}e_h &= \hat{H}(x,V_x,u) - H(x,V_x^*,u^*) \\ &= \hat{V}_x^{\mathrm{T}}\hat{U}(\hat{x},u) + y^{\mathrm{T}}Qy + u^{\mathrm{T}}Ru\end{aligned} \tag{29}$$

In order to make e_h small enough, the weight of the evaluation network is adjusted online to minimize the quadratic function $E = e_h^{\mathrm{T}}e_h/2$. The gradient descent method can be used to evaluate the adjustment rate of the network

$$\begin{aligned}\dot{\hat{W}} &= -\frac{r}{(1+\varphi^{\mathrm{T}}\varphi)^2}\frac{\partial E}{\partial \hat{W}} \\ &= -\frac{r\varphi}{(1+\varphi^{\mathrm{T}}\varphi)^2}(\varphi^{\mathrm{T}}\hat{W} + y^{\mathrm{T}}Qy + u^{\mathrm{T}}Ru)\end{aligned} \tag{30}$$

where $\varphi = (\partial s(\hat{x},u)/\partial\hat{x})\dot{\hat{x}}$.

For the convenience of subsequent analysis, make the following assumptions

Assumption 2. The control output is Lipschitz and satisfies

$$\|u - u^*\| \le L(\hat{x} - x) \tag{31}$$

where L is positive constant.

Theorem 2. For the system (1), if the observer is (3) and the weight update laws of the observer and evaluation network are (7) and (30), respectively, the state quantity of the closed-loop system x, the state estimation error $\tilde{x}$, and the estimated weight error $\tilde{\theta}$ of the observer, and the estimated weight error $\tilde{W} = W - \hat{W}$ of the critic network satisfy UUB.

Proof. Define the following Lyapunov candidate function

$$V_a = V_1 + V^* + V_2 \tag{32}$$

where $V_2 = tr(W^{\mathrm{T}}W)/(r)$. According to equation (1), differentiate V^* with respect to time, we have

$$\dot{V}^* = (V_x^*)^{\mathrm{T}}\dot{x} = (V_x^*)^{\mathrm{T}}U(x,u) \tag{33}$$

According to the optimal control strategy (27), we have

$$V_x^{*\mathrm{T}} = -2u^{*\mathrm{T}}\left((\frac{\partial U(x,u^*)}{\partial u^*})^{\mathrm{T}}\frac{\partial U(x,u^*)}{\partial u^*}\right)^{-1}(\frac{\partial U(x,u^*)}{\partial u^*})^{\mathrm{T}} \tag{34}$$

Because

$$H(x,V_x^*,u^*) = \hat{V}_x^{\mathrm{T}}\hat{U}(\hat{x},u) + y^{\mathrm{T}}Qy + u^{*\mathrm{T}}Ru^* \tag{35}$$

Bring (34) (35) into (33), then

$$\begin{aligned}\dot{V}^* &= 2u^{*\mathrm{T}}\left((\frac{\partial U(x,u^*)}{\partial u^*})^{\mathrm{T}}\frac{\partial U(x,u^*)}{\partial u^*}\right)^{-1}(\frac{\partial U(x,u^*)}{\partial u^*})\\ &\quad (U(x,u^*)-U(x,u)) - y^{\mathrm{T}}Qy - u^{*\mathrm{T}}Ru^*\\ &= 2u^{*\mathrm{T}}N(U(x,u^*)-U(x,u)) - y^{\mathrm{T}}Qy - u^{*\mathrm{T}}Ru^*\end{aligned} \tag{36}$$

where

$$N = \left((\frac{\partial U(x,u^*)}{\partial u^*})^{\mathrm{T}}\frac{\partial U(x,u^*)}{\partial u^*}\right)^{-1}(\frac{\partial U(x,u^*)}{\partial u^*})$$

According to the mean value theorem, equation (36) can be rewritten as

$$\dot{V}^* = 2u^{*\mathrm{T}}N\frac{\partial U(x,u_1)}{\partial u_1}(u^*-u) - y^{\mathrm{T}}Qy - u^{*\mathrm{T}}Ru^* \tag{37}$$

where u_1 is between u^* and u . According to assumption 1, there is a positive constant η such that $\|N \times \partial U(x,u_1)/\partial u_1\| \le \eta$, combined with assumption 2, equation (37) can be rewritten as

$$\dot{V}^* \le 2\eta L\|u^*\|\|\tilde{x}\| - \lambda_{\min}(Q)\|y\|^2 - \lambda_{\min}(R)\|u^*\|^2 \tag{38}$$

According to theorem 1, there is a positive constant such that $\|\tilde{x}\| \le \kappa$. If the following inequality holds

$$\lambda_{\min}(R) > \frac{2\eta L\kappa}{\|u^*\|} \tag{39}$$

Then $\dot{V}^* < 0$.

Differentiate V_2 with respect to time, then

$$\dot{V}_2 = tr(\tilde{W}^{\mathrm{T}}\dot{\tilde{W}})/(2r) \tag{40}$$

When (30) is brought into (40), it can be obtained

$$\dot{V}_2 = \frac{1}{r}tr(W^{\mathrm{T}}\frac{r\varphi}{(1+\varphi^{\mathrm{T}}\varphi)}(-\varphi^{\mathrm{T}}\tilde{W} + \varphi^{\mathrm{T}}W + y^{\mathrm{T}}Qy + u^{\mathrm{T}}Ru)) \tag{41}$$

In combination with (25), we have

$$\dot{V}_2 = \frac{1}{r}tr(\tilde{W}^{\mathrm{T}}\frac{r\varphi}{(1+\varphi^{\mathrm{T}}\varphi)}(-\varphi^{\mathrm{T}}\tilde{W} + \phi)) \tag{42}$$

where $\phi = (\partial\varepsilon/\partial\hat{x})\dot{\hat{x}}$. Using Young's inequality, (42) can be written as

$$\dot{V}_2 \le -\frac{\|\varphi\|^2\|\tilde{W}\|^2}{(1+\varphi^{\mathrm{T}}\varphi)} + (\frac{\|\varphi\|^2\|\tilde{W}\|^2}{2} + \frac{\phi^2}{2})\frac{1}{(1+\varphi^{\mathrm{T}}\varphi)} \tag{43}$$

Therefore, in order to make $\dot{V}_2 < 0$, the weight estimation error of critic network satisfies the following condition

$$\|\tilde{W}\| > \left\|\frac{\phi}{\varphi}\right\| \tag{44}$$

According to Lyapunov theory, when (17) (39) and (44) are satisfied, $\dot{V}_a < 0$. Theorem 2 is proved.

4. SIMULATION RESEARCH

In order to verify the effectiveness of the control strategy, the following nonlinear system is used as the simulation object

$$\begin{cases}\dot{x}_1 = -2x_1 + x_2\\ \dot{x}_2 = -x_1 - (4-\sin^2(x_1))x_2 + u + 0.1u^2\\ y = Cx\end{cases} \tag{45}$$

where the states are $x = [x_1, x_2]^{\mathrm{T}}$, $C = [1,0]$, u is the control input. The initial value of system states are $x(0) = [1,-1.5]^{\mathrm{T}}$. The initial value of observer are $\hat{x}(0) = [0,0.5]^{\mathrm{T}}$. The parameters of observer are

$$A = \begin{bmatrix}0 & 1\\ -6 & -5\end{bmatrix},\quad T = \begin{bmatrix}10\\ -2\end{bmatrix}.$$

The structure of neural network (7) is 3-3-2, and the weight initial value is $\theta(0) = [1,0.5,0.2;-0.2,-0.5,-1]$, the parameters of update laws (18) are $a = 0.1$, $\lambda = 10$. The parameters of performance index are $Q = 1$, $R = 10$. The structure of critic network (24) is 3-4-1, and the initial weight values are $W(0) = [4,-2,3,2]^{\mathrm{T}}$. The parameter of update laws (30) is $r = 10$. The simulation results are shown in Fig. 1 to Fig. 4.

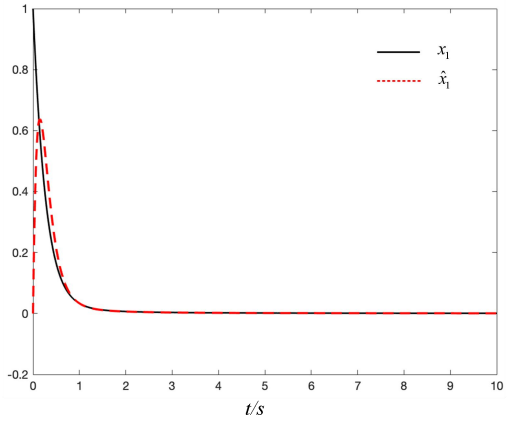


Fig. 1 The nominal system states x_1 and estimating state $\hat{x}_1$.

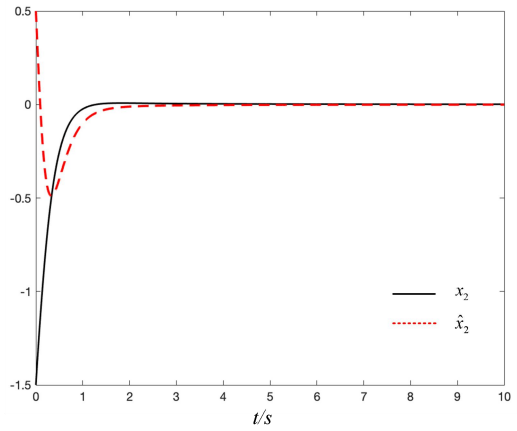


Fig. 2 The nominal system states x_2 and estimating state $\hat{x}_2$

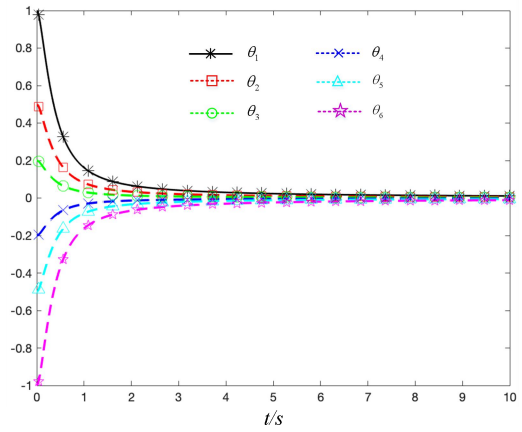


Fig. 3 The estimating weight vector θ of neural network

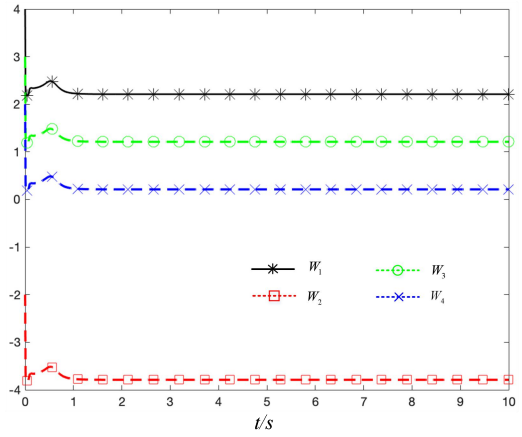


Fig. 4 The estimating weight vector W of critic network

5. CONCLUSION

An optimal control strategy based on neural network is proposed for a class of unknown nonlinear systems. The neural network state observer is used to estimate the state values of the unknown nonlinear system, so as to solve the problem that some state cannot be measured in practical application. Then, the output feedback optimal control strategy based on ADP algorithm is designed through the state variables estimated by the observer. In this paper, the weight update rate of neural network and critic network is obtained by Lyapunov method, and the stability of closed-loop system is proved strictly. Finally, the effectiveness of the controller is proved by simulation

References

[1] Yi J, Chen S, Zhong X, et al. Event-Triggered Globalized Dual Heuristic Programming and Its Application to Networked Control Systems[J]. IEEE Transactions on Industrial Informatics, 2019, 15(3): 1383-1392.

[2] Yang X, He H. Adaptive Critic Designs for Event-Triggered Robust Control of Nonlinear Systems With Unknown Dynamics[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2019, 49(6): 2255-2267.

[3] Wang D, He H, Liu D, et al. Adaptive Critic Nonlinear Robust Control: A Survey[J]. IEEE Transactions on Systems, Man, and Cybernetics, 2017, 47(10): 3429-3451.

[4] F. L. Lewis, D. Vrabie, and V. L. Syrmos, Optimal Control, 3rd ed. New York, NY, USA: Wiley, 2012

[5] R. E. Bellman, Dynamic Programming. Princeton, NJ, USA: Princeton Univ. Press, 1957.

[6] P. J. Werbos, “Beyond regression: New tools for prediction and analysis in the behavioural sciences,” Ph.D. dissertation, Harvard Univ., Cambridge, MA, USA, 1974.

[7] P. J. Werbos, “Advanced forecasting methods for global crisis warning and models of intelligence,” Gener. Syst. Yearbook, vol. 22, no. 12, pp. 25–38, 1977.

[8] Wang D, Liu D, Li H, et al. Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming[J]. Information Sciences, 2014: 167-179.

[9] Fan B, Yang Q, Tang X, et al. Robust ADP Design for Continuous-Time Nonlinear Systems With Output Constraints[J]. IEEE Transactions on Neural Networks, 2018, 29(6): 2127-2138.

[10] G. Wen, C. L. P. Chen and B. Li, "Optimized Formation Control Using Simplified Reinforcement Learning for a Class of Multiagent Systems With Unknown Dynamics," in IEEE Transactions on Industrial Electronics, vol. 67, no. 9, pp. 7879-7888, Sept. 2020, doi: 10.1109/TIE.2019.2946545.

[11] Gao W , Jiang Z P . Adaptive Dynamic Programming and Adaptive Optimal Output Regulation of Linear Systems[J]. Automatic Control, IEEE Transactions on, 2016, 61(12):4164-4169.

[12] Yang X , Liu D , Wei Q . Online approximate optimal control for affine non-linear systems with unknown internal dynamics using adaptive dynamic programming[J]. IET CONTROL THEORY AND APPLICATIONS, 2014, 8(16):1676-1688.

[13] Gao F, Chen W, Li Z, et al. Event-triggered cooperative learning from output feedback control for multi-agent systems[J]. Neurocomputing, 2018: 70-79.

[14] Abu-Khalaf M , Lewis F L . Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach[J]. Automatica, 2005, 41(5):779-791.

Path Planning of Mobile Robot in Complex Environment Based on Genetic Algorithm and Improved Artificial Potential Field Method

Feng Liu*,**,†, Hualing He*, Zhihua Li*,**, Zhi-Hong Guan***
* School of Automation, China University of Geosciences, Wuhan 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
*** College of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
† E-mail: fliu@cug.edu.cn；13986145301@163.com

Abstract

In this paper, a genetic algorithm and an improved artificial potential field method is proposed for path planning of mobile robot in complex environment. A gravitational function is introduced to solve the problem of the target unreachable. For the local minimum point problem, the virtual target point method and Gauss function are introduced as the gravitational potential energy function of the virtual target to guide the robot out of the trap area quickly. The velocity repulsive force is introduced into the velocity potential field, and the artificial potential field is effectively adjusted to ensure its effectiveness in a dynamic environment. The genetic algorithm is combined with the improved post-potential field method to optimize the path length. Finally, the effectiveness of method is proved by simulation results.

Keywords: Path planning, artificial potential field, gravitation function, dynamic environment, genetic algorithms.

1. INTRODUCTION

The task of path planning is to find a collision-free path from the starting position to the target position according to certain evaluation criteria in the environment with obstacles. At present, the methods of robot path planning are divided into two categories: traditional path planning method and artificial intelligence path planning method [1]. Traditional path planning methods mainly include visual graph method, grid method, free space method, artificial potential field method, etc. Artificial intelligence path planning methods mainly include expert system, neural network, fuzzy algorithm, genetic algorithm and some bionic algorithms. Among them, the artificial potential field method is widely used in path planning. Artificial potential field method is a common method for path planning, which is simple and efficient.

The application of artificial potential field method is limited due to the problems such as unreachable target, easy to fall into local optimum and poor adaptability to dynamic environment. Many scholars proposed a variety of improved methods to overcome such drawbacks. By adding escape force [2], the robot can be helped to escape from the local minimum point. However, taking escape measures after the robot falls into the local minimum point will waste a lot of time. Using the idea of walking along the wall [3] to escape from the local minimum point is an intuitive and practical method in a static environment. In [4], the relative velocity factor is introduced into the repulsion potential field function to avoid dynamic obstacles, but the dynamic obstacle avoidance efficiency is not high and the planning path is long. Although these methods improve the defects of the artificial potential field method to some extent, they still have different limitations. In order to solve this problem, this paper improves the gravitational field function, introduces virtual target point and velocity potential field, and overcomes its shortcomings successfully. Finally, it combines with genetic algorithm to improve the planning performance of artificial potential field method.

2. TRADITIONAL ARTIFICIAL POTENTIAL FIELD METHOD

2.1 Principle of traditional artificial potential field method

The concept of artificial potential field was first proposed by Khatib [5] in 1986, using to control a manipulator motion, avoid neighboring obstacles, and plan the arm's motion. It has become a more efficient and mature method in path planning. The basic idea of path planning using artificial potential field method is to simulate the environmental information of mobile robot into virtual potential field, in which the gravitational potential field U_{att} acts on the robot by the target point, and the repulsive potential field U_{rep} is generated by all obstacles. The two potential fields work together to form a compound artificial potential field U. Under the action of gravitational field, the robot moves continuously to the target point. At the same time, the repulsion field generated by the obstacles in the environment acts on the robot to avoid the surrounding obstacles, and the most

mobile robot can safely reach the target point around the obstacles.

The realization of potential field method is as follows. The target point, obstacle and robot are simplified as particles. The robot is subjected to both gravitation from the target point F_{att} and repulsion F_{rep} from one or more obstacles. The artificial potential field function includes gravitational field function and repulsion field function. The specific model is as follows:

$$U_{att}(X) = \frac{1}{2} K_a |X - X_g|^2 \quad (1)$$

The gravitational force from the object point acts on robot in potential field. Potential function being used in the algorithm is shown in Eq. (1), the gravitational coefficient is the coordinate of the robot (x, y), the coordinate of the target point (x_g, y_g), and $|X\text{-}X_g|$ is the relative distance between the robot and the target point, where gravitational force is negative gradient of its potential function, the direction of which is pointing at the object point with corresponding gravitational force as:

$$F_{att}(X) = K_a \left| X - X_g \right| \quad (2)$$

The repulsive force from obstacles acts on robot in virtual force field. Potential function used in the algorithm is:

$$U_{rep}(X) = \begin{cases} \frac{1}{2} K_r (\frac{1}{d} - \frac{1}{d_0})^2, & d \le d_0, \\ 0, & d > d_0. \end{cases} \quad (3)$$

where K_r is the repulsion coefficient, d is the repulsion influence distance of the obstacle and the distance between the target and the obstacle. The repulsive force is negative gradient of its potential function as:

$$F_{rep}(X) = \begin{cases} K_r (\frac{1}{d} - \frac{1}{d_0}) \frac{1}{d^2}, & d \le d_0, \\ 0, & d > d_0. \end{cases} \quad (4)$$

In conclusion, the resultant force of the robot is $F_{total} = F_{att} + F_{rep}$, and the direction of the resultant force determines the motion of the robot.

2.2 Limitations of traditional artificial potential field method

Artificial potential field method is a very efficient local path planning algorithm, but there are some problems in the complex environment. From the above Eqs. (2), (4) show that the gravitation will be smaller in pace with the distance between the obstacle and robot which becomes smaller as the repulsion is contrary. This process does not take into account the information of the distribution of local obstacles. Although the artificial potential field method is fast and effective, it also has the following limitations:

1) When the target is near an obstacle, the target cannot be reached. This problem is also known as GNRON [6].

2) There is a local minimum trap when the force equilibrium point exists.

3) Not suitable for dynamic environment.

3. AN IMPROVED METHOD OF APF METHOD

3.1 Solutions for GNRON problem

The problem of GNRON is that the obstacle is near the target point, which causes the target point is no longer the minimum potential energy point in the whole situation, and finally the robot cannot move to the target point correctly. For the GNRON problem, many experts and scholars have put forward many kinds of methods through continuous research, most of which are to improve the traditional repulsion function, and add the following factors to the repulsive function based on the distance term between robot and target point. The root cause of GNRON problem is that the minimum value of the total potential field function is not at the target point. In order to solve this problem, a new attraction field function is constructed. The improved function of gravitational potential field is as follows.

$$U_{att}(X) = K_a \left| X - X_g \right|^{n_1} + K_a (\frac{1}{a_0^{n_2}} - \frac{1}{(a_0 + \left| X - X_g \right|)^{n_2}}) \quad (5)$$

By calculating the negative gradient of the improved gravitational potential field function in the original gravitational direction, $\boldsymbol{F}_{att}$ can be obtained.

$$F_{att}(X) = n_1 K_a \left| X - X_g \right|^{n_1 - 1} + \frac{K_a n_2}{(a_0 + \left| X - X_g \right|)^{n_2 + 1}} \quad (6)$$

where the parameters of n_1, n_2, and a_0 in the formulas above range from normal numbers larger than 0, and $a_0 < |X\text{-}X_g|$.

When the robot reaches the target point, the distance between the robot and the target point is zero, and the latter one is zero. Therefore, no additional force will be produced and the results will not be affected. Compared with the traditional gravitational function, the new function maintains a strong gravitational force at a distance, and significantly reduces the gravitational force near the target location. The gradient of potential energy and the attraction near the target increase. The combined potential field function of gravity and repulsion is:

$$U_{total}(X) = U_{rep}(X) + U_{att}(X) \quad (7)$$

where $U_{rep}(X)$ represents the repulsion potential field intensity generated by obstacles, $U_{att}(X)$ represents the

potential field intensity generated by improved target points, and $U_{total}(X)$ represents the total potential field intensity.

In the improved combined potential function, at the target position x=0, that is, the robot's target position, the potential function has the minimum value. So the robot can reach the target position.

3.2 Local minimum solution

The local minimum problem is that the robot is forced to balance at this point, so the robot stagnates or lingers at the minimum point instead of moving towards the target point.

In order to solve the local minimum problem, a method of creating virtual target point near the minimum value to generate additional potential field is proposed. The improved algorithm mainly includes two parts: detection of minima and virtual establishment of target point. In this paper, we propose a method to determine the local minimum: when the distance between step n_1 and step n_2 is d, and the range of d is 0 to m(m is a non negative number tending to zero infinitely, and the difference between n_1 and n_2 is about 3 to 10), we consider that the robot is trapped in the local minimum, then we adopt the strategy of establishing a virtual target point.

The selection method of virtual target point is shown in fig.1. When falling into the local minimum, first find the nearest obstacle coordinate to the robot, and then find a point as the virtual target coordinate with the line L_2 which is perpendicular to the line L_1 between the two. When the robot falls into the local minimum point, the force of attraction and repulsion is zero. When the virtual target point is added, the force of attraction and repulsion is not zero. In this way, the robot moves to the original target point under the joint action of the virtual target point and the original target point. When the robot gets rid of the local minimum point and cancels the virtual target point, the robot moves towards the target point under the joint action of the original target point and the obstacle.

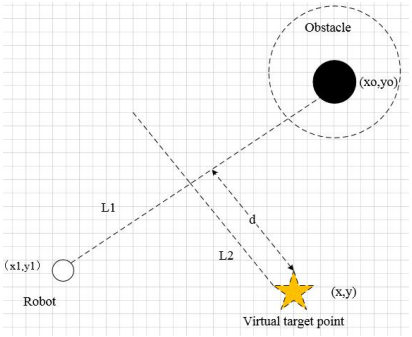


Fig. 1 Schematic diagram of virtual target point

From above diagrams, two equations can be established through the mathematical relationship between them to solve the coordinates of virtual objects. In this paper, the position of the virtual target point is obtained by Eq.7. Among them, (x_1, y_1) is the current position coordinate of robot, (x_0, y_0) is the position coordinate of obstacle that robot receives the maximum repulsion force, (x, y) is the position coordinate of virtual target traction point, and d is optional constant.

$$\begin{cases} x = x_1 - d \\ y = \dfrac{d(x_0 - x_1)}{y_0 - y_1} + y_1 \end{cases} \tag{8}$$

In this paper, gaussian function is selected as the gravitational potential energy function of the virtual target. The expression of the virtual target potential function is as follows:

$$U_{\text{add}}(X) = \begin{cases} \dfrac{K_c}{\sigma} e^{-\frac{(d-d_s)^2}{2\sigma^2}} - \dfrac{K_c}{\sigma} e^{-\frac{d_s^2}{2\sigma^2}}, & d \le 2d_s, \\ 0 & ,d > 2d_s. \end{cases} \tag{9}$$

where K_c, σ and d_s are constants. According to the characteristics of Gauss function, σ determines the sharpness of the curve, that is, the speed of the virtual gravitational potential energy descending, and d determines the offset of the curve, and the gravitational potential energy function reaches the maximum value at the point with the distance of d_s. If the distance between the local minimum point and the virtual target point is d_s, then the maximum value of the potential energy function is taken, because the maximum value of the potential energy function should be taken near the local minimum point, so as to quickly escape from the minimum point, and then quickly return to the target point. According to the nature of gravitational potential energy function, in order to make the robot escape from the local minimum point smoothly and ensure that it will not deviate from the original target point, then the potential energy at the virtual target point should be zero, which can ensure that the virtual target point will not affect the robot to move forward to the original target point.

3.3 Solution of robot dynamic obstacle avoidance

In order to improve the dynamic environment planning ability of the traditional artificial potential field method and improve the obstacle avoidance ability of the robot to adapt to the actual complex environment, the speed of the obstacle is taken into account and the speed repulsion field is established. The velocity repulsion field function is:

$$U_{\text{repv}}(i) = K_{rv} \left| V_o - V_r \right| \sin\boldsymbol{\theta} = K_{rv} V_{or} \sin\boldsymbol{\theta} \tag{10}$$

where V_{or} represents the velocity of the obstacle relative to

the robot, θ represents the angle between the position vector of the robot relative to the obstacle and the relative velocity vector, V_o and V_{or} represents the velocity vector of the obstacle and the robot respectively. Then, the repulsion field function generated by the relative position between the original robot and the obstacle can be expressed as follows:

$$U_{rep}(i)=\begin{cases}\frac{1}{2}\eta(\frac{1}{X-X_{oi}}-\frac{1}{\rho_o})^2+K_{rv}V_{or}\sin\theta, & |X-X_{oi}|<\rho_o,\\ 0, & |X-X_{oi}|\geq\rho_o.\end{cases} \quad (11)$$

There are two components in the modified repulsion function, one of which is affected by distance and the other by velocity. The two components are adjusted in real time by their respective gain coefficients.

Similarly, the relative velocity repulsive force is obtained by taking the negative gradient of the relative velocity repulsive force.

$$F_{repv}(i)=-\nabla U_{repv}(i)=K_{rv}\left|V_{or}\sin\theta\right| \quad (12)$$

The repulsion force direction is perpendicular to the relative velocity vector of the robot relative to the obstacle and away from the obstacle.

The resultant repulsion force of the robot in the repulsion field is as follows:

$$F_{total}(i)=\sum_{i}^{m}F_{rep}(i)+\sum_{i}^{n}(F_{repv}(i)+F_{rep}(i)) \quad (13)$$

where m and n represent the number of static and dynamic obstacles in the environment respectively.

4. The OPTIMIZATION OF GENETIC ALGORITHM

In order to reduce the path length of the mobile robot and improve the ride stability of the mobile robot, genetic algorithm is used to optimize the relevant parameters of the artificial potential field. The parameters to be optimized mainly include step length L and θ direction angle of movement. Using the biological evolution mechanism of genetic algorithm to optimize these two parameters can improve the effect of path planning.

In this paper, the strength of the total potential field received by the robot in the potential field is selected as the fitness function, and the specific steps of optimizing the parameters introduced into the improved artificial potential field method by using genetic algorithm are as follows.

Step 1: Set the initialization conditions. The values of maxgen, popsize, crossover probability Pc and mutation probability P_m of genetic algorithm were set in detail.

Step 2: Determine the encoding mode and length of parameters. In this paper, the two parameters in this paper are encoded with 7-digit binary characters, and the total length of chromosome is 14.

Step 3: Initialize the population. A series of individuals are randomly generated according to the set population size, denoted as $E(i)$, where $i=1,2.....$.

Step 4: Calculate the fitness function value fitness(i) of each individual in the population. In this paper, the strength of the potential field in the robot is chosen as the fitness function. The specific fitness function formula is as follows:

$$fitness\ (i)=\frac{1}{[U_{att}(X)+U_{add}(X)+i=\sum_{i=1}^{n}U_{rep}(X_i)]} \quad (14)$$

In the above formula (14): $U_{att}(X)$ is the strength of gravitational potential field; $U_{add}(X)$ is the strength of additional potential field generated by virtual target point; $U_{rep}(X_i)$ is the strength of repulsive potential field of the ith obstacle; n is the number of obstacles in the environment.

Step 5: Select individuals by fitness proportional selection method. According to fitness (i) of individual i and the proportion $fitness\ (i)/\sum fitness\ (i)$ of total fitness $\sum \text{fitness}\ (i)$, determine the probability of the individual being selected for replication.

Step 6: The next generation population GApopi(K +1) is generated by genetic operation of population GApopi(K +1) using single-point crossover and single-point mutation operators.

Step7: repeat steps 4 to 6 until the parameters do not change for 10 generations or reach the maximum number of iterations set in advance.

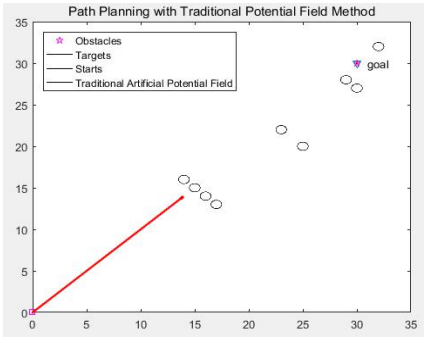
5. EXPERIMENTAL SIMULATION AND ANALYSIS

5.1 Static obstacle avoidance simulation analysis

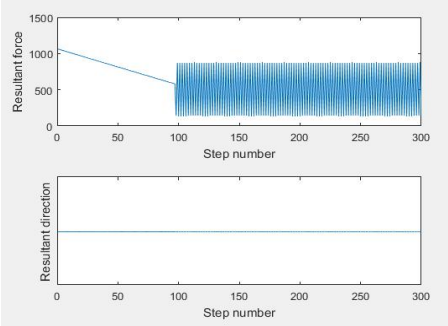
In the static environment, the path planning simulation of the traditional artificial potential field method (APF), the general improved potential field method (IAPF) and the improved artificial potential field method (GA-APF) are compared in the case of target unreachable and local minimum.

In Fig. 2~3, when the local minimum appears, the traditional artificial potential field planning trajectory appears strong jitter near the obstacle. However, the improved algorithm in this paper not only greatly reduces the number of planning path steps, but also the path is relatively smooth.

As can be seen from the comparison between Fig.2 (b) and Fig.3 (b), in the traditional potential field method, the

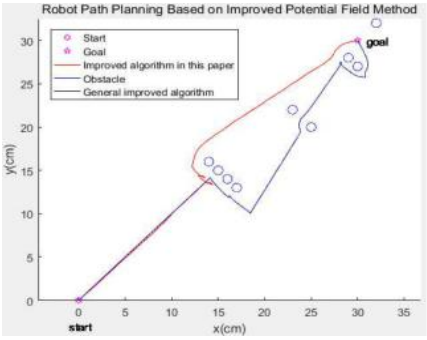


(a) The path planning of improved artificial potential field algorithm

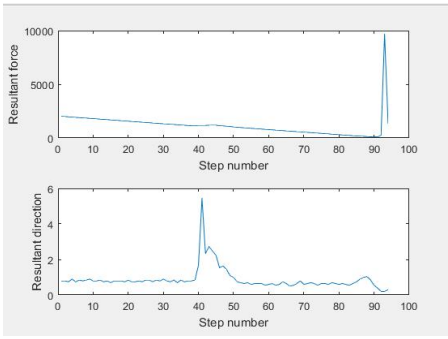


(b) The force size and direction of improved artificial potential field algorithm

Fig.2 The traditional artificial potential field method for path planning.



(a) The path planning of improved artificial potential field algorithm



(b) The force size and direction of improved artificial potential field algorithm

Fig. 3 Improved artificial potential field path planning

direction changes of resultant force subjected to the robot is more frequent and drastic, but in the improved algorithm, the direction change of resultant force is relatively smooth.

At the same time, table 1 shows the path length and iteration time of improved algorithm in this paper (GA-APF), the traditional potential field method (APF) and the general improved algorithm (IAPF) in the case of local minimum. It can be seen from the table that this method can not only make the robot walk out of the local minimum area smoothly, but also save the distance and time.

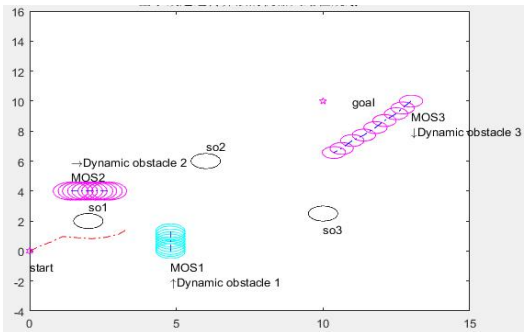
Table 1 Path planning with different methods

model	Length(m)	Iteration times
APF	---	---
IAPF	54.5363	274
GA-APF	45.5017	95

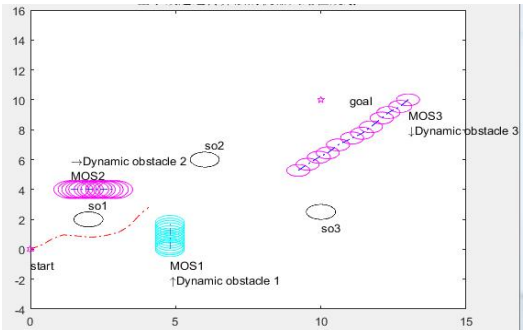
B. Dynamic obstacle avoidance simulation analysis

In the complex dynamic simulation environment shown in the figure below, the process of using the algorithm in this paper to carry out path planning for mobile robots to achieve obstacle avoidance is shown in Fig. 4.

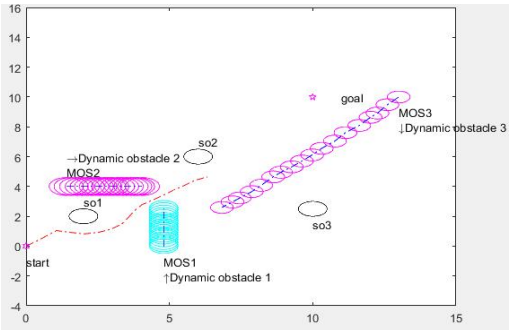
In Figs. 4, the motion directions of three obstacles are set, and the robot can successfully bypass the dynamic and static obstacles. The details are as follows: at t = 9 s, robots will be meet with the first dynamic content, and bypass its move on smoothly; at t = 12 s, into the influence of dynamic objects, the second in under the action of new repulsive force function smoothly around it; at t = 18 s, the speed of the robot head on corresponding to the third dynamic content, under the action of new repuls-



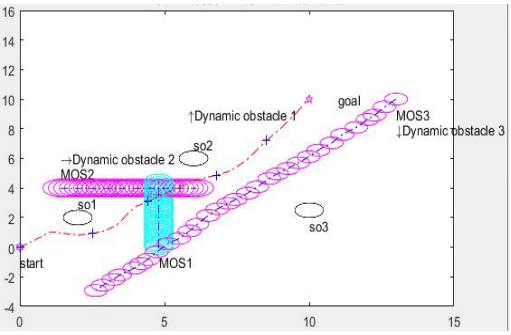
(a) t = 9s



(b) t =12s



(c) t =18s



(d) t =31s

Fig.4 Trajectory of robot at different times in dynamic environment

ive force function smoothly around it; at t = 18 s, the speed of the robot head on corresponding to the third dynamic content, under the action of new potential field function, changed the original course, avoid the collision. At t=31s, the robot has successfully reached the target point. The simulation results show that the improved algorithm can effectively solve the path planning problem of the robot in complex dynamic environment.

6. CONCLUSIONS

To solve the problem of traditional artificial potential field method is easy to fall into local minimum, target cannot reach the issue such as difference of path planning ability and dynamic environment, better adapt to the complex actual environment, this paper made the following several aspects to improve: firstly, change the traditional function model, gravity ensures that the gravity of the robot goal near the end of the dominant position, make the robot can be stable in target; Secondly, when the robot falls into the local minimum, a virtual target point is generated near the trap area, and gaussian function is selected as the virtual potential field function to help the robot escape from the current local minimum quickly. Then the repulsion potential field of the obstacle position is improved and the repulsion potential field of the obstacle speed is increased to avoid the dynamic obstacles. Finally, the genetic algorithm is used to optimize the improved potential field method to get the optimal path. The simulation results prove the feasibility of the improved method.

Acknowledgements

This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61472374, 61503053, 61603358.

References

[1] Zhu D, Yan M, Overview of Mobile Robot Path Planning Technology, Control and Decision-making, 25(07):961-967, 2010.

[2] Li H, Ma B, Chen H, Research on path planning based on artificial potential field method and invading weeds method, Control Engineering, 22(1): 38ł44, 2015.

[3] Borenstein J, Koren Y, Real-time Obstacle Avoidance for Fast Mobile Robots, IEEE Trans. Systems, Man, and Cybernetics, 19(5): 1179-1187, 1989.

[4] Han Y, Liu G, Mobile robot motion planning based on potential field in dynamic environment, Robot, 28(1): 45-49, 2006.

[5] Khatib O, Real-time obstacle avoidance for manipulators and mobilerobors, TheInternational Journal of Robotics Research,5(1): 90-98,1986.

[6] Ge S S, Cui Y J, New potential functions for mobile robot path planning, in IEEE Transactions on Robotics and Automation, 16(5): 615-620, 2000.

A Customer Experience Mapping for Business Innovation Case Description

Masaaki KUNIGAMI*, Takamasa KIKUCHI**, Hiroshi TAKAHASHI**, Takao TERANO***
* Department of Computer Science, School of Computing, Tokyo Institute of Technology
Yokohama, Kanagawa 226-0026, Japan
** Graduate School of Business Administration, Keio University
Yokohama, Kanagawa 223-8526, Japan
*** Chiba University of Commerce
Ichikawa, Chiba 272-8512, Japan

Abstract

We propose a mapping model for describing customer experience transition processes in business innovation cases known as System Experience Boundaries Map (SEBM). Different from other customer experience mapping methods, SEBM focuses on potential boundaries that restrict the customer experiences. SEBM also represents a customer side process of business innovation as a resolution of those restrictions. Used with the Managerial Decision-Making Description Model previously presented, SEBM describes value co-creation processes in actual business cases. We also consider SEBM application to facilitate and log-analysis on business gaming.

Keywords: User Experience Mapping Model, Formal Description Model, Business Case, Business Innovation, Co-Creation Process

1. INTRODUCTION

This paper presents the System Experience Boundary Map (SEBM) as an experience-mapping model for formally describing the changes in customer experience due to new products and services in the case of business innovation. By SEBM, we are going to improve the description about the business innovation in the customer side. Also, in combination with the Managerial Decision-Making Description Model (MDDM) [1], [2], [3], SEBM will contribute to describe the case for business innovation as a process of interaction on both the firm and customer sides.

Here, business innovation means that on the customer side, new products and services are being introduced with qualitative changes in the customer's experience in usage and consumption. SEBM makes it possible to compare the cases of business innovation from the customer side by describing the changes in the customer experience in the business case formally. It also provides a common way to visualize the process of customer side change in business gaming, similar to the business case.

In this paper, "customer experience" means a sequence that is segmented into several stages of a script of usage and consumption behaviors of a customer regarding a certain good or service. On the other hands, "qualitative change of the customer experience" means that a new good or a service resolve the existing limitations / boundary of customer's behavior at a certain stage of the customer experience and redefine customer's script of behavior.

SEBM is close to the user experience mapping method ([4], [5]) that pays attention to changes in the user interface when compared with the methods such as User Experience Map [6], Customer Journey Map [7], and Mental Model Diagram [8]. Meanwhile, SEBM is not oriented toward designing services and user interfaces based on the desired user experience. Also different from the conventional method, SEBM extracts limits and restrictions of the customer behavior from the existing user experience and focuses on the change of the customer behavior due to the introduction of new goods and services. SEBM enables us to visualize the business innovation case including before and after situations and innovative factors from the customer side.

SEBM describes the case of business innovation on the customer side by focusing on the behavior change of the customer. MDDM describes the case of business innovation on the enterprise side as a change in the combination of management objectives and management resources. By combining them, it is possible to express them in parallel with the process of business structure change on the company side. This makes it possible to describe value co-creation [9] in service science as a process of interaction between changes in the business structure of a company and behavior changes on the customer side.

2. METHODOLOGIES

SEBM is a novel table form model that describes the customer's liberation from bounded experience for existing goods and services due to the emergence of goods and services (Fig. 1). SEBM consists of the following three-step procedure: 1) decomposing the characteristics of the experience into two axes of stage and phase (Stage-Aspect Decomposition), 2) extracting each decomposed stage, and potential limits/constraints of the experience in each phase (Boundary Extraction), and 3) describing new behavior by solutions to the limits/constraints of the experience (Solution and Transition).

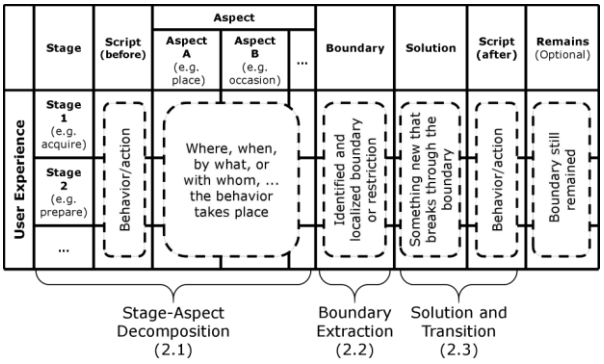


Fig.1 System experience boundary map (SEBM).

SEBM describes the change of customer experience as a table decomposed from a certain viewpoint by these procedures. This experience map enables us to describe when, where, and what kind of constraint exists for a customer, and what kind of experience is generated by eliminating the constraint.

2.1 Stage-Aspect Decomposition

First, the characteristics of the customer experience of goods and services are decomposed into two axes: the Stage and Aspect of realizing the experience. Here the Stage is a division according to the order in which the experience is realized, such as "Acquire," "Prepare," "Enjoy," "Keep," etc. The Aspect is a classification according to the nature of the experience, such as "Occurrence," "Place," "Device," "Subject/Object," etc. that the experience is realized for each stage. Each category is selected from the perspective of the map author based on the contents of the case.

Stage-Aspect Decomposition (Fig. 1, left) arranges the Stages in rows from top to bottom in the order of realization, and columns as selected columns. We decompose the characteristics of the script that represents the experience of the customer on existing goods and services into a table format.

2.2 Boundary Extraction

Regarding the characteristics of the experience decomposed at the "Stage," the boundaries/constraints under the existing goods/services are described based on the contents of the case and the viewpoint of the map author. This makes it possible to explicitly describe which limit/constraint is latent in the consumption behavior of existing goods/services related to the characteristics of the stage or aspect of the experience. (Fig. 1, middle)

In SEBM, since the experience is decomposed in terms of stages and phases when there are multiple limits / constraints, it is possible to identify and draw the potential places of each limit/constraint. Or, even when the viewpoints of the case writer and map writer are different, those differences can be drawn separately.

2.3 Solution and Transition

This describes the changes (relaxation of the limits/constraints) of new experiences for the potential/constraint of the latent experience under the specified existing goods/services. Furthermore, the new experience made possible by it is described as a script (Fig. 1, right).

Here also, even when there are multiple relaxations of the limits/constraints, or when the viewpoints of the case writer and map writer differ, it is possible to specify and draw the changing points.

Additionally, even if the solution is given to only a part of a plurality of potential limits/constraints, the potential places of the remaining limits/constraints are specified in the same format. For this reason, it is possible to describe and compare innovation chains by making multiple SEBMs continuous.

2.4 Connection with MDDM

When combined, SEBM and MDDM are able to describe innovation and value co-creation by customers and companies formally.

Business innovation is described in SEBM as a change to a new script by relaxing the potential limits of the customer experience, whereas in MDDM, it is described as a change in the combination of purpose and means in a company's hierarchical business structure (Fig. 2).

Therefore, SEBM and MDDM have a complementary relationship as a case description model for business innovation. This is not just a joint description, but a bilateral description of the interaction process between the changing company's business and the changing customer's experiences. Depending on the content of the case, it has the potential as a formal description model of

value co-creation between companies and customers.

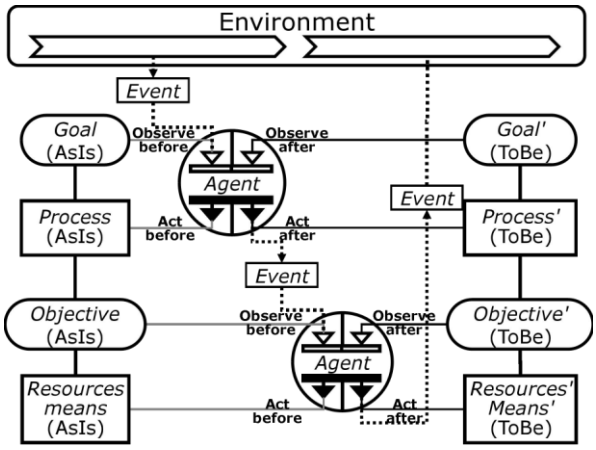


Fig. 2 Managerial decision-making description model (MDDM).

When connecting MDDM and SEBM, the customer's experience and its changes are regarded as changes in the external environment for the company. Then, we take an approach to refine some of the environmental components of MDDM with SEBM. In MDDM, the customer's consumption behavior is described in the environmental component as part of the market landscape. The interaction between the market and management decision-making is expressed by the connection of events from the environmental component to the agent, which is the trigger of decision-making, or the connection of the event to the environmental component that changes from the decision of the agent. Therefore, changes in the customer's experience of making decisions within MDDM are described by replacing some of the environmental components with SEBM (Fig. 3).

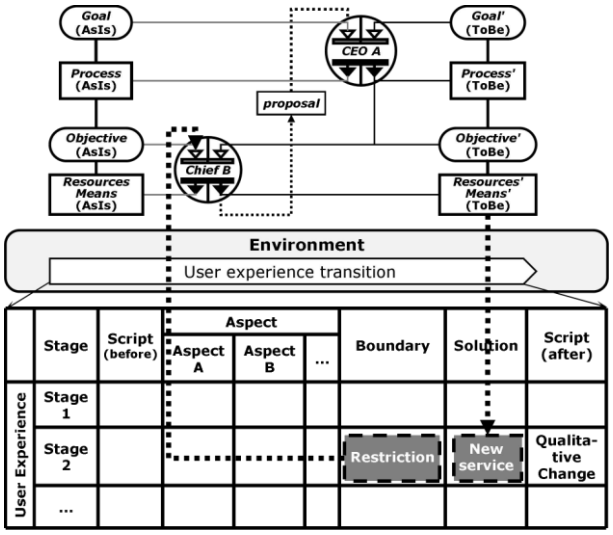


Fig. 3 Description of business innovation by connecting MDDM (top) and SEBM (bottom). The thick dotted lines represent the interactions that occur between corporate decision-making and changes in the customer experience. There are three possible patterns for this connection:

1) To consider the limitations/constraints of the experience at a particular stage/aspect in SEBM as a trigger event for MDDM management decision-making.
2) To consider new goods and services resulting from MDDM's management decision-making as an event and use them as solutions to the limitations/constraints of the experience in SEBM.
3) To grasp the change in experience due to the relaxation of the limit in SEBM as a change in the market collectively and connect it to the trigger event of the decision-making of the agent on the MDDM side.

This connection between SEBM and MDDM suggests that value co-creation could be formalized. For example, when the above-mentioned connections are bidirectionally included in the formal description of a business case, it is an expression of the interaction between changes in the business structure of a company and changes in the behavior of customers.

3. APPLICATION TO ACTUAL BUSINESS CASES

The case of actual business innovation is expressed using SEBM. As a basic example, we will use the same sample case from MDDM [2] and show that this can describe the change of customer experience on the customer side.

We also show that by combining customer experience changes by SEBM and management decision-making by MDDM, the actual business case can be formalized from both the customer and management sides.

3.1 Applying SEBM to Actual Business Cases

For the example, we show SEBM's description of the case of Honda's entry into the North American motorcycle market. This is a classic case in Christensen's work [10], but the information from Honda's website ([11], [12]) is also taken into account to describe from the customer's perspective.

The outline of the case [10] by Christensen is as follows:
・Honda was looking to enter the market for large high-speed highway bikes like Harley Davidson which was popular in the North American market with the advantage of low cost.
・While Kawashima, the manager of American Honda Motors at that time, was developing a local dealer, Honda's large motorcycle was not accepted in the North American market.
・Whereas the Super Cub, which he brought in for business and ran off-road distractions, aroused consumer interest and demand, so Kawashima convinced the Tokyo head office to introduce a lightweight recreational bike to

the North American market.

・Honda changed its strategy and created a new market in North America.

We describe in SEBM the changes in consumer experience before and after the introduction of the Super Cub (Fig. 4). For simplicity, the experience stage is divided into two parts: “purchase” and “usage,” and aspect is divided into “place” and “atmosphere.” According to Honda’s website ([11], [12]), the limitations/constraints of the potential North American customer’s experience were that the large bikes of the time had a very bad impression from society as a defective vehicle with a leather jacket, covered with oil. A store like a garage wasn't the place that the general public would want to visit. Small products other than large bikes for highways were undeveloped (Fig. 4, Boundary column).

In a market with such potential limitations/constraints, the Super Cub gained popularity due to its toy-like ease. Some customers took the Super Cub to a camp or picnic with a pickup truck and ran through the forest. Such usage was not expected at the Japanese head office. Kawashima focused on this and reported it to the head office, and expanded the sales channels to be an inexpensive lightweight recreational bike other than specialized dealers. He launched the “Nicest people on Honda” campaign. As a result, he was able to offer the Super Cub as a vehicle for citizens, such as students and women, different from the conventional motorcycle customers (Fig. 4, Solution column).

	Stage	Script (before)	Aspect		Boundary	Solution	Script (after)
			Place	Atmos-phere			
User Experience	Buy	Go to dealer shop and buy	Dealer -ship	Dirty garage	Hesitation in visiting Expensive	Outdoor outfitters & sports shops Low price	Buy at sports shop Present for teens at Christmas
	Use / ride	Wear black leather jackets Ride motor-cycle on highway	High-way	Outlaw culture High speed cruising	Not any good reputation Highway riding only	Silent engine Ad “Nicest people” Anybody could ride Toy-like familiarity	Go to school Go shopping with skirt Bring to camping, hunting , fishing

Fig. 4 SEBM on customer experience in the case of Honda’s entry into the North American market.

As a result, at the purchase stage, customer behaviors such as easily purchasing Super Cubs at sports equipment stores and making them a Christmas present for teenagers were added. At the use stage, customer behaviors such as running off-roads, and students and women moving casually and fashionably when going to school or going out were added (Fig. 4 Script (after) column).

3.2 Combination with MDDM

Regarding the description of the case including the interaction (value co-creation) between the customer and company sides by the combination of the customer side experience (SEBM) described in Section 2.4 and the management side decision-making (MDDM), we show the cases of Honda and Sony.

Figure 5 shows the combination of SEBM and MDDM in the Honda case. The unexpected use of off-road riding by some customers caused a bottom-up decision to open up the market for lightweight recreational bikes in North America (thick dotted line from bottom to middle of the figure). It also shows that a new experience of daily movement of students and women occurred due to Honda’s strategy shift (thick dotted line from middle to bottom of the figure).

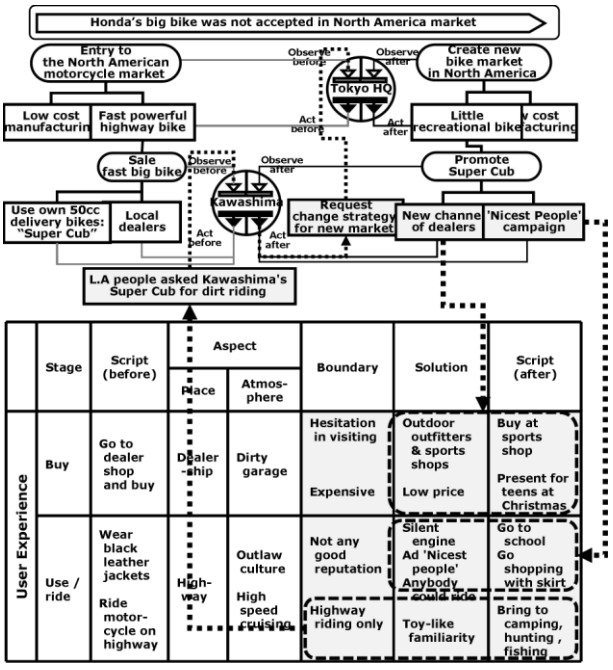
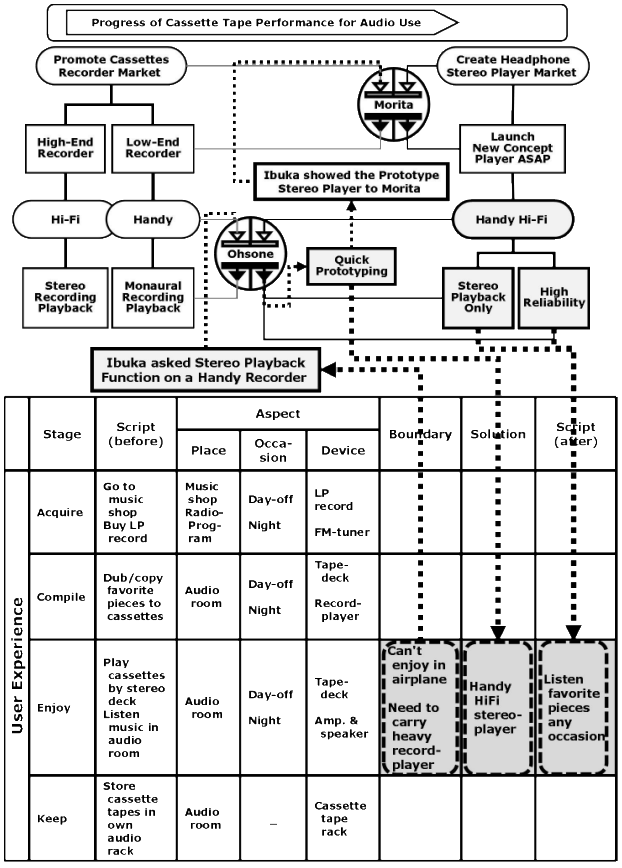


Fig. 5 Combination of MDDM and SEBM in Honda’s case. The thick dotted line represents the interaction between corporate decision-making and customer experience.

As another example, the case of market creation by Sony’s headphone stereo (Walkman) is also described by MDDM and SEBM. The following is an outline of the case based on Sony’s website ([13], [14]). In the late 1970s, Sony was improving the performance of compact cassette tapes and their recorders. Ibuka, the honorary chairman at that time, personally requested Ozone, the director of the tape recorder division, to create a lightweight and compact playback device for listening to music on an airplane. Ozone remodeled a handy type monaural recording/playback machine and prototyped a stereo playback only machine. When Ibuka tried to test this prototype with Morita, who was chairman, Morita

decided to introduce this prototype as a Walkman into the market for young people. As a result, Sony created a new market.

The limit of the experience in this Sony Walkman case was the constraint of the opportunity to enjoy music. Despite the improvements in performance of compact cassette tapes and recording/playback equipment, it became possible to freely record music on compact cassette tapes, but the preconception existed that large and heavy recording/playback equipment must be used for playback.



	Stage	Script (before)	Aspect: Place	Aspect: Occa-sion	Aspect: Device	Boundary	Solution	Script (after)
User Experience	Acquire	Go to music shop Buy LP record	Music shop Radio-Prog-ram	Day-off Night	LP record FM-tuner			
	Compile	Dub/copy favorite pieces to cassettes	Audio room	Day-off Night	Tape-deck Record-player			
	Enjoy	Play cassettes by stereo deck Listen music in audio room	Audio room	Day-off Night	Tape-deck Amp. & speaker	Can't enjoy in airplane Need to carry heavy record-player	Handy HiFi stereo-player	Listen favorite pieces any occasion
	Keep	Store cassette tapes in own audio rack	Audio room	–	Cassette tape rack			

Fig. 6 Combination of MDDM and SEBM in Sony's case. The description of the unchanged part on the right side of SEBM is omitted.

The stereo reproduction-only machine had this limit and solution discovered by the cooperation of Ibuka as a zero-order user and Ozone as a technician. Furthermore, with Ibuka as a user, Morita as a manager, and Ozone as a technician, the Walkman was introduced to the market as a product, enabling a new experience of bringing music (Fig. 6).

As we have seen, describing cases by connecting SEBM and MDDM allows the following two points to be described formally:
1) discovering the solution of the potential experience limit on the customer side by interaction from both the customer and company sides, and 2) creating a new experience.

4. CONSIDERATIONS

Regarding connecting SEBM and MDDM, we consider the possibility of application in business case learning and business gaming learning from the viewpoint of facilitation support and learning log-analysis.

In learning with business case or gaming about business innovation, it is desirable to consider both the organizational decision-making of a company and the change of ownership value, use value, or consumption behavior change in the customer from both sides. However, the content of the teaching material and viewpoint of the learner do not necessarily include both in a well-balanced manner. In such a case, to promptly consider the experience of the customer about teaching materials and learners who tend to lean toward the decision-making process inside the company, conversely, the process of implementation within the company when tending to the marketing and product design tend to occur. To speed up the discussion, preparing SEBM, MDDM, or both may be effective in reducing the load on the facilitator in correcting bias and variations in learning.

In business case learning or gaming, it is difficult to record the log in a form that can be compared between learners. However, if it becomes easy to formally describe the changes in the business structure and customer experience, it will be possible to look back on the learners at the time of after-review and to compare them with the records of past learners. The description including useful findings in such inter-comparison may be a stylized fact extracted from the case or gaming.

5. SUMMARY AND REMARKS

This paper has proposed SEBM that formally describes the changes in the customer experience due to new products and services in the case of business innovation. SEBM can describe the process of business structure change on the enterprise side and the change of customer experience in parallel by combining with MDDM. In addition, it is possible to describe the value co-creation formally as a process of the total action.

SEBM is considered to be promising for application in business case learning and business gaming learning, and we would like to deepen its investigation in the future.

ACKNOWLEDGMENTS

This work is supported in part by the Grant of Foundation for the Fusion Of Science and Technology. The authors would like to thank Enago (www.enago.jp) for the English language proof.

References

[1] M. Kunigami, T. Kikuchi, T. Terano, A Formal Model of Managerial Decision Making for Business Case Description, GEAR2018 Letters of the Special Discussion on Evolutionary Computation and Artificial Intelligence, 2018.

[2] M. Kunigami, T. Kikuchi, T. Terano, A Formal Description Model for Business Innovation Case, JSAI Special Interest Group on Business Informatics, 2018.

[3] T. Kikuchi, M. Kunigami, H. Takahashi, M. Toriyama, T. Terano, "Description of decision making process in actual business cases and virtual cases from organizational agent model using managerial decision-making description model", Inform. Proc. Soc. Jap. Vol. 60, No.10, 2019, pp. 1704–1718.

[4] J. Kalbach, Mapping Experiences: A Complete Guide to Creating Value through Journeys, Blueprints, and Diagrams (1st Ed), O'Reilly Media, 2016.

[5] S. Gibbons, UX Mapping Methods Compared: A Cheat Sheet, Nielsen Norman Group, 2017, available: https://www.nngroup.com/articles/ux-mapping-cheat-sheet/

[6] Adaptive Path: Adaptive Path's Guide to Experience Mapping, Adaptive Path, 2013, available: https://s3-us-west-1.amazonaws.com/adaptivepath/apguide/download/Adaptive_Paths_Guide_to_Experience_Mapping.pdf

[7] G. Bernard, P. Andritsos, "A Process Mining Based Model for Customer Journey Mapping", Proceedings 29th International Conference on Advanced Information Systems Engineering (CAiSE 2017), Vol.1848, 2017 pp.49-56. available: https://serval.unil.ch/en/notice/serval:BIB_6FD456C4AB58

[8] I. Young, K. Mangalam, "Launching Problem Space Research in the Frenzy of Software Production", Interactions, Vol.25, No.1, 2017, pp.66-69. available: https://dl.acm.org/doi/abs/10.1145/3159449

[9] F. Adrian, A.F. Payne, K. Storbacka, P. Frow, "Managing the co-creation of value", Journal of the Academy of Marketing Science. Vol.36, 2008, pp.83-96.

[10] C. M. Christensen, The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail, Harvard Business Review Press, 1997(reprint 2016).

[11] Honda Official Website: 50years-history, Establishing American Honda, 1959, available: https://www.honda.co.jp/50years-history/challenge/1959establishingamericanhonda/index.html

[12] Honda Official Website: The Super Cub's Overseas Advance, available: https://global.honda/products/motorcycles/supercub-anniv/story/vol3.html

[13] Sony Official website: Sony History, Chapter 5 Prompting Compact Cassettes Worldwide, available: https://www.sony.net/SonyInfo/CorporateInfo/History/SonyHistory/2-05.html

[14] Sony Official website: Sony History, Chapter 6 Just Try It, available: https://www.sony.net/SonyInfo/CorporateInfo/History/SonyHistory/2-06.html

Multi-Feature Fusion Based Deep Forest for Hyperspectral Image Classification

Peng Liu[*, **], Xiao-Bo Liu[*, **], Zhi-Hua Cai[*,**], Yu-Lin Qiao[*, **]
* School of Automation, China University of Geosciences,Wuhan, 430074, China
** Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan 430074, China
*** School of Computer Science, China University of Geosciences, Wuhan 430074, China

Abstract

Multi-feature fusion is a useful way to improve the deep learning for hyperspectral image(HSI) classification. But the multi-feature fusion is usually at the decision level of classifier, which causing less link between features or poor extensibility of feature. In this paper, a multi-feature fusion based deepForest for HSI classification is proposed, which uses three deep multi-grained scanning branches in dgcForest to extract and fuse morphological features, saliency features, and edge features, and the fused features are sent into cascade forest in dgcForest for classification. Experimental results indicate that the proposed framework consumes less training time and has better performance on two HSI data sets.

Keywords: DeepForest, Hyperspectral Image Classification, Multi-Feature Fusion

1. INTRODUCTION

To make full use of information contained in hyperspectral data, multi-feature based strategy has been widely applied to current popular method. He et al.[1] proposed a multi-scale 3D deep convolutional neural network to jointly learn both 2D multi-scale spatial feature and 1D spectral feature from HSI data in an end-to-end approach, achieved better results with large-scale data set. However, DNN is hard to fuse multi-feature obtained by different method limited by the structure of neural network. Zhang et al.[2] and Jia et al.[3] chosen morphological profiles, Gabor textural, etc. and used classic classifier like the SVM to predict, improved classification accuracy of classic classifier. But those methods haven't fused features or just fused features on decision level, without considering the correlation among features. Thus there will be a certain potential for finding a better feature connector and classifier.

Deep forest is a new deep model of the alternative DNN proposed by Zhou et al.[4], which is a powerful opponent of DNN on the accuracy and time consuming. dgcForest is an improved version of deep forest, modified for HSI classification, which improved the ability of deep forest to extract deep feature by deepening the original multi-grained scanning forest[5]. Although dgcForest achieved a more satisfactory effect, but its ability to extract image feature in different HSI data sets needs to be improved.

Extended morphological profile(EMP) is a simple and commonly to use feature map that can denoise images[6]. Saliency detection can highlight the spatial scene features[7]. However, saliency detection will weaken the boundary information of each area in HSI. Edge detection can detect the edge feature of feature map, which makes up for the deficiency of saliency detection. Thus the combination of three features has certain rationality.

In this paper, we proposed a multi-feature fusion based deepforest method for hyperspectral image classification. The main contributions of this paper are summarized as follows.

1) Three deep multi-grained scanning branches in dgcForest were used to deeply extract EMP features, saliency features and edge features, which can supplement each other, to make full use of spatial information of HSI.

2) In order to enhanced features, voting fusion method was used to fuse three deeply extracted features, which can make the link between features closer than decision-level fusion method.

2. PROPOSED METHOD

Before using extended morphological profile(EMP), Principal Component Analysis(PCA) needs to be used to reduce the dimension of raw HSI, as well as to reduce redundancy among bands. HSI often contains background information that does not need to be classified. When spatial information is used for classification, the classification accuracy is easily affected by background information. Boolean Map Saliency Detection(BMS) is a simple and efficient visual saliency detection method, which can be used to highlight features of the foreground. We use BMS on EMP features to obtain enhanced features. Saliency detection is hard to

detect a clear outline of object, the edge feature of the ground object will be weakened. Thus an edge detection method called canny edge detection is introduced to detect the edge of EMP feature map.

In the processes of deepforest based multi-feature fusion, three deep multi-grained scanning branches in dgcForest are used to further extract three different features and output two vector blocks with size of $w/2 \times w/2 \times c$, respectively. w is size of sliding window to obtain pixel block sample. c is dimension of vector obtain by each minimum unit in cascade forest[5]. The output vector is soft output and indicates strength of each feature, which can be fuzzy, posterior probabilities, certainty, or possibility values[8]. In our proposed framework, voting fusion strategy was used to fuse three features. Thus, a fused feature (FF) can be obtained by add three soft output (SO). $FF = [F_1, F_1, F_1, \ldots\ldots, F_n]$ and $SO^l = [S_1^l, S_2^l, S_3^l \ldots\ldots S_n^l]$ in which $F_i = S_i^{l1} + S_i^{l2} + S_i^{l3}$. F_i is each degree of support of fused feature and S_i^{l1} is each degree of support of each extracted soft output by three different extraction branches. $n = w/2 \times w/2 \times c$[5]. Them, fused features with constant dimension are obtained.

Finally, the fused vector was sent into cascade forest to obtain final prediction.

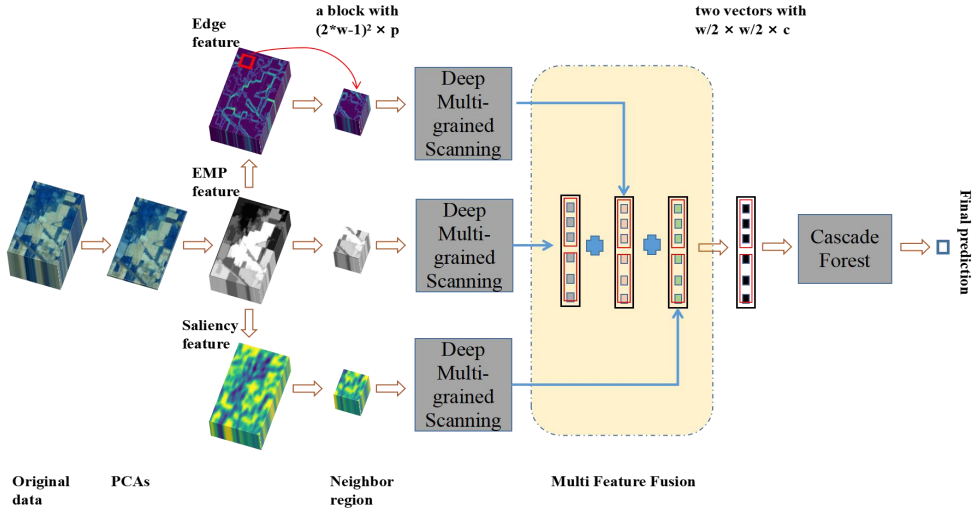


Fig. 1 The flowchart of our proposed method.

3. EXPERIMENTAL RESULTS

This subsection presents the classification accuracy of several states of the art methods and three cases that our method use different type of features. Two HSI data sets are used in experiment. And 10 percent of each data set was chosen as training data set.

Table 1 The detail of two data sets.

Dataset name	Classes	Pixels	bands	resolution	Train data scale
Indian Pine	16	145*145	200	20m	10%(each class)
Salinas	16	512*217	200	3.7m	10%(each class)

EMP, EMP+Saliency, MfdForest means that our algorithm uses different type of feature extraction branch. The accuracy and running time of different algorithms can be seen in Fig 2, Fig 3, Table 2, Table 3. Table 1 shows the detail of two data sets.

In Fig 2 and Fig 3, our algorithm is ahead of other methods in OA in two data sets. Although our method has a lower accuracy than ResNet when using less than three branches in Indian Pines data set, but three branches case still has a best performance. What's more, in the case of our method using different number of extraction branches, the accuracy was increased with more branches, it shows that each feature we chosen and feature fusion method we used has played a significant role. In Table 2 and Table 3, our method also has superiority in OA, AA and Kappa comparing to other methods, and AA has a subtle change with case of using two extraction branches in Indian Pines data set, it shows that saliency detection hasn't improve classification in some classes, which is the point we need to solve later.

In terms of running time, which is the total time of training and testing, although our improved algorithm is slower than dgcForest, the reduction in speed has not be presented as multiple. This effect is mainly achieved by our parallel computing branch. Therefore, there is still an advantage in running time.

4. CONCLUSION

In this paper, we introduce a multi-feature fusion based deep forest to extract and fuse three feature maps of the HSI for improving classification accuracy. For a certain pixel, even the classification improved by certain feature map is not good, but one of the other two features may

has a better effect, so it can work together for each pixel to get the best results. Although we use three branches to increase the size of the model, these three branches structures can be done in parallel, so as not to cause excessive speed reduction. The effectiveness of our method can also be seen from the experimental results.

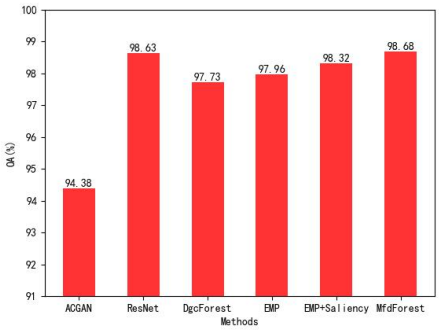


Fig. 2 OA of different methods in Indian Pines.

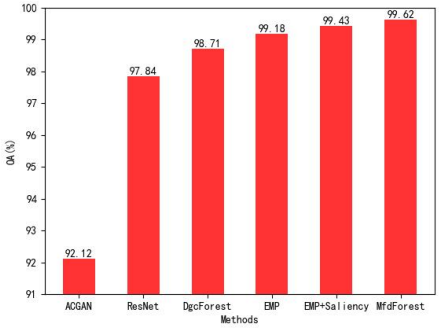


Fig. 3 OA of different methods in Salinas.

Table 2 Accuracy comparison and running time of different algorithms in Indian Pines data set

Method	ACGAN	ResNet	DgcForest	EMP	EMP+Saliency	MfdForest
OA	94.38±1.88	98.63±0.25	97.73±0.23	97.96±0.3	98.32±0.14	**98.68±0.15**
AA	76.29±3.28	90.33±0.29	95.92±0.51	94.62±2.62	94.59±2.02	**96.8±1.54**
K*100	93.58±2.15	98.44±0.28	97.41±0.18	96.23±0.77	97.37±1.01	**98.55±0.71**
Time(s)	836.15	923.48	**94.6**	114.37	139.93	152.11

Table 3 Accuracy comparison and running time of different algorithms in Salinas data set

Method	ACGAN	ResNet	DgcForest	EMP	EMP+Saliency	MfdForest
OA	92.12±1.42	97.84±0.12	98.71±0.11	99.18±0.22	99.43±0.28	**99.62±0.21**
AA	86.59±2.96	98.63±0.27	98.63±0.09	99.18±0.28	99.43±0.49	**99.62±0.32**
K*100	86.59±2.96	97.59±0.13	98.68±0.12	99.09±0.33	99.36±0.24	**99.57±0.15**
Time(s)	1020.11	1861.4	**368.9**	437.21	472.98	531.29

Acknowledgements

This work was supported by National Nature Science Foundation of China (Grant Nos. 61973285, 61873249, 61773355), and National Nature Science Foundation of Hubei Province (Grant No. 2018CFB528).

References

[1] M. He, B. Li, H. Chen, "Multi-scale 3D deep convolutional neural network for hyperspectral image classification," 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2018, pp. 3904-3908.

[2] C. Zhang, M. Han, M. Xu, "Multi-feature classification of hyperspectral image via probabilistic SVM and guided filter," 2018 IEEE International Joint Conference on Neural Networks (IJCNN). IEEE, 2018. doi: 10.1109/IJCNN.2018.8489452.

[3] S. Jia, J. Xian., "Multi-feature-based decision fusion framework for hyperspectral magery classification," 2018 IEEE International Geoscience and Remote Sensing Symposium(IGARSS). IEEE, 2018, pp. 5-8.

[4] Z. H. Zhou., J. Feng, "Deep forest: towards an alternative to deep neural networks," 2017 International Joint Conference on Artificial Intelligence Organization(IGCAI). 2017, pp. 3553-3559.

[5] X. Liu, R. Wang, Z. Cai, et al, "Deep multigrained cascade forest for hyperspectral image classification," 2019 IEEE Transactions on Geoscience and Remote Sensing(TGRS). 2019, val. 99, PP. 1-15.

[6] H. Luo, Y. Y. Tang, X. Yang, et al, "Autoencoder with extended morphological profile for hyperspectral image classification," 2017 IEEE International Conference on Cybernetics(CYBCONF). IEEE, 2017, pp. 293-296.

[7] J. Zhang, S. Sclaroff, "Saliency detection: a boolean map approach," 2013 IEEE International Conference on Computer Vision(ICCV). IEEE, 2013, pp. 153-160.

[8] G. Guo, D. Neagu, X. Huang, and Y. Bi, " An effective combination of multiple classifier for toxicity prediction," 2006 International Conference on Fuzzy System and Knowledge Discovery(FSKD). 2006, pp. 481-490.

Demagnetization Fault Diagnosis of Permanent Magnet Synchronous Motor Considering Inductance Disturbance

Fan Xiao*, Jing He*, Miao Y Zhang**(Corresponding Author)
* College of electrical and information engineering, Hunan University of Technology
Zhuzhou, Hunan, China
** College of railway transportation locomotive and vehicle college, Hunan Railway Professional Technology College, Zhuzhou, Hunan, China

Abstract

Aiming at the problem of demagnetization fault diagnosis of permanent magnet synchronous motor (PMSM) under the condition of inductance change, a demagnetization fault detection method based on adaptive sliding mode observer is proposed. Firstly, The mathematical model of demagnetization fault of PMSM in synchronous rotating coordinate system is established, and the inductance disturbance is analyzed separately. Then, considering the different characteristics of flux linkage fault and inductance disturbance, an adaptive sliding mode observer is proposed, and two different adaptive laws are designed to ensure the accuracy of fault diagnosis and to eliminating the influence caused by inductance disturbance, thus finally achieving the purpose of the robust diagnosis of demagnetization fault.

Keywords: PMSM, demagnetization fault, inductance disturbance, adaptive control, sliding mode control.

1. INTRODUCTION

Permanent magnet synchronous motor (PMSM) has the advantages of simple structure, high efficiency and high power density, and is widely used in machinery manufacturing equipment, industrial robots, computer peripherals, instrumentation, mini-cars and electric bicycles. Traditional AC servo system generally adopts PID control, which has the advantages of simple algorithm, high reliability and convenient adjustment. However, PMSM is a complex object with multivariable, strong coupling, nonlinearity and variable parameters. Although conventional PID control can meet the control requirements in a certain range, it is difficult to meet the requirements of high-performance control when the system parameters change or are affected by external uncertain factors [1].

In recent years, the research on the fault diagnosis of permanent magnet synchronous motor has been highly concerned and studied by scholars at home and abroad. Various disturbance factors that affect the stable operation of the motor will be produced during PMSM operation. The main disturbance factors are load disturbance, motor parameter disturbance and external nonlinear disturbance. The appearance of various disturbance factors will make the permanent magnet magnetic conductivity of permanent magnet motor drop or demagnetize, and the motor will not run stably, and even the motor will be scrapped in serious cases [2]. Therefore, it is particularly important to study the demagnetization fault detection of permanent magnet synchronous motor with disturbance factor. At present, the main fault diagnosis methods are sliding mode control, adaptive control and so on. Sliding mode control has good control performance for nonlinear systems, so it has been widely used in various industrial control objects. And, it is insensitive to the model error of the research object, the change of object parameters and external interference [3], which makes sliding mode control have very important theoretical research significance in the fault diagnosis problem of permanent magnet synchronous motor. Adaptive control can modify the control process according to the change of the dynamic characteristics of the controlled object and its disturbance. Using adaptive estimation algorithm [4], a specified performance index can reach and maintain optimal solution or approximate optimal solution.

At present, many literatures have studied various disturbance factors of permanent magnet synchronous motor [5], among which Cho Y [6] studied space vector control under load disturbance, Z Chang-fan [7] studied vector control under resistance disturbance, and Wu H [8] studied vector control under motor parameter disturbance. Aiming at the adaptive control method, Kim [9] proposed an adaptive speed control scheme under parameter variation. Nguyen A T [10] proposed a simple adaptive driving speed controller under parameter disturbance, Jing L [11] proposed an adaptive sliding mode observer under load and parameter disturbance to reduce speed fluctuation, and Jin-Woo Jung [12] proposed an adaptive

proportional-integral-derivative (PID) speed control scheme for PMSM.

In the above research literatures, either the inductance disturbance is not considered, or the inductance disturbance is assumed to be a constant value change, that is, the change rate of inductance derivative is approximately zero.However, in practice, the temperature or current of the motor changes greatly during the long-term movement, and the inductance changes in various processes. In this paper, aiming at the problem that the traditional sliding mode observer can not accurately detect the demagnetization fault when the inductance is variable and the change rate of inductance derivative is not 0 in PMSM operation, an adaptive sliding mode observer is proposed to detect the demagnetization fault. Considering the different characteristics of flux linkage fault and inductance disturbance, an adaptive sliding mode observer is proposed and two different adaptive laws are designed to ensure the accuracy of fault diagnosis and to eliminate the influence caused by inductance disturbance, thus achieving the purpose of demagnetization fault diagnosis.

2. DESCRIPTION OF THE PROBLEM

In magnetic field oriented coordinate d-q, the mathematical model [13] of demagnetization current of the PMSM is

$$\begin{cases} \dfrac{di_d}{dt} = \dfrac{u_d}{L} - \dfrac{R}{L} i_d + \omega i_q \\ \dfrac{di_q}{dt} = \dfrac{u_q}{L} - \dfrac{R}{L} i_q - \omega i_d - \omega \dfrac{\psi_f}{L} \end{cases} \tag{1}$$

where i_d, i_q are current of $d-q$ axis, u_d, u_q are voltage of $d-q$ axis, R is resistance, ω is angular velocity, L is inductance, ψ_f is the nominal value of flux linkage.

When the demagnetization fault occurs in the motor, the amplitude and direction of the permanent magnet flux vector will change.As shown in the Fig.1, the flux linkage vector of permanent magnet changes from ψ_f to ψ_r. There is a deviation angle γ between the magnetic field direction of the motor and the permanent magnet flux linkage direction. ψ_{rd} is flux linkage of d-axis, ψ_{rq} is flux linkage of q-axis.

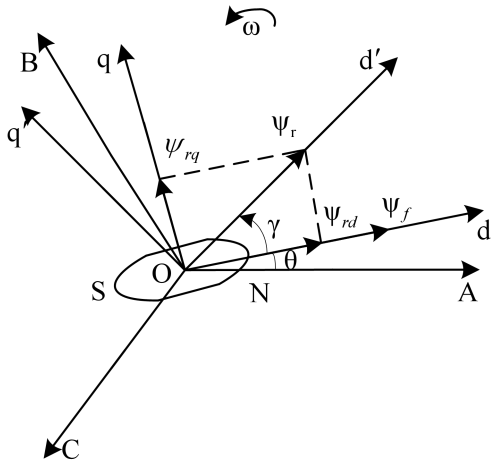


Fig.1 Change of flux linkage of PMSM

Under the inductance disturbance, the demagnetization mathematical model is given as follows

$$\begin{cases} \dfrac{di_d}{dt} = \dfrac{u_d}{L+\Delta L} - \dfrac{R}{L+\Delta L} i_d + \omega i_q + \omega \dfrac{\psi_{rq}}{L+\Delta L} \\ \dfrac{di_q}{dt} = \dfrac{u_q}{L+\Delta L} - \dfrac{R}{L+\Delta L} i_q - \omega i_d - \omega \dfrac{\psi_{rd}}{L+\Delta L} \end{cases} \tag{2}$$

where ΔL is the inductance disturbance, ψ_{rd} is flux linkage of d-axis, ψ_{rq} is flux linkage of q-axis.

We let $\dfrac{1}{L+\Delta L} = \dfrac{1}{L} - \dfrac{1}{m}$, where $m = \dfrac{L^2}{\Delta L} + L$,equation (2) transfers into

$$\begin{cases} \dfrac{di_d}{dt} = \dfrac{u_d}{L} - \dfrac{R}{L} i_d + \omega \dfrac{\psi_{rq}}{L} + \omega i_q \\ -\dfrac{u_d}{m} + \dfrac{R}{m} i_d - \omega \dfrac{\psi_{rq}}{m} \\ \dfrac{di_q}{dt} = \dfrac{u_q}{L} - \dfrac{R}{L} i_q - \omega i_d - \omega \dfrac{\psi_{rd}}{L} \\ -\dfrac{u_q}{m} + \dfrac{R}{m} i_q + \omega \dfrac{\psi_{rd}}{m} \end{cases} \tag{3}$$

The state equation of the PMSM is obtained from equation (3)

$$\begin{cases} \dot{x} = Ax + Bu + Ef_a + d \\ y = Cx \end{cases} \tag{4}$$

where state variables are

$$x = \begin{bmatrix} i_d \\ i_q \end{bmatrix}, u = \begin{bmatrix} u_d \\ u_q \end{bmatrix}, f_a = \begin{bmatrix} \psi_{rd} \\ \psi_{rq} \end{bmatrix}, d = \begin{bmatrix} -\dfrac{u_d}{m} + \dfrac{R}{m} i_d - \omega \dfrac{\psi_{rq}}{m} \\ -\dfrac{u_q}{m} + \dfrac{R}{m} i_q + \omega \dfrac{\psi_{rd}}{m} \end{bmatrix}$$

coefficient matrixes are

$$A = \begin{bmatrix} -\dfrac{R}{L} & \omega \\ -\omega & -\dfrac{R}{L} \end{bmatrix}, \quad B = \begin{bmatrix} \dfrac{1}{L} & 0 \\ 0 & \dfrac{1}{L} \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, E = \begin{bmatrix} 0 & \dfrac{\omega}{L} \\ \dfrac{\omega}{L} & 0 \end{bmatrix}$$

3. DESIGN OF ADAPTIVE SLIDING MODE OBSERVER

For the PMSM system described in (4), an adaptive sliding mode observer is constructed as follows

$$\begin{cases} \dot{\hat{x}} = A\hat{x} + Bu + E\hat{f} - G(\hat{y} - y) + \hat{d} \\ \hat{y} = C\hat{x} \end{cases} \tag{5}$$

where G is to-be-designed matrix, superscript "∧" is the observed value of relevant disturbances.

Observer deviation is defined as, $e(t) = \hat{x}(t) - x(t)$.

Output deviation is defined as, $e_y(t) = \hat{y}(t) - y(t)$

$$\begin{aligned} \dot{e}(t) &= \dot{\hat{x}}(t) - \dot{x}(t) \\ &= (A - LC)e(t) + Ee_{f(t)} + e_{d(t)} \end{aligned} \tag{6}$$

A new variable is defined as

$$\begin{aligned} s(t) &= \dot{e}(t) + \sum e(t) \\ &= (A - LC)e(t) + Ee_f(t) + e_{d(t)} + \sum e(t) \\ s_y(t) &= Cs(t) \end{aligned} \tag{7}$$

where $\sum \in R^{n\times n}, \sum = diag(\sigma,...,\sigma)$ and $\sigma > 0$

According to formula(6) and (7), we obtain

$$\begin{aligned} \dot{s}(t) &= (A - LC)s(t) + E(\dot{e}_{f(t)} + \sum e_{f(t)}) \\ &\quad + \dot{e}_{d(t)} + \sum e_{d(t)} \end{aligned} \tag{8}$$

Fixed value fault shows that the fault is fixed value or slow change, and satisfies $\dot{f}(t) = 0$, so

$$\dot{e}_f(t) = \dot{\hat{f}}(t) \tag{9}$$

Since the disturbance d is variable, $\dot{d} \neq 0$, and

$$\dot{e}_{d(t)} = \dot{\hat{d}} - \dot{d} \tag{10}$$

In order to achieve the optimal value of sliding mode condition, the adaptive control estimation algorithm is designed as follows

$$\dot{\hat{f}}(t) = -\Gamma_1 E^T P e(t) \tag{11}$$

$$\dot{\hat{d}}(t) = -\Gamma_2 P s(t) \tag{12}$$

where $\Gamma_1, \Gamma_2 \in R^{r\times r}$ is the adaptive learning rate, and $\Gamma_1 > 0, \Gamma_2 > 0$. P is to-be-designed matrix.

Lemma 1: For any positive number μ and symmetric positive definite matrix P, the following inequality holds

$$2x^T Py \le \frac{1}{\mu} x^T Px + \mu y^T Py, \qquad x, y \in R^n \tag{13}$$

Following positive definite function is chosen as *Lyapunov* function

$$V(t) = e(t)^T P e(t) + e_{f(t)}{}^T \Gamma_1^{-1} e_{f(t)} + \frac{1}{\sigma} e_{d(t)}{}^T \Gamma_2^{-1} e_{d(t)} \tag{14}$$

By deriving formula (14) and substituting (6), (11) and (12), we obtain

$$\begin{aligned} \dot{V}(t) &= 2e(t)^T P\dot{e}(t) + 2e_{f(t)}{}^T \Gamma_1^{-1} \dot{e}_{f(t)} + 2\frac{1}{\sigma} e_{d(t)}{}^T \Gamma_2^{-1} \dot{e}_{d(t)} \\ &= e(t)^T (P(A - LC) + (A - LC)^T P)e(t) - \\ &\quad 2\frac{1}{\sigma} e_{d(t)}{}^T P(A - LC)e(t) - 2\frac{1}{\sigma} e_{d(t)}{}^T PEe_{f(t)} - \\ &\quad 2\frac{1}{\sigma} e_{d(t)}{}^T Pe_{d(t)} - 2\frac{1}{\sigma} e_{d(t)}{}^T \Gamma_2^{-1} \dot{d}(t) \end{aligned} \tag{15}$$

According to Lemma 1, we obtain

$$\begin{aligned} &-2\frac{1}{\sigma} e_{d(t)}{}^T P(A - LC)e(t) \\ &\le \frac{1}{\sigma\mu_1} e_{d(t)}{}^T Pe_{d(t)} + \frac{\mu_1}{\sigma} e(t)^T (A - LC)^T P(A - LC)e(t) \end{aligned} \tag{16}$$

$$\begin{aligned} &-2\frac{1}{\sigma} e_{d(t)}{}^T PEe_{f(t)} \\ &\le \frac{1}{\sigma\mu_2} e_{d(t)}{}^T Pe_{d(t)} + \frac{\mu_2}{\sigma} e_{f(t)}{}^T E^T PEe_{f(t)} \end{aligned} \tag{17}$$

Since E is a Full column rank, the inverse matrix of E exists and is unique. $E = {}^+(E^T E)^{-1} E^T$

$$\begin{aligned} &-2\frac{1}{\sigma} e_{d(t)}{}^T \Gamma_2^{-1} \dot{d}(t) = -2\frac{1}{\sigma} e_{d(t)}{}^T E^T \left(E^+\right)^T \Gamma_2^{-1} \dot{d}(t) \\ &= -2\frac{1}{\sigma} e_{d(t)}{}^T E^T PP^{-1} \left(E^+\right)^T \Gamma_2^{-1} \dot{d}(t) \\ &\le \frac{1}{\sigma\mu_3} e_{d(t)}{}^T E^T PEe_{d(t)} + \frac{\mu_3}{\sigma} \dot{d}(t)^T \Gamma_2^{-1} E^+ P^{-1} \left(E^+\right)^T \Gamma_2^{-1} \dot{d}(t) \\ &\le \frac{1}{\sigma\mu_3} e_{d(t)}{}^T E^T PEe_{d(t)} + \frac{\mu_3}{\sigma} D^2 \lambda_{max} (\Gamma_2^{-1} E^+ P^{-1} \left(E^+\right)^T \Gamma_2^{-1}) \end{aligned}$$

then

$$\begin{aligned} \dot{V} &\le -M_1 \|e(t)\|^2 - M_2 \left\|e_{f(t)}\right\|^2 - M_3 \left\|e_{d(t)}\right\|^2 + \delta \\ &= Z^T \theta Z + \delta \end{aligned} \tag{18}$$

where

$$Z=\begin{bmatrix} e(t) \\ e_{f(t)} \\ e_{d(t)} \end{bmatrix}, \theta=\begin{bmatrix} -M_1 & & \\ & -M_2 & \\ & & -M_3 \end{bmatrix},$$

$$\delta=\frac{\mu_3}{\sigma}D^2\lambda_{max}(\Gamma_2^{-1}E^+P^{-1}\left(E^+\right)^T\Gamma_2^{-1})$$

Letting $\eta=\lambda_{(min)}(-\theta)$, we obtain $\dot{V}\le-\eta\|Z\|^2+\delta$

When $\|Z\|^2\ge\frac{\delta}{\eta}$, $V\le 0$. That is $\left\langle Z\middle|\|Z\|^2\ge\frac{\delta}{\eta}\right\rangle$, we can make the state estimation error e and the permanent magnet flux linkage estimation error e_f converge to zero.

4. SIMULATION ANALYSIS

PMSM system based on adaptive sliding mode observer method is established in MATLAB/Simulink. The main parameters of PMSM used in the MATLAB/Simulink are are given in Table 1.

Table1 The setting value of motor parameters

Motor parameters	or slow change
Resistance/ / Ω	2.875
Number of pole-pairs	4
Inductance /H	0.0085
The rotor flux linkage/Wb	0.175
The rotor inertia/Kgm	0.008

The parameters of the observer module are as following

$$L=\begin{bmatrix} 8000 & 1300 \\ 2000 & -8000 \end{bmatrix} \quad \Gamma_1=\begin{bmatrix} -0.5 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\Gamma_2=\begin{bmatrix} -10 & 1 \\ 1 & -10 \end{bmatrix} \quad P=\begin{bmatrix} 0.1 & 1 \\ -1 & 5 \end{bmatrix}$$

The load torque disturbance changes from 0Nm to 20Nm at 0.05s. The flux linkage deflects π/9 at 1s. The inductance disturbance changes from 0.0085H to 0.0080H at 1.5s. System simulation time is 0-2s.The following simulations are based on adaptive sliding mode observer and traditional sliding mode observer, and the simulation waveforms are shown in Fig. 2- Fig. 9.

As is shown in Fig. 2, since we choose the control algorithm with id =0, the current id always stays at 0 A regardless of load in 0.05s, the demagnetization fault in 1s, and the inductance change in 1.5s. But at the moment of disturbance and fault, the current jitters and then quickly recovers to 0 A. Therefore, we can know that under the control algorithm with id =0, the d-axis current will be restored to zero regardless of load, fault or disturbance.Comparing fig.2 and fig. 3, it can be obtained that both observers can recover quickly. However, the adaptive sliding mode observer can recover faster than the traditional sliding mode observer, which is obviously better than the traditional sliding mode observer.

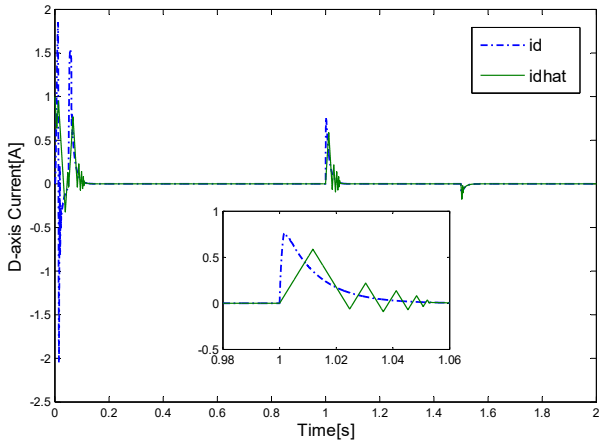


Fig.2 Simulation waveforms of i_d, $\hat{i}_d$ based on adaptive sliding mode observer

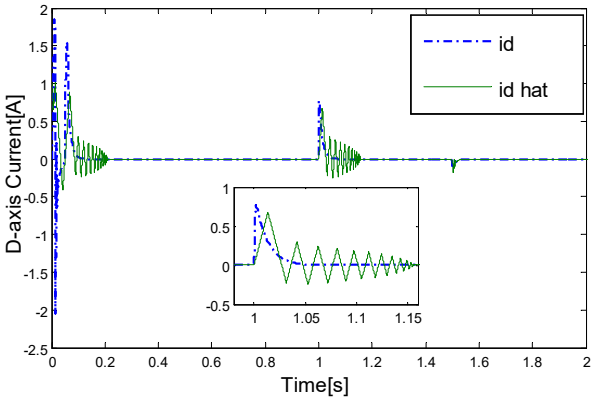


Fig.3 Simulation waveforms of i_d, $\hat{i}_d$ based on traditional sliding mode observer

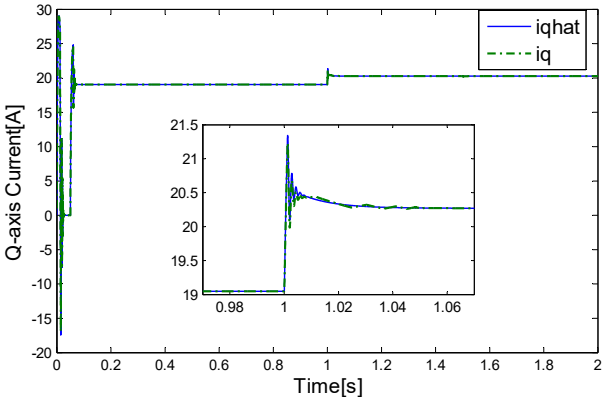


Fig.4 Simulation waveforms of i_q, $\hat{i}_q$ based on adaptive sliding mode observer

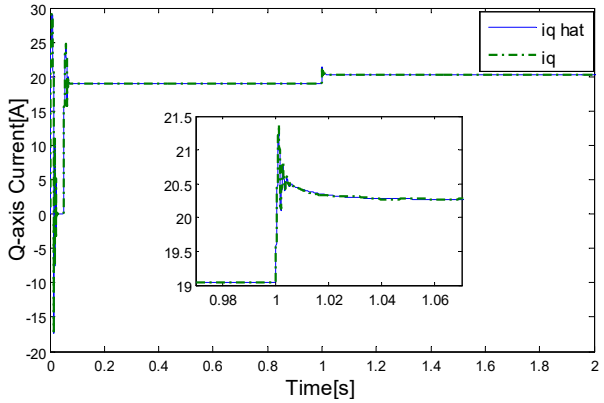


Fig.5 Simulation waveforms of $i_q, \hat{i}_q$ based on traditional sliding mode observer

As is shown in Fig. 4, when the PMSM is loaded at 0.05s, the current i_q jumps from initial value 0A to 19A.When the demagnetization fault occurs at 0.4s and the inductance changes at 1.5s, i_q increases slightly. We can conclude that only the loading makes i_q increase by a large margin, the demagnetization fault has little effect on i_q and the inductance parameter variation of the PMSM basically has nothing to do with i_q. Comparing fig. 5, it can be seen that the adaptive sliding mode observer and the traditional sliding mode observer have the same performance on the q-axis current simulation waveform, and there is no obvious difference.

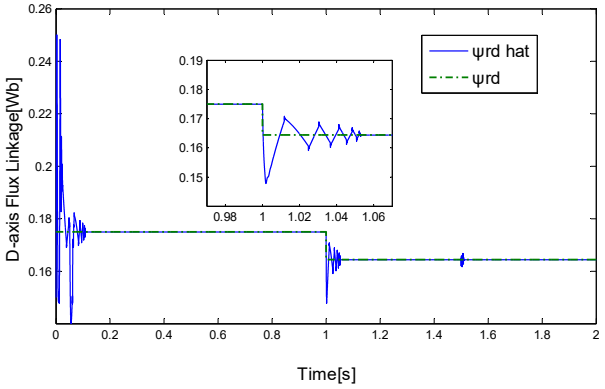


Fig.6 Simulation waveforms of ψ_{rd} , $\hat{\psi}_{rd}$ based on adaptive sliding mode observer

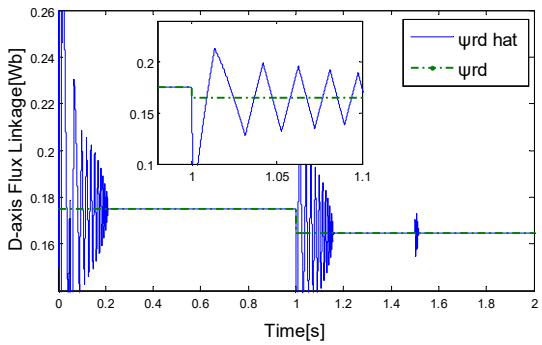


Fig.7 Simulation waveforms of ψ_{rd} , $\hat{\psi}_{rd}$ based on traditional sliding mode observer

As is shown in Fig. 6, when the demagnetization occurs at 1s, ψ_{rd} decreases from 0.175Wb to 0.170Wb, $\hat{\psi}_{rd}$ shakes and then quickly keeps up with ψ_{rd}. It shakes when the inductance disturbance occurs at 1.5s. From Fig. 6 and Fig. 7, it can be seen that compared with the traditional sliding mode observer, the flux linkage observation based on the adaptive sliding mode observer can track the set value more accurately, and the amplitude is smaller and the speed is faster.

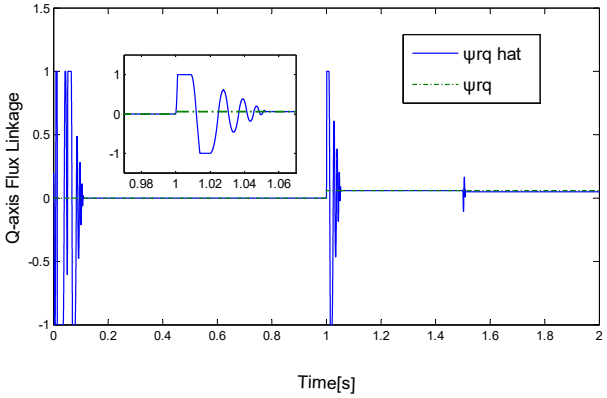


Fig.8 Simulation waveforms of ψ_{rq} , $\hat{\psi}_{rq}$ based on adaptive sliding mode observer

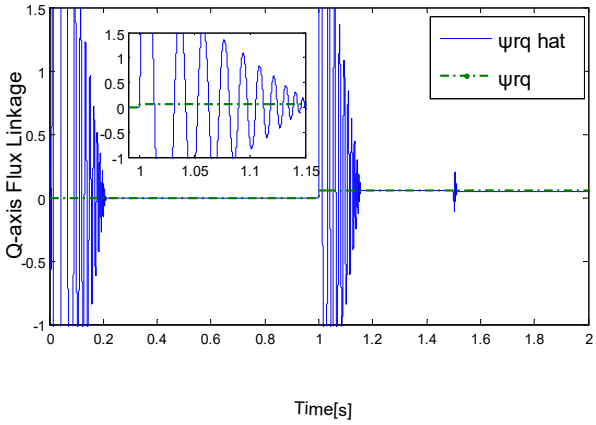


Fig.9 Simulation waveforms of ψ_{rq} , $\hat{\psi}_{rq}$ based on traditional sliding mode observer

As is shown in Fig. 5, when the demagnetization occurs at 1s, ψ increases from 0 to 0.06Wb, $\hat{\psi}_{rq}$ shakes and then quickly keeps up with ψ_{rq} . It takes when the inductance disturbance occurs at 1.5s. In Fig. 9, it can be clearly seen that $\hat{\psi}_{rq}$ shakes to a great extent after demagnetization fault, and although $\hat{\psi}_{rq}$ finall keeps up with ψ_{rq} .i t takes more time and has a larger amplitude.

Matlab simulation results verify the feasibility and effic-tiveness of theoretical method.

5. CONCLUSION

In this paper, aiming at the problem that the inductance is variable and the change rate of inductance derivative is not 0 in PMSM operation, an adaptive sliding mode observer based demagnetization fault diagnosis method for PMSM is proposed.Major contributions:1. The mathematical model of demagnetization fault of PMSM in synchronous rotating coordinate system is established considering the change of flux linkage and inductance. 2. Considering the characteristics of inductance disturbance, an adaptive sliding mode observer is designed, and two adaptive estimation algorithms of flux linkage and inductance are designed to realize the demagnetization fault detection of PMSM. 3. The Matlab simulation results verify the effectiveness of this method, and provide reference value for practical engineering application. However, this paper only analyzes the demagnetization fault of PMSM under inductance disturbance. In practical application, the working conditions are more complex and changeable, so it is necessary to study the influence of various disturbance factors on the motor, and how to diagnose the fault quickly.

Acknowledgements

This work was supported by the National Key R&D Program of China (No. 2018YFD0400705).

REFERENCES

[1] DUAN Fangbin, TAN Guangxing,FENG Chuchu FENG Chuchu and TIAN Junnan,Optimal Sliding Mode Control for Permanent Magnet Synchronous Motor, Electric Machines and Control Application, 2019,46(2):6.

[2] Vinson G, Combacau M, Prado T,et al. Permanent magnets synchronous machines faults detection and identification[C]., 2012:3925-3930.

[3] S Yong-lu, K Zhao-yi. Adaptive Sliding Mode Variable Structure Control for a New Hyper Chaos System[J]. JOURNAL OF CHONGQING INSTITUTE OF TECHNOLOGY,2010, 24(11): 109-112, 126(in Chinese).

[4] Design of Adaptive Fuzzy PID Controller for Speed Control of BLDC Motor[J]

[5] A novel current vector decomposition controller design for six-phase permanent magnet synchronous motor[J]

[6] Cho Y, Kim D, Lee K,et al. Torque ripple reduction and fast torque response strategy of direct torque control for permanent-magnet synchronous motor[C]., 2013:1-6.

[7] Z Chang-fan, Z Miao-ying, Z Fa-ming,et al. A cascade observer to detect demagnetization faults for PMSM[J]. Electric Machines and Control,2017, 21(2): 45-54(in Chinese).

[8] Research and Design of Permanent Magnet Synchronous Motor Vector Control for Electric Vehicle[J].Zhang C F, Wu H, He J, et al. Consensus Tracking for Multi-Motor System via Observer Based Variable Structure Approach[J]. Journal of the Franklin Institute, 2015, 352(8): 3366-3377.

[9] Kim, Seok-Kyoon. Robust adaptive speed regulator with self-tuning law for surfaced-mounted permanent magnet synchronous motor[J]. Control Engineering Practice, 2017, 61:55-71.

[10] Nguyen A T , Rafaq M S , Choi H H , et al. A Model Reference Adaptive Control Based Speed Controller for a Surface-Mounted Permanent Magnet Synchronous Motor Drive[J]. IEEE Transactions on Industrial Electronics, 2018, 65(12):9399-9409.

[11] Jing L , Hong-Wen L I , Yong-Ting D . Current adaptive sliding mode control based on disturbance observer for permanent magnet synchronous motor[J]. Optics & Precision Engineering, 2017, 25(5):1229-1241.

[12] Jung J W , Leu V Q , Do T D , et al. Adaptive PID Speed Control Design for Permanent Magnet Synchronous Motor Drives[J]. IEEE Transactions on Power Electronics, 2014, 30(2):900-908.

[13] Changfan Zhang, Gongping Wu, Fei Rong, Jianghua Feng, Lin Jia, Jing He, Shoudao Huang. Robust fault-tolerant predictive current control for permanent magnet synchronous motors considering demagnetization fault[J]. IEEE Transactions on Industrial Electronics, 2018, 65(7): 5324-5334.

An Improved Approach for Detection and Pose Estimation of Texture-Less Objects

Jian Peng *, Ya Su *
* School of Automation, China University of Geosciences, Wuhan, China
* Hubei key Laboratory of Advanced Control and Intelligent Automation for Complex Systems, Wuhan, China

Abstract

At present, the detection and pose estimation for texture-less objects still face some challenges, such as occlusion, background clutter, and object depth changes. Among the many methods of texture-less object detection and pose estimation, LineMOD algorithm is one of the most representative algorithms, but it has the problem of being sensitive to the depth of the template. This paper improves the problem of LineMOD's sensitivity to the depth of the template, and proposes a multi-scale template training method during the template training stage. When performing template matching, the test image is first divided into several regions, and then according to the depth of each test image area, the trained template at a similar depth is selected.Thus, the depth of the template used by the algorithm in template matching is similar to the depth of the target object without traversing all templates, which ensures the recognition accuracy.In addition, the object roughing technique is proposed in this paper. The algorithm will avoid a lot of useless matching operations and further improve the speed of the algorithm. Experiments show that the improved strategy of LineMOD algorithm in this paper can effectively solve the problem of the template depth sensitivity of the algorithm.

Keywords: Computer Vision; Object Detection and Pose Estimation; LineMOD Algorithm

1. INTRODUCTION

Object detection and pose estimation are important components in computer vision systems and research problems in the field of computer vision. Research on object detection and pose estimation algorithms is also crucial to the field of robotics and augmented reality. In order to improve the degree of automation and production efficiency, many factories use a large number of robots instead of manual workers [1]. For example, in the mutual fields of augmented reality and object sorting, robots are already the most important part. Augmented reality is a hot research field in recent years. Its purpose is to fuse and interact the virtual world with the real world. The realization of augmented reality technology is based on object detection and pose estimation, and on this basis, a virtual world is established and enabled to interact with the real world. Although robots have high demands in the industry, there are still some technical problems in the robot system that need to be effectively solved, such as the robot grabbing scattered and disordered objects. A good vision system is the prerequisite for the robot to complete any operation, and the robot's vision system mainly includes the recognition and attitude estimation technology of the target object.

Due to the wide application of object detection and pose estimation, in recent years, many scholars have conducted in-depth research on it and proposed many excellent algorithms. However, most of these algorithms are for the detection and pose estimation of textured objects. There are still many problems to be solved for the detection and pose estimation of texture-less objects. In the industry, there are a large number of applications for the detection and pose estimation of texture-less objects. Because of the rich pattern information on the surface of textured objects, effective features can be extracted from these rich patterns, so the detection and pose estimation of textured objects is relatively simple. Texture-less objects cannot extract effective features from their surfaces because their surfaces are smooth and non-textured. Therefore, this paper mainly studies the detection and pose estimation of texture-less objects. Based on the LineMOD algorithm,an improved strategy is proposed for the depth sensitivity of its template, which enables it to perform object detection and attitude estimation quickly and accurately in scenes with a wide range of object depth variation.LineMOD method was proposed by hinterstoiser [2] in 2011. It mainly solves the problem of real-time detection and positioning of 3D objects in complex background. It uses the information of rgbd, and can deal with the situation without texture, and does not need lengthy training time.

2. RELATED WORK

At present, object detection and pose estimation techniques are mainly divided into four categories: methods based on local features, methods based on deep learning, methods based on point pair and methods based on template matching. This section will summarize and explain the advantages and disadvantages of these four categories of methods.

2.1 Methods Based On Local Features

Local features have been widely used in object detection.

Common local features are SIFT [3], SURF [4], ORB [5]. These features are carefully designed, and have some unique advantages, such as rotation, lighting and scale invariance, so these features have good robustness when dealing with scenes with occlusion and lighting changes. However, these features are only suitable for the recognition of textured objects, and the recognition effect for texture-less objects is not good. In many fields, such as industrial robot operation and augmented reality technology,operation objects are often texture-less artifacts, so these local feature-based methods are not applicable to actual industrial scenes.

2.2 Methods Based On Deep Learning

Since deep learning technology can break through the limitations of some traditional methods, deep learning technology has been widely used in the field of object detection and recognition. In 2D object detection and recognition, more representative methods include CNN [6], RCNN, faster-RCNN [7], YOLO [8] [9], SSD [10]. On the basis of these 2D object detection and recognition methods, object detection and 6D pose estimation techniques have been gradually developed, which are mainly divided into two categories: SSD-6D and YOLO-6D. SSD-6D [11] is composed of two basic network structures, namely SSD network and automatic encoding network. Among them, the SSD network takes separate RGB images as input, which is mainly responsible for object recognition and pose estimation; while the automatic coding network takes the CAD model of the target object as the input, mainly to extract the high-dimensional features of the target object. When the SSD-6D network is working, the SSD network first outputs the area where the target object in the test image is located, and then obtains the target object's posture by comparing the features extracted from the area with the features extracted by the automatic encoding network. The YOLO-6D[12] network takes RGB images as input and outputs the 3D bounding box and classification prediction of the target object in the test image. After obtaining the 3D bounding box of the object, the posture of the target object can be calculated by the PNP algorithm [13]. The advantage of the method based on deep learning is that after the network training is completed, the speed of object recognition and pose estimation can be faster and the accuracy can be higher.

The disadvantage is that the hardware requirements are higher, and each time a new target object is added, the network needs to be retrained, which greatly limits the industrial application of this type of method.

2.3 Methods Based On Point Pair

Drost et al. proposed a method based on matching directional point pairs between the point cloud of the test scene and the object model [14], which has become one of the classic 6D pose estimation methods. Prior to this, the accuracy and speed of the global-based method did not meet the satisfactory requirements, and was mainly limited to the classification and recognition of certain specific objects; in contrast, the local matching method based on local invariant features was proved very effective, but the production of local invariant features depends largely on the quality of the acquired data and model data. Compared with these methods, the method adopts the idea of global modeling and local matching. During training, the points of the model are sampled, and similar features are grouped together and stored in a hash table; during the test, random reference points are found in the scene, similar model point pairs are searched in the hash table, and a fast voting method similar to Hof transform is used to vote for each matching point pair, the peak value in the accumulator is extracted as the candidate of pose, and the best pose is finally selected through ICP optimization. Compared with the local matching method that requires dense local information, this method uses a sparse set of directional points to represent the model and scene data. In this way, the recognition speed can be significantly improved without reducing the recognition rate.

2.4. Method Based On Template Matching

LineMOD algorithm [15] is the most representative template matching algorithm and the algorithm with the best effect. The LineMOD algorithm renders the RGB image and depth image of the target object from different perspectives, and then extracts the gradient direction and the surface normal direction from these images as features to make a template. Unlike traditional template matching methods, in the template trained by the LineMOD algorithm, its features are discretized, so not all feature points participate in the template matching operation, so this will greatly reduce the computational complexity of the algorithm. In the detection phase, the template and the test image are tested for similarity using a sliding window. If the similarity is higher than the set threshold, it indicates that there is a target object in the test image, and the pose of the corresponding template can be regarded as the initial pose of the target object. After the initial posture of the target object is obtained, the precise posture of the target object can be further solved by the ICP algorithm [16]. The LineMOD algorithm has the advantage of fast speed and high accuracy of object pose estimation, and when a new target object needs to be added, training a new target object template is very fast.Just render the template image of the target object and then extract and save the template features.This feature makes it have a unique advantage in actual industrial production. However, the disadvantages of LineMOD algorithm are quite obvious, because it is a

template matching algorithm, soit is sensitive to occlusion and template depth.

Rigas et al. [17][18] proposed the LCHF method, which is different from the LineMOD method to extract the template from the entire template image. This method divides the template image into several small blocks, and then extracts the template from each image block. In order to improve the algorithm speed, the random forest is trained by comparing the similarity between templates, and the template is stored orderly, which greatly improves the matching efficiency.It is difficult to master the size and number of the template image blocks in this method and its implementation is complicated. Haoruo Zhang et al. [19] proposed a cascading template matching method, and in order to solve the problem of sensitivity to template scale, scale-independent technology was proposed. First, the template under fixed depth was trained. In the test stage, the scene image was scaled to the size of the template image according to the relationship between scene depth and template depth, and then the template matching operation was carried out. However, there is a zooming operation in this method. If the zooming is serious, the image information may be lost seriously.

In summary, considering the requirements of industrial reliability, real-time performance, and rapid training of newly added objects, the LineMOD algorithm is still the most suitable method among these methods, but the template depth sensitivity problem of this algorithm has not yet been obtained. This paper proposes an improved solution to the depth-sensitive problem of the template of LineMOD algorithm. Compared with literature [19], this method does not require serious image scaling. Experiments show that the improved strategy in this paper enables the LineMOD algorithm to perform object detection and attitude estimation quickly and accurately in scenes with a wide range of object depth variation.

3. PROPOSED METHOD

In order to improve the recognition accuracy, the LineMOD algorithm needs to train the template at various depths of the target object. Since the LineMOD algorithm needs to traverse all the templates of the target object during template matching, the speed of the algorithm will decrease as the depth of the object template increases. If the depth of the target object in the scene varies in a large range, the number of templates will be greatly increased by training the templates at various depths of the target object, and the speed of the LineMOD algorithm will inevitably decrease.Aiming at this problem of LineMOD algorithm, this paper improves the template invocation mode of this algorithm when template matching, and proposes a template invocation object strategy based on Scene-Patch.The strategy mainly includes three key technologies: multi-scale template training method, scene image region division based on depth map, and rough object positioning, which enables the algorithm to select the templatespecifically when the template depth typeincreases.Therefore, the algorithm can still quickly identify the target object in the scene when the depth of the target object changes in a wide range.

3.1. Multi-Scale Template Training

The templates trained at multiple depths are called multi-scale templates. Since the template depth is a discrete quantity, it is impossible to train the templates at all depths when training multi-scale templates. Therefore, it is necessary to determine the step size of the depth change and the depth range to be covered when training the template. The step size of the depth change can be set according to the size of the target object. In practical application, the distance between the target object and the camera is a variable within a certain range.In the actual scene, the maximum and minimum depth of the target object in the scene image can be determined, thus the depth range that the template needs to cover can be determined.Thetraining method of the multi-scale template is as follows.

Determine the maximum and minimum depth that th-e target object can reach in the actual scene, denoted as $m = (D_{\max} - D_{\min}) / r$ $D_{\max}$ and $D_{\min}$, then the de-pth range covered by t-he template is $[D_{\min}, D_{\max}]$. The radius of the circumscribing sphere of the target object is denoted as r, and several depth layers ar e set in steps size of r, and the depth of eachdepth layer is denoted as D_i ,where $i \in (1,2,\cdots,m)$, $D_{\min} \le D_i \le D_{\max}$.

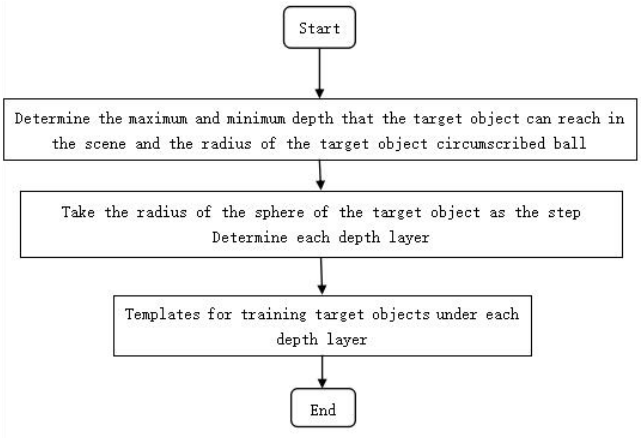


Fig.1 Multi-scale template training flowchart

Table 1. The pseudo code for coarse positioning of an object

Object coarsepositioning
Input: the scene image area block W and its corresponding depth D, template T with depth D
Output: there may be a set $P\{\ \}$ of target object positions in area W
Parameter: Sampling step size of detection point S, R, η
1: $(w_1, w_2, \cdots w_n) \leftarrow (W, s)$ Determine the detection pointsin the area block W
2: $(N_{\max}, N_{\min}) \leftarrow (T, \eta)$ Calculate the maximum andminimum number of deep edges in thetemplate
3: for $i \leftarrow 1$ to n
4: $l_i \leftarrow (w_i, R)$ Calculate the imaging size of the target object at the detection point
5: $w_i \leftarrow (l_i, D, w_i)$ Zoom the image block at the detection point
6: Calculate the number of depth edges num_i of the imageblock w_i
7: if $80\% \cdot N_{\min} \le num_i \le 150\% \cdot N_{\max}$ then
8: $P\{\ \} = P\{\ \} + r_i$ r_i is the position of the detection point w_i
9: end for

Train the template sequences at the corresponding depths of all the depth layers, then a total of m template sequences can be obtained.The depth of the template in each template sequence is the same, and the depth is equal to the depth D_i of the corresponding depth layer. The process of multi-scale template training is shown in Fig 1.

3.2. Scene Image Region Division Based On Depth Map

The purpose of the scene image area division based on the depth map is to divide the scene image into several areas according to the depth value of each pixel of the scene image, and the depth value in each area is within an approximate range. In the subsequent template matching operation in the image area of each scene, the training template at that depth can be targeted according to the depth of the corresponding depth layer in the image area of each scene, and only match the template with the training template at this depth.In this way, it does not need to traverse all templates of the target object, but also ensures the normal use of linear storage acceleration technology.The scene area division strategy based on the scene depth image is described as follows.

Set the initial points at the same interval on the scene depth map, denoted as c_i, $i \in (1,2,\cdots,n)$, the size of n depends on the size of the scene image, and the radius of the circumscribing sphere of the target object is denoted as r. Taking each initial point as the center, spread the range of the surrounding area in a breadth-first search strategy. When the maximum depth value $\beta_{\max}$ and the minimum depth value $\beta_{\min}$ in the area satisfy equation (1), the diffusion is stopped. At this time, the area to which each initial point diffuses during the diffusion process is a divided scene image area, denoted as $C_i, i \in (1,2,\cdots,n)$.

$$\beta_{\max} - \beta_{\min} \ge 2r \tag{1}$$

In this way, the scene image is divided into several regions with approximately equal depths. The color image is divided in the same way according to the division result of the depth image. After the scene image is divided into several areas C_i with substantially the same depth, there is a lot of overlap between the image areas. If template matching is performed directly on these image areas, many repeated matching operations will be generated ,which greatly reduces the speed of the algorithm. Therefore, in order to integrate each region block of the scene image, it is also necessary to merge the region blocks of the scene image. When merging scene image area blocks, first calculate the average depth of each scene area block as d^i_{avg}, attach each scene image area block to a depth layer that minimizes the difference between d^i_{avg} and D_i .Then the scene image blocks attached to the same depth layer are merged to form a new scene image block. In each new scene image area block, there may be multiple small area blocks that are not connected to each other. If the scene images are not connected, subsequent template matching operations cannot be performed. Therefore, it is necessary to separate the unconnected area blocks in each new scene image area. After separation, the final scene image area blocks are recorded as $W_i, i \in (1,2,\cdots,m)$. Each scene image area block corresponds to the depth layer to which it is attached,and its depth is represented by the depth of the depth layer. When template matching is carried out on each scene image area block later, the template under this depth can be directly called. The scene RGB image is divided in the same way according to the division result of the depth image.

3.3. Coarse Positioning Of Objects

After the scene image is divided into the area blocks with the depth difference within a certain range according to its depth value, many of these area blocks obviously do not have the target object in them. Inspired by literature [19], this section proposes a filtering method based on depth edge detection. While filtering out scene image area blocks that obviously do not contain the target object, it is possible to locate possible target objects in the remaining scene image area blocks s position.

In a certain area block W of the scene image, the detection points are set with a fixed step, each detection point is denoted as w_i , $i \in (1,2,\cdots,n)$, where n represents the number of detection points, and its size is determined by the size of the area block and the sampling step of the detection point, the actual depth of each detection point is denoted as z_i , $i \in (1,2\cdots,n)$. According to the depth D corresponding to the area block W , the template of the target object trained at this depth is retrieved. The size l_i of the target object imaged at the actual depth of the detection point w_i can be obtained from equation (2):

$$l_i = f \cdot (R / z_i) \quad (2)$$

Where f is the focal length of the camera and R is the diameter of the circumscribing sphere of the target object. When the detection point w_i is on the target object in the scene, l_i is just the side length of the largest 2D bounding box of the target object. Similarly, the size L of the template image with the depth D can be obtained from equation (3).

$$L = f \cdot (R / D) \quad (3)$$

The test image block with the detection point w_i as the center and l_i as the side length is scaled to the same size as the template in the proportional relationship of equation (4). The main purpose of this operation is to scale the imaging of the target object at the depth of each inspection point in the scene image to the size imaged by the template at the depth D , in preparation for the coarse positioning of the object. Although there are image zoom operations here, the depth values contained in the scene image area blocks are within a certain range, and the corresponding depths are not much different from the depth D of the called template, even if there is an image zoom operation,the scale is not too large, and the impact of image scaling can be ignored.

$$l_i / L = D / z_i \quad (4)$$

After scaling the test image block at the detection point w_i to the same size as the template image, in order to simplify the symbol marking, the test image block corresponding to each detection point is still marked with w_i , $i \in (1,2\cdots,n)$. The coarse localization technique in this section is based on depth edge detection. In fact, the depth edge of the image is calculated by the Soble operator. The depth edge appears at pixels where the calculation result of the Soble operator is greater than the set threshold. In this paper, the threshold is set to $30\%\eta$, where η is the side length of the maximum bounding box of the target object. The maximum and minimum values of the number of depth edges of the template at depth D are calculated as $N_{\max}$ and $N_{\min}$, respectively. If the number of depth edges num_i in the image block w_i satisfies the relationship (5), it indicates that there may be a target object at the detection point w_i .

$$80\% \cdot N_{\min} \le num_i \le 150\% \cdot N_{\max} \quad (5)$$

The detection points that satisfy the relationship (5) are called target detection points. In this paper, only the area blocks of the scene image that contain the most target detection points are retained, and the other area blocks are directly discarded. Since the position of the target detection point in the scene image is known, the possible position of the target object in the test image is also determined.Subsequently,the template matching operation will only be performed at the target detection point in the reserved scene image area block. Avoid a lot of useless matching, thereby improving the speed of object recognition. Taking the scene image area block W as an example, the pseudo code for coarse positioning of an object is as shown in Table1.

3.4. Invocation Method Of Multi-Scale Template

After the coarse positioning technology of the object, the target detection point in the scene image can be located, and only the test image area block containing the most target detection points is retained.However, the target object may exist at the detection point of the target, so it is necessary to accurately identify the object through template matching later.When accurate object recognition is performed, the template matching operation will be performed only at the target detection points of the remaining scene image area blocks to identify the target object. When performing template matching at the target detection point, the template trained at the depth can be retrieved according to the depth of the depth layer corresponding to the scene image area block, and template matching is only performed with the template trained at the depth, without traversal all templates of target objects, which will greatly improve the speed of the algorithm. Therefore, the template invocation strategy based on scenarios-patch enables the LineMOD algorit-

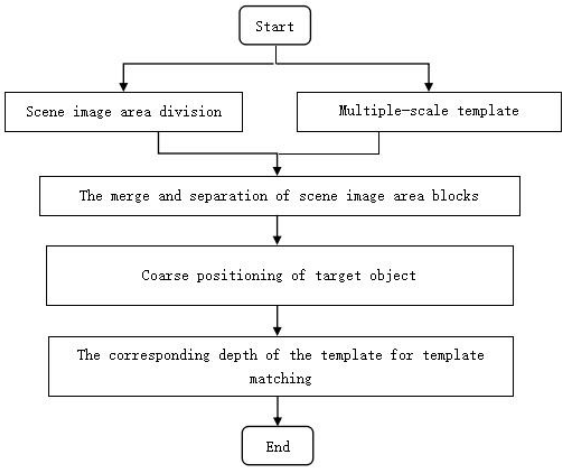


Fig.2 Flowchart of template invocation method based on Scene-Patch

hm to be used in scenes with a wide range of depth changes of target objects.In the case that the depth of the template is always similar to the depth of the target object in the scene, the target object in the scene can be quickly identified, so as to solve the problem that the LineMOD algorithm is sensitive to the depth of the template. The flow of the template invocation method based on Scene-Patch is shown in Fig 2.

4. EXPERIMENTS

4.1. Introduction To The Data Set

In this paper, we will test the improved strategy of this article on the LineMOD dataset. This dataset was created by Hinterstoisser et al. and has become the most commonly used dataset for evaluating the performance of object pose estimation algorithms [20]. The LineMOD data set contains 15 kinds of 3D models of non-textured objects. During the experiment, the template images can be directly rendered from the 3D models of these objects, and then the templates are trained from the template images. In addition, each object in the data set has more than a thousand test image sequences, and each test image carries information such as the pose of the target object in the test image, the distance between the target object and the camera, and camera parameters. The distance between the target object and the camera in the test image is called the depth of the target object in the test image, and the depth of the target object in the test image of this data set varies from 65 cm-115cm.

4.2. Multi-Scale Template Training

The biggest difference between this paper and the original LineMOD algorithm when training the template is the difference in step size of depth change.This paper uses the radius of the circumscribed spherel of the target object as the depth change step, and the original LineMOD algorithm takes a fixed value of 0.1m as the depth change step. The other parameter settings when training the template are respectively: the azimuth step length when acquiring the template image is 15°, the pitch angle range is -45°-45°, the step length is 10°, and the in-plane rotation range is -45°-45 °, the rotation angle step is 10°. The total template number of each object in the LineMOD data set trained in this paper is compared with the original LineMOD algorithm as shown in Figure 4. Object numbers 1-15 in the figure correspond to the monkey, vise, drill, camera, watering can, electric iron, and table lamp, telephones, toy cats, punching machines, toy ducks, drinking cups, bowls, egg boxes and glue. Since the original LineMOD algorithm uses a fixed depth step when training templates, the number of templates trained for each object is 11,664. It can be seen from Figure 3 that the number of templates of most objects trained in this paper is greater than the number of templates trained by the original LineMOD algorithm. As can be seen from Table 2, the improved LineMOD algorithm can effectively solve the depth sensitivity problem, and the object recognition rate has a certain improvement compared with the original algorithm.

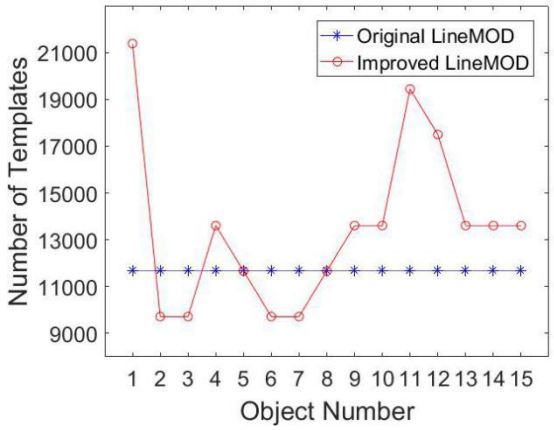


Fig.3 The number of templates trained by the LineMOD algorithm for each object in the LineMOD data set before and after improvement

4.3. Test Results And Analysis

This article first tests the accuracy of the improved LineMOD algorithm for object pose estimation. When conducting an object pose estimation experiment based on the LineMOD algorithm, first of all, it is necessary to determine the method for evaluating the correctness of the estimated pose. This article uses the evaluation method proposed in [15] to judge whether a pose estimation result is correct.

Suppose an object model M , which is composed of n points, denoted as $\{m_1,\ldots,m_n\}$ .The object pose estimated by the algorithm is denoted as $(R,\ t)$, and the real pose of the target object in the scene image is denoted as $(\bar{R},\bar{t})$, then the calculation formula of accuracy s of estimation pose is shown in Equation (6).

$$s = avg\left\|(R\cdot m+t)-(\bar{R}\cdot m+\bar{t})\right\| \quad (6)$$

As for the target object with symmetrical geometric structure or rotating structure, the same template image may be obtained from different perspectives, so the calculation method of accuracy s of pose estimation needs to be modified, as shown in Formula (7).

$$s = avg\ \min\left\|(R\cdot m_1+t)-(\bar{R}\cdot m_2+\bar{t})\right\| \quad (7)$$

((

If the accuracy of pose estimation s satisfies equation (8), it indicates that pose estimation is correct.

$$k_m d > s \quad (8)$$

In Equation (8), k_m is the control coefficient, which is set as 0.1 in this paper, and is the diameter of the target object's packet catching ball.

Table 2. The recognition rate (%) of the LineMOD algorithm before and after the improvement of the pose estimation of each object in the LineMOD data set

Target object	Target object radius(cm)	Number of depths		The total number of template		The recognition rate	
		Original LineMOD	Improved LineMOD	Original LineMOD	Improved LineMOD	Original LineMOD	Improved LineMOD
Little monkey	5.1	6	11	11664	21384	98.5	98.6
Vice	12.4	6	5	11664	9720	99.1	99.3
Driller	13.1	6	5	11664	9720	98.2	98.1
Camera	8.6	6	7	11664	13608	99.3	99.5
Watering can	10.0	6	6	11664	11664	98.7	98.9
Electric iron	13.9	6	5	11664	9720	98.3	99.0
Table Lamp	14.1	6	5	11664	9720	99.0	98.6
Telephone	10.6	6	6	11664	11664	97.2	98.4
Toy cat	7.7	6	7	11664	13608	99.5	99.6
Hole puncher	7.3	6	7	11664	13608	97.6	98.3
Toy duck	5.5	6	10	11664	19440	98.1	98.0
Drinking glass	6.2	6	9	11664	17496	98.6	97.5
Bowl	8.4	6	7	11664	13608	99.8	98.6
Egg box	8.2	6	7	11664	13608	99.6	99.8
Glue	8.8	6	7	11664	13608	97.3	98.1

Finally, the accuracy of pose estimation of the target object is taken as the evaluation standard. The accuracy of pose estimation is the ratio between the correct number of test images estimated by the target object and all the test images of the target object.The improved LineMOD algorithm visualizes the pose estimation of the target object in some test images in the data set as shown in Figure 4, where the calculated pose of the target object in each picture is expressed in the form of a three-dimensional coordinate axis.

Before and after improvement, LineMOD algorithm estimates the correct rate of pose estimation of each object in LineMOD data set as shown in Table 3. The improved LineMOD algorithm has an average pose estimation accuracy rate of 95.8% on the LineMOD data set, while the original LineMOD algorithm has an average pose estimation accuracy rate of 95.3% on the LineMOD data set. The improved LineMOD algorithm improves the average pose estimation accuracy of the dataset by 0.5% compared with the original LineMOD algorithm.

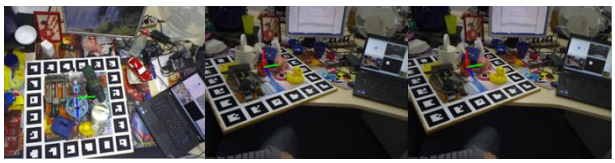


Fig.4 shows the effect of improved LineMOD algorithm pose estimation

In terms of speed, in the environment where the computer hardware is configured with an Intel core i7 quad-core processor and 8G of running memory, the average time required for the original LineMOD algorithm to complete the object pose estimation of a test image of the LineMOD data set is 0.2s ,the average time required for the improved LineMOD algorithm to complete the object pose estimation for a pair of test images in the LineMOD data set is 0.15s. The depth of the target object in the test image of the LineMOD dataset is not constant, and the variation range is 65cm-115cm. Therefore, it is necessary to train templates at various depths to ensure the accuracy of the algorithm. It can be seen from Section 4.2 that the improved LineMOD algorithm has trained more types of templates for most objects in the LineMOD data set. The

Table 3. The accuracy rate (%) and time of the LineMOD algorithm before and after the improvement of the pose estimation of each object in the LineMOD data set.

	The accuracy rate (%)		The time(s)	
Target object (number of test images)	Original LineMOD	Improved LineMOD	Original LineMOD	Improved LineMOD
Little monkey (1235)	94.6	95.3	0.199	0.149
Vice (1214)	97.3	98.1	0.194	0.146
Driller (1187)	91.5	92.6	0.191	0.143
Camera (1200)	96.1	98.5	0.192	0.149
Watering can (1195)	95.6	96.7	0.191	0.146
Electric iron (1151)	95.9	98.4	0.189	0.138
Table Lamp (1226)	94.5	94.3	0.202	0.147
Telephone (1224)	93.1	95.1	0.205	0.148
Toy cat (1178)	98.6	97.9	0.192	0.141
Hole puncher (1236)	93.2	92.6	0.211	0.155
Toy duck (1253)	94.5	96.0	0.215	0.158
Drinking glass (1239)	96.8	95.8	0.198	0.157
Bowl (1232)	99.0	94.8	0.203	0.148
Egg box (1252)	99.2	98.2	0.219	0.169
Glue(1219)	90.6	93.0	0.199	0.156
The average (18241)	95.3	95.8	0.2	0.15

number of templates trained is greater than the number of templates trained by the original LineMOD algorithm. When the number of templates increases, the improved LineMOD algorithm's pose estimation speed is faster than the original LineMOD algorithm.This is because the original LineMOD algorithm needs to traverse all templates of the target object when it performs pose estimation for the target object, and the improved LineMOD algorithm can specifically call the template at the corresponding depth for template matching, which greatly improves the speed of the algorithm, so the speed of the improved LineMOD algorithm will not be affected when the depth of the target object template increases.Furthermore, when the depth of the target object in the scene varies in a large range and more templates need to be trained, the improved LineMOD algorithm can still quickly estimate the pose of the target object, thus solving the problem that the original LineMOD algorithm is sensitive to the depth of the template.

5. CONCLUSIONS

The LineMOD algorithm can realize the recognition and pose estimation of texture-less objects in a messy background. It can adapt to the needs of different scenes by adding different templates, and has the advantages of fast speed and high precision.At present, it is still widely used in the field of pose estimation of texture-less objects in industry. However, this algorithm has the problem of depth sensitivity to the template, which makes it difficult to achieve satisfactory results in some special industrial scenarios.

This paper proposes corresponding improvement schemes for the defects of the LineMOD algorithm, improves the template invocation methodof the algorithm, and proposes a template invocation strategy based on Scene-Patch, which can make the LineMOD algorithm in scenes where the depth of the target object varies widely and the depth of the target object in the scene is always the same as the depth of the template,the target object can still be quickly and accurately identified. Experiments show that this strategy can solve the problem of LineMOD algorithm's sensitivity to the depth of the template.

References

[1] Caudell T P, Mizell D W. Augmented reality: An application of heads-updisplay technology to manual manufacturing processe[C].Proceedings ofthe twenty-fifth Hawaii international conference on system sciences. IEEE,1992, 2: 659-669.

[2] Hinterstoisser S, Holzer S, Cagniart C, et al. Multimodaltemplates for real-time detection of texture-less objects in heavily clutteredscenes[C]. International Conference on Computer Vision. IEEE Computer Society,2011:858-865.

[3] LoweDG.Object recognitionfrom local scale-invariant feature[C].In Proceedings of international conference on computer vision, Corfu, Greece, pp. 1999, 99(2): 1150-1157.

[4] Sun R , Qian J , Jose R H , et al. A Flexiblea-nd Efficient Real-Time ORB-Based Full-HD Im-age Feature Extraction Accelerator[J]. IEEE Transactions on Very Large Scale Integration (VLSI)

Systems, 2019, PP(99):1-11.
[5] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative toSIFT or SUR[J].International Conference on Computer Vision,Barcelona.2011:2564-2571.
[6] Yun R , Changren Z , Shunping X . Object De -tection Based on Fast/Faster RCNN Employing Fully Convolutional Architectures[J]. Mathematical Problems in Engineering,2018,(2018-1-9), 2018, 2018:1-7.
[7] Ren S, He K, Girshick R, et al. Faster r-cnn: T -owards real-time objectdetection with region pro -posal network[C].Advances in neural informationprocessing systems. 2015,39(6): 91-99.
[8] Redmon J, Farhadi A. YOLO9000: better, faster, stronge[C].Proceedingsof the IEEE conference on computer vision and pattern recognition. 2017:7263-7271.
[9] Redmon J, Divvala S, Girshick R, et al. You on -ly look once: Unified, real-time object detection [C].Proceedings of the IEEE conference on computervision and pattern recognition. 2016: 779-788.
[10] Liu W, Anguelov D, Erhan D, et al. Ssd: Singl- e shot multiboxdetecto[C].European conference o -n computer vision. Springer, Cham,2016: 21-37.
[11] Kehl W, Manhardt F, Tombari F, et al. SSD-6D: Making RGB-based 3Ddetection and 6D pose es -timation great agai[C].Proceedings of the IEEE International Conference on Computer Vision. 2017: 1521-1529.
[12] Tekin B, Sinha S N, Fua P. Real-time seamless single shot 6d object posepredictio[C].Proceedings of the IEEE Conference on Computer Vision andPattern Recognition. 2018: 292-301.
[13] Lepetit V , Moreno-Noguer F , Fua P . EPnP: An AccurateO(n) Solution to the PnP Problem[J]. international journal of computer vision, 2009, 81(2):155-166.
[14] Drost B, Ulrich M, Navab N, et al. Model glob -ally, match locally: Efficient androbust 3D obje -ct recognition[C]. Computer Vision and Pattern Recognition. IEEE,2010:998-1005.
[15] Hinterstoisser S, Cagniart C, Ilic S, et al. Gradi -entresponse maps for real-time detection of te-x tureless objects.[J]. IEEETransactions on Pattern Analysis & Machine Intelligence, 2012,34(5):876 -888.
[16] Lee J , Lee M , Kang S S , et al. Real-time 3 D Pose Estimation of SmallRing-Shaped Bin-pic -king Objects using Deep Learning and ICP Alg -orithm[J]. Journal of Institute of Control, 2019, 25(9):760-769.
[17] J.A. Hartigan, M.A. Wong. A K-means clusterin-galgorithm[J]. Appl Stat, 2013, 28(1):100-108
[18] Guo Y, Bennamoun M, Sohel F, et al. A Compr-ehensive Performance Evaluation of 3D Local Feature Descriptors[J]. International Journal of C-omputer Vision, 2016, 116(1):66-89.
[19] Zhang H, Cao Q. Texture-less object detection a -nd 6D pose estimation in RGB-D images[J]. R-obotics & Autonomous Systems, 2017, 95:64-79.
[20] Aldoma A, Tombari F, Rusu R B, et al. OUR-CVFH-Oriented, Unique and Repeatable Clustere -d Viewpoint Feature Histogram for Object Reco -gnition and 6DOF Pose Estimation[J]. 2012, 7476:113-122.

Positioning Method of Dulcimer Keys Based on Binocular Vision

De Tang*,**, Ziying Zhang***, Xin Chen*,**,†, Zhe Xiao*,**, Mengxi Qin*,**
* School of Automation, China University of Geosciences, Wuhan, 430074, P. R. China
** Hubei key Laboratory of Advanced Control and Intelligent Automation for Complex Systems
*** School of Arts and Communication, China University of Geosciences, Wuhan, 430074, P. R. China
†Corresponding author: chenxin@cug.edu.cn (Xin Chen)

Abstract

As a major branch of musical robots, the dulcimer robot has the advantages of speed and flexibility in performance. Before the robot performs, the coordinates of the dulcimer keys need to be transmitted to the robot for completing the positioning. However, the traditional manual positioning method is inefficient. A visual positioning method based on binocular vision is proposed in this paper on account of solving this problem, which can obtain the 3D coordinates of the dulcimer keys in a short time. To solve the problem of identifying the dulcimer keys from the complex background, a grayscale-based template matching method is used to identify them. To solve the spatial positioning of the dulcimer keys, a binocular stereo vision model is used to complete the three-dimensional reconstruction of them. The experimental results show that the visual positioning method proposed in this paper can effectively identify and locate the dulcimer keys.

Keywords: Music robot, Dulcimer keys, Template matching, Binocular stereo vision, Visual positioning.

1. INTRODUCTION

With the continuous development of science and technology, more and more humanoid robots have been developed to serve and meet the various needs of human beings [1]. As an important branch of future robots, musical robots are also frequently used in daily life. The dulcimer performance robot is a typical music robot. Before the robot performs, it needs to obtain and transmit the 3D coordinate information of the dulcimer phonemes to the robot. Only in this way can the robot hammer the dulcimer correctly. However, the traditional manual positioning method for the dulcimer phonemes is low in efficiency. Generally, this method takes 3-5 minutes to locate a phoneme, and the long-term positioning work severely hinders the development of the music robots. Therefore, a positioning method with high efficiency and high precision is of great significance to solve this problem.

It can be clearly seen from Fig. 1 that the dulcimer is consisted of many dulcimer keys and strings, and its structure is very complicated. Where, the strings, weakly textured objects, are composed of a series of thin and dense steel wire, and it is difficult to extract their features. The dulcimer keys are a number of white convex cone protruding outwards. Compared to the strings, the size of dulcimer keys is larger and easier to recognize. Therefore, it plays an important role in distinguishing phonemes. Based on the above analysis, the recognition and positioning of dulcimer phonemes is transformed into the recognition and positioning of dulcimer keys.

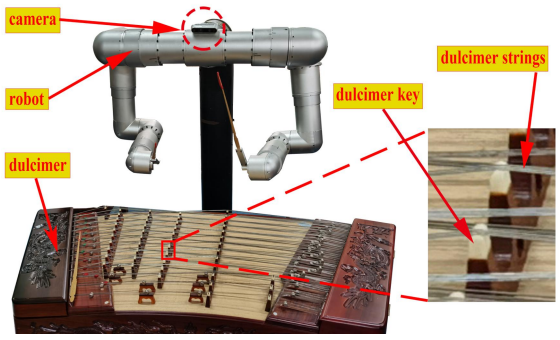


Fig. 1 Robot and dulcimer

However, recognizing of dulcimer keys is not easy. On the one hand, the RGB image obtained by camera is greatly affected by the ambient light due to the dulcimer key is white. Its color will change as the ambient light changes. On the other hand, the captured image may be interfered by background, and any object similar to the shape of the element on the dulcimer may cause incorrect recognition. Another important issue is that the recognition algorithm must be efficient enough to meet the real-time performance requirements of a music robot.

Although there is no special literature on the research of recognition for the dulcimer keys until now, many methods for object recognition and positioning have been proposed in a large amount of literature. These methods are broadly divided into three types: color-based method [2, 3], shape-based method [4, 5], grayscale-based method [6, 7].

The main principle of the color-based method is to map

the RGB (Red, Green, Blue) color space to the HSV color space [8]. Although the advantage of this method is simple and fast, the color is greatly disturbed by external environmental factors and the robustness is poor. The principle of the shape-based method is to detect target edges, such as sudden points, to obtain target edge information, and then match these geometric elements. This method can greatly reduce the amount of data and exclude irrelevant pairs of data [9]. Although this method is faster, its disadvantage is that it is greatly affected by lighting factors. The last method uses the detailed information of the image grays and gradient to extract feature points. Then, the feature points are matched according to the relative position information [10]. The advantage of this method is that it can effectively adapt to various lighting changes and other occlusions. Thus, grayscale-based template matching method can be used for recognition of dulcimer keys.

To locate dulcimer keys from vision, their 3D information need be obtained. Although the traditional monocular camera is widely used [11], it is difficult to directly obtain the 3D information of the target. In recent years, binocular stereo vision has been widely used in industry and other fields due to its advantages of high accuracy, real-time, and relatively low cost [12, 13]. Binocular stereo vision is an important form of machine vision, which is very suitable for positioning the dulcimer keys.

A positioning system is designed in this paper, which can be used to locate the dulcimer keys with high accuracy in close-range measurement. Firstly, the RGB-D image of dulcimer, pixel-aligned RGB and depth image, is obtained by using binocular vision. Then, these images are preprocessed, which helps subsequent algorithms to process these images. Afterwards, the template matching method is used to identify the dulcimer keys, and binocular vision is used to locate them. Finally, the 3D coordinates of the dulcimer keys relative to the camera is obtained, and the positioning is accomplished. This system uses binocular vision to obtain static RGB-D images. The advantage is that it can reduce the difficulty of the algorithm to process these images. In addition, the image can also be calculated off-line to avoid continuous occupation of the binocular camera.

The rest of this paper is organized as follows. In the section 2, the system framework for positioning dulcimer keys will be given. The section 3 is the image processing part, including the template matching algorithm for identifying dulcimer keys and the binocular stereo vision system for positioning them. The section 4 is the experimental part, which includes the result of the experiment and the effectiveness analysis of the method. The section 5 gives a brief summary of the full text.

2. SYSTEM FRAMEWORK

The positioning method proposed in this paper mainly focus on visual positioning of dulcimer keys, which can obtain their high-precision 3D coordinates in a short time. It consists mainly of three steps: image acquisition and preprocessing, dulcimer keys recognition and dulcimer keys positioning. The specific procedure of this system is shown in Fig. 2.

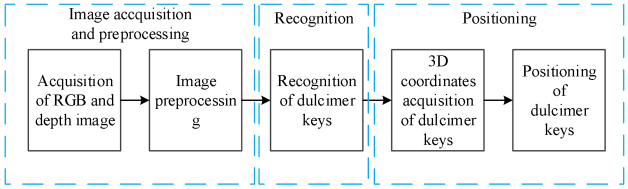


Fig. 2 The procedure of the positioning system

In the first step, the binocular vision model is used to obtain the RGB-D image of dulcimer. Then, these images are preprocessed in order to obtained high quality images that are easy to be processed by the algorithm. In the next two steps, the key issue is how to recognize and locate weakly textured dulcimer keys from the complex dulcimer surface. The grayscale-based template matching method is used to obtain the pixel coordinate of the dulcimer keys and the binocular stereo vision is used to complete the 3D reconstruction of them. Finally, the 3D coordinates of these keys can be obtained.

3. IMAGE PROCESSING

This section contains the specific process of recognition and positioning the dulcimer keys.

3.1 Images acquisition and pre-processing

The binocular camera can obtain depth information of the object by using two cameras [14]. The mathematical model of active binocular vision is shown in Fig. 3. As shown in this figure, Z_C and Z_C' are the optical axes of the left and right cameras, respectively. The two optical axes are parallel within the error range. The distance between the optical axes is T, C_l and C_r are the centers of the left and right camera sensor chips, respectively. The f is the focal length of the camera, and $P_X(x, y, f)$ and $P_{X'}(x', y', f')$ are 2D coordinates of the laser point P imaged by the left and right cameras.

Parallax $d(d = x - x')$ is the pixel difference between two imaging points. The mathematical relationship bet-ween the depth Z and the disparity d can be obtained from the similar triangle theorem as shown in Eq. (1).

$$Z = \frac{T \times f}{d} \tag{1}$$

Based on the above theory, the RGB-D image containing color and depth information can be obtained. These two images are used to complete positioning for the dulcimer keys.

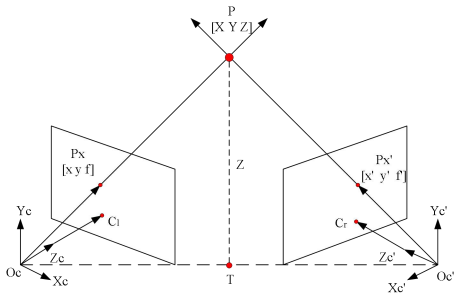


Fig. 3 Mathematical model of active binocular camera

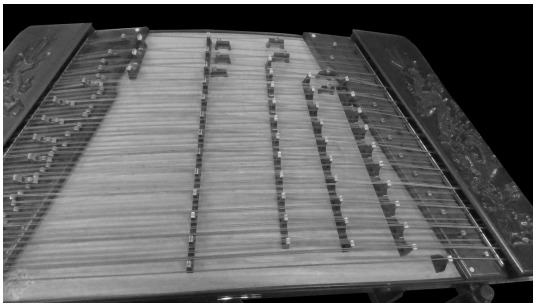
The RGB image carry a lot of color information, and some image processing algorithms are difficult to directly process the color images. Therefore, it is necessary to preprocess the image. The image pre-processing step consists of the RGB image segmentation and graying process. First, the image is segmented with frame-difference method, which is considered the fastest method with the best results. Then we further process the image to obtain a grayscale image, which can improve the processing speed of the algorithm and avoid irrelevant background interference.

The principle of frame-difference method is that to perform difference operation on images. Assuming the two frames are f_k and f_{k+1}, and their pixel values are $f_k(x,y)$ and $f_{k+1}(x,y)$. The difference image binary threshold is T, the difference image $D(x,y)$ is shown in Eq. (2).

$$D(x,y) = \begin{cases} 1, \left| f_{k+1}(x,y) - f_k(x,y) \right| > T \\ 0, \text{ others} \end{cases} \tag{2}$$

The result of extracting dulcimer by frame-difference method is shown in Fig. 4.

(a) Before



(b) After

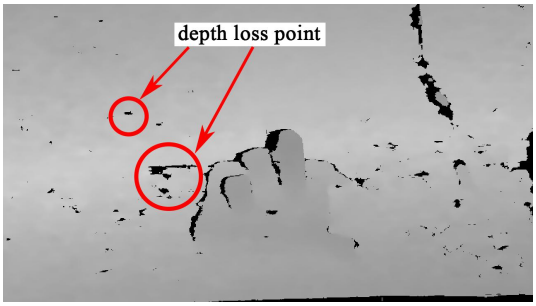
Fig. 4 The effect of frame-difference method

As shown in Fig. 4, the method can complete the extraction of dulcimer in complex environment well. Compared to the original image Fig. 4(a), the Fig. 4(b) is an 8-bit single-channel grayscale image containing only the dulcimer. Therefore, the background can be filtered out with frame-difference method.

The depth images carry a lot of depth information about the dulcimer keys. However, the binocular camera uses a video stream to transmit, the depth frames are saved will not be the same anytime. Part of the depth information that is not collected in the previous frame may be collected in the next few frames. As shown in Fig. 5(a), the black area in the figure is the depth loss area. To solve the problem, this paper proposes a multi-frame synthesis method, whose steps are as follows:

(1) Save several (approximately 20) depth images of target from the same perspective.
(2) Iterate through all the pixels in the head frame and return the coordinates of the pixel holes to the next depth frame.
(3) Fill the head frame with the next frame depth information at the coordinates.
(4) Loop through the above steps until all images are called.

It can be clearly seen from Fig. 5(b) that the synthesized depth image is obviously better than the original image. Because of its less loss of depth information, the error of 3D reconstruction is reduced.



(a) Original image

(b) Processed image

Fig. 5 The effect of multi-frame synthesis method

Most useless information is filtered out through image preprocessing, and high-quality depth images are obtained. These images are used in subsequent steps to identify and locate the dulcimer keys.

3.2 Recognition of dulcimer keys

As can be seen from Fig. 1, the white dulcimer keys are similar in color to the dulcimer surface, and the image is easily affected by the external light. Many scholars have proposed some target recognition methods, which can be broadly divided into color-based method [2], shape-based method [5], grayscale-based method [6]. Compared with the previous two methods, the grayscale-based method is faster and real-time in the recognition of dulcimer keys.

Based on the pre-processed RGB image obtained in the previous step, the image contains brightness information only. In this section, a grayscale-based template matching method can be used to identify the dulcimer keys and further obtain the pixel coordinates of these keys.

The matching process is shown in Fig. 6. Suppose T and I represents a template image of size $w \times h$ and an image to be matched, respectively. The coordinate (x', y') is the starting coordinate of the upper left corner of the T.

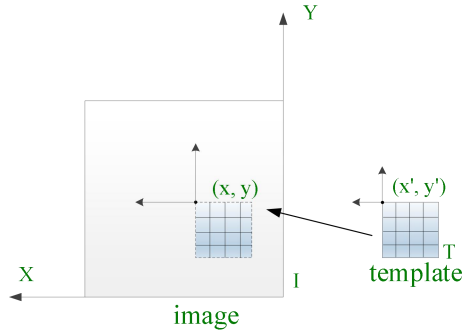


Fig. 6 The process of template matching

According to certain rules, to slide the template image T into the image I until all pixels are matched. For the single-channel gray image I, NSD (Normalized square difference, NSD) method is used to calculate the score. This method is simple in calculation and high in precision [15], which meets the real-time requirement of recognition algorithm for dulcimer keys. The matching score for each pixel is given based on Eq. (3). When the matching score $R(x, y)$ meets the requirements, the coordinates of the dulcimer keys in the image I are recorded.

$$R(x,y)=\frac{\sum_{x',y'}\left[T\left(x',y'\right)-I\left(x+x',y+y'\right)\right]^2}{\sqrt{\sum_{x',y'}T\left(x',y'\right)^2}\times\sqrt{\sum_{x',y'}I\left(x+x',y+y'\right)^2}} \quad (3)$$

The closer $R(x, y)$ is to 0, the higher the similarity is, or the lower it is. Marking the dulcimer keys recognized and recording the pixel coordinates of it in order. These coordinates are used to obtain the 3D coordinates of the keys in the next step.

3.3 Positioning of dulcimer keys

After obtaining the pixel coordinates of dulcimer keys, the binocular vision is used to complete the 3D reconstruction of these keys. The RGB-D image obtained by camera are shown as Fig. 7.

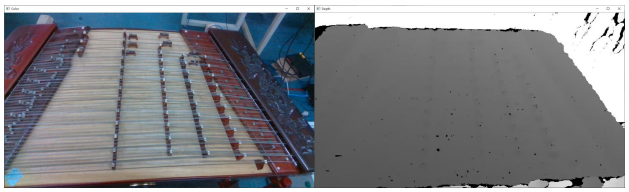


Fig. 7 The RGB-D image of dulcimer

After obtaining the depth information of dulcimer key Z from depth image, the 2D coordinates of this key are converted into 3D coordinates. The transformation process is shown in Fig. 8.

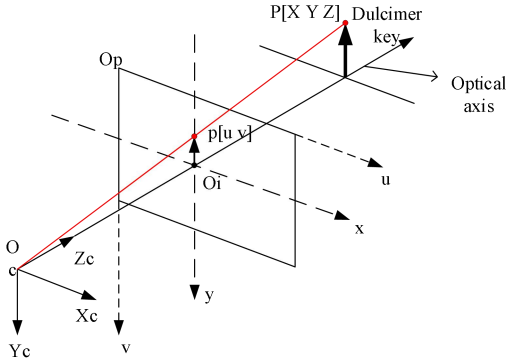


Fig. 8 Coordinate transformation

It can be seen from Fig. 8 that the figure involves the transformation of three coordinates, which are the image coordinate system O_i-xy, the pixel coordinate system O_p-uv and the camera coordinate system O_C-$X_C Y_C Z_C$.

The point $p(u,v)$ is the dulcimer key in the coordinate system O_p-uv, and the point $P(X,Y,Z)$ is it in O_C-$X_CY_CZ_C$. The transformational relation between the pixel coordinate system and the image coordinate system can be obtained from Eq. (4).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{4}$$

Where, dx and dy are the pixel sizes on x and y axis, separately. The relationship between the pixel coordinate system and the camera coordinate system can be obtained from Fig. 3 and Fig. 8 as shown in Eq. (5):

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} 1/dx & 0 & u_0 \\ 0 & 1/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{5}$$

Thus, we have completed the 3D reconstruction for dulcimer keys, and obtained the 3D coordinate information of these keys relative to the camera. The positioning of dulcimer keys can be achieved effectively.

4. EXPERIMENT AND ANALYSIS

This section includes the positioning experiment and comparison experiment. Then, the experimental results and analysis are given.

4.1 Experimental environment

The binocular vision equipment used in this experiment is the Realsense D435 depth camera designed by Intel. Its related parameters are shown in Table 1.

Table 1. Camera parameters

Parameters	Parameter Values
Resolution (pixels)	1280×720
Focal length	1.93 mm
Base line	50 mm
Pixel size	2.04 um
Field of view	H 69.4°; V 42.5°; D 77°

This experiment was run on Windows 10 Pro operating system with an Intel Core i7 CPU, 16 GB RAM, and the program running platform was Visual Studio 2019 Community.

4.2 Positioning experiment

This experiment uses the RGB-D image obtained by binocular camera to recognition and positioning of dulcimer keys. Firstly, the template matching method is used to identify the dulcimer keys. Then, the depth image is used to carry out 3D reconstruction of the them. Finally, the positioning of the dulcimer keys is completed.

The positioning experiment results are shown in Fig. 9. It can be seen from the figure that all the 48 dulcimer keys in the field of vision have been identified and positioned, and the 3D coordinates of each key have been obtained. And then the point-cloud of dulcimer keys are shown in Fig. 10.

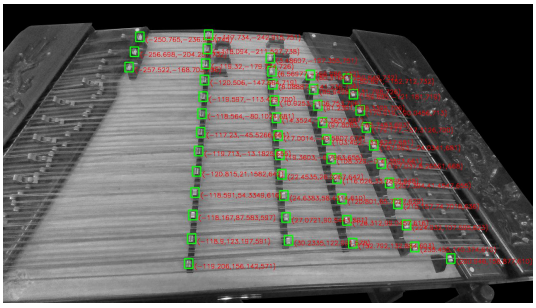


Fig. 9 Results of dulcimer keys positioning

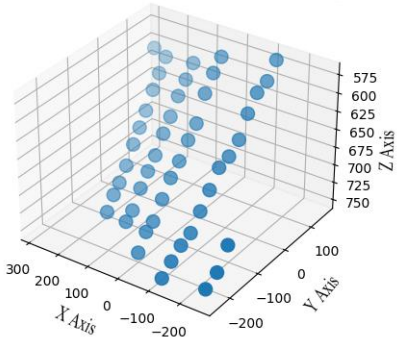


Fig. 10 Point-cloud of dulcimer keys

It can be seen from the point cloud that the positions of these dulcimer keys are basically consistent with their actual positions.

In order to verify the accuracy of 3D reconstruction, the distance information is used as the evaluation index in this experiment. Suppose the dulcimer is divided into two areas A and B evenly from the center. The distances were measured 3 times from the A and B areas, and the results are shown in Table 2.

Table 2. Results of distance test

Area	Measured(mm)	Actual(mm)	Error rate(%)
A	745	740	0.68
A	706	705	0.14
A	597	595	0.34
B	732	730	0.27
B	668	670	0.30
B	603	600	0.50

It can be clearly seen from Fig. 9 and Table 2 that based on the positioning method, dulcimer keys within the range of 500-800 mm can be located. In addition, the distance error rate is less than 0.8 % . The 10 dulcimer keys were selected randomly and the distance between

them was measured on account of testing of the 3D reconstruction effectiveness. The 5 groups of measurement data are shown in Table 3.

Table 3. Results of distance test

Group	Measured(mm)	Actual(mm)	Error(mm)
1	33	33.9	0.9
2	36	35.7	0.3
3	76	78.9	2.9
4	130	129.5	0.5
5	154	155.2	1.2

It can be seen from Table 3 that the maximum error the measurement data is no more than 3 *mm* . This error is allowed in this dulcimer performance system. The experimental results show that the positioning method of dulcimer keys has high accuracy, which basically meets the requirements of positioning.

4.3 Comparison of other methods

In order to further illustrate that our method is faster than the traditional manual positioning method, the time required by the two positioning methods is selected as the evaluation index.

The three points P_A , P_B and P_C on the dulcimer are located by our method and manual positioning method. The time required for each method was recorded, as shown in Table 4.

Our method includes all the time of recognition, positioning and inverse solution of manipulator, and the average time of this algorithm is only 5.4 seconds. Compared with the traditional manual positioning method, our method takes less time and has high efficiency.

Table 4. Comparison results of positioning time.

Method	P_A	P_B	P_C	Avg.
Our method	5.3*s*	6.2*s*	4.8*s*	5.4*s*
Manual method	190*s*	170*s*	170*s*	176*s*

Our method includes all the time of recognition, positioning and inverse solution of manipulator, and the average time of this algorithm is only 5.4 seconds. Compared with the traditional manual positioning method, our method takes less time and has high efficiency.

Another result of comparison experiment is given on account of illustrating the 3D reconstruction effect of our method is better. By comparing the number of invalid 3D points with Wang [9], the results of the two experiments are shown in Table 5.

It can be seen from Table 5 that the depth image obtained by our method has fewer invalid points, and our method has higher 3D reconstruction accuracy.

Table 5. Comparison results of with other methods.

Method	Valid points	Invalid points	Evaluation
Our method (Test.No.1)	908710	12890	Less depth loss
Wang [9] (Test. No.1)	904131	17468	More depth loss
Our method (Test.No.1)	921437	163	Less depth loss
Wang [9] (Test. No.1)	914679	6921	More depth loss

Obviously, the visual positioning method proposed in this paper is more effective in positioning the dulcimer keys. Therefore, the visual positioning method proposed in this paper provides a solution to the problem of low efficiency of traditional manual positioning.

5. CONCLUSION

In order to solve the problem of low efficiency when manually positioning the dulcimer keys, this paper proposes a visual positioning method based on binocular vision. In our method, grayscale-based template matching method is used to complete recognition of the dulcimer keys. Then binocular vision technology is used to measure the 3D information of the dulcimer keys. Relevant experiments show that within a measuring range of 800 *mm* , the distance measurement error is less than 0.8 % . And then, the time required for positioning is far less than the manual positioning method. Relying on the high accuracy of binocular stereo vision, the positioning method proposed in this paper can be effectively applied to the visual servo system of music robots for developing more intelligent robots.

Acknowledgements

This work is supported by the Hubei Provincial Natural Science Foundation of China under Grant 2019CFB581, Grant 2017CFA030 and Grant 2015CFA010, the National Natural Science Foundation of China under Grants 61873248, the major research project of China University of Geosciences (Wuhan) under Grant CUG180701 and the 111 project under Grant B17040.

References

[1] A. Sullivan and M. Bers, “Dancing robots: integrating art, music, and robotics in Singapore’s early childhood centers”, International Journal of Technology and Design Education, Vol. 28, No.2, 2018, pp. 325-346.

[2] A. Pourreza and K. Kiani, “A partial-duplicate image retrieval method using color-based SIFT”, 24th Iranian Conference on Electrical Engineering, 2016, pp. 1410-1415.

[3] C. H. Sung and M. J. Chung, “Dense scene 3D reconstruction using color-based sampling with fusion of image and sparse laser”, 17th Korea-Japan Joint Workshop on Frontiers of Computer Vision, 2011, pp. 1-6.

[4] Y. Yu, H. Cao, Z. Wang, et al, “Texture-and-Shape based active contour model for insulator segmentation”, IEEE Access, Vol. 7, 2019, pp. 78706-78714.

[5] X. Jian, X. Chen, Z. Xiao, et al, “Bolt positioning method based on active binocular vision”, 38th Chinese Control Conference, 2019, pp. 7057-7062.

[6] J. Xia, “Template matching algorithm based on gradient search”, International Conference on Mechatronics and Control, 2014, pp. 1472-1475.

[7] X. Fu and H. Gao, “Gray-based news video text extraction approach”, 5th International Conference on Computer Sciences and Convergence Information Technology, 2010, pp. 208-211.

[8] S. Li and G. Guo, “The application of improved HSV color space model in image processing”, 2nd International Conference on Future Computer and Communication, 2010, pp. V2-10-V2-13.

[9] F. Wang, X. Chen, C. Tan, et al, “Hexagon-Shaped screw recognition and positioning system based on binocular vision”, 37th Chinese Control Conference, 2018, pp. 5481-5486.

[10] J. Quan, S. Quan, Y. Shi and Z. Xue, “A fast license plate segmentation and recognition method based on the modified template matching”, International Congress on Image and Signal Processing, 2009, pp. 1-6.

[11] J. Liu, W. Liu and L. Gao, et al, “Detection and localization of underwater targets based on monocular vision”, 2nd International Conference on Advanced Robotics and Mechatronics, 2017, pp. 100-105.

[12] R. Xiang, H. Jiang, Y. Ying, “Recognition of clustered tomatoes based on binocular stereo vision”, Computers and Electronics in Agriculture, Vol. 106, 2014, pp. 75-90.

[13] X. Gong, X, Chen, et al, “Positioning method of insulator sheds based on depth information”, 39th Chinese Control Conference, 2020, pp. 3765-3770.

[14] J. Sun, Y. Ma, H. Yang and X. Zhu, “Camera calibration and its application of binocular stereo vision based on artificial neural network”, 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, 2016, pp. 761-765.

[15] Z. Xia, X. Yang, F. Meng and S. Wang, “Research of similarity measures for scene matching algorithm”, Congress on Engineering and Technology, 2012, pp. 1-5.

Disaster Relation Diagram Based on a Disaster Causation Database Extracted from Japanese Newspaper Articles

Fumihiro SAKAHIRA *, U HIROI**
* IoE Business Department, KOZO KEIKAKU ENGINEERING Inc.
Nakano-ku, TOKYO, Japan
** Graduate School of Engineering, the University of Tokyo
Bunkyo-ku, TOKYO, Japan

Abstract

A new method for creating a relation diagram of events that occur during disasters by extracting causal knowledge from Japanese newspaper articles and designing a causal network is proposed herein. Machine learning discriminant models were created for both succession expressions with causation and conventional cue phrases to extract causal sentences. We found that causal sentences can be extracted with a certain degree of accuracy from disaster articles. We were also able to create a causal network using sentences as nodes and links. The relation diagram using our new method has extracted events and causal knowledge that were unavailable in the disaster relation diagram designed using conventional methods.

Keywords: Disaster Relation Diagram, Causal Knowledge, Causal Network, Machine Learning, Japanese Newspaper Articles.

1. INTRODUCTION

A new method for creating a relation diagram of events that occur during disasters by extracting causal knowledge from Japanese newspaper articles and designing a causal network is proposed herein.

During a disaster, many events occur in a chain of events. The relation diagram, which organizes the chain of events, is an effective method for disaster prevention planning and education because organizing past disasters allows us to understand in advance the chain of events of future disasters. Most disaster relation diagrams are prepared by several people based on the experiences of those involved in the disaster and extensive expert knowledge; for example, the relation diagram of the 1995 South Hyogo Prefecture Earthquake [1], a large-scale disaster that occurred in Japan.

However, seemingly unimportant events not perceived by involved parties and experts may result in the omission of events that later have a significant negative impact. It is therefore desirable to extract the widest possible range of events and their chains, including any possible event and causal knowledge in the early stages of developing a relationship diagram.

Then, this study's novelty produces a relation diagram for the most comprehensive range of disaster events omitted by conventional methods. To realize this, the study was undertaken in two steps: First, the automatic extraction of causal sentences about disasters from Japanese newspaper articles using machine learning and the creation of a causal database of disaster causation and second, the creation of a causal network using the disaster causation database.

2. RELATED WORK

2.1 Creation of Disaster Relation Diagram

As mentioned above, traditionally, disaster relation diagrams were manually designed based on the experiences of those who have experienced disasters and their extensive expert knowledge. For example, the relation diagram of the 1995 South Hyogo Prefecture Earthquake was designed by group work, applying the KJ method [1]. The KJ method organizes data by describing the data on cards, grouping the cards, and illustrating them. However, because this method relies on the knowledge of the group members, we cannot exclude the possibility that causal knowledge omitted from their knowledge might not be reflected in the relation diagram. Therefore, we reduce this possibility as much as possible by experimenting with mechanical methods for extracting exhaustive causal knowledge about disasters from newspaper articles.

2.2 Extraction of Disaster Causal Knowledge

The extraction of causal sentences has traditionally been an important task in natural language processing research, in particular, the method of extracting causal sentences in English using cue phrases [2][3]. These cue phrases in Japanese are expressions that directly link cause and effect, such as "を背景に (wo haikei ni: on the backdrop of)" and "のために (no tame ni: because of)." Sakaji et al. [4][5] have a proven track record for the automatic

extraction of causal sentences from Japanese newspaper articles using machine learning. They used syntactic patterns based on cue phrases to automatically extract causal relationships for both single and multiple sentences. They also used syntactic features to predict the presence or absence of causal relationships through machine learning, which allows for the extraction of less frequently occurring causal relationships [6]. However, these studies were primarily concerned with economics articles, which are often described in clear causal terms, and it is unclear whether their method can be applied to disaster-related causal relationships. Therefore, the current study evaluated the applicability of using Sakaji et al.'s method [4][5] for disaster articles with respect to cue phrases. However, many disaster-related causes and effects are not clearly described. An example of this is causal succession. For instance, there are no cue phrases in the sentence "高速道路やビルが倒壊して、道路をふさぎ消防車も入れなかった(kosoku doro ya biru ga tokaisite, doro wo fusagi shobosha mo hairenakkata: The highway and buildings collapsed, blocking the road and preventing fire trucks from getting in)." However, we can interpret "高速道路やビルが倒壊して(kosoku doro ya biru ga tokaisite: The highway and buildings collapsed)" as the cause and "道路をふさぎ (doro wo fusagi: blocking the road)" as the effect. Although such representations are mentioned by Sakaji et al. [4], they were outside the scope of their study. Therefore, regarding the succession expressions with causation in Sakahira et al. [7], in this study, we examined and evaluated a new method for extracting succession expressions with causation.

2.3 Creation of causal networks

The causal chains represented by notes and links by Sato et al. [8] and Aono et al. [9] are less readable because they include only a few keywords and, therefore, require inference between words to be understood. In contrast, Ishii et al. [10] used the subject-verb-object (SVO) syntax to improve readability. However, a causal relationship was not always written in the SVO syntax. For example, although the phrase "地震による火災 (zishin ni yoru kasai: fire caused by the earthquake)" does not use SVO syntax, it does contain a causal relationship. However, advances in natural language processing techniques have made it relatively easy to compute similarities between sentences rather than using a few keywords. For example, Nishimura et al. [11] used the similarity of word embeddings in sentences to create a causal network. Therefore, in the current study, similar to Nishimura et al., we created a causal network by determining the similarity between the expression of the effect part of a sentence and the causal part of the expression of another sentence based on the cosine similarity of the word embeddings in the sentences.

2.4 Others

Other text mining studies of Japanese disaster articles have been limited to using word co-occurrence networks to extract words to characterize the attributes of victims and the stages of disasters [12][13] and, as such, have not directly contributed to the creation of disaster relation diagrams.

3. STUDY ISSUES

A method for designing the relation diagram for a most comprehensive range of disaster events omitted by conventional methods is proposed herein by undertaking the following:

- Evaluation of the applicability of Sakaji et al.'s method [4][5] for disaster articles using cue phrases.
- Examination and evaluation of a new method for the extraction of succession expressions with causation.
- Examination of whether word embeddings can be used to create causal networks using the sentences themselves as nodes and links.
- We compare the disaster relation diagram of the 1995 South Hyogo Prefecture Earthquake [1] created using conventional methods with our new method.

4. METHODS

4.1 Newspaper articles about the disaster

In this study, we surveyed disasters that occurred in the month after the 1995 South Hyogo Prefecture Earthquake. The data were collected from the *Asahi Shimbun'*s 1995 edition of their newspaper article text data collection [14].

4.2 A discriminant model of causal sentences using cue phrases

To create a discriminant model for causal sentences using cue phrases, we followed the method of Sakaji et al. [4][5], who used 35 different cue phrases and syntactic features [5]. Using syntactic features is an effective way to extract causal relationships occurring less frequently, according to Sakaji et al. [6]. We built the model by annotating 1,092 sentences, approximately one-third of the 3,281 extracted sentences, to determine the presence or absence of a causal relationship, of which 70% were used for training and the remaining 30% used for testing. The presence or absence of causality in 1,092 sentences were used as the objective variable. The following syntactic and semantic features were adopted as explanatory variables (Table 1), according to Sakaji et al. [4][5]: 1) Pairs of particles hanging over the main and base clauses, 2) Superordinate concepts of nouns in clauses hanging on the main and base clauses, 3) Parts of

speech in the morphological analysis before a clue phrase, 4) Thirty-five different cue phrases, 5) Morpheme unigrams, and 6) Morpheme bigram. However, regarding 2) Superordinate concepts of nouns in clauses hanging on the main and base clauses, we used extended language ontology from Japanese WordNet [15] for semantics rather than the Nihongo Goi Taikei-A Japanese Lexicon [16] for convenience. The training models were created using a support vector machine (SVM) and random forest, respectively, and their accuracy was compared.

To obtain the syntactic and semantic features, CaboCha [17] and MeCab [18] were used as a syntactic analyzer and a morphological analyzer, respectively, and morphological analyses were performed on phrases separated by clauses.

Table 1. Syntactic and semantic features used as explanatory variables for the cue phrases

Syntactic features	1) Pairs of particles hanging over the main and base clauses
Semantic features	2) Superordinate concepts of nouns in clauses hanging on the main and base clauses (from Japanese WordNet)
Other features	3) Parts of speech in the morphological analysis just before a clue phrase
	4) Thirty-five different cue phrases
	5) Morpheme unigrams (frequency of two or more)
	6) Morpheme bigram

4.3 A discriminant model of causal sentences using succession expressions with causation

Eight patterns of succession expressions with causation were identified from a newspaper article dated January 17, 1995 (Table 2) (Sakahira et al. [7]).

Table 2. Causal succession expression patterns

No.	Morpheme
1	Verb formed by adding “する(*suru*: do)” to a noun + “して、(*shite*) “
2	Verb formed by adding “する(*suru*: do)” to a noun + “し、(*shi*) “
3	Verb formed by adding “する(*suru*: do)” to a noun + “しており、(*shiteori*) “
4	Verb formed by adding “する(*suru*: do)” to a noun + “され、(*sare*) “
5	Verb formed by adding “する(*suru*: do)” to a noun + “せず、(*sezu*: not) “
6	Conjunctive forms of verbs + “て、(*te*)”
7	Conjunctive forms of verbs + “、”
8	Imperfective form verbs + “ず、(*zu*)”

We created a discriminant model for causal sentences using succession expressions, as with a discriminant model for causal sentences using cue phrases using syntactic features to predict the presence or absence of causal relationships through machine learning. We built the model by annotating 633 sentences, or approximately one-third of the 1,893 extracted sentences, to determine the presence or absence of a causal relationship, of which 70% were used for training and the remaining 30% used for testing. The presence or absence of causality in 633 sentences was used as the objective variable, and the following syntactic and semantic features were adopted as explanatory variables (Table 3), with some changes regarding the discriminant model for causal sentences using cue phrases. The principal difference between cue phrases and succession expressions from causation models is adopting 3) match or mismatch of the superordinate concepts of the words of the subject before and after a succession expression with causation, based on Sakahira et al. [7]. The training models were created using SVM and random forest, respectively, and their accuracy was compared.

Table 3. Model features using succession expressions with causation

Syntactic features	1) Pairs of particles hanging over the main and base clauses
Semantic features	2) Superordinate concepts of nouns in clauses hanging on the main and base clauses (from Japanese WordNet)
	3) Match or mismatch of the superordinate concepts of the words of the subject before and after a succession expression with causation (from Japanese WordNet)
	4) Superordinate concepts of verbs contained in succession expressions with causation (from Japanese WordNet)
Other features	5) Parts of speech in the morphological analysis just before a succession expression with causation
	6) Thirty-five different cue phrases
	7) Morpheme unigrams (frequency of two or more)
	8) Morpheme bigram

4.4 Creation of a causal database and a causal network

For the causal sentences extracted by each discriminant model, a database was created by dividing them into cause and effect parts. For sentences containing clue phrases, we divided them into cause and effect parts based on the syntactic patterns of Sakaji et al. [5]. For sentences containing a causal succession expression, we deemed the sentence before the causal succession expression as the causal part and after it as the effect part.

For the causal network, a single causal relation consisted of two event nodes: the starting node was the cause and the ending node was the effect. We additionally updated the causal network by adding an endpoint node beyond that node, with that endpoint node as the starting point. Specifically, the cause part, which had high similarity to the search sentence or phrase of a starting node, was searched for in the database. Then, the effect part, set with a cause part having a high similarity, was made a candidate for an endpoint node. The endpoint node was determined by manually selecting it from among the candidate endpoint nodes. This endpoint node was then used as a starting node, and the candidate endpoint nodes were searched for beyond it. This process was repeated any number of times to create a causal network.

For the calculation of similarity, cosine similarity was calculated using the mean of the word embeddings in the sentence or phrase. Having obtained the word embeddings, the cosine similarity was calculated using Ginza [19], a Japanese natural language processing library.

5. RESULTS AND DISCUSSION

5.1 Discriminant for causal sentences using cue phrases

Table 4 shows the test results of the causal sentence discriminant model for cue phrases trained using SVM and random forest. The averages are the weighted averages of the number of causal and non-causal sentences for precision, recall, and F-score, and the total is the number of data. The average SVM results were Precision (0.72), Recall (0.71), and F-score (0.71). However, the average random forest results were Precision (0.71), Recall (0.72), and F-score (0.71). The average accuracy of the random forest has an equally high value as SVM; however, in terms of the Recall value of causal sentences, the SVM has a higher value than the random forest. This study emphasized the importance of exhaustively extracting causal knowledge to create a relation diagram. Therefore, we focused on the small number of causal sentences excluded (i.e., the recall value of causal sentences). Therefore, we adopt the SVM results, where 0.69 of 122 sentences contained a causal relationship in the test data, and 82 could be extracted using the discriminant model. Compared with Sakaji et al.'s experimental results obtained without additional training data, the SVM values were both slightly lower (Precision: 0.802, Recall: 0.753, and F-score: 0.777) [4] and higher (Precision: 0.34, Recall: 0.71, and F-score: 0.46) [5]. Therefore, Sakaji et al.'s method can be applied to disaster articles for the discriminant of causal sentences using cue phrases.

Table 4. Discriminant results for causal sentences using cue phrases

Model	Sentence	Precision	Recall	F-score	Number
SVM	Non-causal	0.80	0.72	0.76	206
	Causal	0.60	0.69	0.64	122
	Average/ Total	0.72	0.71	0.71	328
Random Forest	Non-causal	0.76	0.81	0.78	206
	Causal	0.63	0.57	0.60	122
	Average/ Total	0.71	0.72	0.71	328

5.2 Discriminant for causal sentences using succession expressions with causation

Table 5 shows the evaluation results of the discriminant model of causal sentences of succession expressions with causation trained by the SVM and random forest (for the explanation, see Table 4). The average SVM results were Precision (0.75), Recall (0.76), and F-score (0.76). However, the average random forest results were Precision (0.62), Recall (0.73), and F-score (0.68). The average accuracy of the random forest has a slightly lower value than SVM; however, in terms of the Recall value of causal sentences, the SVM has a significantly higher value than the random forest. As mentioned above, this study emphasized the importance of exhaustively extracting causal knowledge to create a relation diagram. Therefore, we focused on the recall value of causal sentences. Therefore, we adopt the SVM results. These values were higher than those obtained by the discriminant model of causal sentences for cue phrases. Therefore, for disaster articles, the discriminant of causal sentences using succession expressions with causation method was found to be as accurate as the cue phrases method. However, the recall value of causal sentence was 0.53, meaning that in the test data, 51 sentences truly contained a causal relationship, of which we were able to extract 27 sentences using the discriminant model. This value is lower than that obtained using cue phrases. The improvement of this value is an issue for future studies.

Table 5. Discriminant results for causal sentences using cue phrases

Model	Sentence	Precision	Recall	F-score	Number
SVM	Non-causal	0.83	0.84	0.84	139
	Causal	0.55	0.53	0.54	51
	Average/ Total	0.75	0.76	0.76	190
Random Forest	Non-causal	0.76	0.92	0.83	139
	Causal	0.48	0.20	0.28	51
	Average/ Total	0.62	0.73	0.68	190

5.3 Creation of the causal database and the causal network

A causal database was created for all newspaper articles published during the month after the 1995 South Hyogo Prefecture Earthquake based on sentences extracted using the discriminant model of causal sentences using cue phrases and the discriminant model of causal sentences using succession expressions with causation. Table 6 presents an example of the results using the causal database to search for the causal part of a sentence with "地震の発生(*zishin no haxtusei*: earthquake occurrence)" as the first search sentence and displaying the cause part of the sentence with the effect part in the order of similarity. However, as also shown in Table 6, the search results contained some clauses that were not causal sentences (i.e., No. 1) or whose meaning could not be read (i.e., Nos. 6, 7, and 10). The former was due to the accuracy of the discriminant model, while the latter may be due to sentences containing clauses whose meanings could not be read from a single sentence. These are issues that need to be addressed in future studies.

Table 6. Top 10 results of the causal database search for "地震の発生 (*zishin no haxtusei*: earthquake occurrence)"

No.	Cause	Effect	Similarity
1	地震発生直後から (*zishin haxtusei chokugo kara*: right after the earthquake)	食料やトイレ、寝る場所など避難者の世話に追われ続ける (*shokuryo ya toire, nerubasho nado hinansha no sewa ni owaretudukeru*: continuing to take care of the evacuees, including food, toilets and places to sleep)	0.93
2	地震で(*zishin de*: in the earthquake)	日本のビルや新幹線、高速道路などが、あれほどひどく壊れる (*nihon no biru ya shinkansen, kosokudoro nado ga, arehodo kowareru*: Japanese buildings, bullet trains, highways, etc. will be destroyed so badly)	0.92
3	地震による(*zishin ni yoru*: due to the earthquake)	火災(*kasai*: fire)	0.91
4	地震による	停電(*teiden*: power outage)	0.91
5	地震による	津波や台風での高波 (*tsunami ya taifu deno takanami*: tsunami and typhoon surge)	0.91
6	地震による	打撃(*dageki*: strike)	0.91
7	地震による	ゆがみ(*yugami*: distortion)	0.91
8	地震による	断水(*dansui*: water outage)	0.91
9	地震による	強い揺れ(*tsuyoi yure*: strong shaking)	0.91
10	地震による	損壊(*sonkai*: damage)	0.91

In the current study, when creating the causal network based on the search results, we manually decided whether or not to adopt them as endpoint nodes. In this way, the causal network was created by extracting the effect part of the causal sentence from the search results as the endpoint node for the causal sentence to be employed in the causal network, and thereafter, the endpoint node was used as the starting node to search for other similar causal parts. To compare the differences between the causal network using our method and the conventional relation diagram, the causal network was created under the same number of conditions that the relation diagram for the 1995 South Hyogo Prefecture Earthquake [1] had 14 child nodes linked to "地震の発生(zishin no haxtusei: earthquake occurrence)" in the first hierarchy and the maximum number of child nodes from each node was about three in subsequent hierarchies. Table 7 presents an example of our selection of the top 3, based on our judgment of the search results of the extracted sentence "断水(dansui: water outage)" (see Table 6). In particular, No. 2 in Table 7 was extracted as a sentence. Tables 6 and 7 demonstrate that causal relationships with similar causal parts had a relatively high degree of similarity to the search statements. Therefore, although further study is required, a causal network can be created for disaster articles, even from sentences that contain causal expressions.

Table 7. Top 3 selected search results for "断水(dansui: water outage)"

No.	Cause	Effect	Similarity
1	断水で(*dansui de*: because of the water outage)	洗濯や食事の用意も大変だ(*sentaku ya shokuji no yoi mo taihen da*: it's a lot of laundry and food)	0.74
2	地震では、水道管が壊れ(*zishin deha, suidokan ga kowar*e: in the earthquake, the water pipes broke)	消火栓が使えなくなる(*shokasen ga tsukaenakunaru*: the fire hydrant will not work)	0.64
3	断水による(*dansui ni yoru*: due to the water outage)	トイレの問題(*toire no mondai*: toilet problems)	0.61

The resulting causal network is presented in Fig. 1. For convenience, we have limited the number of levels to three. Each node is an event, and the arrowed lines indicate the beginning point as the cause and the endpoint as the effect. The gray-filled nodes indicate events that were not in the manually created relation diagram [1]. The bold arrowed lines indicate causation that was not in the manually created relation diagram [1]. The expressions of these nodes have been translated into English after we summarized and simplified their contents. Fig. 1 demonstrates that we were able to extract events and causal knowledge that were not found in the manually generated relation diagram.

In particular, the following events were not found in the conventional relation diagram and could be extracted using this study's method, namely "Loosing of the ground," "Order employees to go home," "Continuous power outages due to the inability to perform construction," "Falling demand for gasoline," "Investigating the impact of exports on the destination country," "Shares in Hanshin area companies plummet," "Guiding students to examine," "Issuance of deficit bonds," "Impossible to publish a newspaper," "Severe damage to manufactures," "Inquiries about foreign investor activity," "Strengthening cooperation between local governments and the self-defense forces," "Difficulty in securing a place for the body to be laid to rest," "Difficulty in holding local elections," "Desire for your own private toilet," "Not drinking water, not eating anything," "Construction of container yards in other ports," "Substitution of Tokyo and Yokohama ports because of the fullness of the port of Nagoya," "Fare increases," "Domino effect of disruptions in the transportation of necessary supplies and parts," "Stricken relatives rushing in," "Scrapping plans to operate the new JR bullet train," "Considering risk diversification, such as restructuring the overseas network," "Cautious about discussing tax hikes early," "No prospect of restoring cable TV viewing," "Request to industry associations to prevent price increases," "Expansion of purchases from other regions and overseas," "No prospect of full recovery," "Decrease in income from electricity rates at the power company," and "Continuous current account deficit."

Furthermore, there are differences in the causes of "Stopping plant production, shutting down production lines, shutting down operations and reducing production." While the conventional relation diagram [1] identified "Damage to building structures" as the cause, the causal network in this study extracted "Logistical disruption" and "Power outage" as the cause.

Based on these results, we have shown that our new method can be expected to effectively extract the widest possible range of possible disaster events in the early stages of developing a relationship diagram.

6. CONCLUSION AND FUTURE WORK

Our method should effectively extract the widest possible range of disaster events during the early stages of developing a relationship diagram because we could extract events and causal relationships that could not be found in the conventional relation diagram [1]. In this process, we found that Sakaji et al.'s [4][5] cue phase method could be applied to disaster articles about the 1995 South Hyogo Prefecture Earthquake published in a newspaper for the discriminant of causal sentences. We additionally found that our new method that uses succession expressions with causation was as accurate as using cue phrases. We then examined whether word embeddings can be used to create causal networks using sentences as nodes and links. We found that causal sentences with similar causal parts had a high degree of

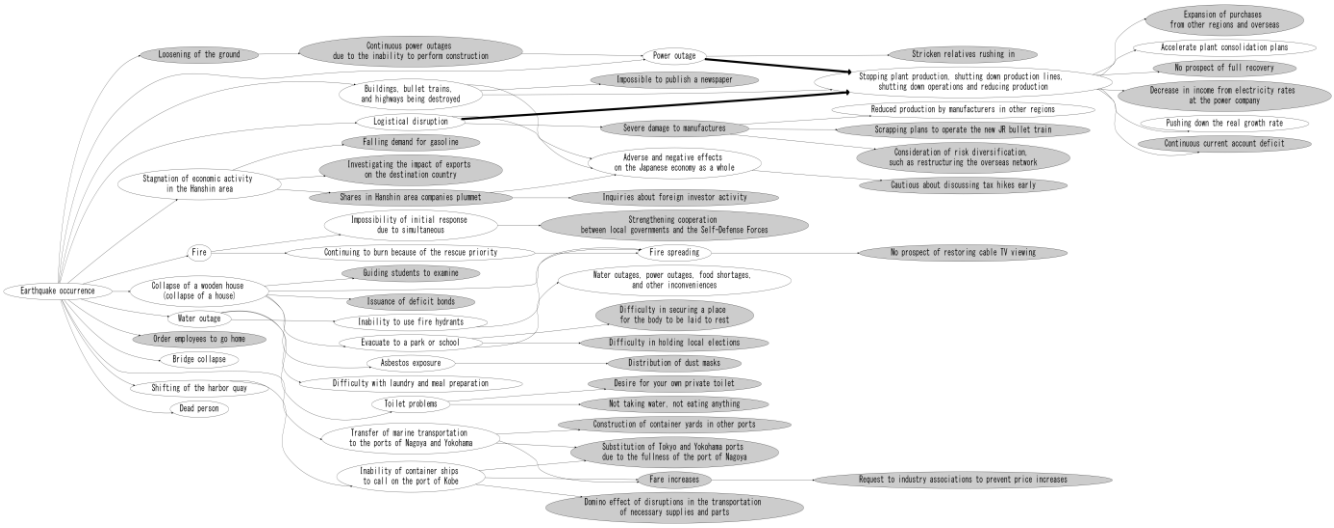


Fig.1 Part of the causal network that was created for the 1995 South Hyogo Prefecture Earthquake (in English). Each node is an event, and the arrowed lines show the beginning point, which is the cause, and the endpoint, which is the effect. Events and causations excluded in the conventional relation diagram [1] are shown as gray-filled nodes and bold arrowed lines, respectively. The English expressions in the nodes were translated from Japanese after the authors summarized and simplified their contents.

similarity to the search sentence. Therefore, a causal network can be created from disaster articles using sentences that contain causal expressions, although there remain outstanding issues, as discussed below.

However, we did find causal knowledge that was not included in the causal database created using the discriminant model but was present in the conventional relation diagram [1]. There are three possible causes for this. First, not all causal knowledge could be extracted because the discriminant model is not 100% accurate. Second, the causal knowledge was described by expressions other than the cue phrases or the succession expressions with causation. Third, some causal knowledge was not mentioned in the newspaper articles because it was not recognized in the first month after the event, but it was discovered by later verification. Thus, while this study's method can be used to extract events and causal knowledge that are not found in the conventional relation diagrams, it may not be able to extract important causal knowledge. Therefore, this study's method and the manual generation of a relation diagram are complementary; each is a supplementary measure.

The study's technical challenges include improving the accuracy of the discriminant model and improving the accuracy of the search for creating causal networks. As for improving the accuracy of the discriminant model, as well as improving this study's model, it may be possible to extract cases in which causal knowledge is described by expressions other than the cue phrases used here and by succession expressions with causation using bidirectional encoder representations from transformers models (i.e. BERT) [20] (see Niki et al. [21]). As for improving the accuracy of the search for the creation of causal networks, it was found that while causal knowledge with similar causal parts had high similarity to the search sentences, unhelpful causal knowledge was also extracted with high similarity at the same time. Therefore, in the current study, each search result was adopted as an endpoint node after the authors manually determined whether it was useful or not. The cause of this is related to the learning model, from where the word embeddings were obtained, but it may also have been influenced by there being many phrases in the newspaper articles that modified them in addition to describing causal events; the word embeddings in these phrases were noisy and had a negative impact on the search results. This problem was not solved by Nishimura et al. [11], who used a similar technique; therefore the causal network was created by manually selecting the nodes. In the future, we would like to consider a method for searching for this problem by extracting only words or phrases that are important for the causal knowledge in newspaper articles.

Finally, the current study's method allowed us to extract events and causal knowledge that were not found in the conventional relation diagram, indicating the possibility of creating more extensive relation diagrams. Future improvements in the accuracy of and comparisons with other disaster reports will make it possible to classify not only different events between disasters but also cases where the same event has a different causal relationship. It is hoped that such arrangements will be useful for planning disaster management policies and disaster management education. In addition, it is expected that a disaster causation database can be developed as part of a system to support the decision-making of disaster management personnel by presenting the next linked events in real time from the events occurring at the time of a disaster.

Acknowledgments

We would like to thank Professor Takao Terano for his useful and constructive suggestions. This work was supported by JSPS KAKENHI (Grant Number 20H02410) and the Japan Science and Technology Agency PRESTO.

References

[1] Kajimatoshibosaikenkyukai, "Zishinbousai to anzentoshi", 1996. (in Japanese)

[2] C. Khoo, J. Kornfilt, R. N. Oddy and S. H. Myaeng, "Automatic Extraction of Cause-Effect Information from Newspaper Text without Knowledge-based Inferencing", Literary & Linguistic Computing, Vol. 13, No. 4, 1998, pp. 177-186.

[3] R. Girju. "Automatic Detection of Causal Relations for Question Answering" in ACL Workshop on Multilingual Summarization and Question Answering, 2003, pp. 76-83.

[4] H. Sakaji and S. Masuyama, "A Method of Extracting Sentences Including Causal Relations from Newspaper Articles", IEICE Trans. Inf.& Syst. (Japanese edition), Vol. J94-D, No. 8, 2011, pp. 1496-1506. (in Japanese)

[5] H. Sakaji, H. Sakai and S. Masuyama, "An Extraction Method of Causal Knowledge from Newspaper Corpus", J. Fac. Sci. Tech., Seikei Univ., Vol. 51, No. 2, 2014, pp. 23-28. (in Japanese with English abstract)

[6] H. Sakaji, R. Murono, H. Sakai, J. Bennett and K. Izumi, "Discovery of Rare Causal Knowledge from Financial Statement Summaries", in 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 602-608.

[7] F. Sakahira and U Hiroi, “Study on the Method of Creating a Real-Time Disaster Relation Diagram”, in Proceedings of the 21st Conference Japan Society for Disaster Information Studies, 2019, pp. 96-97. (in Japanese)

[8] T. Sato and M. Horita, “Assessing the Plausibility of Inference Based on Automated Construction of Causal Networks Using Web-mining”, Sociotechnica, Vol. 4, 2006, pp. 66-74. (in Japanese)

[9] H. Aono and M. Ota, “Construction of a Causal Network and Acquisition of Causal Knowledges by Searching Factors”, IPSJ SIG Technical Report, Vol. 2009-DBS-149, No. 9, 2009, pp. 1-8. (in Japanese with English abstract)

[10] H. Ishii, Q. MA, and M. Yoshikawa, “Causal Network Construction using SVO Structure”, IPSJ SIG Technical Report, Vol. 2009-DBS-149, No. 10, 2009, pp. 1-8. (in Japanese with English abstract)

[11] K. Nishimura, H. Sakaji and K. Izumi, “Creation of Causal Relation Network using Semantic Similarity”, in The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018, pp. 1-4. (in Japanese with English abstract)

[12] T. Fujimori, M. Koyama and J. Kiyono, “A Study of Circumstances of Disaster Victims according to Multiple Attributes using Text Mining Method for Newspaper Articles Related to The 2011 Great East Japan Earthquake”, Journal of social safety science, No. 23, 2014, pp. 55-64. (in Japanese with English abstract)

[13] H. Kato, N. Nojima, M. Koyama and K. Tanaka, “Text Mining of Newspaper Articles Related to Lifeline Damage in the 2016 Kumamoto Earthquake: Comparison of Reginal and National Daily Newspapers”, Journal of Japan Society of Civil Engineers, Ser. A1 (Structural Engineering & Earthquake Engineering (SE/EE)), Vol. 75. No. 4, 2019, pp. I_443-I_453. (in Japanese)

[14] The Asahi Shimbun Company, available: http://www.asahi.com/information/cd/gakujutsu.html/

[15] The National Institute of Information and Communications Technology, available: http://compling.hss.ntu.edu.sg/wnja/

[16] A. Kobayashi, S. Masuyama and S. Sekine, “A Method for Automatic Ontology Construction Using Wikipedia”, IEICE Trans. Inf.& Syst. (Japanese edition), Vol. J93-D, No. 12, 2010, pp. 2597-2609. (in Japanese)

[17] CaboCha: Yet another Japanese Dependency Structure Analyzer, available: https://taku910.github.io/cabocha/

[18] MeCab: Yet another Part-of-Speech and Morphological Analyzer, available: https://taku910.github.io/mecab/

[19] GiNZA-Japanese NLP Library, available: https://megagonlabs.github.io/ginza/

[20] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, in Proceedings of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol. 1, 2019, pages 4171–4186.

[21] Y. Niki, H. Sakaji, K. Izumi and H. Matsushima, “Further Pretraining BERT for Causality Existence Classification in Financial Domain”, in the 34th Annual Conference of the Japanese Society for Artificial Intelligence, 2020, pp. 1-4. (in Japanese with English abstract)

Control of Hydraulic Sea floor Drill Rig Magazines Based on Finite-State Machine

Junxiang Wang* **, Hao Sun* Zhengdong Zhu* **, Ying Zhou* **, Lan Jiang* **, Li Yuan* **
* School of Automation, China University of Geosciences
Wuhan 430074, P. R. China
**Hubei Key Laboratory of Advanced Control and Intelligent Automation for Complex Systems
Wuhan 430074, P. R. China

Abstract

In order to make the hydraulic sea floor drill rig magazines to automatic management, an electro-magnetic valve control approach based on finite-state machine is proposed. According to the persistence and repeatability of the magazine rotational movement characteristic, the magazine rotating is divided into four states by acting time. With the state of swinging, catching and releasing of the loading arms, an information database for cores storage will refresh. The finite-state machine is used to control electromagnetic valve, and a state-flow module structure is built in the controller, making magazine to act to achieve the drill pipe storage. Finally, the experiment using the magazine prototype verified the feasibility of this method.

Keywords: Finite-state machine; Sea floor drill rig; Core barrel magazine

1. INTRODUCTION

As a kind of drilling equipment that works completely on the bottom of the sea, sea floor drilling rig has become indispensable and important for a variety of research targets in marine sciences including gas hydrates, ore formation, and paleoclimate. To expend the drilling depth for explore, many drilling rigs use the barrel magazine to store the drill pipes. The British Geological Survey operates two single barrel drill rigs. The 5 m rock drill was developed [1]. In 2006 the BGS developed a multi-barrel rig that could drill to a depth of 15 m and has now been upgraded for a drilling depth of 50 m [2].

When the sea floor drill rig is working at maximum of 2000 m blow the sea level by the length and strength of the umbilical. The rig is powered by hydraulic pumps that are driven with electric motors. The environment where sea floor drill rig works is complicated, it lacks sensors to feedback the rig's state. The conventional control method is based on the display of video cameras that mounted on the rig. It cases the delay of operating.

Finite-state machine (FSM) is a mathematical model that the number of states is limited and the state is transferred by events or conditions [3-5]. FSM was originally used for video game and film fields, then developed for robot program architecture. It is an abstract machine that can be in one of a finite number of states.

The conventional control method for magazine management requires the operators to control the two hydraulic feed cylinders sequentially, which costs time and concentration. And without a visible display of the information of the magazine. The operators need to record on paper, that makes it difficult to remember which pipe has cores in the magazine. In the Generally, the key in the control of magazine rotation is to control the electromagnetic valves correctly [6-7]. The component of valves movement is countable and certain just satisfy the require of the FSM. As the magazine is at a specific action at a certain time. The actions are not interfered. In this paper, a finite-state machine approach was proposed to achieve the automation of the drilling magazine. The magazine's movement were divided into four states. By the states transferred, the magazine finished the automatic rotation. A database was established for the information management including the continues rotating, the information for the existence of the drill cores and the serial number of the present position refreshing.

2. MECHANICAL STRUCTRURE

The barrel magazine management system is shown as the Fig. 1. There are three parts of the system: magazine, drilling mast and loading arms. All of them are driven by hydraulic feed cylinders. The barrel magazine has 10 storing holes, and each hole can store 2 drill pipes. While the rig is drilling on the sea floor,

the loading arms swing to inner position to grab a single pipe and move to the middle position. The drill head provides rotation and the necessary torque for screwing the drill pipe together and for the rotary drilling. This is used in combination with a stationary foot clamp to hold the drill string, and a rotating clamp on the drill head to assemble and break down the drill string. Then the loading arms swing to the middle position from the outer position to grab the drill pipe and put it back to magazine. The returned pipe is filled with the cores.

The barrel magazine is driven by two hydraulic feed cylinders which control the rotation and position of the magazine respectively. They are on the top of the magazine, which are vertical and horizontal respectively. The horizontal cylinder rotates the magazine and the vertical cylinder fixes the magazine. The cylinders are controlled by the three position four-way directional control electromagnetic valves. To rotate the magazine to next position, the operator should accomplish the four movement in sequence. A loading arm is used to grab the required pipe as well as for putting it back into the magazines for storage.

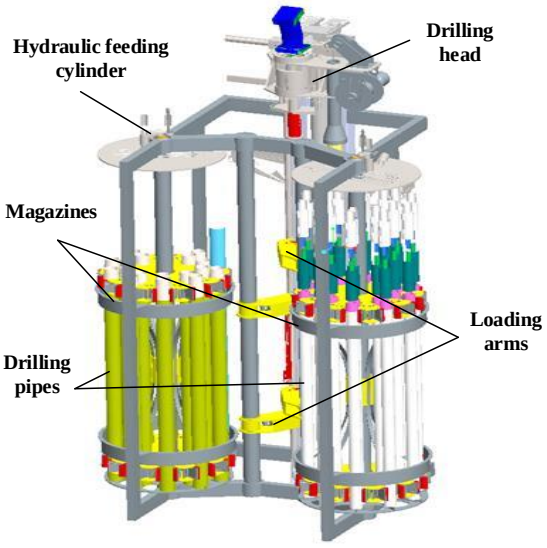


Fig. 1 The 3D module of drilling magazine

3. INFORMATION RECORDER

Conventionally, the only approach to record the magazine state is by the video signal. But as the rig's vibration, the video may be not clear. It will be difficult for them to record the information. So, the information recorder should be established to support recording the state information. At beginning of drilling process, the magazine is full of the hollow pipes. As the outer pipe is grabbed and taken by loading arm, the inner pipe will be pushed to the outer position for the next catching. This is used in combination with a stationary foot clamp to hold the drill string, and a rotating clamp on the drill head to assemble and break down the drill string. Until the drilling depth reaches the target depth, the drilling pipes are put back with drilling cores. The later drill core which is put back to the storing hole will push the former drill core to the inner position. An information database was set up to record the magazine state and the existence of the drill cores. The detail content was shown in Table. 1.

The drilling process can be visible and simplified by setting the database and the display of the magazine information. The database refreshes as the movement of the magazine and loading arm. The two loading arms have three positions and one common position which called middle position. When the loading arm is on the grabbing position, and it is moving to middle position from inner position, that means the hollow pipe is catch for drilling [8].

Table 1. Content of the database

Content	State
Loading arm position	Out/Mid/Inn
Loading arm state	Grabbing/Loosing
The present No. of the magazine	1-10
The existence of hollow pipes	Yes/No
The existence of drill cores	Yes/No

4. FSM CONTROL METHOD

In this paper, a finite-state machine (FSM) is used in a barrel drilling magazine control for the drilling process, due to the hydraulic cylinders of magazine has some limit state (left limit state, right limit state and stop). Each state of the cylinder behavior keeps similar in rotating. The control system of magazine always stays in a fixed state at any period. Besides, the trigger condition between two states is fixed too. Such a regular nature is consistent with the application condition of FSM [9]. According to the mathematical definition of FSM, the corresponding relation between magazine management and FSM's five elements (Q, Σ, δ, $q0$, F) of mathematical definition is described as below:

1. Q is a finite set of all states, which are corresponding to the hydraulic feed cylinder state, plugging,

unplugging, clockwise turning, counterclockwise turning. The Q is represented as formula (1).

$$Q_i = (Q_1\,[plugging], Q_2\,[unplugging], Q_3\,[clockwise\ turning], Q_4\,[counterclockwise\ turning]) \quad (1)$$

2. Σ is a finite set of all input symbols, which are corresponding to the signals of rotation (clockwise and counterclockwise). The Σ is represented as formula (2).

$$\Sigma_i = (\Sigma_1\,[clockwise], \Sigma_2\,[counterclockwise]) \quad (2)$$

3. δ: $Q \times \Sigma \rightarrow Q$ is the transfer function between different rotating states.
4. $q0$ is the starting state, defined as clockwise turning and unplugging in this paper.
5. F is accepting state sets. In this paper, $F = Q$. The F is represented as formula (3).

$$F_i = Q_i \times \delta(q0, \Sigma) \quad (3)$$

The control principle diagram of magazine is shown in the Fig.2. A state-flow module structure is built in the FSM controller, in which the transition conditions between different states are determined according to the starting signal and lasting time [10]. There are two starting signals "Turn clockwise" and "Turn counterclockwise". The initial state is "unplugging and clockwise". As the signal "Turn clockwise (Tu=1)" releases, the magazine turns to "Counterclockwise2". Lasting for 3 seconds, the hydraulic feed cylinder reaches the limit position. Then the magazine turns to "Counterclockwise1". Another 3 seconds, the plugging action finishes. After 4 states acting once, the magazine rotates to the next hole. When the starting signal is "Turn counterclockwise", the movement is similar. The state-flow diagram is created, as shown in Fig. 3.

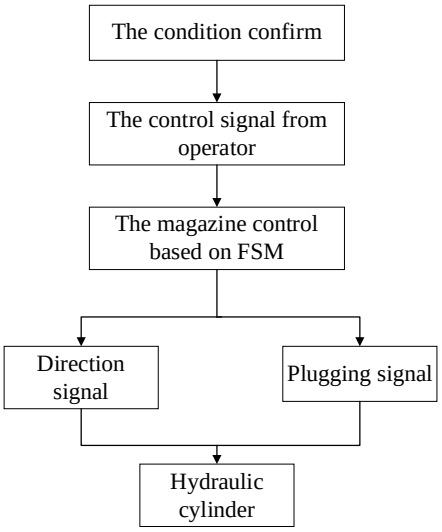


Fig. 2 The control principle diagram of magazine

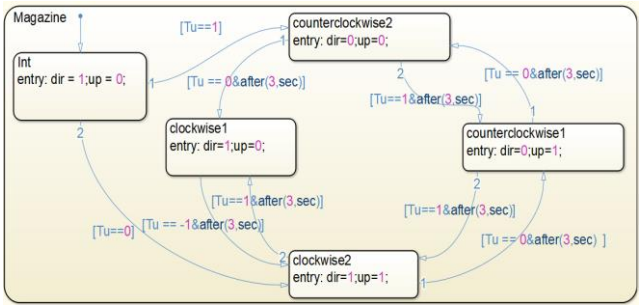


Fig. 3 State-flow diagram of magazine management

Table. 2 States of state-flow module

Dir/Up	0	1
0	Counterclockwise1	Counterclockwise2
1	Clockwise1	Clockwise2

5. SIMULATION AND EXPERIMENT

In the experiment, a step signal "1" presents the order signal "clockwise". On the contrary, the signal "0" present the order "counterclockwise". The state-flow module was run by MATLAB and got the output control signal in the Fig. 4. As the figure shows, in the first 0.2 second, the magazine turned to the initial state. Then the "clockwise" process got into the "counterclockwise1", "counterclockwise2", "clockwise2" and "clockwise1" states in sequence. Every state lasted for 3 seconds. It shows the magazine had finished the clockwise process. In the same way, the counterclockwise process finished in the different sequence.

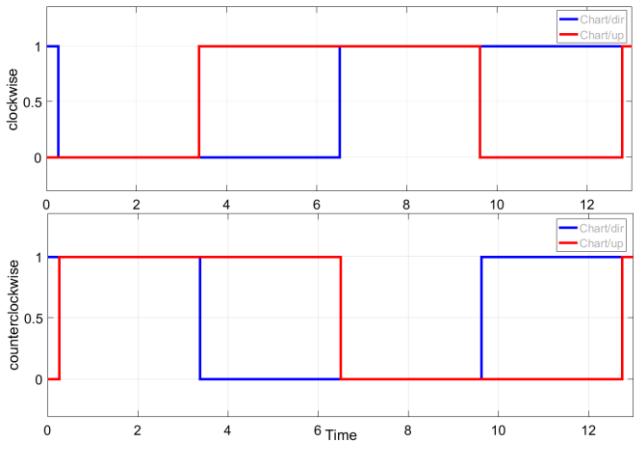


Fig. 4 The output control signal when the orders are "clockwise" and "counterclockwise"

Then a hydraulic powered drilling magazine is used (Fig. 5), and the control signals are provided by the host computer. The control models of four states are saved in the database of PLC. The relationship between hydraulic feed cylinder movement and

drilling magazine rotation has been described above. We operated as the actual drilling process to grab and put back the pipes, and rotate the magazine by the FSM. As a comparison, we control the hydraulic feed cylinder for single action to achieve the drilling process. Totally, 20 times of experiment are carried out for the drill rig. The average magazine rotation errors and time are shown in Table 3. As can be seen, the FSM control can reduce operating time up to 30% and achieve the automation of the magazine management. However, there are still something need to be improved, such as reducing the error rate and improving the condition feedback methods to ensure the action of the cylinder movement. The results of this study can guide the further optimization of control design for drilling magazines.

Fig. 5 The Experimental Drilling Rig

Table. 3 The statistics of the experiment

Control method	Average time	Error rate
Single cylinder control	18.3s	85%
FSM control	12.6s	100%

4. CONCLUSION

This paper analyzed the regularity of the barrel magazine under different command of the direction and times of rotation. And the magazine was divided into four states by hydraulic feed cylinder. A standard database was set up to provide the information and direction control signals sequence of the drilling magazine. An FSM model was established for achieving the automation of the drilling magazine management. The experiments verified the feasibility of this method.

References

[1] Freudenthal, T.Wefer, G. Drilling cores on the sea floor with the remote-controlled sea floor drilling rig MeBo, Geoscientific Instrumentation Methods and Data Systems, Vol. 2, No. 2, 2013, pp. 329-337.

[2] Spagnoli, G. Freudenthal, T. Underwater Drilling Rig for Offshore Geotechnical Explorations for Oil & Gas Structures, Oil Gas-European Magazine, Vol. 39, No. 4, 2013, pp. 185.

[3] Petrenko, A. Acm, Towards Testing from Finite State Machines with Symbolic Inputs and Outputs, 21st Acm/IEEE International Conference on Model Driven Engineering Languages and Systems.

[4] Hong, G. Bai, W. B. Yu, X. A kind of crane loading and unloading simulation based on finite-state Machine, Proceedings of the 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering, Vol. 39, 2015, pp. 318-322.

[5] Ebas, N. A. Hamzah, N. S. A. Jacob, K. Othman, F. S. Rusiman, M. S. Ahmad, N. Abdul-Kahar, R. Algebraic Properties of Finite Switchboard State Machine, Proceeding of the 25th National Symposium on Mathematical Sciences, 2018.

[6] Zhou, E. T. Yang, W. L. Lin, J. Z. Predictive Control of Hydraulic Winch Motion Control, 2009 2nd IEEE International Conference on Computer Science and Information Technology, Vol 4.

[7] Wei, G. Ji, L. Bracket Electro-hydraulic Control System Based on PLC Hydraulic Research, Modern Technologies in Materials, Mechanics and Intelligent Systems, Vol. 1049, 2014, pp. 1042-1-47.

[8] Gromov, M. L. Yevtushenko, N. V. Synthesis of Distinguishing Test Cases for Timed Finite State Machines, Programming and Computer Software, Vol. 36 No. 4, 2010, pp. 216-224.

[9] Xu, B. Shen, J. Liu, S. H. Su, Q.Zhang, J. H. Research and Development of Electro-hydraulic Control Valves Oriented to Industry 4.0: A Review, Chinese Journal of Mechanical Engineering, Vol. 33 No. 1, 2020.

[10] Sun HX, Wei X, Lin T (2012) Research of model transformation approaches based on Finite State Machine. Comput Technol Dev 22(2):10–17

Distributed Consensus Control for General Linear Multi-Agent Systems via a Dynamic Event-Triggered Strategy

Yifei Li*, Xiangdong Liu*, Changkun Du*, Haikuo Liu*, Pingli Lu* **, Ning Dong*
* School of Automation, Beijing Institute of Technology, Beijing 100081, China
** Corresponding author, e-mail: pinglilu@bit.edu.cn

Abstract

This paper puts forward a distributed dynamic event-triggered strategy to solve the distributed event-triggered consensus problem of linear multi-agent systems under directed graphs. Based on dynamic triggering function, each agent can reach consensus asymptotically. Different from existing static triggering schemes, the proposed dynamic triggering scheme, where an internal dynamic variable is involved, results in larger inter-event times and also leads to less communication overheads among agents, which is conducive to guaranteeing that Zeno behavior is excluded for each agent. In addition, under the proposed strategy, neither controller updates nor triggering threshold detections require continuous communication. Finally, the effectiveness of the theoretical analysis is demonstrated by numerical simulations.

Keywords: Multi-agent systems, dynamic event-triggered strategy, distributed control.

1. INTRODUCTION

In the past decades, the problem of coordinated control of multi-agent systems (MASs) has received increasing attention ([1], [2]). Modeling of MASs originates from social animals such as insects, fish, and birds, which benefits the accomplishment of the tasks that are difficult for individuals. MASs also exist in many engineering fields, such as distributed optimization ([3]), wireless sensor networks ([4]), mobile robot collaboration ([5]), drone/satellite formation flight ([6], [7]) and so on. Thus, MASs have enabled many scholars around the world to do a lot of research work in this field in highly interconnected present world (see [8] - [11] and the references therein).

In general, designing a suitable control strategy to make all agents' states reach a common quantity related to certain control performance is a pivotal issue on consensus of MASs. Most of the aforementioned works on consensus problems are obtained under the assumption that continuous communication among agents is available. However, it is difficult to maintain such an environment with continuous communication and unnecessary communication also leads to a waste of energy. To this end, distributed controllers with intermittent communication have been studied recently. In [12], a periodic sampling control protocol was studied, where the sampling is triggered after a fixed time interval. However, the controller still update periodically even after the control target has been achieved. In many cases, due to the limitations of energy supply or communication bandwidth, it is desired that information exchanges among agents only occur at some discrete and non-periodic sampling points, so event-triggered control schemes have been introduced. By viewing a triggered event as a moment when a certain measurement error exceeds a pre-designed threshold, the communication among agents is required under the event-triggered control strategy only when an event is triggered. The event-triggered consensus control problems for MASs with single- or double-integrator dynamics were investigated in [13] - [15]. Subsequently, plenty of scholars conducted research on event-triggered consensus control of general linear MASs. Event-triggered consensus control protocols were designed for general linear MASs over undirected and directed graphs in [16] and [17], respectively. However, in the above-mentioned works, event-triggered functions still need to continuously access to neighbors' state information. To solve this problem, considering general linear MASs under directed graphs, [18] designed a triggering threshold based on exponential function and [19] proposed a triggering threshold related to the sum of relative state estimations from itself and its neighbors for each agent.

It should be pointed out that all aforementioned results are obtained under the framework of static event-triggered control mechanisms. However, a new class of dynamic event-triggered control mechanisms, where internal dynamic variables are involved, have several merits with respect to the commonly studied static one including the significant larger inter-event time, which is beneficial to prevent Zeno behavior in practical application. Therefore, it has been widely investigated that the dynamic event-triggered control method has been used to solve consensus problems of MASs in recent years. [20] occupied internal dynamic variables in event-triggered control for nonlinear systems. [21] improved the form of the dynamic event-triggered mechanism in a distributed manner and extended it to a single-integrator MAS. Based on this work, [22] used internal dynamic variables in self-triggered control to

overcome the drawback of continuous sensing and listening of the triggering. In [23], a dynamic event-triggered communication mechanism was used to address a distributed resource-efficient formation control problem of a networked MAS with general linear system dynamics. In [24], the dynamic average consensus problem was solved by the proposed dynamic event-triggered algorithm, which was robust to network disruptions. However, all interaction topologies among agents in the above-mentioned works are assumed as undirected graphs. In fact, it's very meaningful and practical to investigate the consensus problem over directed graphs. Motived by the previous works, we discuss the dynamic event-triggered consensus problem for general linear MASs on directed graphs.

In a word, we will discuss the dynamic event-triggered consensus problem of general linear MASs under directed graphs and exhibit Zeno behavior in this paper. The principal contributions are summarized as follows:

- Compared to the static event-triggered mechanism, such as [17] - [19], the dynamic event-triggered function with an internal dynamic variable proposed in this paper yields the larger triggering intervals, which benefits the exclusion of Zeno behavior in practice. Moreover, the communication instants are reduced significantly which also saves the communication energy greatly.
- Different from most of the existing works on dynamic triggering mechanisms, which mainly focus on the integrator-type dynamics and undirected graphs ([20] - [24]), this paper investigates the consensus problem with dynamic event-triggered strategy for general linear MASs on directed graphs, which in turn poses more challenges in the consensus stability analysis and Zeno behavior exclusion due to the more general agents' models and more complex communication topologies.
- The issue that continuous access to neighbors' states is still required in agent's own triggering detection is ignored in many existing works on both static and dynamic triggering mechanisms (see [13], [16], [17], [20]), which brings about a paradox to the original purpose of saving communication energy by introducing the event-triggered strategy. In this endeavor, this paper aims to avoid the continuous communication in not only the controller update but also the triggering detection, which poses more challenges in the triggering function design under the framework of the dynamic event-triggered strategy.

The rest of this paper is organized as follows. Some preliminaries including useful knowledge and the dynamics are introduced in Section 2. Section 3 presents the main result and Section 4 illustrates the result through simulation examples. Section 5 concludes the paper.

Notation: Let R be the set of real numbers and $R^{m\times n}$ be the set of $\mathrm{m}\times\mathrm{n}$ real matrices, respectively. $\mathcal{I}$ is a set of positive integers and $\mathcal{I}_N=\{1,2,\dots,N\}$. 0_{N} and 1_{N} mean the $N\times 1$ column vector of all zeros and ones, respectively. For a vector $x\in \mathrm{R}^{\mathrm{n}}$, $x>0$ means that every entry $x_i>0$ with $i=1,2,\dots,n$. For a symmetric matrix P, $P>0$ means that P is positive definite and $\lambda_{\max}(P)$ $\left(\lambda_{\min}(P)\right)$ means the maximum (minimum) eigenvalues of P. The superscript T and the symbol $\otimes$ represent the transpose for matrices and the Kronecker product, respectively. Denote $|\cdot|$ as the Euclidean norm for vectors and the induced 2-norm for matrices.

2. PRELIMINARIES

In this section, we introduce some definitions in algebraic graph theory and the considered MAS briefly.

2.1 Graph Theory

Consider a group of MASs with N agents. A directed graph $\mathcal{G}=(\mathcal{V},\mathcal{E},\mathcal{A})$ consists of a nonempty finite node set $\mathcal{V}=\{v_1,\dots,v_N\}$, an edge set $\mathcal{E}\in(\mathcal{V}\times\mathcal{V})$ and a weighted adjacency matrix $\mathcal{A}=[a_{ij}]\in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$. The edge $(v_i,v_j)\in\mathcal{E}$ indicates that the node v_j can receive information from the node $\mathrm{v_i}$ or the node v_i can broadcast information to the node v_j. The neighbor set of node v_i is denoted by $\mathcal{N}_i=\{v_j\in\mathcal{V}:(v_j,v_i)\in\mathcal{E}\}$. The adjacency matrix $\mathcal{A}$ of a directed graph is given by $a_{ii}=0$, $a_{ij}>0$ if $(v_j,v_i)\in\mathcal{E}$, and $a_{ij}=0$ otherwise. The Laplacian matrix of $\mathcal{G}$ is defined as $\mathcal{L}=[l_{ij}]\in \mathrm{R}^{\mathrm{N}\times\mathrm{N}}$, where $l_{ii}=\sum_{\mathrm{j}=1}^{\mathrm{N}}a_{ij}$, $l_{ij}=-a_{ij}$ with $i\neq j$. If between any pair of distinct nodes v_i and v_j in a directed graph $\mathcal{G}$, there exists a directed path from v_i to v_j, $i,j=1,2,\dots,N$, $\mathcal{G}$ is strongly connected. For the purpose of drawing forth our main result, we need the following assumptions and lemmas.

Assumption1: The directed graph $\mathcal{G}$ is strongly connected.

Lemma1: ([25]) The general algebraic connectivity of a strongly connected graph $\mathcal{G}$ associated with the Laplacian matrix $\mathcal{L}$ is defined by $a(\mathcal{L})=\min\limits_{r^Tx=0,x\neq 0}\frac{x^T\tilde{\mathcal{L}}x}{x^TRx}$,

where $\tilde{\mathcal{L}}=\frac{1}{2}(R\mathcal{L}+\mathcal{L}R^T)$, $R=\mathrm{diag}(r_1,\dots,r_N)$ with $\mathbf{r}=(r_1,\dots,r_N)^T$ satisfying $\mathbf{r}^T\mathcal{L}=\mathbf{0_N}$ and $\sum_{i=1}^{N}r_i=1$.

Lemma2: ([25]) Given any $x,y\in R^N$, Young's inequality

states that for any $\psi > 0$, $\mathrm{x^T y} \le \frac{\mathrm{x^T x}}{2\psi} + \frac{\psi \mathrm{y^T y}}{2}$.

Lemma3: ([26]) For a strongly connected directed graph $\mathcal{G}$, zero is a simple eigenvalue of $\mathcal{L}$ with the corresponding right eigenvector $\mathbf{1}_\mathrm{N}$, that is $\mathcal{L}\mathbf{1}_\mathrm{N} = 0$, and there exists a positive vector $\mathbf{r} = (r_1, \dots, r_N)^T$ satisfy $\mathbf{r^T 1}_\mathrm{N} = 1$ such that $\mathbf{r^T}\mathcal{L} = \mathbf{0}_N$.

2.2 Multi-Agent System Model

Consider a linear MAS consisting of N identical agents, indexed by $1, \dots, N$. The dynamics of the ith agent is described by

$$\dot{x}_i(t) = Ax_i(t) + Bu_i(t), \quad i \in \mathcal{I}_\mathcal{N}, \tag{1}$$

where $x_i(t) \in \mathrm{R}^n$ and $u_i(t) \in \mathrm{R}^p$ are the agent state and the control input, respectively. $A \in \mathrm{R}^{n\times n}$ and $B \in \mathrm{R}^{n\times p}$.

Assumption2: The matrix pair (A, B) in (1) is stabilizable.

The objective of this paper is to design a distributed dynamic event-triggered consensus strategy for each agent such that the states of all agents achieve consensus while avoiding Zeno behavior.

3. MAIN RESULT

In this section, a dynamic event-triggered control strategy will be proposed to deal with consensus problems for the linear MASs under strongly connected directed graphs without Zeno behavior.

A dynamic event-triggered consensus control protocol is proposed for each agent as follows:

$$u_i(t) = cKz_i(t), \tag{2}$$

where $c > 0$, and the feedback gain matrix $K \in \mathrm{R}^{p\times n}$ is chosen by $K = -B^T P$ with a positive matrix P to be decided. Moreover,

$$z_i(t) = \sum_{j\in\mathcal{N}_i} a_{ij}\left(e^{A\left(t-t^i_{k_i}\right)}x_i\left(t^i_{k_i}\right) - e^{A\left(t-t^j_{k_j}\right)}x_j\left(t^j_{k_j}\right)\right),$$

where $\mathrm{k_i} \in \mathcal{I}_\mathcal{N}$, a_{ij} is the ijth entry of the adjacency matrix $\mathcal{A}$, $t^i_{k_i}$ and $x_i\left(t^i_{k_i}\right)$ are the latest event-triggered time and the latest broadcast state of agent i, respectively. For the sake of simplicity, we denote $e^{A\left(t-t^i_{k_i}\right)}x_i\left(t^i_{k_i}\right)$ as $\overline{x_i}(t)$, so the measurement error is defined as

$$e_i(t) = e^{A\left(t-t^i_{k_i}\right)}x_i\left(t^i_{k_i}\right) - x_i(t) = \overline{x_i}(t) - x_i(t).$$

Then, we define $\delta(t) = x(t) - (1_N r^T \otimes I_n)x(t) = (M \otimes I_n)x(t)$ as the disagreement vector with $M = \left(I_N - 1_N r^T\right)$.

Therefore, the closed-loop form of the MAS and the disagreement vector can be expressed as

$$\dot{x}(t) = (I_N \otimes A + c\mathcal{L} \otimes BK)x(t) + (c\mathcal{L} \otimes BK)e(t), \tag{3}$$

$$\dot{\delta}(t) = (I_N \otimes A + c\mathcal{L} \otimes BK)\delta(t) + (c\mathcal{L} \otimes BK)e(t), \tag{4}$$

where $x(t) = [x_1^T(t), \dots, x_N^T(t)]^T$, $e(t) = [e_1^T(t), \dots, e_N^T(t)]^T$, $\delta(t) = [\delta_1^T(t), \delta_2^T(t), \dots, \delta_N^T(t)]$.

Now, we introduce an internal dynamic variable satisfying

$$\dot{\eta}_i(t) = -\mu_i\eta_i(t) + \xi_i\left(\frac{\epsilon}{\alpha^*}|z_i(t)|^2 - |e_i(t)|^2\right), \tag{5}$$

Where $i \in \mathcal{I}_\mathcal{N}$, $\eta(0) > 0$ and the parameters satisfy $\mu_i > 0$, $\epsilon > 0$, $\alpha^* > \kappa|\mathcal{L}|^2$ and $\xi_i \in [0, \gamma]$ with

$$\kappa = \left(1 - \frac{1}{\alpha}\right)\|M\|^2 + \frac{c}{\beta}\lambda_{max}(\mathcal{L}^T R^2 \mathcal{L} \otimes PBB^T P) + 2c\|M^T R\mathcal{L} \otimes PBB^T P\|,$$

where R defined in Lemma1. Moreover, α, β are the Young inequality parameters satisfying $0 < \alpha < 1$ and $\beta > 0$.

Moreover, inspired by [22], we assume that the first triggering time $t^i_1 = 0$, so the triggering times $\{\mathrm{t^i_k}\}_{\mathrm{k=2}}^{\infty}$ is determined by

$$t^i_{k_i+1} = \max_{r\ge t^i_{k_i}}\left\{\left(\|e_i\|^2 \le \frac{\epsilon}{\alpha^*}\|z_i(t)\|^2 + \frac{\eta_i(t)}{\theta_i}, \forall t\left[t^i_{k_i}, r\right]\right\}, \tag{6}$$

where $\theta_i > \frac{\kappa - \xi_i}{\mu_i}$.

Remark1: The proposed mechanism (6) is called the dynamic event-triggered mechanism since it involves an internal variable $\eta_\mathrm{i}(\mathrm{t})$. If setting $\eta_\mathrm{i}(\mathrm{t}) \equiv 0$, we can get the static event-triggered mechanism (7). In addition, it can also be seen a limit case of the dynamic triggering mechanism (6) when the parameter θ_i goes larger enough.

$$\mathrm{t^i_{k_i+1}} = \max_{\mathrm{r\ge t^i_{k_i}}}\{\|e_i\| \le \sqrt{\frac{\epsilon}{\alpha^*}}\|z_i(t)\|, \forall t \in \left[t^i_{k_i}, r\right]\}. \tag{7}$$

Compared with (6), (7) does not involve any extra dynamic variables except $x_i(t), \overline{x_i}(t)$ and $\overline{x_j}(t), j \in \mathcal{N}_i$.

Next, we present the following theorem to cope with the dynamic consensus problem.

Theorem1: Consider the linear MAS (1) and suppose that Assumptions 1 and 2 hold. Under the proposed distributed dynamic event-triggered consensus protocol composed of controller (2) and the dynamic triggering mechanism (6) the event-triggered consensus problem can be solved for any initial states if the parameters c, ϵ, α, β are selected such that

$$\frac{c\beta}{1-\alpha}\lambda_{max}(PBB^T P) + \frac{\epsilon}{1-\alpha} + \lambda_{max}(R \otimes Q) < 0,$$

where $Q = PA + A^T P - 2ca(\mathcal{L})PBB^T P < 0$. In addition, Zeno behavior can be excluded.

Proof: According to (5) and (6), one has $\dot{\eta}_i(t) \ge -\left(\mu_i + \frac{\xi_i}{\theta_i}\right)\eta_i(t)$. So it is easy to get

$$\eta_i(t) > \eta_i(0)e^{-\left(\mu_i + \frac{\xi_i}{\theta_i}\right)t} > 0. \tag{8}$$

Therefore, considering the dynamic triggering function, we choose the Lyapunov function as follows.

$$W = \delta^T(R \otimes P)\delta + \sum_{i=1}^{N} \eta_i(t). \quad (9)$$

The time derivative of W along the closed-loop system (4) is given by

$$\dot{W} = \delta^T(R \otimes (PA + A^TP) - c(R\mathcal{L} + \mathcal{L}^T R) \otimes PBB^TP) - 2\delta^T(cR\mathcal{L} \otimes PBB^TP)e + \sum_{i=1}^{N} \dot{\eta}_i(t) \quad (10)$$

It follows from Lemma 1 and Lemma 2 and substituting (6) into (11) yields that

$$\dot{W} \le [\lambda_{max}(R \otimes Q)(1-\alpha) + c\beta\lambda_{max}(PBB^TP)]\|\delta\|^2 + \sum_{i=1}^{N}(\kappa - \xi_i)\|e_i(t)\|^2 + \sum_{i=1}^{N}\xi_i \frac{\epsilon}{\alpha^*}\|z_i(t)\|^2 - \sum_{i=1}^{N}\mu_i\eta_i(t)$$

Since the fact that $z = (\mathcal{L} \otimes I_n)\bar{x} = (\mathcal{L} \otimes I_n)\bar{\delta}$, we have $\|z\|^2 \le \|\mathcal{L}\|^2\|\bar{\delta}\|^2$. Then, according to the triggering function (6) and (12) can be rewritten as

$$\dot{W} \le [\lambda_{max}(R \otimes Q)(1-\alpha) + c\beta\lambda_{max}(PBB^TP) + \epsilon]\|\delta\|^2 - \sum_{i=1}^{N}\left(\mu_i - \frac{\kappa - \xi_i}{\theta_i}\right)\eta_i(t) < 0 \quad (11)$$

Therefore, we can conclude that the disagreement vector $\delta \to \mathbf{0}$ as $t \to \infty$, which means the MAS (1) can achieve the consensus asymptotically.

Now, we prove that Zeno behavior is strictly ruled out for each agent. Firstly, suppose that Zeno behavior is existed, which implies that there exists an agent i, such that $lim\, 1_{k_i \to +\infty} t_{k_i}^i = T_0$, where T_0 is a positive constant.

Let $\varepsilon_0 = \frac{1}{2\|A\|} ln\left(\frac{1}{\varpi}\sqrt{\frac{\eta_i(0)}{\theta_i}}e^{-\frac{1}{2}\left(\mu_i + \frac{\xi_i}{\theta_i}\right)T_0} + 1\right) > 0$, where $\varpi = \frac{\|A\|}{c\rho\|BK\|}$.

Then according to the property of limits, there exists a positive integer $N(\varepsilon_0)$ such that

$$t_{k_i}^i \in [T_0 - \varepsilon_0, T_0], \forall k_i \ge N(\varepsilon_0). \quad (12)$$

Noting that (8) holds, we can conclude that one sufficient condition to guarantee that the inequality in (6) holds is

$$\|e_i(t)\| \le \sqrt{\frac{\eta_i(0)}{\theta_i}}e^{-\frac{1}{2}\left(\sigma_i + \frac{\xi_i}{\theta_i}\right)t}$$

It follows from the fact that the interval between two consecutive triggering events is bounded, so it is apparent that $e^{A\left(t - t_{k_i}^i\right)}$ is bounded for $\forall t \in \left[t_{k_i}^i, t_{k_i+1}^i\right)$. In light of (3), it is not challenging to verify that $(\mathbf{r}^T \otimes e^{-At})x$ is an invariant quantity. Therefore, deriving from $x_i = \delta_i + \sum_{j=1}^{N} r_i x_j$, we obtain that $x(t)$ is finite for any finite t. Thus, we can get that for $\forall t \in \left[t_{k_i}^i, t_{k_i+1}^i\right)$, the triggering error $e(t)$ is also bounded. According to the fact $z = (\mathcal{L} \otimes I_n)(x + e) = (\mathcal{L} \otimes I_n)(\delta + e)$, we know z is bounded. Thus, we use ρ to denote the upper bound of $|z_i(t)|$. Based on (1), it can be obtained that

$$\dot{e}_i(t) = Ae_i(t) - cBKz_i(t). \quad (13)$$

Next, based on the fact that the measurement error is reset to zero once an event is triggered, the solution of (13) follows that $\|e_i(t)\| \le \frac{c\|BK\|}{\|A\|}\rho\left(e^{\|A\|\left(t - t_{k_i}^i\right)} - 1\right)$.

Thus, it can be concluded that one sufficient condition to guarantee that the above inequality holds is

$$\frac{c\|BK\|}{\|A\|}\rho\left(e^{\|A\|\left(t - t_{k_i}^i\right)} - 1\right) \le \sqrt{\frac{\eta_i(0)}{\theta_i}}e^{-\frac{1}{2}\left(\mu_i + \frac{\xi_i}{\theta_i}\right)t}.$$

Now suppose that $t_{N(\varepsilon_0+1)}^i$ denote the next triggering time determined by (6). Then one gets

$$t_{N(\varepsilon_0+1)}^i - t_{N(\varepsilon_0)}^i \ge \frac{1}{\|A\|} ln\left(\frac{1}{\varpi}\sqrt{\frac{\eta_i(0)}{\theta_i}}e^{-\frac{1}{2}\left(\mu_i + \frac{\xi_i}{\theta_i}\right)T_0} + 1\right) = 2\varepsilon_0,$$

which contradicts to (12). Therefore, Zeno behavior is excluded for each agent.

Thus, the proof is accomplished.

Remark 2: For arbitrary (A, B) satisfying **Assumption 2** one can always find a positive definite matrix P such that $Q < 0$ holds. Moreover, the existence of parameters in **Theorem 1** can be guaranteed by selecting parameters more conservatively off-line as long as they satisfy their bounds. Therefore, the proposed dynamic triggering scheme is implementable.

4. SIMULATION EXAMPLE

In this section, we demonstrate the theoretical result by the following numerical example and make the comparison between the dynamic event-triggered control strategy and traditional static one. Consider a group of 6 agents with general linear dynamics (1) with $A = [0\ 7; -1\ 1]$, $B = [2;\ 1]$. The directed graph is shown in Fig.1.

The initial states are given by $x_1(0) = [-30;\ 20], x_2(0) = [15;\ 60], x_3(0) = [10;\ -26], x_4(t) = [3;\ 43], x_5(t) = [13;\ 65], x_6(t) = [28; -30]$. Choose the feedback gain matrix $K = [-0.9850\ \ -0.7367]$. And the parameters are selected as $c = 3.7788, \alpha^* = 10$, $\epsilon = 1$. Moreover, according to the dynamic triggering scheme (6), we choose $\eta_1(0) = \eta_3(0) = \eta_4(0) = \eta_6(0) = 8$, $\eta_2(0) = 3, \eta_5(0) = 5$, $\mu_i = \xi_i = 60$, and $\theta_i = 2$, where $i = 1,2, \ldots 6$.

The simulation results are shown in Fig.2-Fig.5. Fig.2 depicts state evolutions under the dynamic triggering scheme and the static one, respectively. The trajectories of dynamic variable $\eta_i(t)$ is present in Fig.3. Fig.4 shows the evolution of each agent's triggering error with respect to the threshold under the proposed dynamic triggering scheme. These two figures imply that the dynamic variable, tracking errors and corresponding thresholds all converge to zero asymptotically. The corresponding triggering instants under dynamic and static triggering schemes are shown in Fig.5. For a clearer comparison, we record the

triggering numbers for each agent with the dynamic and static triggering schemes in Table 1, which can be obtained that the triggering numbers are greatly reduced under the proposed dynamic triggering scheme.

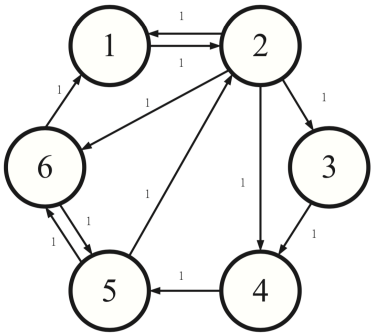
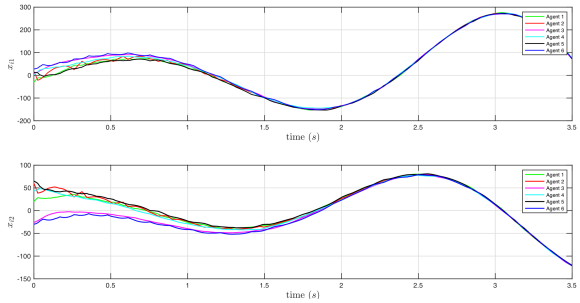
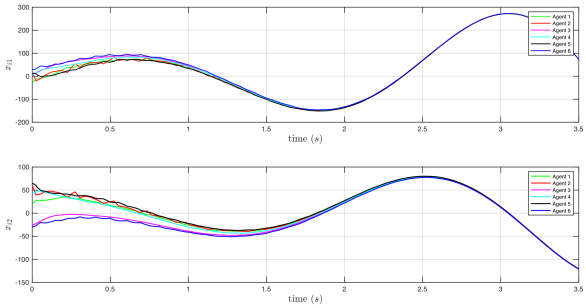


Fig. 1 The communication graph among the agents.



(a) State evolutions of MAS under dynamic triggering scheme (6).



(b) State evolutions of MAS under static triggering scheme (7).

Fig.2 State evolutions under two event-triggered schemes

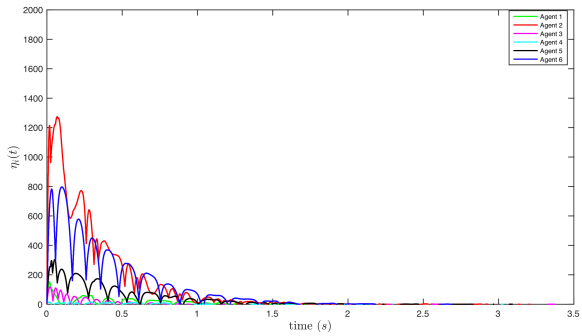


Fig.3 The dynamic variable $\eta_i(t)$ given in (5) for dynamic trigging scheme (6).

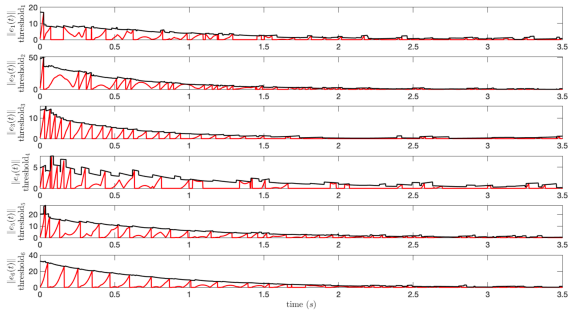
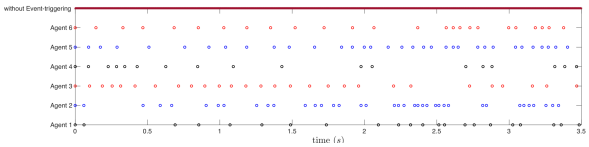
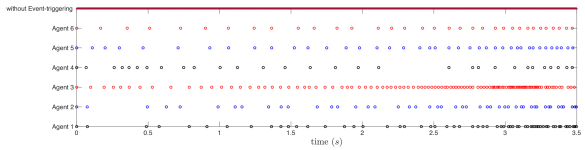


Fig.4 Triggering errors and thresholds for each agent under dynamic triggering scheme (6).



(a) Triggering times under dynamic triggering scheme (6).



(b) Triggering times under static triggering scheme (7).

Fig.5 Triggering times for each agent under two schemes and comparison with the case without event-triggered strategy

Table 1. Triggering numbers on each agent under different triggering schemes.

Agent index i	1	2	3	4	5	6
Dynamic triggering scheme	22	38	28	18	34	25
Static triggering scheme	59	47	109	30	40	34

4. CONCLUSION

The consensus problem for linear MASs under directed graphs has been investigated. A distributed dynamic event-triggered control strategy along with a dynamic triggering function has been addressed. Under the proposed dynamic event-triggered control strategy, there is no need for continuous communication in either controller update or triggering condition monitoring. In addition, each agent does not exhibit Zeno behavior by proving the event time interval is strictly positive. Our future research will be devoted to extending the event-triggered consensus problem for heterogenous MASs under directed graphs. At the same time, it is also necessary to further consider the case when MASs contain some uncertainties or external

disturbances.

References

[1] R. Olfati-Saber and R. M. Murray. Consensus problems in networks of agents with switching topology and time- delays. IEEE Transactions on Automatic Control 2004; 49(9): 1520-1533.

[2] W. Ren and R. W. Beard. Distributed consensus in multi- vehicle cooperative control. London: Springer-Verlag; 2008.

[3] H. Li, C. Huang, G. Chen, X. Liao, and T. Huang. Distributed Consensus Optimization in Multiagent Net- works With Time-Varying Directed Topologies and Quantized Communication. IEEE Transactions on Cybernetics 2017; 47(8): 2044 - 2057.

[4] D. Ding, Z. Wang, D. W. C. Ho, and G. Wei. Observer-Based Event-Triggering Consensus Control for Multi-agent Systems With Lossy Sensors and Cyber-Attacks. IEEE Transactions on Cybernetics 2017; 47(8): 1936- 1947.

[5] Z. Meng, Z. Lin, and W. Ren. Robust cooperative tracking for multiple non-identical second-order nonlinear systems. Automatica 2013; 49(8): 2363-2372.

[6] B. D. O. Anderson, C. Yu, and S. Dasgupta. Control of a three-coleader formation in the plane. Systems & Control Letters 2007; 56(9-10): 573-578.

[7] J. A. Marshall, M. E. Broucke, and B. A. Francis. Pursuit formations of unicycles. Automatica 2006; 42(1): 3-12.

[8] J. Huang, L. Chen, and X. Xie. Distributed Event-triggered Consensus Control for Heterogeneous Multi- agent Systems under Fixed and Switching Topologies. International Journal of Control Automation and Systems 2019; 17(8): 1945-1956.

[9] H. Wang, W. Yu, and W. Ren. Distributed Adaptive Finite-Time Consensus for Second-Order Multiagent Systems With Mismatched Disturbances Under Directed Networks. IEEE Transactions on Cybernetics early access; DOI: 10.1109/TCYB.2019.2903218, 1-12.

[10] L. Ding, Q. L. Han, and X. Ge. An overview of recent advances in event-triggered consensus of multiagent systems. IEEE Transactions on Cybernetics 2018; 48(4): 1110-1123.

[11] X. Tan, J. Cao, and X. Li. Consensus of Leader-Following Multiagent Systems: A Distributed Event-Triggered Impulsive Control Strategy. IEEE Transactions on Cybernetics 2018; 49(3): 792-801.

[12] G. Wen, Z. Duan, and W. Yu. Consensus of multi-agent systems with nonlinear dynamics and sampled-data information: A delayed-input approach. International Journal of Robust and Nonlinear Control 2013; 23(6): 602-619.

[13] D. V. Dimarogonas, E. Frazzoli, and K. H. Johansson. Distributed Event-Triggered Control for Multi-Agent Systems. IEEE Transactions on Automatic Control 2012; 57(5): 1291-1297.

[14] G. S. Seyboth, D. V. Dimarogonas, and K. H. Johansson. Event-based broadcasting for multi-agent average con- sensus. Automatica 2013; 49(1): 245-252.

[15] C. Nowzari and J. Cortes. Distributed event-triggered coordination for average consensus on weight-balanced digraphs. Automatica 2016; 68(1): 237-244.

[16] Z. Zhang, F. Hao, and L. Zhang. Consensus of linear multi- agent systems via event-triggered control. International Journal of Control 2014; 87(6): 1243-1251.

[17] W. Zhu, Z. P. Jiang, and G. Feng. Event-based consensus of multi-agent systems with general linear models. Automatica 2014; 50(2): 552-558.

[18] D. Yang, W. Ren, and X. Liu. Decentralized event-triggered consensus for linear multi-agent systems under general directed graphs. Automatica 2016; 69: 242-249.

[19] X. Liu, C. Du, and H. Liu. Distribute event-triggered consensus control with fully continuous communication free for general linear multi-agent systems under directed graph. International Journal of Robust and Nonlinear Control 2018; 28(1): 132-143.

[20] Girard, Antoine. Dynamic Triggering Mechanisms for Event-Triggered Control. IEEE Transactions on Automatic Control 2015; 60(7): 1992-1997.

[21] X. Yi, K. Liu, and D. V. Dimarogonas. Distributed dynamic event-triggered control for multi-agent systems. In 56th IEEE Conference on Decision and Control (CDC) 2017; 6683-6698.

[22] X. Yi, K. Liu, D. V. Dimarogonas, and K. H. Johansson. Dynamic Event-triggered and Self-Triggered Control for Multi-agent systems. IEEE Transactions on Automatic Control 2019; 64(8): 3300-3307.

[23] X. Ge and Q. Han. Distributed formation control of net- worked multi-agent systems using a dynamic event- triggered communication mechanism. IEEE Transactions on Industrial Electronics 2017; 64(10): 8118-8127.

[24] J. George, X. Yi, and T. Yang. Distributed Robust Dynamic Average Consensus with Dynamic Event-Triggered Communication. In: 2018 IEEE Conference on Decision and Control; 2018; 434-439.

[25] G. H. Hardy, J. L. Littlewood, and G. Polya. Inequalities. Cambridge University, 1952.

[26] W. Yu, G. Chen, and M. Cao. Second-Order Consensus for Multiagent Systems with Directed Topologies and Nonlinear Dynamics. IEEE Transactions on Systems Man and Cybernetics-Part B: Cybernetics 2010; 40(3): 881-891.

Distributed Dynamic Event-Triggered Output Tracking for Heterogeneous Multi-Agent Systems

Changkun Du*, Haikuo Liu*, Yougang Bian**, Pingli Lu*,†, Xiangdong Liu*
* School of Automation, Beijing Institute of Technology, Beijing 100081, China
** State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body,
College of Mechanical and Vehicle Engineering,
Hunan University, Changsha 410082, China
† Correspinding author: pinglilu@bit.edu.cn

Abstract

This paper considers the event-triggered output tracking problem for heterogeneous multi-agent systems with general linear dynamics and an input-bounded leader. The leader considered here has a nonzero and time-varying control input which is unavailable to any followers. In order to achieve output tracking, a hierarchical framework is designed based on the internal model principle. Based on this treatment, a novel distributed dynamic event-triggered control protocol is further proposed such that the output errors between the followers and the leader can converge to zero asymptotically. Different from the existing static event-triggered mechanism, an internal dynamic variable is involved in the proposed dynamic triggering law, which benefits the exclusion of the Zeno behavior. In addition, no continuous communication is needed in either controller updates or triggering detection, which offers more feasibilities in real applications. Finally, numerical simulations are given to illustrate the effectiveness of the theoretical results.

Keywords: Heterogeneous multi-agent systems, output tracking, dynamic event-triggered mechanism, input-bounded leader.

1. INTRODUCTION

Along with the technology development on network communication, embedded computing, environment perception/sensing, and artificial intelligence, the coordination of multi-agent systems (MASs) has received remarkable attention over the past twenty years due to its considerable potential in cyber-physical systems. As a typical issue in the coordination of MASs, the consensus problems including the leaderless consensus case [1]-[3] and leader-following tracking case [4]-[7] have been widely studied. In fact, to perform the common objective efficiently, a leader (real/virtual agent) is usually assigned for the MASs in practice.

The above-mentioned works on the coordinated tracking are mainly focused on homogeneous MASs. However, the inherent differences among each individual agents lead to dynamics heterogeneities in MASs. Therefore, the output tracking problem for heterogeneous MASs has emerged [8]-[13]. In the output tracking problem for heterogeneous MASs, compared with the scenario where the leader has no external input, it is worth pointing out that the nonzero input to the leader has both theoretical and practical meanings. First, the resulting problem is quite similar to the leaderless case when the leader has no external input. Second, nonzero control actions might be implemented on the leader in many circumstances to achieve desired objectives, e.g., to achieve acceleration or deceleration in the platoon control of connected vehicles [14]. Thus, different from [8]-[11], the output tracking problem for heterogeneous MASs with a leader of nonzero input was considered in [12], [13].

One key point in the coordination of MASs is the information exchange among agents. Note that the above-mentioned results are obtained under the assumption that continuous communication among agents is available. However, it is difficult to maintain an abundant communication resource supply and an ideal communication environment. Therefore, as an energy-saving communication fashion, event-triggered strategies have been introduced into the control of MASs where communication overheads can be significantly reduced and unnecessary communication energy can be saved. Thus, it enforces researchers investigate the event-triggered output consensus/tracking problem for heterogeneous MASs. The event-triggered output consensus problem was investigated in [15]. Considering fixed and switching topologies, the event-triggered based output consensus problem was studied in [16]. In addition. cooperative output regulation with the event-triggered strategy was studied in [17]. The Zeno behavior, which means that infinite events are triggered in a finite time-interval, is very important in event-triggered control, and should be excluded to guarantee the feasibility of event-triggered control strategies.

Compared with the so-called static event-triggered

mechanism studied in [17]-[20], dynamic event-triggered mechanisms, where an internal dynamic variable is introduced into the triggering law design in order to rule out the Zeno behavior easily, were studied in [21]-[23]. Note that the above-mentioned results on the dynamic event-triggered mechanism [21]-[23] are designed for homogeneous MASs with integrator-type dynamics. Nevertheless, heterogeneous MASs are of more obvious physical significance in real applications.

To address the event-triggered output tracking problem for heterogeneous MASs with general linear dynamics, a hierarchical framework including a homogeneous internal model layer and a heterogeneous follower layer is designed based on the internal model principle. This treatment not only allows us to convert the output tracking problem of heterogeneous MASs into the leader-following tracking problem of homogeneous MASs and the tracking problem of single agent, but also enables us to design the dynamic event-triggered mechanism in the designed separate homogeneous layer. Along this line, a distributed dynamic event-triggered control protocol with only local information is proposed such that the outputs of followers can asymptotically track that of the leader. The contribution of this paper is threefold:

1) Different from the results in [8]-[13] which are obtained under continuous communication, the communication overheads can be reduced by the proposed dynamic event-triggered mechanism. Additionally, compared with the static event-triggered mechanism reported in [15]-[20], an internal dynamic variable is involved in the proposed triggering law design, which benefits the exclusion of the Zeno behavior.

2) Compared with most of the existing results on the dynamic event-triggered mechanism focused on homogeneous MASs with integrator-type dynamics [21]-[23], the proposed dynamic triggering law can address the event-triggered based output tracking control for heterogeneous MASs with general linear dynamics. Note that this setting in turn poses more challenges in the tracking stability analysis and the Zeno behavior exclusion due to the more generalities in agents' models. Moreover, the limitation of continuously accessing to neighbors' states in agent's own triggering detection is removed in the proposed dynamic event-triggered mechanism, which offers easier implementations in real applications.

3) Different from [8]-[11], [17] where the leader is assumed to have no external input, this paper addresses the output tracking problem for general linear heterogeneous MASs with an input-bounded leader. Additionally, the input to the leader is not known to the followers.

The rest of the paper is organized as follows. Preliminaries and problem statement are introduced in Section 2. The main result is shown in Section 3. Simulations and conclusions are shown in Section 4 and 5, respectively.

2. PRELIMINARIES AND PROBLEM FORMULATION

2.1 Preliminaries

Let $\mathbf{R}$, $\mathbf{R}^N$, and $\mathbf{R}^{N\times M}$ be the set of real numbers, $N\times 1$ real vectors, and $N\times M$ real matrices, respectively. Let $\mathbf{0}_N$ and $\mathbf{0}_{N\times M}$ be the $N\times 1$ column vector and $N\times M$ matrix of all zeros, respectively. I_N represents the identity matrix of dimension N. We denote by $\lambda_{\max}(\cdot)$ ($\lambda_{\min}(\cdot)$) the maximum (minimum) eigenvalue of a symmetric matrix. The superscript T means the transpose for real matrices. $\mathrm{diag}\{\cdot\}$ represents a diagonal matrix. Let $\|\cdot\|$ denote the Euclidean norm for vectors and the induced 2-norm for matrices. $\otimes$ denotes the Kronecker product. For a real symmetric matrix, let $P>0$ ($P<0$) denote that P is positive (negative)-definite. $\mathrm{sig}(x)=\mathrm{sign}(x)|x|$ when $x\in\mathbf{R}$ and $\mathrm{sig}(\cdot)$ operates on it componentwise when $x\in\mathbf{R}^n$.

A weighted graph denoted by $\mathcal{G}=(\mathcal{V}(\mathcal{G}),\mathcal{E}(\mathcal{G}))$ with an agent set $\mathcal{V}(\mathcal{G})=\{v_0,v_1,\cdots,v_N\}$ and an edge set $\mathcal{E}(\mathcal{G})\subseteq\mathcal{V}\times\mathcal{V}$ is used to describe the interaction among agents in leader-following heterogeneous MASs. A link $(v_i,v_j)\in\mathcal{E}(\mathcal{G})$ means that agent v_i can receive the information from agent v_j. When $(v_i,v_j)\in\mathcal{E}(\mathcal{G})$ implies $(v_j,v_i)\in\mathcal{E}(\mathcal{G})$, the graph $\mathcal{G}$ is undirected. The adjacency matrix $\mathcal{A}=[a_{ij}]$ associated with $\mathcal{G}$ is defined as $a_{ii}=0$ and $a_{ij}>0$ if $(v_i,v_j)\in\mathcal{E}(\mathcal{G})$. An induced subgraph $\mathcal{G}_s$ is a graph such that $\mathcal{V}(\mathcal{G}_s)\subset\mathcal{V}(\mathcal{G})$ and $(v_i,v_j)\in\mathcal{V}(\mathcal{G}_s)$ if and only if $(v_i,v_j)\in\mathcal{V}(\mathcal{G})$. An undirected subgraph $\mathcal{G}_s$ with the node set $\{v_1,\cdots,v_N\}$ is used to represent follower agents. The Laplacian matrix of a graph is defined by $\mathcal{L}=[l_{ij}]$ with $l_{ii}=\sum_{j\neq i}a_{ij}$ and . $l_{ij}=-a_{ij}$. Let $\mathcal{L}_s$ denote the Laplacian matrix associated with the subgraph $\mathcal{G}_s$. Since the followers have no influence over the leader, we have $a_{0i}=0$, for $i=1,\cdots,N$. A pinning matrix $\mathcal{P}=\mathrm{diag}\{a_{10},\cdots,a_{N0}\}$ is defined to describe the followers' availability to the leader. More specifically, $a_{i0}>0$ if follower i can receive the leader's information, otherwise $a_{i0}=0$.

Lemma 2.1: [24] Let $\mathcal{H}=\mathcal{L}_s+\mathcal{P}$ and $\overline{\mathcal{G}}$ be the augmented graph associated with $\mathcal{H}$. If the subgraph $\mathcal{G}_s$ is undirected and $\overline{\mathcal{G}}$ is connected, then $\mathcal{H}$ is symmetric and positive-definite.

2.2 Problem formulation

Consider a heterogeneous MAS composed of N followers and one leader. The dynamics of agents are as follows.

$$\begin{aligned} \dot{x}_i(t) &= A_i x_i(t) + B_i u_i(t), \\ y_i(t) &= C_i x_i(t), i \in \mathcal{F} \cup \{0\}, \end{aligned} \tag{1}$$

where $\mathcal{F} = \{1, 2, \cdots, N\}$ represents the set of followers. $x_i(t) \in \mathbf{R}^{n_i}$, $u_i(t) \in \mathbf{R}^{p_i}$, and $y_i(t) \in \mathbf{R}^{q}$ are the states, control inputs, and outputs of followers, respectively. $A_i \in \mathbf{R}^{n_i \times n_i}$, $B_i \in \mathbf{R}^{n_i \times p_i}$, and $C_i \in \mathbf{R}^{q \times n_i}$ are the system matrix, input matrix, and output matrix of followers, respectively. $x_0(t) \in \mathbf{R}^{n}$ and $y_0(t) \in \mathbf{R}^{q}$ are the leader's states and outputs, respectively. $u_0(t) \in \mathbf{R}^{p}$ is the leader's unknown but bounded control input satisfying $\| u_0(t) \| < \gamma$. In addition, $A_0 \in \mathbf{R}^{n \times n}$, $B_0 \in \mathbf{R}^{n \times p}$, and $C_0 \in \mathbf{R}^{q \times n}$.

Assumption 2.1: The leader's information can be obtained by only a subset of followers and the graph $\overline{\mathcal{G}}$ is connected.

Assumption 2.2: The matrix pairs (A_i, B_i) and (C_i, A_i) in (1) are stabilizable and detectable, respectively.

Assumption 2.3: Suppose A_0 satisfies

$$\operatorname{rank}\begin{pmatrix} \begin{bmatrix} A_i - \lambda I & B_i \\ C_i & \mathbf{0} \end{bmatrix} \end{pmatrix} = n_i + q, i \in \mathcal{F}, \forall \lambda \in \sigma(A_0).$$

Remark 1: Different from the leader without external control input studied in [8]-[11], [17], the leader considered in this paper has a nonzero time-varying control input and the input to the leader is unavailable to any followers.

The objective is to design a distributed dynamic event-triggered control protocol with only local information to solve the output tracking problem of heterogeneous MASs.

Lemma 2.2: [25]
For any $a, b \in \mathbf{R}^N$ and any $\Omega \in \mathbf{R}^{N \times N} > 0$, the inequality $a^T b \le \frac{a^T \Omega^{-1} a}{2} + \frac{b^T \Omega b}{2}$ holds.

3. MAIN RESULT

In this section, a distributed control protocol via the dynamic event-triggered strategy is proposed such that the output tracking of heterogeneous MASs can be achieved.

Based on the internal model principle, the following distributed dynamic event-triggered control strategy is adopted for each agent. The following control strategy is composed of the internal model (2a), the observer (2c), the controller (2d), and the dynamic triggering law (5).

$$\dot{\zeta}_i(t) = A_0 \zeta_i(t) - c_1 B_0 B_0^T P z_i(t) - c_2 \operatorname{sig}(P z_i(t)), \tag{2a}$$

$$z_i(t) = \sum_{j \in N_i} a_{ij} (\overline{\zeta}_i(t) - \overline{\zeta}_j(t)) + a_{i0} (\overline{\zeta}_i(t) - \overline{x}_0(t)), \tag{2b}$$

$$\hat{\dot{x}}_i(t) = A_i \hat{x}_i(t) + B_i u_i(t) + F_i (y_i(t) - C_i \hat{x}_i(t)), \tag{2c}$$

$$u_i(t) = K_i (\hat{x}_i(t) - \Pi_i \zeta_i(t)) + \Gamma_i \zeta_i(t). \tag{2d}$$

where $\zeta_i(t) \in \mathbf{R}^n$ is the internal model state, $c_1 > 1/\lambda_{\min}(\mathcal{H})$ with $\mathcal{H}$ defined in Lemma 2.1, $c_2 > \frac{\sqrt{N} \| \mathcal{H} \otimes P B_0 \| \gamma}{\sqrt{\lambda_{\min}(\mathcal{H} \otimes P)}}$, and $P \in \mathbf{R}^{n \times n} > 0$ is to be calculated as follows. By solving the linear matrix inequality $X A_0^T + A_0^T X - 2 B_0 B_0^T < 0$, there must exist a positive-definite solution X since the matrix pair (A_0, B_0) is stabilizable. Then, we can calculate that $P = X^{-1}$. $\hat{x}_i(t) \in \mathbf{R}^{n_i}$ is the estimated state of $x_i(t)$. $F_i \in \mathbf{R}^{n_i \times q_i}$ and $K_i \in \mathbf{R}^{p_i \times n_i}$ are the feedback gain matrices to be selected such that $A_i - F_i C_i$ and $A_i + B_i K_i$ are Hurwitz, respectively. $z_i(t)$ is the estimated disagreement error, where a_{ii} is the j th entry of the adjacency matrix associated with graph j is the i th diagonal entry of the pinning matrix $\mathcal{P}$. $\overline{\zeta}_i(t) = e^{A_0 (t - t^i_{k_i})} \zeta_i(t^i_{k_i})$ where $\zeta_i(t^i_{k_i})$ is the most recent broadcast internal model state of agent i with $t^i_{k_i}$, $k_i = 1, 2, \cdots$, being the latest triggering instant. Especially, $\overline{x}_0(t) = x_0(t^i_{k_0}) = x_0(t)$. In addition, the matrices Π_i and Γ_i satisfy that

$$A_i \Pi_i + B_i \Gamma_i = \Pi_i A_0, C_i \Pi_i = C_0 \tag{3}$$

The above equations are also called regulator equations in output regulation theory [26].

Remark 2: There must exist matrices Π_i and Γ_i satisfying (3) since (C_0, A_0) is observable and Assumption 2.3 holds.

The measurement error for each agent is defined as

$$e_i(t) = \overline{\zeta}_i(t) - \zeta_i(t), i \in \mathcal{F}. \tag{4}$$

Inspired by [21] and [23], the following dynamic triggering function is proposed.

$$\begin{aligned} f_i(t, e_i(t), z_i(t), \varphi_i(t)) = {} & \alpha_1^* \| e_i(t) \|^2 + \alpha_2^* \| e_i(t) \| \\ & - \epsilon \| z_i(t) \|^2 - \theta_i \varphi_i(t), i \in \mathcal{F}, \end{aligned} \tag{5}$$

where $\alpha_1^* > \max\{\rho(1 - \frac{1}{\cdot})\lambda_{\max}(Q) + c_1 \lambda_{\max}(\mathcal{H}^2)$ $\| P B_0 B_0^T P \| (\frac{\| P B_0 B_0^T P \|}{\alpha_2} + 2) + c_2 \frac{\lambda_{\max}(\mathcal{H}^2 \otimes P^2)}{\alpha_3}, 2\epsilon \| H \|^2 \}$, and $\alpha_2^* > 2 c_2 \sqrt{\frac{\lambda_{\min}(\mathcal{H} \otimes P)}{N}}$ with $0 < \rho < 1$, $\epsilon > 0$, $0 < \alpha_1 < 1$, $\alpha_2 > 0$, $\alpha_3 > 0$, and $Q = \mathcal{H} \otimes (P A_0 + A_0^T P - 2 c_1 \lambda_{\min}(\mathcal{H}) P B_0 B_0^T P)$. Note that Q is a negative-definite matrix. In addition, $0 < \theta_i < \kappa_i$ with κ_i defined in (6). In the proposed dynamic triggering function (5), $\alpha_1^* \| e_i(t) \|^2 + \alpha_2^* \| e_i(t) \|$ is the

triggering error, and $\epsilon\| z_i(t)\|^2 + \theta_i\varphi_i(t)$ is the dynamic triggering threshold where the proposed internal dynamic variable $\varphi_i(t)$ to agent i is designed as follows.

$$\begin{aligned}\dot{\varphi}_i(t) = &-\kappa_i\varphi_i(t) + \\ &\varepsilon_i(\epsilon\| z_i(t)\|^2 - \alpha_1^*\| e_i(t)\|^2 - \alpha_2^*\| e_i(t)\|), i \in \mathcal{F},\end{aligned} \tag{6}$$

Remark 3: Different from the static triggering functions, the proposed triggering function (5) can be referred as a dynamic triggering function since the an extra dynamic variables $\varphi_i(t)$ is involved in the triggering function design. This treatment offers the larger triggering intervals, which benefits the exclusion of the Zeno behaviors in practice [23].

The the tracking error $o_i(t)$ between the internal model and the leader is defined as

$$o_i(t) = \zeta_i(t) - x_0(t), i \in \mathcal{F}. \tag{7}$$

With the stack vector $\zeta(t) = [\zeta_1^T(t), \cdots, \zeta_N^T(t)]^T$, $z(t) = [z_1^T(t), \cdots, z_N^T(t)]^T$, $o(t) = [o_1^T(t), \cdots, o_N^T(t)]^T$, and $e(t) = [e_1^T(t), \cdots, e_N^T(t)]^T$, (2a) and (2b) can rewritten in the following compact form according to (7).

$$\begin{aligned}\dot{\zeta}(t) = &(I_N \otimes A_0)\zeta(t) - c_1(I_N \otimes B_0B_0^TP)z(t) \\ &- c_2\text{sig}((I_N \otimes P)z(t)),\end{aligned} \tag{8a}$$

$$z(t) = (\mathcal{H} \otimes I_n)(o(t) + e(t)), \tag{8b}$$

where the facts that $\sum_{j\in\mathcal{N}_i} a_{ij}(\bar{\zeta}_i(t) - \bar{\zeta}_j(t)) = \sum_{j\in\mathcal{N}_i} a_{ij}((\bar{\zeta}_i(t) - \bar{x}_0(t)) - (\bar{\zeta}_i(t) - \bar{x}_0(t))$ and $\bar{x}_0(t) = x_0(t)$ have been used.

Invoking (1), (7), and (8), one has

$$\begin{aligned}\dot{o}(t) = &(I_N \otimes A_0)o(t) - c_1(I_N \otimes B_0B_0^TP)z(t) \\ &- c_2\text{sig}((I_N \otimes P)z(t)) - (I_N \otimes B_0)\tilde{u}_0(t),\end{aligned} \tag{9}$$

where $\zeta_0(t) = [x_0^T(t), x_0^T(t), \cdots, x_0^T(t)]^T$, $\tilde{u}_0(t) = [u_0^T(t), u_0^T(t), \cdots, u_0^T(t)]^T$.

Theorem 3.1: Suppose Assumptions 2.1-2.3 hold. Under the proposed control strategy (2) and (5), the event-triggered output tracking problem for heterogeneous MASs (1) can be solved for all initial conditions if the parameters are selected such that $(c_1\alpha_2 + c_2\alpha_3\lambda_{\max}(P^2) + 2\epsilon)/(\rho(1-\alpha_1)\lambda_{\min}(\mathcal{H}^{-2})) < -\lambda_{\max}(Q)$ holds.

Proof: Consider the following two steps to prove the achievement of the event-triggered output tracking. 1) we will prove that the tracking error between the internal model and the leader can asymptotically converge to zero; 2) we will prove that the tacking error between the internal model and each individual follower can asymptotically converge to zero. The index t is .

Step 1: Note that $\dot{\varphi}_i(t) \geq -(\kappa_i + \varepsilon_i\theta_i)\varphi_i(t)$ according to (5) and (6). Therefore, $\varphi_i(t) \geq e^{-(\kappa_i+\varepsilon_i\theta_i)t}\varphi_i(0) > 0$ since $\varphi_i(0) > 0$. Take into account the following Lyapunov function candidate for (7).

$$V = V_1 + V_2, \tag{10}$$

where $V_1 = o^T(\mathcal{H} \otimes P)o$ and $V_2 = \sum_{i-1}^N \varphi_i$. Taking the time derivative of V_1 in (10) along (9) yields

$$\begin{aligned}\dot{V}_1 = &\ 2o^T(\mathcal{H} \otimes P)\dot{o} \\ = &\ 2o^T(\mathcal{H} \otimes PA_0)o - 2o^T(\mathcal{H}^2 \otimes c_1PB_0B_0^TP)o \\ &-2o^T(\mathcal{H}^2 \otimes c_1PB_0B_0^TP)e - 2o^T(\mathcal{H} \otimes PB_0)\tilde{u}_0 \\ &-2c_2o^T(\mathcal{H} \otimes P)\text{sig}((I_N \otimes P)z),\end{aligned} \tag{11}$$

where we have used (8b). In what follows, each term in (11) will be analyzed.

Analyzing the first two terms in (11), one has

$$\begin{aligned}&2o^T(\mathcal{H} \otimes PA_0)o - 2o^T(\mathcal{H}^2 \otimes c_1PB_0B_0^TP)o \\ = &\ o^T(\mathcal{H} \otimes (PA_0 + A_0^TP))o - 2o^T(\mathcal{H}^2 \otimes c_1PB_0B_0^TP)o \\ \leq &\ o^T(\mathcal{H} \otimes (PA_0 + A_0^TP - 2c_1\lambda_{\min}(\mathcal{H})PB_0B_0^TP))o \\ \leq &\ \lambda_{\max}(Q)o^To,\end{aligned} \tag{12}$$

where Lemma 2.1 has been used. Note that $Q < 0$ according to above analysis. By recalling the fact that $o = (\mathcal{H}^{-1} \otimes I_n)z - e$, (12) can be further bounded as follows.

$$\begin{aligned}&\lambda_{\max}(Q)o^To \\ = &\ \rho\lambda_{\max}(Q)o^To + (1-\rho)\lambda_{\max}(Q)o^To \\ = &\ \rho\lambda_{\max}(Q)[z^T(\mathcal{H}^{-2} \otimes I_n)z - 2z^T(\mathcal{H}^{-1} \otimes I_n)e + e^Te] \\ &+(1-\rho)\lambda_{\max}(Q)o^To \\ \leq &\ \rho(1-\alpha_1)\lambda_{\min}(\mathcal{H}^{-2})\lambda_{\max}(Q)z^Tz \\ &+\rho(1-\frac{1}{\alpha_1})\lambda_{\max}(Q)e^Te + (1-\rho)\lambda_{\max}(Q)o^To,\end{aligned} \tag{13}$$

where the fact that $2z^T(\mathcal{H}^{-1} \otimes I_n)e \leq \alpha_1 z^T(\mathcal{H}^{-2} \otimes I_n)z + \frac{1}{\alpha_1}e^Te$ has been used. Additionally, note that $0 < \alpha_1 < 1$ and $0 < \rho < 1$.

It follows from (8b) and Lemma 2.2 that the third term in (11) can be handled as

$$\begin{aligned}&-2o^T(\mathcal{H}^2 \otimes c_1PB_0B_0^TP)e \\ &\leq c_1\alpha_2z^Tz + c_1\lambda_{\max}(\mathcal{H}^2)\| PB_0B_0^TP\|(\frac{\| PB_0B_0^TP\|}{\alpha_2} + 2)e^Te.\end{aligned} \tag{14}$$

The forth term of (11) can be bounded as

$$-2o^T(\mathcal{H} \otimes PB_0)\tilde{u}_0 \leq 2\| \mathcal{H} \otimes PB_0\| \ \| \tilde{u}_0\| \ \| o\| \tag{15}$$

We cope with the last term of (11) as follows.

$$-2c_2 o^T(\mathcal{H}\otimes P)\mathrm{sig}((I_N\otimes P)z)$$
$$=-2c_2[\|(\mathcal{H}\otimes P)(o+e)\|-e^T(\mathcal{H}\otimes P)\mathrm{sig}((I_N\otimes P)z)].$$

Note that $\|(\mathcal{H}\otimes P)(o+e)\|=\sqrt{(o+e)^T(\mathcal{H}\otimes P)(o+e)}\geq \sqrt{\lambda_{\min}(\mathcal{H}\otimes P)}\|o+e\|\geq\sqrt{\lambda_{\min}(\mathcal{H}\otimes P)}(\|o\|-\|e\|)$ and $2c_2e^T(\mathcal{H}\otimes P)\mathrm{sig}((I_N\otimes P)z)\leq c_2(\alpha_3\lambda_{\max}(P^2)z^Tz+\frac{\lambda_{\max}(\mathcal{H}^2\otimes P^2)}{\alpha_3})e^Te$. Therefore, one has

$$\begin{aligned}&-2c_2o^T(\mathcal{H}\otimes P)\mathrm{sig}((I_N\otimes P)z)\\ \leq\ &-2c_2\sqrt{\lambda_{\min}(\mathcal{H}\otimes P)}\|o\|+2c_2\sqrt{\lambda_{\min}(\mathcal{H}\otimes P)}\|e\|\\ &+c_2\alpha_3\lambda_{\max}(P^2)z^Tz+c_2\frac{\lambda_{\max}(\mathcal{H}^2\otimes P^2)}{\alpha_3}e^Te\end{aligned}\quad(16)$$

Taking the time derivative of V_2 in (10) along (6) gives

$$\dot V_2=-\sum_{i=1}^N\kappa_i\varphi_i+\sum_{i=1}^N\varepsilon_i(\epsilon\|z_i\|^2-\alpha_1^*\|e_i\|^2-\alpha_2^*\|e\|)\quad(17)$$

Invoking (12)-(17), one can obtain the following result according to the proposed dynamic triggering law (5).

$$\begin{aligned}\dot V\leq\ &[\rho(1-\alpha_1)\lambda_{\min}(\mathcal{H}^{-2})\lambda_{\max}(Q)+c_1\alpha_2\\ &+c_2\alpha_3\lambda_{\max}(P^2)]\|z\|^2+\beta_1\|e\|^2+\beta_2\|e\|\\ &-\sum_{i=1}^N\kappa_i\varphi_i+\sum_{i=1}^N\varepsilon_i(\epsilon\|z_i\|^2-\alpha_1^*\|e_i\|^2-\alpha_2^*\|e_i\|)\\ &+(2\sqrt{N}\|\mathcal{H}\otimes PB_0\|\gamma-2c_2\sqrt{\lambda_{\min}(\mathcal{H}\otimes P)})\|o\|\\ &-\sum_{i=1}^N\kappa_i\varphi_i+(1-\rho)\lambda_{\max}(Q)o^To\\ \leq\ &[\rho(1-\alpha_1)\lambda_{\min}(\mathcal{H}^{-2})\lambda_{\max}(Q)+c_1\alpha_2\\ &+c_2\alpha_3\lambda_{\max}(P^2)+2\epsilon]\|z\|^2+\sum_{i=1}^N(\theta_i-\kappa_i)\varphi_i\\ &+(1-\rho)\lambda_{\max}(Q)o^To,\end{aligned}\quad(18)$$

where $\beta_1=\rho(1-\frac{1}{\cdot})\lambda_{\max}(Q)+c_1\lambda_{\max}(\mathcal{H}^2)\|PB_0B_0^TP\|(\frac{\|PB_0B_0^TP\|}{\alpha_2}+2)+c_2\frac{\lambda_{\max}(\mathcal{H}^2\otimes P^2)}{\alpha_3}$ and $\beta_2=2c_2\sqrt{\lambda_{\min}(\mathcal{H}\otimes P)}$. Recalling the fact that $\varphi_i(t)\geq \mathrm{S}e^{-(\kappa_i+\varepsilon_i\theta_i)t}\varphi_i(0)>0$, (18) can be rewritten as

$$\dot V\leq(1-\rho)\lambda_{\max}(Q)o^To+\sum_{i=1}^N(\theta_i-\kappa_i)\varphi_i<0.\quad(19)$$

It follows from (10) and (19) that $o_i\to\mathbf{0}_n$ as $t\to\infty$, which implies that $\zeta_i\to x_0$ as $t\to\infty$. In the following, we will prove that the tacking error between the internal model and each individual follower can asymptotically converge to zero based on the analysis in Step 1..

Step 2: Before moving on, we define the observer error as $\varpi_i(t)=x_i(t)-\hat x_i(t)$. It follows from (1) and (2c) that

$$\dot\varpi_i(t)=\dot x_i(t)-\dot{\hat x}_i(t)=(A_i-F_iC_i)\varpi_i(t),i\in\mathcal{F}.\quad(20)$$

Since $A_i-F_iC_i$ are Hurwitz, it follows from (20) that the observer error asymptotically converges to zero, namely, $\varpi_i\to\mathbf{0}_{n_i}$ as $t\to\infty$.

Define the tacking error between the internal model and each individual follower as $\vartheta_i(t)=x_i(t)-\Pi_i\zeta_i(t)$. According to (1), (2a), (2d), and (3) that

$$\begin{aligned}\dot\vartheta_i(t)=\ &\dot x_i(t)-\Pi\dot\zeta_i(t)\\ =\ &(A_i+B_iK_i)\vartheta_i(t)-B_iK_i\varpi_i+c_1\Pi_iB_0B_0^TPz_i(t)\\ &+c_2\Pi_i\mathrm{sig}(Pz_i(t)),i\in\mathcal{F}.\end{aligned}\quad(21)$$

Note that z_i and ϖ_i converge to zero as time goes to infinity. Therefore, the tacking error between the internal model and each individual follower can asymptotically converge to zero since $A_i+B_iK_i$ are Hurwitz according to (21), namely, $\vartheta_i\to\mathbf{0}_n$ as $t\to\infty$. As a result, one has $x_i-\Pi_i\zeta_i\to\mathbf{0}_n$ as $t\to\infty$. Note that $o_i\to\mathbf{0}_n$ as $t\to\infty$, which implies $\zeta_i\to x_0$ as $t\to\infty$ by recalling the result mentioned in Step 1. Therefore, one has $x_i-\Pi_ix_0\to\mathbf{0}_n$ as $t\to\infty$, which indicates $C_ix_i-C_i\Pi_ix_0\to\mathbf{0}_a$ as $t\to\infty$. It follows from (3) that $y_i\to y_0$ as $t\to\infty$. Therefore, the output tracking of heterogeneous MASs can be addressed.

Next, the Zeno behavior will be analyzed. Partly inspired by [23], the Zeno behavior is ruled out by contradiction. Suppose the Zeno behavior does exist. Thus, for some follower i, $\lim_{k_i\to\infty}t_{k_i}^i=T_0$ with T_0 be a positive constant Let

$$\tau_0=\frac{1}{2\|A_0\|}\ln\left(\frac{\sqrt{(\alpha_2^*)^2+4\alpha_1^*\theta_ie^{-(\kappa_i+\varepsilon_i\theta_i)T_0}\varphi_i(0)}-\alpha_2^*}{2\alpha_1^*\varrho_*}+1\right),$$

where $\varrho_*=\frac{\varrho}{\|A_0\|}$ and $^*=\|P\|(c_1\|B_0\|^2+c_2)$ with ϱ be the upper bound of $\|z_i(t)\|$. Note that $\tau_0>0$. Then according to the property of limits, there exists a positive integer $k_i(\tau_0)$ such that

$$t_{k_i}^i\in[T_0-\tau_0,T_0],\forall k_i\geq k_i(\tau_0).\quad(22)$$

Consider any two consecutive triggered events for follower i, $t\in[t_{k_i}^i,t_{k_i+1}^i)$. According to (8b), one has $\|z\|\leq\|\mathcal{H}\|(\|o\|+\|e\|)$. It follows from (5) that $\alpha_1^*\|e\|^2+\sqrt{N}\alpha_2^*\|e\|\leq\epsilon\|z\|^2+\sum_{i=1}^N\theta_i\varphi_i(t)$. Thus, one has $\alpha_1^*\|e\|^2+\sqrt{N}\alpha_2^*\|e\|\leq 2\epsilon\|\mathcal{H}\|^2(\|o\|^2+\|e\|^2)+\sum_{i=1}^N\theta_i\varphi_i(t)$, which indicates $(\alpha_1^*-2\epsilon\|\mathcal{H}\|^2)\|e\|^2+\sqrt{N}\alpha_2^*\|e\|\leq 2\epsilon\|\mathcal{H}\|^2\|o\|^2+\sum_{i=1}^N\theta_i\varphi_i(t)$. Invoking (10) and (19), it is apparent that $\|o(t)\|$ and $\|\varphi_i(t)\|$ are bounded. Then $\|e(t)\|$ is bounded according to the above analysis. Thus, $\|z(t)\|$ is bounded by recalling (8b). Therefore, $\|z_i(t)\|$ is bounded. As mentioned above, we denote its upper bounded as ϱ. It follows from (4) that $\dot e_i(t)=A_0e_i(t)+c_1B_0B_0^TPz_i(t)+c_2\mathrm{sig}(Pz_i(t))$. Thus, we can obtain $\|e_i(t)\|\leq{}^*\int_{t_{k_i}^i}^t\|e^{A_0(t-s)}\|\mathrm{d}s\leq{}^*\int_{t_{k_i}^i}^t\|e^{\|A_0\|(t-s)}\|\mathrm{d}s\leq\varrho_*(e^{\|A_0\|(t-t_{k_i}^i)}-1)$, where $s\in[t_{k_i}^i,t_{k_i+1}^i)$. Thus, according to

(5) and the fact that $\varphi_i(t) \geq e^{-(\kappa_i+\varepsilon_i\theta_i)t}\varphi_i(0) > 0$, when an event is triggered, the inter-event time interval must be greater than or equal to the solution of the following equality.

$$\alpha_1^*\phi(t)^2 + \alpha_2^*\phi(t) = e^{-(\kappa_i+\varepsilon_i\theta_i)t}\varphi_i(0), \qquad (23)$$

where $\phi(t) = \| \varrho_*(e^{\| A_0 \|(t-t_{k_i}^i)} - 1) \|$. Therefore, by calculating (23), one has

$$t_{k_i(\tau_0)+1}^i - t_{k_i(\tau_0)}^i \geq \frac{1}{\| A_0 \|}\ln\left(\frac{\sqrt{(\alpha_2^*)^2 + 4\alpha_1^*\theta_i e^{-(\kappa_i+\varepsilon_i\theta_i)t_{k_i(\tau_0)+1}^i}\varphi_i(0)} - \alpha_2^*}{2\alpha_1^*\varrho_*} + 1\right)$$
$$\geq \frac{1}{\| A_0 \|}\ln\left(\frac{\sqrt{(\alpha_2^*)^2 + 4\alpha_1^*\theta_i e^{-(\kappa_i+\varepsilon_i\theta_i)T_0}\varphi_i(0)} - \alpha_2^*}{2\alpha_1^*\varrho_*} + 1\right) = 2\tau_0$$

Which contradicts to (22). Thus, Zeno behavior is excluded.

■

4. NUMERICAL SIMULATION

In this section, we demonstrate the theoretical result by the following numerical simulation. Consider an MAS composed of 6 followers and 1 leader described by (1) with

$$A_0 = \begin{bmatrix} 0 & 7 \\ -1 & 1 \end{bmatrix}, B_0 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, C_0 = \begin{bmatrix} 1 & 0 \end{bmatrix},$$
$$A_i = \begin{bmatrix} 0 & 7 & 0 \\ -1 & 1 & c_i \\ 0 & -d_i & -a_i \end{bmatrix}, B_i = \begin{bmatrix} 0 \\ 0 \\ b_i \end{bmatrix}, C_i = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},$$

where a_i, b_i, c_i, and d_i, $i \in \mathcal{F}$ are positive constants. The system parameters $\{a_i, b_i, c_i, d_i\}$ for each agent i are chosen as $\{1,1,1,0\}$, $\{10,2,1,0\}$, $\{2,1,1,10\}$, $\{2,1,1,1\}$, $\{1,1,1,0\}$, and $\{10,2,1,0\}$. The control input in the leader is given by $u_0 = 0.01\sin(t)$. Consider the Laplacian matrix of the followers as $\mathcal{L}_s = 0.5*[1.1-0.5000-0.6;-0.51.3-0.8000;0-0.81.5-0.\ 700;.\ 00-0.71.8-1.10;000-1.12.1-1;-0.6000-11.6]$ .The pinning matrix $\mathcal{P}$ is given by $\mathcal{P} = \text{diag}\{0.5\ 0\ 1.5\ 1\ 0\ 0\}$. We choose the feedback gain matrix K_i and F_i for each agent i as $K_1 = [-27\ -37\ -9]$, $K_2 = [-13.5\ -18.50]$, $K_3 = [-27\ -27\ -8]$, $K_4 = [-27\ -36\ -7]$, $K_5 = [-27\ -37\ -9]$, $K_6 = [-13.5\ -18.50]$ and $F_1 = [8\ 19\ 8]^T$, $F_2 = [-137\ 343]^T$, $F_3 = [73\ -4]^T$, $F_4 = [712\ -4]$, $F_5 = [81\ 98]^T$, $F_6 = [-137\ -343]^T$ such that all the matrices $A_i + B_iK_i$, and $A_i - F_iC_i$, $i \in \mathcal{F}$, are Hurwitz, respectively. Calculate $P = [0.9216\ -0.3318; -0.3318\ 1.2439]$ and

$$\Pi_i = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \Gamma_i = \begin{bmatrix} 0 & \frac{d_i}{b_i} \end{bmatrix}, i \in \mathcal{F}.$$

We choose $c_1 = 8.4$, $c_2 = 0.25$, and other appropriate parameters. The leader's state and the internal model states evolved under the proposed output tracking control protocol are presented in Fig. 1. It is shown that all the internal model states can track the leader from Fig. 1. The outputs of all followers and the leader are shown in Fig. 2, which indicates that the output tracking problem for general linear heterogeneous MASs can be solved under the proposed method. Comparing Fig. 1 with Fig. 2, it can be seen that the convergence rate of the tracking error between the internal model and the leader may be faster than that of the tracking error between the internal model and each individual follower since the hierarchical framework is designed based on the internal model principle. The triggering errors and dynamic thresholds are shown in Fig. 3. The corresponding triggering instants are shown in Fig. 4. Compared with the output tracking control for heterogeneous MASs under continuous communication, the proposed strategy can achieve output tracking with energy reduction in communication.

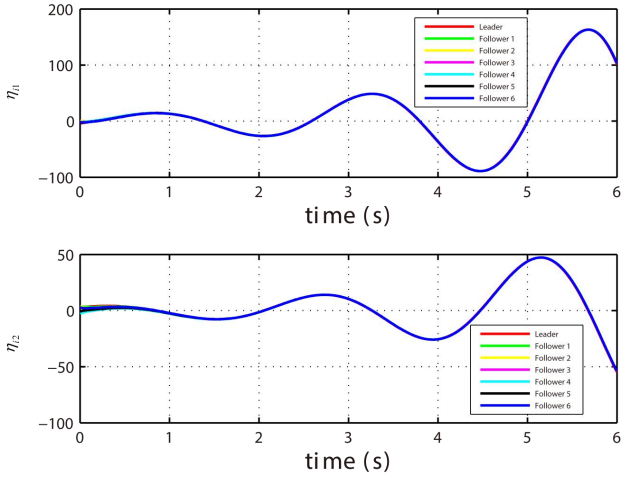


Fig. 1. The state evolution of the leader and the internal model.

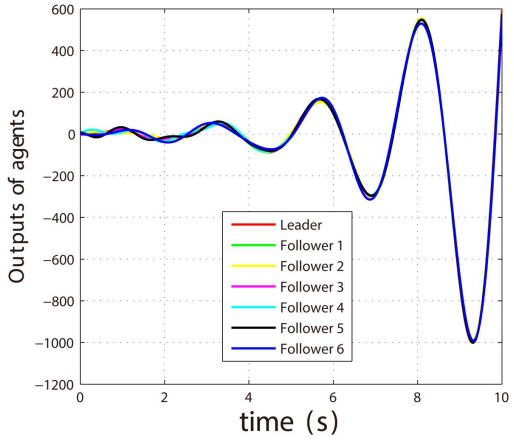


Fig. 2. The outputs of the leader and followers.

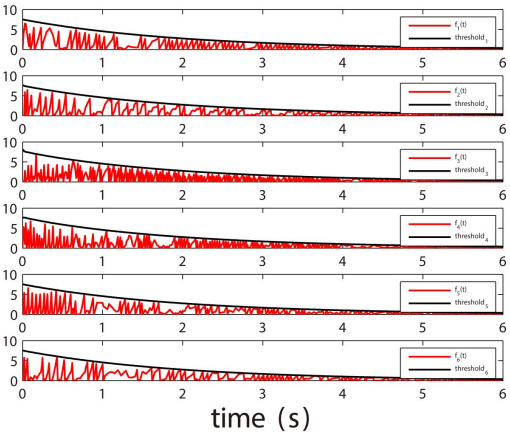


Fig. 3. The triggering errors and thresholds under the proposed dynamic triggering mechanism.

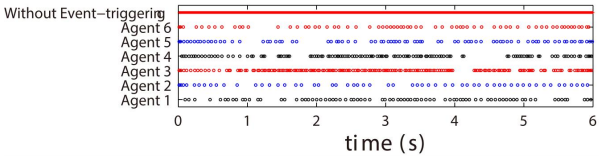


Fig. 4. The triggering instants under the proposed dynamic triggering mechanism.

5. CONCLUSION

This paper has studied the event-triggered output tracking problem for general linear heterogeneous MASs with a leader of nonzero and time-varying input. We have proposed a distributed dynamic event-triggered control protocol to achieve output tracking without the need for continuous communication in either controller updates or triggering detection. Moreover, the Zeno behavior has been excluded to guarantee the feasibility of the proposed control protocol.

References

[1] R. Olfati-Saber, J. A. Fax, and R. M. Murray, " Consensus and cooperation in networked multi-agent systems," Proceedings of the IEEE, vol. 95, no. 1, pp. 215-233, 2007.

[2] W. Ren, R. W. Beard, and E. M. Atkins, "Information consensus in multivehicle cooperative control", IEEE Control Systems, vol. 27, no. 2, pp. 71-82, 2007.

[3] Y. Cao and W. Ren, " Distributed coordinated tracking with reduced interaction via a variable structure approach " IEEE Transactions on Automatic Control, vol. 57, no. 1, pp. 33-48, 2012.

[4] Z. Meng, Z. Lin, and W. Ren, "Robust cooperative tracking for multiple non-identical second-order nonlinear systems," Automatica, Vol. 49, No. 8, pp. 2363–2372, 2013.

[5] Z. Li, X. Liu, W. Ren, and L. Xie, "Distributed tracking control for linear multiagent systemswith a leader of bounded unknown input," IEEE Transactions on Automatic Control, vol. 58, no. 2, pp. 518–523, 2013.

[6] J. Fu and J. Wang, "Observer-based finite-time coordinated tracking for general linear multi-agent systems," Automatica, vol. 66, no. 4, pp. 231–237, 2016.

[7] M. Ye, B. D. O. Anderson, and C. Yu, "Leader tracking of eulerlagrange agents on directed switching networks using a modelindependent algorithm," IEEE Transactions on Control of Network Systems, vol. 6, no. 2, pp. 561–571, 2019.

[8] Y. Su and J. Huang, "Cooperative output regulation of linear multiagent systems," IEEE Transactions on Automatic Control, vol. 57, no. 4, pp. 1062–1066, 2012.

[9] Y. Su, Y. Hong, and J. Huang, "A general result on the robust cooperative output regulation for linear uncertain multi-agent systems," IEEE Transactions on Automatic Control, vol. 58, no. 5, pp. 1275–1279, 2013.

[10] H. F. Grip, A. Saberi, and A. A. Stoorvogel, "On the existence of virtual exosystems for synchronized linear networks," Automatica, vol. 49, no. 10, pp. 3145–3148, 2013.

[11] H. Hong, X. Wang, and Z.-P. Jiang, "Distributed output regulation of leader-follower multi-agent systems," International Journal of Robust and Nonlinear Control, vol. 23, no. 1, pp. 48–66, 2013.

[12] W. Dong, F. D. L. Torre, and Y. Xing, "Distributed output tracking control of heterogeneous linear agents," In Proceedings of 2014 American Control Conference, pp. 4677–4682, 2014.

[13] J. Back and J. S. Kim, "Output feedback practical coordinated tracking of uncertain heterogeneous multi-agent systems under switching network topology," IEEE Transactions on Automatic Control, vol. 62, no. 12, pp. 6399–6406, 2017.

[14] Y. Bian, Y. Zheng, W. Ren, S. E. Li, J. Wang, and K. Li, "Reducing time headway for platooning of connected vehicles via V2V communication," Transportation Research Part C: Emerging Technologies, vol. 102, pp. 87–105, 2019.

[15] W. Hu, L. Liu, and G. Feng, "Output consensus of heterogeneous linear multi-agent systems by distributed event-triggered/self-triggered strategy," IEEE Transactions on Cybernetics, vol. 45, no. 8, pp. 1914–1924, 2017.

[16] X. Liu, H. Liu, C. Du, P. Lu, and W. Gao, "Event-triggered output consensus for

heterogeneous multi-agent systems with fixed and switching directed topologies," International Journal of Robust and Nonlinear Control, vol. 29, no. 14, pp. 4681–4699, 2019.

[17] W. Hu and L. Liu, "Cooperative output regulation of heterogeneous linear multi-agent systems by event-triggered control," IEEE Transactions on Cybernetics, vol. 47, no. 1, pp. 105–116, 2017.

[18] D. Yang, W. Ren, X. Liu, and W. Chen, "Decentralized event-triggered consensus for linear multi-agent systems under general directed graphs," Automatica, vol. 69, no. 7, pp. 242–249, 2016.

[19] X. Liu, C. Du, P. Lu, and D. Yang, "Distributed event-triggered feedback consensus control with state-dependent threshold for general linear multi-agent systems," International Journal of Robust and Nonlinear Control, vol. 27, no. 15, pp. 2589–2609, 2017.

[20] X. Liu, C. Du, H. Liu, and P. Lu, "Distributed event-triggered consensus control with fully continuous communication free for general linear multi-agent systems under directed graph," International Journal of Robust and Nonlinear Control, vol. 28, no. 1, pp. 132–143, 2018.

[21] A. Girard, "Dynamic triggering mechanisms for event-triggered control," IEEE Transactions on Automatic Control, vol. 60, no. 7, pp. 1992–1997, .

[22] X. Yi, K. Liu, and D. V. Dimarogonas, "Distributed dynamic eventtriggered control for multi-agent systems," In Proceedings of 56th IEEE Conference on Decision and Control, pp. 6683–6688, 2017.

[23] X. Yi, K. Liu, D. V. Dimarogonas, and K. H. Johansson, "Dynamic event-triggered and self-triggered control for multi-agent systems," IEEE Transactions on Automatic Control, vol. 64, no. 8, pp. 3300– 3307, 2018.

[24] H. Zhang, Z. Li, Z. Qu, and F. L. Lewis, "On constructing Lyapunov functions for multi-agent systems," Automatica, vol. 58, no. 8, pp. 39– 42, 2015.

[25] G. H. Hardy, J. E. Littlewood, and G. Polya, Inequalities. Cambridge, UK: Cambridge University Press, 1952.

[26] J. Huang, Nonlinear Output Regulation: Theory and Applications Philadelphia, PA, 2004.

Distributed Output Feedback Consensus Control of Multiple Lur'e Systems Based on Event-triggered Mechanism

Jianjun Sun*, Haikuo Liu*, Changkun Du*, Xiangdong Liu*, Zhen Chen*, Pingli Lu* **
* School of Automation, Beijing Institute of Technology
Beijing 100081, China
** Corresponding author, e-mail: pinglilu@bit.edu.cn

Abstract

In this paper, the output feedback consensus control problem is investigated for a class of multi-agent systems with Lur'e non-linear dynamics over directed topology. Based on the distributed observer, a novel distributed event-triggered consensus protocol is proposed. The proposed event-triggered protocol removes the limitations of continuous communication and saves the on-board resources, that is, the continuous communication is avoided for both controller updating and event-triggered condition monitoring. Furthermore, the con-sensus is proved by Lyapunov method and Zeno behavior is excluded. Finally, simulation results attests to the properties of the proposed protocols.

Keywords: Consensus control, Event-triggered strategy, Multiple Lur'e nonlinear system, Directed topology, Output Feedback

1. IMPORTANT INFORMATION

In recent years, various applications in distributed co-ordination of multi-agent systems (MASs) take increasing attention, such as spacecraft formation flying, cooperative surveillance and sensor networks [1], [2], [3]. As a basic problem in MASs, the consensus control problem is to design a control protocol to make all the agents' states converge to a same value. Based on consensus, the agents in network can reach an agreement. With the increasing complexity of the network, the centralized method depended on the global information is difficult to satisfy the practical needs. Hence, the distribute control is investigated, since it only need local information for each agent to make decision, which means lower resource consumption. Much research on distributed consensus control has been conducted in [4], [5] and references therein.

The dynamics of each agent is an important factor for the designing of consensus control protocols. Most existing works are focus on the MASs with linear dynamics, such as single-integrator network [4], double-integrator network [5], general linear network [6]. Comparing with linear dynamics, the nonlinear dynamics maybe more valuable for the practical application. For instance, aircrafts, flexible robotic arms, and several chaotic systems, including Chua's circuit, can be represented in the nonlinear type. In this paper, the typical Lur'e nonlinear dynamics, which contains a nonlinear feedback loop with certain slope or sector condition, will be considered for the consensus problem. A class of robotic arms and several chaotic systems can usually be represented by Lur'e type [7]. There are some works investigated the consensus problem of multiple Lur'e systems. Ref. [8] proved that the network of multiple Lur'e systems has a desirable unbounded synchronized region based on designed control protocol. For the uncertain multiple Lur'e systems, an adaptive consensus tracking protocol was designed in [9] to solve the robust consensus tracking problem. Consider the directed interaction networks, the multiple Lur'e systems are divided into the leader's layer and the followers' layer to de- sign the control protocol and solve the consensus problem in [10]. Ref. [11] proposed a fully distributed adaptive protocol to solve the tracking consensus problem for multiple Lur'e systems over a directed graph contained a directed spanning tree. Despite all these works, a fundamental problem of the consensus of multiple Lur'e systems, that is designing distributed output feedback consensus protocols, has not been researched comprehensively. Considering the full state information of Lur'e systems is quite difficult to obtain, it is preferred to use output feedback control in practice. Ref. [12] proposed a distributed controller to solve the consensus problem of multiple Lur'e systems only used relative output information.

However, one common limitation existing in the aforementioned papers is that the control protocols are continuous- time consensus algorithms. The computation ability, adaptivity and robustness are always limited for each agent in the MASs, the continuous-time algorithm will rise sharply computation and information transmission among agents and even result in whole system crash. The event-triggered mechanism, as fully intermittent algorithm, only require each agent communicate with its neighbors and updates its actuators when needed. Therefore, the event-triggered method is

used to design the consensus control protocols in several works. Ref. [13] improved the performance of multi-agent systems by propose an event-triggered scheduler and show how the scheduler relax the traditional periodic execution. In [14] and [15], the authors propose a self-triggered algorithm for the consensus control of integrator dynamic multi-agent system, which further reduce the calculation and save the on- board resources. Considering the Lur'e systems networks, the authors of [16] design an event-triggered protocol to reduce the energy consumption and the updates frequency. It should be noticed, under the above protocol in [16], the triggering detection of each agent still need its neighbors' state information continuously.

Motivated by above facts stated, we will investigate the distributed event-triggered consensus protocol based on out- put feedback for multiple Lur'e systems under directed topology. First, a distributed state observer is designed to estimate the state for each agent. Secondly, with the estimated state and state-dependent event-triggered threshold, we propose a novel distributed output feedback event-triggered consensus controller for each agent with zero final consensus error. Then, we prove that there is no Zeno behavior under the proposed protocol, that is, the event would not be triggered continuously under the state-dependent triggering conditions. The primary contributions are summarized as follows:

- The event-triggered output feedback controller with fully intermittent communication is proposed in this paper. Compared with continuous-time output feed- back strategy in [12], the distributed event-triggered output feedback consensus protocol in this paper can effectively reduce the on-board resources consumption. Compared with the event-triggered strategy in [16] which needs continuous triggering condition monitoring, in this paper, neither the controller updates nor the triggering detection need the continuous information transmission among the network.
- Compared with the state feedback strategy [16], in this paper, the output feedback strategy is investigated. In practical, when the state is immeasurable, the output feedback strategy will work better than state feedback strategy. Meanwhile, the nonlinearity of Lur'e system and the introduction of observer will bring challenge for the elimination of Zeno behavior.

2. PREPARATION OF MANUSCRIPTS

2.1 Notations

Let $\mathbf{R}^{m\times n}$ be the set of $m\times n$ real matrices, and $\mathbf{R}^n$ be the n-dimensional real column vector space, respectively. Let $\mathbf{1}_N(\mathbf{0}_N)$ be appropriate column vector of ones with scalar $1(0)$ as special cases.$diag(a_1,\dots,a_n)$ represents a diagonal matrix composed by a_i . $\lambda_{\max}(P)(\lambda_{\min}(P))$ denotes the maximum(minimum) eigenvalue of matrix P. $\|\cdot\|$ represents the Euclidean norm for vectors and the induced 2-norm for matrices. A symmetric matrix $P>0(\geq 0)$ denotes P is positive(semi-positive) definite. $A\otimes B$ denotes the Kronecker product of matrix A and B. $\mathcal{I}_N$ is a set of positive integers, i.e. $\mathcal{I}_N=1,2,\dots,N$.

2.2 Algebraic graph theory

The interconnection topology of N agents can be described as a weighted directed graph $\mathcal{G}(\mathcal{A})=(\mathcal{V},\mathcal{E},\mathcal{A})$, where $\mathcal{V}=(v_1,\dots,v_N)$ is a set of nodes, $\mathcal{E}=\mathcal{V}\times\mathcal{V}$ is a set of edges and $\mathcal{A}=[a_{ij}]\in\mathbf{R}^{N\times N}$ is the adjacency matrix, respectively. The entry $a_{ij}>0$ if $(v_j,v_i)\in\mathcal{E}$ and $a_{ij}=0$ otherwise. Correspondingly, the Laplacian matrix $\mathcal{L}=[l_{ij}]\in\mathbf{R}^{N\times N}$ is defined by $l_{ij}=\sum_{k=1}^{N}a_{ik}$ if $j=i$ and $l_{ij}=-a_{ij}$ otherwise. The directed path that connects x_i and x_j in the graph $\mathcal{G}$ is a sequence of distinct nodes $x_{i_0},x_{i_1},x_{i_2},\dots,x_{i_m}$, where $x_{i_0}=x_i,x_{i_m}=x_j$ and $(x_{i_r},x_{i_{r+1}})\in\mathcal{E},0\leq r\leq m-1$.

Definition 1:(Strongly connected graph) A directed graph is a strongly connected graph if and only if there is a directed path between any pair of different nodes.

Definition 2:(General algebraic connectivity) For a strongly connected graph $\mathcal{G}$ with Laplacian matrix $\mathcal{L}$, the general algebraic connectivity is defined by

$$\alpha(\mathcal{L})=\min_{\mathbf{r}^Tx=0,x\neq 0}\frac{x^T\bar{\mathcal{L}}x}{x^TRx},$$

Where $\bar{\mathcal{L}}=\frac{1}{2}(R\mathcal{L}+\mathcal{L}R^{\mathrm{T}}),R=diag(r_1,\dots,r_N),\mathbf{r}=[r_1,\dots,r_N]^{\mathrm{T}}$, with $r_i>0,\mathbf{r}^T\mathcal{L}=\mathbf{0}_N^T$ and $\sum_{i=1}^{N}r_i=1$.

Definition 3:(Irreducible matrix) A matrix P is termed irreducible if there does not exist a permutation matrix X such that X^TPX is block triangular. Otherwise, P is termed reducible.

Lemma 1:[18] An $n\times n$ matrix A is irreducible if and only if its corresponding network $\mathcal{G}$ is strongly connected.

Lemma 2:[19] If the Laplacian matrix $\mathcal{L}$ is irreducible, there will exist a vector $\mathbf{r}=[r_1,\dots,r_N]^T,\mathbf{r}>0,\mathbf{1}_N^T\mathbf{r}=1$ such that $\mathbf{r}^T\mathcal{L}=\mathbf{0}_N^T$.

Lemma 3:[20] **(Young's inequality)** Given any $x,y\in\mathbf{R}^n$, for $\forall\epsilon>0$, the following inequality holds.

$$x^Ty\leq\frac{x^Tx}{2\epsilon}+\frac{\epsilon y^Ty}{2}.$$

2.3 *Problem statement*

In this paper, we consider a network of N Lur'e nonlinear dynamic described as

$$\begin{aligned} \dot{x}_i(t) &= Ax_i(t) + Bu_i(t) + E\varphi(y_i(t)), \\ y_i(t) &= Cx_i(t), i \in \mathcal{I}_N, \end{aligned} \tag{1}$$

where $x_i \in \mathbf{R}^n$ is the state to be synchronized, $y_i \in \mathbf{R}^m$ is the output and $u_i \in \mathbf{R}^p$ is the control input. $\varphi(y_i(t)) = [\varphi_1(y_{i1}(t)), \varphi_2(y_{i2}(t)), \dots, \varphi_m(y_{im}(t))]^T \in \mathbf{R}^m$ represents a nonlinear feedback loop belonging to the slope $[0, \Delta]$ with $\Delta = diag\{\delta_1, \delta_2, \dots, \delta_m\}$, i.e.,

$$\begin{aligned} &0 \le \frac{\varphi_k(b) - \varphi_k(a)}{b - a} \le \delta_k, \quad \varphi_k(0) = 0, \\ &\forall b \ne a \in \mathbf{R}, \quad k = 1, \dots, m. \end{aligned} \tag{2}$$

and A, B, C and E are system matrices with compatible dimensions.

Considering the multiple Lur'e systems described by (1), the consensus problem is said to be achieved if and only if for any initial conditions, one has

$$\lim_{t \to \infty} \| x_i(t) - x_j(t) \| = 0, \quad \forall i, j \in \mathcal{I}_N. \tag{3}$$

The objective of this paper is to design the proper controller and event-triggered protocol for each agent such that consensus condition (3) holds.

Throughout this paper, unless otherwise specified, the MASs (1) are assumed to satisfy the following assumption.

Assumption1: The matrix pair (A, B, C) in (1) is stabilizable and detectable.

Assumption2: The communication topology $\mathcal{G}$ is a strongly connected topology.

3. MAIN RESULT

In this section, we will show the main result about distributed event-triggered output feedback consensus control of multiple Lur'e systems. Firstly, we design a distributed observer to estimate the state of each agent. Then, according to the estimated state, we design a state-dependent event-triggered strategy and give the distributed control protocol. Finally, we obtain two theorems, which prove that the consensus of multi-agent systems based on designed protocol and the elimination of Zeno behavior.

3.1 *Distributed controller based on observer*

Considering that it is difficult to get the full states value of the system directly, the state observer is often used to estimate the system states. For the multiple Lur'e systems, the author of proposed an observer based on relative error to estimate consensus error of MASs. The observer can be expressed as

$$\dot{\hat{e}} = [I_N \otimes (A + FC)]\hat{e}(t) - (\mathcal{L} \otimes F)y(t) + (\mathcal{L} \otimes B)u(t)$$

where $\hat{e}$ is the estimated value of relative error, $F \in \mathbf{R}^{p \times n}$ is the observer gain matrix. And the observer needs continuously information of neighbors. We further find $\hat{e}(t) = (\mathcal{L} \otimes I_n)\hat{x}(t)$ where $\hat{x}$ is estimated state and $\hat{x}$ satisfies traditional state observer equation, that is

$$\dot{\hat{x}} = [I_N \otimes (A + FC)]\hat{x}(t) - (I_N \otimes FC)x(t) + (I_N \otimes B)u(t).$$

Based on the above observer, the estimated state of each agent only depends on its own output information. In this paper, the observer is applied to the distributed event-triggered consensus control strategy, and the control value u_i for agent i is described as the following equation:

$$\begin{aligned} &\dot{\hat{x}}_i(t) = A\hat{x}_i(t) + Bu_i(t) + F(y_i(t) - C\hat{x}_i(t)), \\ &u_i(t) = cK \sum_{j \in \mathcal{N}_i} a_{ij} \left(e^{A(t - t_k^i)}\hat{x}_i(t_k^i) - e^{A(t - t_k^j)}\hat{x}_j(t_k^j)\right), \\ &t \in [t_k^i, t_{k+1}^i), i \in \mathcal{I}_N \end{aligned} \tag{4}$$

where $\hat{x}_i$ is the estimated state of agent i. $K \in \mathbf{R}^{p \times n}$ is the feedback gain matrix, $c > 0$ is the coupling gain, and a_{ij} is the ij-th entry of the adjacency matrix $\mathcal{A}$. t_k^i is the latest triggering moment of agent i, and $x_i(t_k^i)$ is the last broadcast state of agent i. The exponential matrix $e^{A(t - t_k^i)}$ is to make the state estimation more accurate. From formula (4), we can see that agent only transfer the observe state to its neighbor intermittently, so the controller can avoid continuously updating the control. Next we define state measurement error of agent i as:

$$\xi_i(t) = x_i(t) - \hat{x}_i(t)$$

and the estimation error as:

$$\hat{e}_i(t) = e^{A(t - t_k^i)}\hat{x}_i(t_k^i) - \hat{x}_i(t).$$

Then, taking note of formula (4) and formula (1), we can get the closed-loop system of agent i as:

$$\begin{aligned} \dot{\xi}_i(t) = &\quad (A - FC)\xi_i(t) + E\varphi(y_i(t)), \\ \dot{x}_i(t) = &\quad Ax_i(t) + cBK \sum_{j \in \mathcal{N}_i} a_{ij} \left(e^{A(t - t_k^i)}\hat{x}_i(t_k^i)\right. \\ &\left. - e^{A(t - t_k^j)}\hat{x}_j(t_k^j)\right) + E\varphi(y_i(t)). \end{aligned}$$

Denote

$$\xi(t) = [\xi_1^T(t), \dots, \xi_N^T(t)]^T,$$
$$\hat{e}(t) = [\hat{e}_1^T(t), \dots, \hat{e}_N^T(t)]^T,$$
$$\Phi(y(t)) = [\phi(y_1(t)), \dots, \phi(y_N(t))].$$

According controller (4), the multi-agent closed-loop systemcan be calculated as

$$\begin{aligned}\dot{\xi}(t) = &\ [I_N \otimes (A - FC)]\xi(t) + (I_N \otimes E)\Phi(y(t)),\\ \dot{x}(t) = &\ (I_N \otimes A + \mathcal{L} \otimes cBK)x(t)\\ &+(\mathcal{L} \otimes cBK)(\hat{e}(t) - \xi(t)) + (I_N \otimes E)\Phi(y(t)).\end{aligned}$$

3.2 *Event-triggered threshold based on observed state*

Next, we can see from intuition that the estimation error can not be infinitely large in the control process, or it will lead to state divergence, which can also be reflected in the proof of consistency. So we design the state-dependent event-triggered threshold as

$$\| \hat{e}_i(t) \|^2 \leq \frac{\eta}{\omega} \| \hat{z}_i(t) \|^2 \qquad (5)$$

with

$$\hat{z}_i(t) = \sum_{j \in \mathcal{N}_i} a_{ij} \left(e^{A\left(t - t_k^i\right)} \hat{x}_i\left(t_k^i\right) - e^{A\left(t - t_k^j\right)} \hat{x}_j\left(t_k^j\right) \right)$$

$$0 < \eta < -H, w \geq \left|\left|\mathcal{L}\right|\right|^2 \Omega,$$

and

$$H = -2\beta_2 \lambda_{min}(R)\left(1 - \frac{1}{\epsilon}\right) + 4\left|\left|\Delta\right|\right|^2 + 2c\left|\left|\mathcal{L} \otimes P_2 B B^T P_2\right|\right|^2 + 2\left|\left|C^T C\right|\right|^2$$

$$H = c - 2\beta_2 \lambda_{min}(R)(1 - \epsilon)\left|\left|\mathcal{M}\right|\right|^2 + 4\left|\left|\mathcal{M} \otimes \Delta \mathrm{C}\right|\right|^2 + 2c\left|\left|\mathcal{L} \otimes P_2 B B^T P_2\right|\right|^2 + 2\left|\left|C^T C\right|\right|^2$$

where $\beta_2 > 0$ is a constant and $\epsilon > 1$ is the Young's inequality parameter. R is defined as Definition 2. The selection of ϵ needs to satisfy $H < 0, \Omega > 0\$$. P_2 is a positive matrix and $\mathcal{M} = I_N - \mathbf{1}_N \mathbf{r}^T$.

Therefore, we can get the event-triggered function of agent i from formula (5):

$$f_i\left(t, \hat{e}_i(t)\right) = \| \hat{e}_i(t) \|^2 - \frac{\eta}{\omega} \| \hat{z}_i(t) \|^2, i \in \mathcal{I}_N. \qquad (6)$$

When $f_i(t, \hat{e}_i(t)) > 0$, the event of agent i will be triggered and t_k^i will be recorded as current time t. Then the agent i will updates its controller and broadcast its information to its neighbors. From above description, we can see that the communication between agents occurs only at triggering times.

From controller (4) and event-triggered function (6), we can see that the information transfer between agents is only observe value.

3.3 *Consensus and Zeno behavior analysis*

In this section, we transfer the consensus problem to the stability problem by using disagreement vector $\delta(t) = [\delta_1(t)^T, \dots, \delta_N(t)^T]^T$, and $\delta_i(t) = x_i(t) - \sum_{j=1}^N r_j x_j(t)$. And considering the state information of the system not available, we define the new disagreement vector

$$\hat{\delta}_i(t) = \hat{x}_i(t) - \sum_{i=1}^N r_j \hat{x}_j(t)$$

and

$$\hat{\delta}(t) = (\mathcal{M} \otimes I_n)\hat{x}(t)$$
$$\bar{\hat{\delta}}(t) = (\mathcal{M} \otimes I_n)\bar{\hat{x}}(t)$$

where $\bar{\hat{x}}(t) = [e^{A(t - t_k^i)}\hat{x}(t_k^i)^T, \dots, e^{A(t - t_k^N)}\hat{x}(t_k^N)^T]$. Then we take $\hat{\xi}(t) = \delta(t) - \hat{\delta}(t) = (\mathcal{M} \otimes I_n)\xi(t)$ as disagreement error vector, and get

$$\begin{aligned}\dot{\hat{\delta}}(t) = &\ (I_N \otimes A + \mathcal{L} \otimes cBK)\hat{\delta}(t) + (\mathcal{L} \otimes cBK)\hat{e}(t)\\ &+(I_N \otimes FC)\hat{\xi}(t), \qquad (7)\\ \dot{\hat{\xi}}(t) = &\ [I_N \otimes (A - FC)]\hat{\xi}(t) + (\mathcal{M} \otimes E)\Phi(y(t)).\end{aligned}$$

Considering the definition of above variable, if $\hat{\delta}(t)$ and $\hat{\xi}(t)$ tend to zero, then $\delta(t)$ tends to zero that is $x_i(t) - x_j(t), i \in \mathcal{I}_N$ tend to zero. It can be seen from the above description that we transform the consensus problem based on output feedback of system (1) into the stability problem of system(7).

Next, we give the corresponding theorems and proof.

Theorem1: Consider the multiple Lur'e systems (1) satisfying Assumptions 1, 2and adopting the event-triggered control protocol (4) and (5). The disagreement vector $\delta(t)$ converges to zero for all initial conditions when $c > 0, c(\alpha(\mathcal{L})) \geq 1$ and taking $K = -\frac{1}{2} B^T P_2, F = P_1^{-1} C^T$, where P_1, P_2 are the solution of following algebraic linear matrix inequations :

$$P_1 A + A^T P_1 - 2C^T C + \beta_1 I < 0,$$
$$P_2 A + A^T P_2 - P_2 B B^T P_2 + \beta_2 I < 0$$

where $\beta_1 \geq \frac{1}{\lambda_{\min}(R)}(1 + \| R^2 \otimes P_1 E E^T P_1 + I_N \otimes 2\Delta^2 C^T C \|)$ and $\beta_2 > 0$.

We take the Lyapunov function candidate as following:

$$V(t) = \hat{\xi}^T(t)(R \otimes P_1)\hat{\xi}(t) + \hat{\delta}^T(t)(R \otimes P_2)\hat{\delta}(t).$$

The derivative of $V(t)$ along the trajectory of (7) is

$$\begin{aligned}\dot{V} = & 2\hat{\xi}^T(R\otimes P_1)\dot{\hat{\xi}} + 2\hat{\delta}^T(R\otimes P_2)\dot{\hat{\delta}} \\ = & \hat{\xi}^T[R\otimes[P_1(A-FC)+(A-FC)^TP_1]]\hat{\xi} \\ & +2\hat{\xi}^T(R\mathcal{M}\otimes P_1E)\Phi(y)+\hat{\delta}^T[R\otimes(P_2A \\ & +A^TP_2)+R\mathcal{L}\otimes cP_2BK+\mathcal{L}^TR \\ & \otimes cK^TB^TP_2]\hat{\delta}+2\hat{\delta}^T(\mathcal{L}\otimes cP_2BK)\hat{e} \\ & +2\hat{\delta}^T(I_N\otimes P_2FC)\hat{\xi}.\end{aligned}$$

According to Lemma 3 and techniques in matrix theory and inequalities, one has

$$\dot{V} \le \sum_{i=1}^{N} H\left\|\bar{\delta}_i\right\|^2 + \Omega\left|\left|\hat{e}_i\right|\right|^2 + \Pi\left\|\hat{\xi}_i\right\|^2$$

where $\Pi = -\beta_1\lambda_{\min}(R) + 1 + \| R^2\otimes P_1EE^TP_1 + I_N\otimes 2C^T\Delta^2C \|$. By choosing β_1 such that $\beta_1 \ge \frac{1}{\lambda_{\min}(R)}(1+\| R^2\otimes P_1EE^TP_1 + I_N\otimes 2\Delta^2C^TC \|)$, one has $\Pi < 0$. Therefore,

$$\dot{V} \le \sum_{i=1}^{N} H\left\|\bar{\delta}_i\right\|^2 + \Omega\left|\left|\hat{e}_i\right|\right|^2$$

Then considering event-triggered threshold satisfying (5) and $\omega \ge \| \mathcal{L} \|^2 \Omega$, $\frac{\|\hat{z}_i\|^2}{\|\mathcal{L}\|^2} \le \| \bar{\delta}_i \|^2$, one has

$$\| \hat{e}_i \|^2 \le \frac{\eta}{\omega}\| \hat{z}_i \|^2 \le \frac{\eta}{\Omega\| \mathcal{L} \|^2}\| \hat{z}_i \|^2 \le \frac{\eta}{\Omega}\| \bar{\delta}_i \|^2.$$

Furthermore, since $0 < \eta < -H$, we have

$$\dot{V} \le \sum_{i=1}^{N}(n+H)\left\|\bar{\delta}_i\right\|^2 < 0$$

By using Lyapunov stability theorem, we can obtain that the new disagreement vector $\hat{\delta}$ and disagreement error vector $\hat{\xi}$ asymptotically converge to zero, that is $\hat{\delta}\to 0, \delta-\hat{\delta}\to 0$ as $t\to\infty$. It implies that $\delta\to 0$ as $t\to\infty$ and $\lim_{t\to\infty}\| x_i(t)-x_j(t) \| = 0, \forall i,j\in\mathcal{I}_N$. In summary, the consensus proof of Lur'e MASs (1) using the control protocol based on output feedback is completed.

And we give the following theorem excludes the Zeno behavior.

Theorem2: Considering the multiple Lur'e systems (1) under the event-triggered control protocol (4) and (6), the time interval between arbitrary two triggering of agent i is larger than a positive lower bounded, or bounded from below by a positive constant, that is no Zeno behavior.

For each $[t_k^i, t_{k+1}^i)$, we denote $\mathcal{K}(t) = \frac{\|\hat{e}(t)\|}{\|\hat{z}(t)\|}$, and derive

$$\dot{\mathcal{K}}(t) = \frac{d}{dt}\left(\frac{\| \hat{e}(t) \|}{\| \hat{z}(t) \|}\right) \le \frac{\| \dot{\hat{e}}(t) \|}{\| \hat{z}(t) \|} + \mathcal{K}(t)\frac{\| \dot{\hat{z}}(t) \|}{\| \hat{z}(t) \|}. \quad (8)$$

First, for $\| \dot{\hat{e}}(t) \|/\| \hat{z}(t) \|$ in formula (8), we have $\dot{\hat{e}}(t) = (I_N\otimes A)\hat{e}(t) - (I_N\otimes cBK)\hat{z}(t) - (I_N\otimes FC)\xi(t)$. By denoting $\mathcal{F} = \| I_N\otimes A \|, \mathcal{H} = \| I_N\otimes cBK \|$ and $\mathcal{R} = \| I_N\otimes FC \|$, we obtain that

$$\| \dot{\hat{e}}(t) \| \le \mathcal{F}\| \hat{e}(t) \| + \mathcal{H}\| \hat{z}(t) \| + \mathcal{R}\| \xi(t) \|. \quad (9)$$

Consider that $\| \xi(t) \|$ is bounded and $\| \hat{z}(t) \| > 0, i.e. \exists\epsilon > 0, s.t. \| \hat{z}(t) \| \ge \epsilon$ before the MASs achieve consensus, $\frac{\|\xi(t)\|}{\|\hat{z}(t)\|} \le M$ where $M > 0$. We derive inequality (9) as

$$\frac{\| \dot{\hat{e}}(t) \|}{\| \hat{z}(t) \|} \le \mathcal{F}\frac{\| \hat{e}(t) \|}{\| \hat{z}(t) \|} + \mathcal{R}M + \mathcal{H}. \quad (10)$$

Next, for $\mathcal{K}(t)(\| \dot{\hat{z}}(t) \|/\| \hat{z}(t) \|)$ in formula (8), we have $\dot{\hat{z}} = (I_N\otimes A)\hat{z}(t)$ since $\hat{z}(t) = (\mathcal{L}\otimes I_n)\bar{\hat{x}}(t)$ and $\dot{\bar{\hat{x}}}(t) = (I_N\otimes A)\bar{\hat{x}}(t)$. And we get

$$\frac{\| \dot{\hat{z}}(t) \|}{\| \hat{z}(t) \|} \le \| I_N\otimes A \|. \quad (11)$$

Finally, substituting (10) and (11) into (8), we have

$$\dot{\mathcal{K}}(t) \le \mathcal{P}\mathcal{K}(t) + \mathcal{Q},$$

where $\mathcal{P} = \mathcal{F} + \| I_N\otimes A \|, \mathcal{Q} = \mathcal{R}M + \mathcal{H}$. So $\mathcal{K}(t)$ satisfies the bound $\mathcal{K} \le \Psi(t,\Psi), t\in[t_k^i, t_{k+1}^i)$, where Ψ is the solution of

$$\dot{\Psi}(t) = \mathcal{P}\Psi(t) + \mathcal{Q}. \quad (12)$$

Because $e_i(t_k^i) = 0, \mathcal{K}(t_k^i) = 0$, we set the initial condition $\Psi_0 = \Psi(t_k^i) = 0$ and get the solution $\Psi(t,\Psi_0) = \frac{\mathcal{Q}}{\mathcal{P}}e^{\mathcal{P}t} - \frac{\mathcal{Q}}{\mathcal{P}}$ of formula (12). Consider the event-triggered function (6) and the event-triggered time intervals have lower bound τ that satisfies $\Psi(\tau,0) = \frac{\mathcal{Q}}{\mathcal{P}}e^{\mathcal{P}\tau} - \frac{\mathcal{Q}}{\mathcal{P}}$. We can obtain $\tau = \frac{1}{\mathcal{P}}\ln(1+\frac{\mathcal{P}\mathcal{D}}{\mathcal{Q}}) > 0$ by solving $\Psi(\tau,0) = \mathcal{D}$ where $\mathcal{D} = \sqrt{\frac{\eta}{\omega}}$. So Zeno behavior is excluded.

4. CONCLUSION

In this section, a numerical example is performed to verify the effectiveness of the proposed distributed event-triggered protocol. Considering a group of six agents described by Chua's circuits, as a typical chaotic system, its synchronization problem has been widely studied. The dimensionless state equations of Chua's circuit can be transformed into the Lur'e form. Based on the method in , we give explicitly the Lur'e systems (1) as:

$$A = \begin{bmatrix}-1.25 & 5 & 0\\ 1 & -1 & 1\\ 0 & -6 & -0.01\end{bmatrix}, B = \begin{bmatrix}1\\0\\0\end{bmatrix}, E = \begin{bmatrix}-1.25\\0\\0\end{bmatrix},$$

$$C = [1\ 0\ 0];$$

$$\varphi(y_i) = \frac{1}{2}(|\, x_{i1}-1\,| - |\, x_{i1}+1\,|).$$

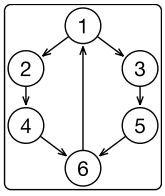


Fig. 1 The directed communication topology

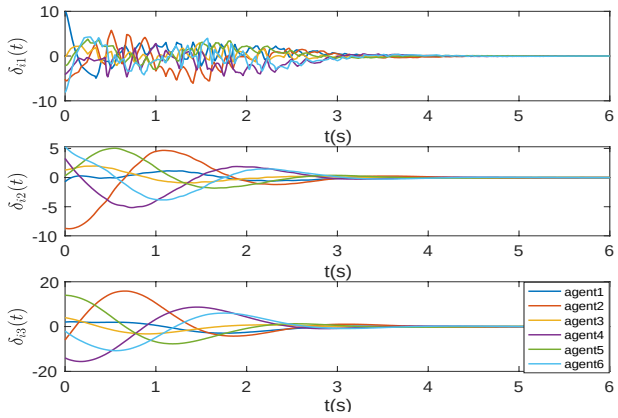


Fig. 2 State errors

The communication topology is given in Fig. 1, which satisfies Assumption 2.

And we give the designed parameters of the distributed event-triggered protocol based on output feedback. According to Theorem 1, we proceed by solving the algebraic Riccati equation, and get feedback matrixes as:

$$K = [-12.5341 \quad -22.7707 \quad 4.1313],$$
$$F = [3.8907; 2.9683; -1.5483].$$

For agent $i = 1,2,\ldots,6$, the initial states are chosen as $x_1(0) = [12; -4; 1], x_2(0) = [-3; -12; -7], x_3(0) = [2; -2; 3], x_4(0) = [-2; 0; -15], x_5(0) = [0; -3; 13], x_6(0) = [-6; 2; -3]$. The parameters in the event-triggered function (6) are calculated as $\eta = 51.6373$ and $\omega = 7354$. Take the initial value of observers as $\hat{x}_i = [0; 0; 0]$. The state errors is shown in Fig. 2, which explain that the consensus is achieved under the designed distributed output feedback control protocol (4) and (6). And the event-triggered function based on output feedback is shown in Fig. 3, which implies the event trigger mechanism (6) and excludes the Zeno behavior. As can be seen from above simulation, the protocol based on output feedback can eliminate the acquiring of full states for consensus problem based on event-triggered mechanism.

5. CONCLUSION

In this paper, we present a novel distributed event-triggered consensus protocol for Lur'e type Contrasting with the current research of Lur'e multi-agent systems, we consider a complex and practical

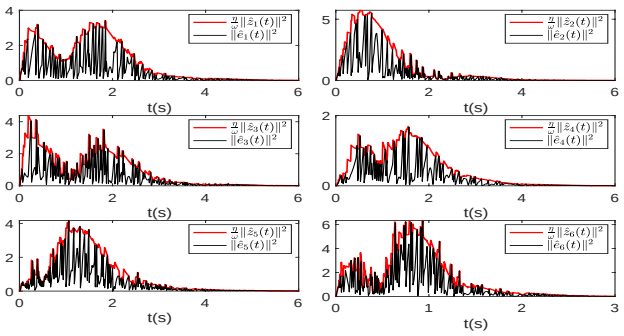


Fig. 3 Event-triggered function

environment, and improve the trigger strategy to reduce the limitation of on-board resources. Finally, we provide numerical simulation that demonstrate our theoretical results.

References

[1] J. Fax and R. Murray, Information Flow and Cooperative Control of vehicle Formations, *IEEE Transactions on Automatic Control*, vol. 49, 2004, pp 1465–1476.

[2] W. Ren, R. Beard and E. Atkins, Information Consensus in Multivehicle Cooperative Control, *IEEE Control Systems Magazine*, vol. 27, 2007, pp 71–82.

[3] H. Yan, F. Qian, H. Zhang, F. Yang, and G. Guo, H_∞ Fault Detection for Networked Mechanical Spring-mass Systems with Incomplete Information, *IEEE Transactions on Automatic Control*, vol. 63, 2016, pp 5622–5631.

[4] R. Olfati-Saber and R. Murray, Consensus Problems in Networks of Agents with Switching Topology and Time-delays, *IEEE Transactions on Automatic Control*, vol. 49, 2004, pp 1520–1533.

[5] W. Ren, On Consensus Algorithms for Double-integrator Dynamics, *IEEE Transactions on Automatic Control*, vol. 53 2008, pp 1503–1509.

[6] J. Wang, D. Cheng. D, X. Hu, Consensus of Multi-agent Linear Dynamic Systems, *Asian Journal of Control*, vol. 10, 2008, pp 144-155.

[7] F. Zhang, H. L. Trentelman, J. M. Scherpen, Fully Distributed Robust Synchronization of Networked Lur'e Systems with Incremental Nonlinearities. *Automatica*, vol. 50, 2014, pp 2515–2526.

[8] Z. Li, Z. Duan, G. Chen, Global Synchronised Regions of Linearly Coupled Lur'e Systems, *International Journal of Control*, vol. 84, 2011, pp 216–227.

[9] Y. Zhao, Z. Duan, G. Wen, and G. Chen, Robust Consensus Tracking of Multi-agent Systems with Uncertain Lur'e-type Non-linear Dynamics, *Control Theory & Applications Iet*, vol. 7, 2013, pp 1249–1260.

Uncovering Users' Decisions through Serious Game Playing with A Formal Description Method

Akinobu SAKATA *, Takamasa KIKUCHI**, Ryuichi OKUMURA***, Masaaki KUNIGAMI***,
Atsushi YOSHIKAWA***, Masayuki YAMAMURA***, Takao TERANO ****
* Department of Computational Intelligence and System Science, Tokyo Institute of Technology
Yokohama, Kanagawa, Japan
** Graduate School of Business Administration, Keio University
Yokohama, Kanagawa, Japan
*** Department of Computer Science, Tokyo Institute of Technology
Yokohama, Kanagawa, Japan
**** Chiba University of Commerce,
Ichikawa, Chiba, Japan

Abstract

The purpose of this study is to verify that virtual cases involving players' changes in awareness during the gaming process can be described with the Managerial Decision-making Description Model (MDDM). Previous studies proposed a method to measure and evaluate players' cognition and judgment during gaming. Based on this, we developed a game system with a function to detect players' changes in awareness. Your Life to Come Game (YLCG), which we originally developed, runs on this system. We checked whether it was possible to formally describe virtual cases in which players experienced changes in awareness during the game and found that the formal description language of the MDDM had this capability.

Keywords: simulation and gaming, serious games, business simulations, formal description method, MDDM, decision making, game performance, virtual case.

1. INTRODUCTION

The purpose of this study is to show that it is possible to formally describe virtual cases involving a player's change in awareness in gaming. We defined a virtual case as a situation that consists of players' cognition and decision-making in a game experience, as opposed to a business case that consists of decision-making in real business. A change in awareness is defined as a change in the priority of a player's play objectives. Play objectives are what players aim to achieve in gaming.

Gaming is a traditional method that originated in military training exercises [1][2]. In recent years, gaming for the business sector has received a lot of attention [3]. In this type of gaming, players compete as individuals or a team to achieve a predetermined goal, following the facilitator's instructions and prescribed rules [3]. In this independent experience, players learn by themselves.

Protocol analysis [4][5][6] has generally been used to measure and evaluate players' cognition and judgment during game playing (e.g., [7], [8], and [9]) . This procedure involves collecting, transcribing, and analyzing voice data emitted by players during the gaming process. Since protocol analysis requires time to compile, analyze, and evaluate the data, it may not be suitable for applications in which game facilitators use the results of the analysis in the middle of a course of education or training with gaming.

The "performance sheet" (PS) [9] developed by Koshiyama et al. is suitable to overcome the problems of these traditional methods. Koshiyama et al. introduced the PS to an existing business simulation game. During the game, each player records his or her own perceptions and judgments on a PS. The information recorded by the player in the PS is the recognition of the target state, state variables and control variables. Then, the researchers compared the histories of players' cognitions and judgments recorded in the PSs with those revealed by protocol analysis. The results showed that the PS could be a good alternative to protocol analysis. Specifically, they found that it is possible to know the history in players' cognitions and judgments and detect changes in them, and to compare them between players.

In this study, we developed a game system with a function to record players' decisions and play goal priorities during the gaming process, based on the works of Koshiyama et al.. We also developed a serious game called Your Life to Come Game (YLCG) that runs on the game system. The game system runs in the PC

environment. In addition, we extracted a virtual case in which a player had a change in

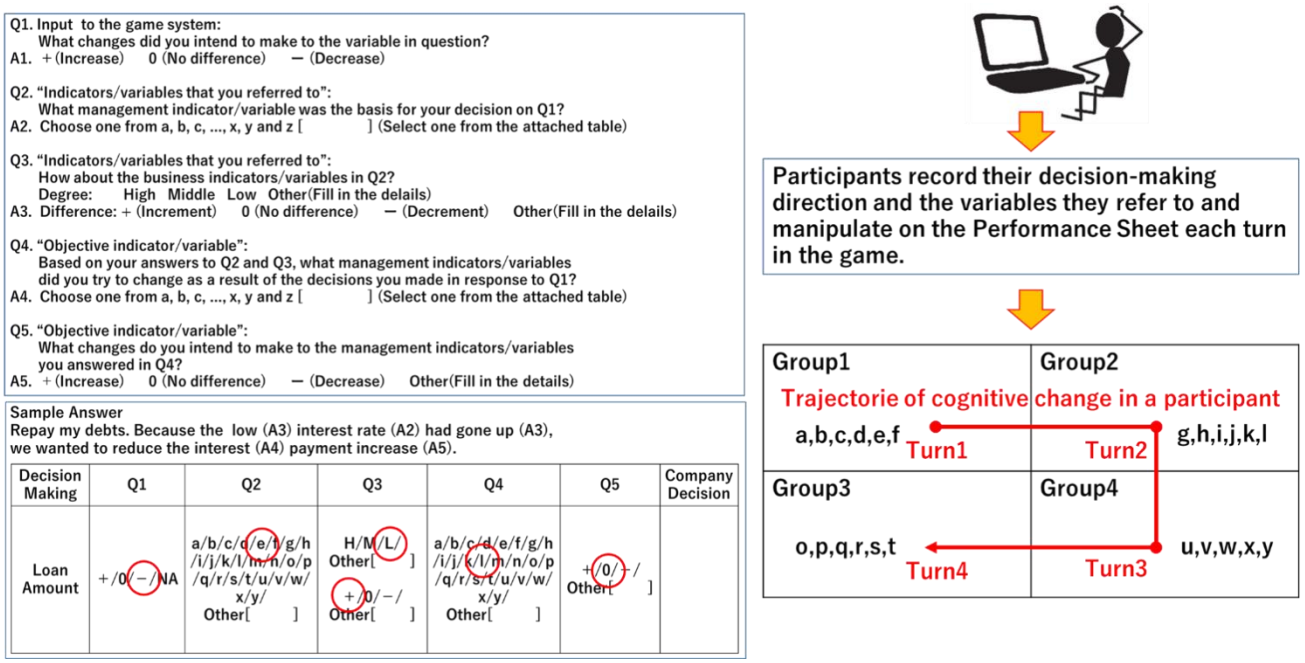


Fig.1 An Overview of Visualization of Cognition and Judgment in Business Games.
The player records his or her perceptions and decisions on the PS each turn. Each of the variables in the game referenced by the player belongs to one group. If the groups are connected by arrows to each other in the order in which variables are referred to, it is easy to visually understand how their cognition changes over time.The figure was described with reference to [9].

awareness, as seen from the play logs collected in the serious game experiment. We then attempted to describe the virtual case using the Managerial Decision-making Description Model (MDDM) [10][11]. The results confirmed that MDDM adequately described the player's change in awareness.

The formal description of the virtual cases generated from the game playlog has the advantages described below. Nakano et al. developed a business game based on real business cases, and showed that virtual cases similar to real business cases can be generated by gaming [12]. Kikuchi et al. showed that real business cases and hypothetical cases generated by Agent-Based Model (ABM) can be formally described and compared, respectively [13]. Based on the work of Nakano et al. and Kikuchi et al. it is not only easier to visually understand virtual cases if we can formally describe virtual cases generated from gaming playlogs, but also to compare virtual cases generated by real cases and ABMs with virtual cases generated from gaming playlogs. A simple understanding and comparison of cases generated by various means could support game facilitators working in gaming and debriefings.

The structure of this paper is as the follows. Section 2 describes previous methods of measuring player's cognition and judgment in the gaming process and MDDM, a formal and comparable model for representing agents' decisions in business cases. Section 3 describes the experimental method of this study. Section 4 describes the results of the experiments. Here, a decision-diagram created using MDDM is presented, which is a virtual case containing player's change in awareness generated from gaming playlogs. In addition, a text describing the participant's perceptions of his own cognition and decision-making, which were obtained from the interviews with the participant conducted after the gaming, is also presented. In Section 5, we analyze the results of the experiments. Section 6 summarizes this study.

2. RELATED WORK

2.1 Methods for Measuring Players' Cognition and Judgment in the Gaming Process

Many researchers used protocol analysis [4][5][6] to measure and evaluate players' cognition during the gaming process (e.g. [7], [8], and [9]). However, there is a problem in that it is difficult to provide players (learners) with instruction based on the results of protocol analysis in educational activities or trainings because the protocol analysis takes time and may not be completed before the end of gaming.

To overcome this challenge, Koshiyama et al. introduced PS into Simulation and Gaming, which has a function to record players' cognition during gaming and visualize it

as the game progresses [9]. Koshiyama et al.'s approach using PS to visualize cognition, shown in Fig. 1, is to

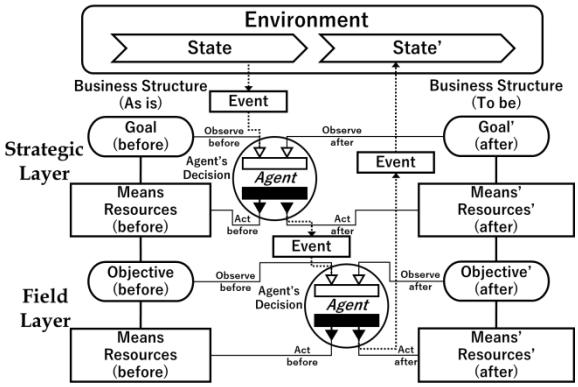


Fig.2 An example of case described by MDDM.

The MDDM represents the managerial decision-making as a decision diagram with the three components, the Environment (top), the business structure (right and left side) and the agent's decision (four terminal elements between the Business Structures)

consider the game as a player's control problem and record how the player perceives the target state, state variables, observed variables, and control variables (i.e., the concept of the problem) at any given time. Koshiyama et al. showed that PS-based method enables experimenters to understand players' cognitions and judgments, detect changes in them, and compare them between players by contrasting the results of the PS-based analysis with the results of protocol analysis. The PS-based analysis method can also be used to visualize changes in players' cognition and decision-making over time so that experimenters can visually understand them.

2.2 Managerial Decision-Making Description Model (MDDM)

Kunigami et al. proposed the MDDM as a formal and comparable model for representing agents' decisions in business cases to respond to changes in the business structure of an organization [10][11]. This model consists of three elements: the Business Structure Component, Environment Component, and Agent's Decision Element. By placing these elements in a frame and linking them together according to their relationships, we describe the decision-making associated with changes in the organization's business structure as a "decision diagram."

Here, the Business Structure Component consists of symbols of "objectives" and "means" in each layer of the organization (Strategic Layer, Middle Layer, Field Layer, etc.). The Environment Structure Component consists of state symbols and their changes over time, both inside and outside the organization. It also places "event" symbols generated from changes in state. The decision-making element of the agent is represented by a device with four terminals. The top two terminals connect the objects that the agent observes. The bottom two terminals are connected to the target on which the agent acts.

The model can be described by the following four features:

(a) The multi-layered structure of the organizational business, and its transition.
(b) The focus (or bounded scope) of the agent's observations and actions.
(c) The agent's position corresponding to each layer in the business structure.
(d) The chronological order and the causality of the agents' decisions.

It has been pointed out that MDDM has the potential to represent the simulation logs of the actual business case and the agent model in a certain common format. In Fig. 2, the MDDM represents management's decision-making as a decision diagram with three elements: the Environment Component (top), the Business Structure Component (left and right), and the Agent's Decision Element (the four end elements between the Business Structure Components).

3. METHODS

In this section, we describe our experimental method. Section 3.1 provides an overview of the serious game YLCG. Next, Section 3.2 introduces the execution and development environment of the game system. Section 3.3 provides information on the experiment participants, and Section 3.4 describes the experiment procedure.

3.1 Your Life to Come Game (YLCG)

This section describes the details of YLCG, a turn-based serious game. In the game, a player takes on the role of a Japanese businessperson and experiences a virtual life in an environment provided by the game system. YLCG requires the player to use player-specific resources (i.e., time, money, and abilities) for various purposes on each turn. The quantity of resources used by the player is assigned by the game system to the MATH model described below. As a result, the player's resource and status information change.

3.1.1 MATH model

YLCG incorporates the MATH model (See Fig. 3) so that the game system makes players experience the consumption and acquisition/loss of resources during their virtual life. For this study, we adopted a simplified MATH model that excluded the health component. The individual equations corresponding to the simplified

MATH model are listed in (1)–(12); the variables of the MATH model are summarized in Table 1.

$$\begin{bmatrix} M_{(t+1)} \\ A_{(t+1)} \end{bmatrix} = \begin{bmatrix} M_{(t)} + \Delta M_{(t)} \\ A_{(t)} + \Delta A_{(t)} \end{bmatrix} \quad \#\#(1)$$

Table 1. The parameters used in the MATH model.

Parameters	Description	Parameters	Description
t_{Ework}	Time spent for work	ε_{Invest}	Normalized random number
t_{Learn}	Time spent on developing skills for the job	ε_{Trust}	Normalized random number
t_{Invest}	Time spent on investments	μ_{Invest}	Average return on investment
m_{Invest}	Funds to be spent on investment	μ_{Trust}	Average return on investment trust
m_{Trust}	Funds to be spent on investment trust	σ_{Invest}	Standard deviation of return on investment
m_{Learn}	Cost of developing job skills	σ_{Trust}	Standard deviation of return on investment trust
c_{Ework}	Effectiveness of growth in work capacity per time spent on the job	Δt	Time available in one turn
c_{Learn}	Effectiveness per time spent on capacity developing skills for the job	ΔA_{Invest}	Amount of change in the ability to invest per turn
c_{Invest}	Effectiveness per time spent on capacity building of investment	ΔA_{Ework}	Amount of change in work capacity per turn
W_{Ework}	Pay per hour spent on the job	Δm_{Invest}	Funds recouped from one turn of investment.
A_{Invest}	Investment ability	Δm_{Ework}	Wages earned from one turn of work.
A_{Ework}	Work ability	Δm_{Trust}	Funds recouped from one turn of investment trust.

Table 2. Detailed Description of Events.

Once an Environment Event (EE) is triggered, it automatically modifies the equations and parameters that make up the MATH model, or changes the status information of the player. Additionally, once Ordinary Event (OE) and Extraordinary Event (ExOE)) is triggered by a player, the game system modifies MATH model parameters and equations, or changes the player's status information.

Event	Type	Occurrence condition	Details
Job Hunting	OE	This event occurs every turn.	Players can choose from the following occupations: university students, freelancers, investors, and company employees.
Marriage Hunting	ExOE	Player's attribute is set to unmarried	Players have a 20% chance of getting married. When the player's attribute becomes a married player, the maximum amount of time resources that can be used in one turn is reduced from 100% to 90%.
Financial Chance	ExOE	This event has a 20 % chance of occurring.	When you perform this event, there is a 5% chance to increase your savings by 5 times and a 95% chance to increase your savings by 1/5.
Childbirth	EE	If the player is married, this event has a 1/3 chance of occurring.	When a birth event occurs, the number of children is automatically increased by one in the player's attributes.
Financial Crisis	EE	This event has a 10% chance of occurring.	In the event of a financial crisis, the return on investments is automatically increased by 0.05 times and the return on mutual funds is increased by 0.25 times.

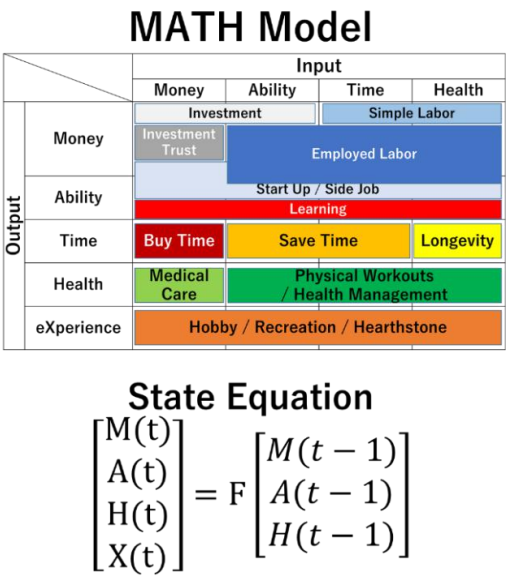


Fig.3 MATH model.

The MATH model represents the phenomenon that a player uses his or her resources (money, ability, time, and health) for various purposes in each turn, and gains (or loses) new resources as a result.

$$A_{(t)} = \begin{bmatrix} A_{Ework(t)} \\ A_{Invest(t)} \end{bmatrix} \quad (2)$$

$$t_{Ework\,(t)} + t_{Invest(t)} + t_{Learn\,(t)} \le \Delta t \quad \#(3)$$

$$m_{Learn\,(t)} m_{Invest\,(t)} + m_{Trust\,(t)} + m_{Learn\,(t)} \le M_{(t)} \quad \#(4)$$

$$\Delta A_{Ework\,(t)} = \left(c_{Ework} \times t_{Ework\,(t)} + c_{Learn} \times \sqrt{m_{Learn\,(t)}} \times t_{Learn\,(t)}\right) \times A_{Ework\,(t)} \quad (5)$$

$$\Delta A_{Invest\,(t)} = c_{Invest} \times \sqrt{t_{Invest\,(t)}} \times A_{Invest\,(t)} \quad \#(6)$$

$$\Delta M_{(t)} = \Delta m_{Ework\,(t)} + \Delta m_{Ework\,(t)} + \Delta m_{Invest\,(t)} + \Delta m_{Trust\,(t)} - m_{Invest\,(t)} - m_{Trust\,(t)} - m_{Learn\,(t)} \quad (7)$$

$$\Delta m_{Ework\,(t)} = W_{Ework} \times \left(1 + \frac{A_{Ework\,(t)} - A_{Ework\,(0)}}{A_{Ework\,(0)}}\right) \times t_{Ework\,(t)} \quad (8)$$

$$\Delta m_{Invest\,(t)} = \varepsilon\left(\mu_{Invest}, \sigma_{Invest\,(t)}\right) \times m_{Invest\,(t)} \quad \#(9)$$

$$\Delta m_{Trust(t)} = \varepsilon\left(\mu_{Trust}, \sigma_{Trust(t)}\right) \times m_{Trust\,(t)} \quad \#(10)$$

$$\sigma_{Invest\,(t)} = \sigma_{Invest\,(0)} \times \left(1 + \frac{A_{Invest\,(t)} - A_{Invest\,(0)}}{A_{Invest\,(0)}}\right) \quad \#(11)$$

$$\sigma_{Trust\,(t)} = \sigma_{Trust\,(0)} \quad \#(12)$$

3.1.2 The Steps to Play YLCG

First, when it is the player's turn, an environmental event (EE) occurs stochastically. When an EE occurs, the MATH model calculations are corrected for each type of EE. Next, players must consider the correspondence between ordinary events (OE) and extraordinary events (ExOE). An OE and ExOE is an event that a player is allowed to process once per turn, unconditionally. An ExOE is generated by the game system when the player meets certain conditions. The game system allows a player to process ExOE only once per turn (see Table 2 for details on each event). Third, a player is ordered to use his or her resources. Fourth, after the player completes the resource allocation task, the game system presents him or her with multiple play objectives and instructs the player to prioritize them. Finally, the player's state is updated according to the MATH model built into the game system, and the turn transitions.

3.1.3 Players' Decision-making about Using Resources

When playing YLCG, players must allocate their unique resources (i.e., time and money) to a total of six items. The items are Money and Time for Stock Investments, Time for Mutual Funds, Time for Work, and Money and Time for Learning. Each turn, the player allocates an amount to be spent on each item from his or her own savings and then allocates time to each item within a range of 0% to 100%.

3.1.4 Prioritizing Play Objectives

At the end of each turn, the game system presents the player with some pre-prepared play objectives and asks him or her to prioritize them. Table 3 shows the options of the play objectives registered in the game system. If the player decides that none of the play objectives presented by the system are appropriate, he or she may add new, original ones. If a change in the order of the play objectives is observed, it is assumed that the player has had a change in awareness. Figure 4 shows a screenshot of the player deciding the priority of the play objectives.

Table 3. Options of play objectives.

No.	Options of play objectives
A	Securing a stable source of income.
B	Acquiring knowledge and skills that are useful on the job.
C	Earning a high income.

3.1.5 Visualizing Players' Process

Each player can see the history of the values of savings, investment capacity, and work capacity on a line chart.

3.1.6 Parameters of YLCG

The values of the parameters used within the YLCG are listed in Table 4.

3.2 Game System

YLCG can be played on a PC running Windows 10. The game system was developed using Unity 2019.4.6f1. The programming language used in the system development with Unity is C#. Unity was selected because it is expected that players will be able to play serious games on non-

Fig. 4 Examples of the game screen of YLCG (Left: Standard screen, Right: The screen in prioritizing play objectives.).
The left column shows the player's various status information (from top to bottom: amount of money saved, investment ability, working ability, occupation, marital status, number of children). On the left side of the screen, events that occur at each turn (from the top, job hunting, marriage hunting, childbirth, profit-telling, and financial crisis) are lined up. At the bottom center of the screen, there are six forms for players to input their decisions (money and time). The player enters numbers into the form using the keyboard.

Table 4. Constants used in the MATH model.
The player is given a parameter set that corresponds to the task he has chosen. Different jobs have different values of the parameters related to the reward of the job.

Job	c_{Ework}	c_{Learn}	c_{Invest}	c_{WEwork}	μ_{Invest}	μ_{Trust}	σ_{Invest}	σ_{Trust}
University Students	0.4	0.28	2.5E-6	1	1.3	1.04	1.3	0.104
Employee	0.7	0.28	2.5E-6	1	1.3	1.04	1.3	0.104
Investor	0.4	0.28	2.5E-6	1	1.3	1.04	1.3	0.104
Part-time jobber	0.4	0.28	2.5E-6	1	1.3	1.04	1.3	0.104

Table 5. History of the participant's decisions and priority of play objectives. ("#" indicates that the value of the distributed resources and the priority of the play objectives has changed from the previous turn.).

Turn	Decisions						Priority of play objectives		
No.	t_{Invest}	t_{Learn}	t_{Ework}	m_{Invest}	m_{Trust}	m_{Ework}	No. 1	No. 2	No. 3
1	0	0	0	0	0	0	A	B	C
2	0	30	10	0	0	1	A	B	C
3	5	30	20	0	2	1	A	B	C
4	10	10	70	0	3	3	A	B	C
5	10	10	70	0	4	3	A	B	C
6	10	10	70	0	2	1	A	B	C
7	10	10	70	0	2	0	A	B	C
8	10	5	70	0	2	0	A	B	C
9	15	5	70	0	**3#**	0	**B#**	**A#**	C
10	15	5	70	0	**4#**	0	B	**C#**	**A#**

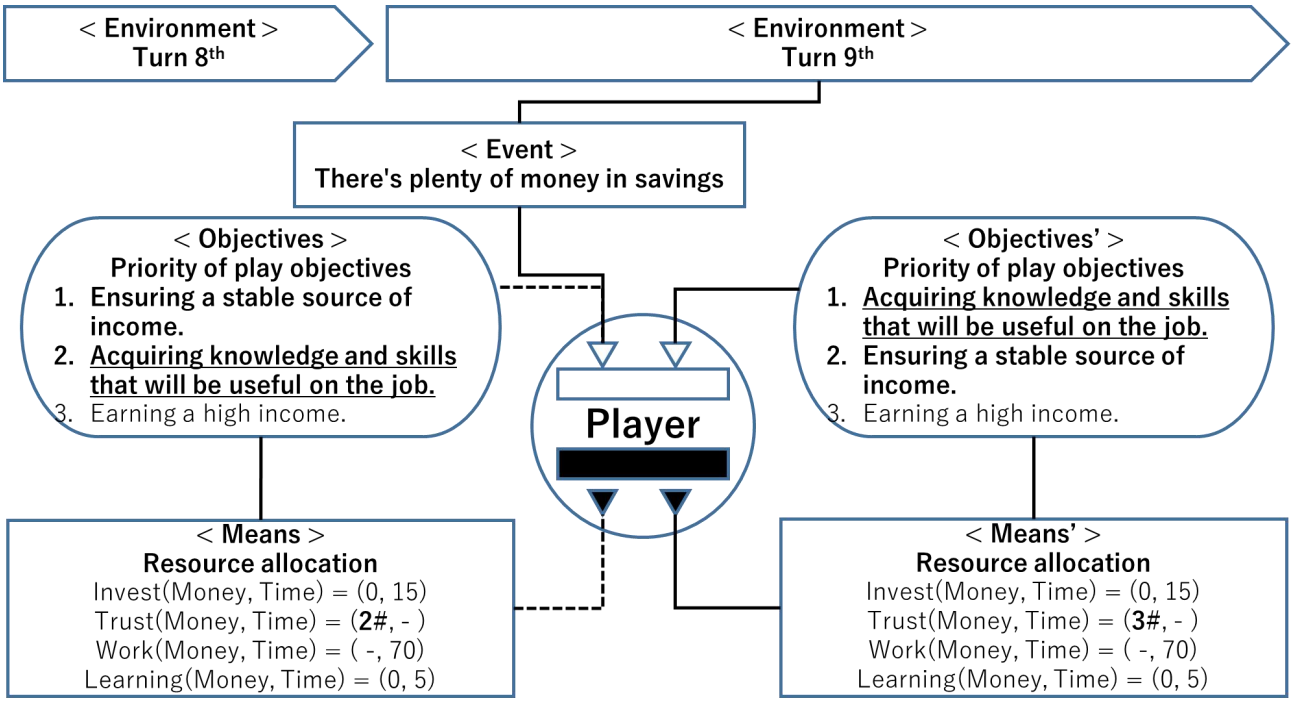


Fig. 5 The above figure was transcribed in the decision diagram from the YLCG playlog and interviews.
Prioritized play goals are described in the "objective" symbol and the results of resource allocation are described in the "means" symbol. From the debriefing interviews, the motivation for the change in decision-making is described in the "event" symbol.

The Question from the Experimenter to the Participant>
From turn 8 to 9, the "Acquiring knowledge and skills that will be useful on the job" moved to the number one priority. In our last interview, you said that in turn 9 you thought, "Now that I have some extra money in savings, I'm going to try some things. Honestly, if the importance of the play goal of "Acquiring knowledge and skills that will be useful on the job" becomes more important, I think we would invest time and money in learning and increasing our capabilities, but You did not ... Why is that?

The Answer from the Participant to the Experimenter>
I thought it was important to learn within the game. But I wasn't sure what benefit I would get from investing my time and money in learning in this game. I was hesitant to invest in learning because while I was playing the game, I didn't realize that there was a way to check what benefit I would get from investing in learning. Additionally, In general, I believe that learning computers is different from learning investments. In the former case, the more you study, the more results you get, but even if you spend a lot of time and money on your investment and study diligently, you won't necessarily succeed. If I thought it was useless to do it in the real world, I didn't want to do it in a game either.

Fig.6 The participant was interviewed on the next day of the experiment, and the answers are summarized by the experimenter as shown above.

Windows operating systems (such as Linux and Mac OS X) in the future.

3.3 Participants

The participant were one Japanese businesspersons.

3.4 Procedure

First, the participants were given an explanation of how to operate the game system and play YLCG. They were briefed on the player status information, the various types of events and how to handle them, the various resources and resource distribution, and how to prioritize play goals and play objectives. Then, the game was played for 10 turns per player. At the end, we asked each participant to explain why his or her changes in awareness occurred.

4. RESULTS

As the example virtual case, a participant's decisions and play goal priorities for each turn are listed in Table 5. The play logs shows that the order of priority of play objects, which did not change through Turn 8, changed on Turn 9. We considered this phenomenon as a change in awareness and created a decision-making diagram in the MDDM (see Fig. 5). The "objectives" symbol was described based on the priorities of the play objectives, and the "means" symbol was described based on the content of the resource allocation. The "event" symbol was based on the results of the interviews with the participants during the post-game debriefing.

On Turn 9, the player changed Play Objective B (Acquiring knowledge and skills that will be useful on the job) from second to first priority. In contrast, the content of the player's decisions regarding resource allocation changed only slightly in terms of the amount of investment trust. The participant was interviewed a second time in order to analyze and clarify the relationship between his change in awareness and judgment as described above (see Fig. 6.

The second interview revealed the following. The participant recognized that there was potential for some benefit in playing YLCG by developing his capacity to invest in learning. However, in playing the game, he did not discover the significance of using his resources for investment learning. Additionally, the participant recognized that taking a lot of time to learn about investing does not make sense in the real world. Therefore, the participant's attitudes about investment in the real world influenced his decisions about resource allocation in the game.

5. DISCUSSION

The first interviews with the player revealed that on Turn 9, he changed his perception of the amount of money he could afford to save. This corresponds to the content of the "event" symbol in the decision diagram. Then, the update of the player's recognition triggered a change in the priority of the play objectives. This corresponds to a change in the content of the "objective" symbol. A change in the content of the "means" symbol would reflect the player's perception that he was now able to do what he could have not done before because he had more money to save.

On the other hand, the second interview with the participant revealed that his change in awareness of the increasing importance of learning was not necessarily reflected in his decision-making during gaming. It seems to have been difficult for the players to understand the structure of the MATH model for developing his capability within the limited play time. In addition, the players' real-life experiences and common sense influenced their decisions for resource allocation. A similar phenomenon was reported by Nakano et al. in their study. Nakano et al. point out that the presence or absence of business experience related with business simulation can make a difference in the gaming experience [12]. This indicates that prior knowledge and beliefs about the problems represented in the game may affect players' perceptions and decisions during gaming.

As discussed, the experimental results showed that it may be possible to formally describe virtual cases in which a player changes his or her awareness during gaming with the MDDM. This suggests that using the MDDM to create a decision diagram of individual players' cognitions and judgments during the gaming process may help the players themselves, experimenters, and other observers to visually and easily understand the players' actions and the intentions behind them. This possibility will need to be tested in the future.

The scope of application of the MDDM would be not limited to a game played by a single player. We also consider that this work can be applied to games in which multiple players participate in repeated interactions. These gaming simulations are designed to allow players to refer to each other's decisions and mid-game performance (e.g., [14], [15], and [16]). In such a situation where players observe each other, one player may experiences insights and changes in his or her cognition and judgment because he or she observes other players' actions and results. A decision diagram described with the MDDM may be useful to understand virtual cases involving cognitive and judgmental changes that occur as a result of player-to-player interactions such as the above. To do this, we need to find evidence that the MDDM's "event" symbol is associated with changes in the decisions of other players.

6. CONCLUSION

In this study, we developed an original serious game, YLCG. Furthermore, we implemented a function in the game system to record the history of players' decisions and play objectives during the gaming process. Next, in the game experiment, we extracted data on the players' changes in awareness and created decision diagrams using the MDDM. As a result, we showed that it was possible to describe a case in which a player's change in awareness with the formal description language, the MDDM.

In the future, we will verify whether our system and the MDDM can formally describe virtual cases that include players' cognition and decisions detected using PS. Additionally, we focused on gaming in which a single player participated in the game in this study; however, in the future we plan to show that the use of MDDM can be effective in gaming in which multiple players participate.

References

[1] R. Smith, The long history of gaming in military training, Simulation & Gaming, Vol. 41, No. 1, 2010, pp. 6–19.

[2] P. P. Perla, and E. McGrady, Why war gaming works, Naval War College Review, Vol. 64, No. 3, 2011, pp. 111–130.

[3] T. Terano, Learning business decisions through cases and games. J. Soc. In strum. Control Eng. Vol. 46, No. 1, 2007, pp. 44–50. (in Japanese)

[4] K. A. Ericsson, and H. A. Simon, Verbal reports as data, Psychological review, Vol. 87, No. 3, 1980, p. 215.

[5] K. A. Ericsson and H. A. Simon, Protocol analysis: Verbal reports as data, the MIT Press, 1984.

[6] K. A. Ericsson and H. A. Simon, Protocol analysis, MIT Press, Cambridge, MA, 1993.

[7] T. Nonaka, K. Miki, R. Odajima and H. Mizuyama, Analysis of Dynamic Decision Making Underpinning Supply Chain Re-Silience: A Serious Game Approach, IFAC-Papers OnLine, Vol. 49, No. 19, 2016, pp. 474 – 479.

[8] Z. Feng, V. A. Gonz alez, M. Trotter, M. Spearpoint, J. Thomas, D. Ellis and R. Lovreglio, How People Make Decisions during Earthquakes and Post-Earthquake Evacuation: Using Verbal Protocol Analysis in Immersive Virtual Reality, Safety Science, Vol.129, 2020, p. 104837.

[9] O. Koshiyama, M. Kunigami, A. Yoshikawa and T. Terano, Analyzing Behaviors of Business Game Learners Using a Modified Performance Sheet, Simulation & Gaming, Vol. 2, No.2, 2011, pp. 86–95, (in Japanese).

[10] M. Kunigami, T. Kikuchi and T. Terano, A Formal Model of Managerial Decision Making for Business Case Description, In F. Koch, A. Yoshikawa, S. Wang, and T. Terano, (Eds.) Evolutionary Computing and Artificial Intelligence. GEAR 2018, Communications in Computer and Information Science 999, 2019, pp. 21–26, Springer.

[11] M. Kunigami, T. Kikuchi, H. Takahashi and T. Terano, A Formal, Descriptive Model for the Business Case of Managerial Decision-Making, In Agents and Multi-Agent Systems: Technologies and Applications 2020, 2020, pp. 355-365, Springer, Singapore.

[12] K. Nakano and T. Terano, Integrating a Learning System with Case Method and Business Gaming, Simulation & Gaming, Vol. 16, No. 1, 2006, pp. 13–27 (in Japanese).

[13] T. Kikuchi, M. Kunigami, H. Takahashi and T. Terano, Explaining Log Data of Agent-Simulation Results with Managerial Decision-Making Description Model, Simulation & Gaming, Vol. 29, No. 1, 2019, pp. 36–48 (in Japanese).

[14] M. Critelli, D. I. Schwartz and S. Gold, Serious socialgames: designing a business simulation game, In 2012 IEEE Inter-national Games Innovation Conference, 2012, pp. 1–4, IEEE.

[15] Y. Shinoda, M. Ryoke, T. Terano and Y. Nakamori, Design of a software agent for business gaming simulation, Journal of Systems Science and Systems Engineering, Vol. 15, No. 1, 2006, pp. 83–94.

[16] K. Nakano, S. Matsuyama and T. Terano, Researchon a learning system toward integration of case method and businessgaming, In Agent-Based Approaches in Economic and Social Com-plex Systems IV, 2007, pp. 21–30, Springer.

A controller for Quantized System under DoS Attacks with UDP-Like Protocol

Wenjie Liu*, Jian Sun*,**, Jie Chen*,***
* State Key Lab of Intelligent Control and Decision of Complex Systems
School of Automation, Beijing Institute of Technology
Beijing 100081, China
** Beijing Institute of Technology Chongqing Innovation Center,
Chongqing, 401120, China
*** Tongji University,
Shanghai 200092, China

Abstract

This paper addresses the problem of stabilizing a linear time-invariant system with quantized signals under UDP-like protocol as well as in the presence of Denial-of-Service (DoS) attacks. A controller that can inform the encoder of the attacks by zero input signals is put forth by considering network phenomena (i.e., constrained bandwidth and DoS attacks) only at the output channel (i.e., the sensor to controller channel), which guarantees the synchronization between the encoder and the decoder without additional acknowledgements from the controller and can thus stabilize the system. The main technical steps behind are as two folds, the first one is by equipping a predictor-based controller with deadbeat controller gain; and the second one lies in the method of designing the sampling period. A numerical example is given to validate the effectiveness of the proposed method.

Keywords: Denial-of-Service (DoS), quantization, user datagram protocol (UDP), acknowledgement-based protocol, deadbeat controller.

1. INTRODUCTION

Driven by advances in computing and networking technologies, nowadays, a majority of modern systems (e.g., [1]-[4]) transmit their measurements and control data through Internet or wireless communication. However, despite of the advantage in flexibility, the use of the networks increases the vulnerability of the control systems and makes them prone to cyber threats. Evidences have shown that malicious attacks can disrupt the nominal performances of the modern applications and consequently inflict serious loss of people's lives and properties, thus the topic of enhancing the resilience of the existing systems in the presence of cyber-attacks has lately triggered considerable attention.

Several cyber-attacks have been investigated, including replay attacks [5], false-data injection attacks [6], and Denial-of-Service (DoS) attacks [7], what we are interested in the present work is with DoS attacks. Generally, DoS attacks are launched to induce jamming in the communication channels, and cause packet losses. Amidst the large body of research on DoS attacks (e.g., [8], [9]), particularly relevant here is the work on resilient control under such attacks, covered in [10]-[20], and references therein. Differences of these papers mainly lie in their assumptions on DoS attacks and in the stability properties they aim to. For instance, [10] characterized DoS attacks by constraining the attacker's energy, whereas [11] described DoS attacks in the form of pulse-width modulated signals. Recently, inspired by average dwell time approach, a generalized attack model was proposed by [12]. Assumptions on this model only constrain attackers on their attack frequency and duration, and the analysis is proceeded by considering the original system as a two-mode switched system, i.e. one mode is the system in the absence of DoS attacks, and the other is in the presence of DoS attacks. Building on this framework and propagating from the switched systems method, studies have been made ranging from controller designing for systems with single output channel [13] to systems with multiple output channels [14], and from nonlinear systems [15] to distributed systems [16]. [17] extended this attack model by allowing not only deterministic strategies but also stochastic ones in the generation of malicious attacks. [21] utilized the model in [17] to characterize a state-dependent jamming attack.

On the other hand, digital sensors, digital controllers and data links with limited data rate are typical in numerous implementations of modern systems, and they all induce some degree of quantization. Therefore, how to design suitable encoding schemes for different systems is quite interesting from theoretical as well as practical point of view. By approximating reachable sets of signals that need to be quantized, the so-called “zooming-in" and “zooming-out" encoding method that can stabilize the continuous time-invariant linear systems is first brought

up by [22]. Due to its simple structure, extensions are developed in several directions, see, e.g. [23]-[33]. Problems of quantization dealing with single-mode systems have been thoroughly studied. To name a few, [23] and [24] stabilized the systems with quantized output measurements. Although the class of observers discussed in [23] and [24] including Luenberger observers as well as pseudo-inverse observers, the condition for the Luenberger observer case is difficult to verify. Therefore, [25] provided a easily verifiable condition by designing a different encoding scheme for Luenberger observer. Another line, pioneered by [26], is to control switched systems under limited bit-rate. Difficulty presented in this setting is how to approximate the reachable sets when switching signals are partially known. For example, assumed that the active mode of the switched signal is known at each sampling time, [26] stabilized the system by extending [27], and [28] derived the ultimate bound of the state trajectories adopting the method developed in [29].

In view of the commonality of the technical tools employed for the analysis of system under DoS attacks and for control design with limited transmission capacity, a marriage of these two research areas is quite nature. In particular, the switched systems approach and the reachable set approximation method will play an essential role in our analysis. In addition, in this context, since the encoder and the decoder are allocated in the different sides of the network, asynchronization may happen due to DoS attacks. For concreteness, if a DoS attack occurs, the decoder is notified by zero packet arrivals at the latest transmission instant, and then its quantization range is updated by the rule used in the presence of DoS attacks; the encoder, however, remains utilizing the quantization range updating rule in the absence of DoS attacks since the input signal, generated by predictor-based controller, does not affect by network phenomena, namely the control input is nonzero. Therefore, synchronization between the encoder and the decoder, namely both the encoder and the decoder should be aware of whether the communication channel is blocked due to the DoS attacks or not, is a prerequisite before designing the encoding schemes for stability. [34] and [35] handled this issue by using an acknowledgement-based protocol, e.g., transmission control protocol (TCP-like protocol). Nevertheless, in the real-time systems, protocols without acknowledgements (ACKs), e.g., user datagram protocol (UDP-like protocol), are more preferable due to the fact that the implementation of this transport protocol is simpler and the additional energy and time consumption for the acknowledgement signals transmission is avoided [36].

For simplicity, we refer to the limited bandwidth and DoS attacks as *network phenomena* in the rest of the paper. This work generalizes the main result of [35] with the UDP-like communication protocol. To avoid confusion, in the following, let the *sampling period* denote the time interval between two consecutive signals from the controller to the actuator, and the *transmission period* refer to the time interval between two consecutive output signals. We consider network phenomena at the sensor to controller channel (output channel). To synchronize the encoder and the decoder without ACKs, we put forth a controller that can make DoS attacks detectable from the control input, under conditions on the controller gain as well as the sampling period. Sufficient condition for stability is derived and analysis shows that the stability condition in [35] can be recovered by using our controller as an axillary detector. In a nutshell, the main contributions of the present work are:

c1) By designing the controller gain and the special relationship between the transmission period and the sampling period, the encoder can detect DoS attacks from the input signals, namely the control inputs turned into zero when DoS attacks occur, and thus can synchronize with the decoder. Therefore, the acknowledgement messages are no more needed, and UDP-like communication protocol is enough for the system;

c2) Assume that the output channel has constrained bandwidth and is subject to DoS attacks. In addition to the controller, an encoding scheme that can stabilize the system under the condition on the quantization level of the output signal and DoS attacks is proposed. Compared with the stability condition using the TCP-like protocol in [35], robustness under DoS attacks is degraded by such controller under the UDP-like protocol, therefore, a modified controller structure is provided to recover the robustness.

The remainder of this paper is organized as follows. Section 2 introduces the preliminaries and the problem formulation. Main results of the paper is presented in Section 3. A numerical example is showcased to substantiate the effectiveness of the proposed method in Section 4. Finally, Section 5 ends the paper with conclusions.

Notation: We denote the set of real numbers by $\mathbb{R}$, and $\mathbb{Z}$ the set of integers. Given $\alpha \in R$ or $\alpha \in Z$, let $R_{>\alpha}(R_{\geq\alpha})$ or $Z_{>\alpha}(Z_{\geq\alpha})$ denote the set of real numbers of integers greater than (greater than or equal to) α. We let N denote the set of natural number and define $N_0 := N \cup \{0\}$. For a vector $v = [v_1, v_2, \cdots, v_n]^T \in R^n$ we denote its maximum norm by $|v| := max\{|v_1|, \cdots, |v_n|\}$ and the corresponding induced norm of a matrix $M \in R^{m\times n}$ by $\|M\| := sup\{|Mv| : v \in R^n, |v| = 1\}$.

2. PRELIMINARIES AND PROBLEM FORMULATION

In this section, we first introduce the closed-loop dynamics and then impose assumptions on the duration and the frequency of DoS attacks.

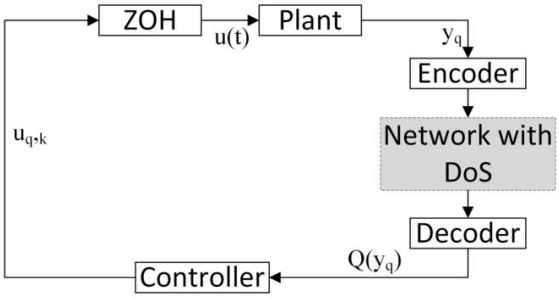


Table 1. The networked control architecture with only output channel connected by network.

2.1 System formulation

The linear continuous-time dynamical system we are to stabilize is as follows:

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) \end{cases} \tag{1}$$

where $x(t) \in \mathbb{R}^{n_x}, u(t) \in \mathbb{R}^{n_u}$, and $y(t) \in \mathbb{R}^{n_y}$ are the state, the input and the output of the plant, respectively. The pair (A, B) is stabilizable and (C, A) is observable. This plant is connected with an observer-based controller through a time-driven encoder and zero-order hold (ZOH). In the present paper, we assume that only the output channel suffers from network phenomena, see Table 1. In addition, the transmission period and the sampling period are denoted by Δ and δ, respectively. Define

$$x_{q,k} := x(q\Delta + k\delta), \quad y_{q,k} := y(q\Delta + k\delta) \tag{2}$$

for every $q \in Z_{\geq 0}$ and $k = 0, \cdots, \frac{\Delta}{b}, b \in N_1$. For notational brevity, let x_q denote $x_{q,0}$. Set

$$A_d := e^{A\delta}, \quad B_d := \int_0^{\delta} e^{As} B \; ds. \tag{3}$$

Throughout this paper, we impose the following assumptions on the system, involving the relationship between the transmission period and the sampling period, and the bound of the initial state.

Assumption1. *(Transmission and sampling period): The transmission period and the sampling period satisfy $\Delta = b\delta, b \in Z_{\geq 1}$.In another word, the transmission period is a multiple of the sampling period, and if $b = 1$, then the transmission attempts are synchronized with the sampling attempts.*

The initial state bound, denote by E_{st}, can be obtained by the zooming-out method procedure in [27], and this method has already extended to the DoS corrupted network by [35]. Therefore, we here assume that the initial bound is known.

Assumption2. *(Initial state bound): A constant $E_{st} > 0$ satisfying $|x_0| \leq E_{st}$ is known.*

2.2 Denial-of-Service attack

We refer to DoS attacks as attacks that prevent transmissions over the communication network. Considering a general DoS model proposed in [12], which constrains the attacker's action in time by only posing limitations on the frequency and the duration of DoS attacks. Since the communication between the plant and the controller happens periodically, it is reasonable to assume that attacks only happen when the plant and the controller is attempting to exchange packets. This can be considered as a model for reactive jamming discussed in [37], and hence instead of the original continuous-time DoS attacks model in [12], we adopt the discrete-time version of this model as mentioned in [32].

First, based on the concept of average dwell-time [38], DoS frequency, denoted by $\Phi_f(k)$, is the limitation on the number of DoS off/on transitions that happen at transmission instants on a time interval $[0,k\Delta)$. For simplicity of the notation, we omit Δ and use $[0,k)$.

Assumption3. *(DoS frequency [35, As 2.2]): For time interval $[0,k)$, there exist constants $\Pi_f \in R_{\geq 0}$ and $v_f \in R_{\geq 2}$ such that*

$$\Phi_f(k) \leq \Pi_f + \frac{k}{v_f} \tag{4}$$

for all $k \in Z_{\geq 0}$.

Second, let $\Phi_d(k)$ denote the DoS duration, which represents the number of time-steps (the length of each step is the transmission interval Δ) that attacks launch on the interval $[0, k\Delta)$, and we also omit Δ.

Assumption3. *(DoS duration [35, As 2.1]): For time interval $[0,k)$, there exist constants $\Pi_d \in R_{\geq 0}$ and $v_d \in Z_{\geq 1}$ such that*

$$\Phi_d(k) \leq \Pi_d + \frac{k}{v_d} \tag{5}$$

for all $k \in \mathbb{Z}_{\geq 0}$.

Remark 1: In Assumption 3, $v_f\Delta$ can be regarded as the average dwell-time between two consecutive DoS attacks launched time-steps. Similarly, Assumption 4 indicates that, the average duration time of DoS attacks dose not exceed a certain proportion of the whole-time domain, as specified by $\frac{1}{v_d}$. Constants Π_f and Π_d are chatter bounds. $v_d \geq 1$ and $v_f \geq 2$ work together to constrain that DoS attacks can neither occur infinitely fast nor last infinitely long.

3. MAIN RESULTS

In the scenario, where only the output channel has constrained bandwidth and is subject to DoS attacks, the decoder is able to recover the output signal form the index received from the encoder if, and only if they share the same quantization range, and hence they should change their update schemes at the same time. Therefore, since the decoder and the encoder are employed on the different sides of the network, without acknowledgements from the decoder, the encoder does not know whether DoS attacks occur or the decoder changes its update scheme or not. This causes severe asynchronization between the encoder and the decoder, and consequently instability of the system. In the following, we first demonstrate that the state diverges due to the asynchronization between the encoder and the decoder under the same controller as in [35], yet without acknowledgements from the controller (i.e., UDP-like protocol). Then, by designing the controller gain and the sampling period, we explain that our controller can guarantee the synchronization between the decoder and the encoder with UDP-like protocol.

3.1 Instability analysis

Consider the controller with Luenberger observer and the sampling period is same as the transmission period. This structure is discussed in [35] and the only difference is that instead of the TCP-like protocol, we here employ the UDP-like protocol. Under this type of protocol, observer in the encoder and the controller side asynchronized when DoS attacks occur, which causes the differences on both of the quantization centers and the quantization ranges. Let $\hat{x}_k$ and $\tilde{x}_k$ denote the estimated state at the controller side and the encoder side, respectively, and we have

$$\begin{cases} \hat{x}_{k+1} = A_d\hat{x}_k + Bu_k + L_k(\widehat{Q}(y_k) - \hat{y}_k) \\ u_k = K\hat{x}_k \end{cases} \quad (6)$$

and

$$\begin{cases} \tilde{x}_{k+1} = A_d\tilde{x}_k + B\tilde{u}_k + L(Q(y_k) - \tilde{y}_k) \\ \tilde{u}_k = K\tilde{x}_k \end{cases} \quad (7)$$

Assume that a DoS attack occurs at $k_1\delta$ and no attacks happen before or after $k_1\delta$, that is,

$$\begin{cases} 0, k = k_1 \\ L, k \neq k_1 \end{cases}.$$

Let $\{s_r\}_{r\in N_0}$ denote the sequence of successful transmission instants. Let $E_{e,k}$ and $E_{d,k}$ denote the estimated error of the state at the encoder side and the decoder side, respectively, and we obtain

$$E_{d,k+1} := \begin{cases} \theta_a E_{d,k}, k\delta \neq s_r \\ \theta_0 E_{d,k}, (k-1)\delta \neq s_r, k\delta = s_r \\ \theta_{na} E_{d,k}, (k-1)\delta = s_r, k\delta = s_r \end{cases}$$

and

$$E_{e,k+1} := \begin{cases} \theta_{na} E_{e,k}, k\delta > 0 \\ \theta_0 E_{e,k}, k\delta = 0 \end{cases}$$

and the quantization signal can be represented as

$$\begin{aligned} \widehat{Q}(y_k) &= \hat{y}_k + Q_k^{num}\frac{E_{d,k}}{N} \\ Q(y_k) &= \tilde{y}_k + Q_k^{num}\frac{E_{e,k}}{N} \end{aligned} \quad (8)$$

Since no attacks happen before $k_1\delta$, $\widehat{Q}(y_k) = Q(y_k), \forall k \le k_1$, and we further deduce that,

$$\begin{aligned} &\hat{x}_{k_1} = \tilde{x}_{k_1} \quad (9) \\ &\hat{x}_{k_1+1} = (A_d + B_dK)\hat{x}_{k_1} \\ &\tilde{x}_{k_1+1} = (A_d + B_dK)\tilde{x}_{k_1} + LQ_{k_1}^{num}\frac{E_{e,k_1}}{N} \\ &\hat{x}_{k_1+2} = (A_d + B_dK)^2\hat{x}_{k_1} + LQ_{k_1+1}^{num}\frac{E_{d,k_1+1}}{N} \\ &\tilde{x}_{k_1+2} = (A_d + B_dK)^2\tilde{x}_{k_1} + LQ_{k_1+1}^{num}\frac{E_{e,k_1+1}}{N} + (A_d \\ &\qquad + B_dK)LQ_{k_1}^{num}\frac{E_{e,k_1}}{N}. \end{aligned}$$

To reach a contradiction, assuming the sequence $\{E_{e,k}, k \ge k_1\}$ satisfies $E_{e,k} > |x_k - \tilde{x}_k|, \forall k \ge k_1$. Since $E_{e,k+1} = \theta_{na}E_{e,k}, k > 0$, sequence $\{|x_k - \tilde{x}_k|\}$ is decreasing. Let $\bar{A} := A_d + B_dK$ and $\widetilde{A} = A_d - LC$. Combining (8) and (9), we arrive at

$$\begin{aligned} |x_{k_1+1} - \tilde{x}_{k_1+1}| &= |\widetilde{A}(x_{k_1} - \tilde{x}_{k_1}) - L(Q(y_{k_1}) - y_{k_1})| \\ &\le E_{e,k_1+1} =: \widetilde{E}_{e,k_1+1}. \end{aligned}$$

Similarly,

$$\begin{aligned} |x_{k_1+2} - \tilde{x}_{k_1+2}| &= |\widetilde{A}(x_{k_1+1} - \tilde{x}_{k_1+1}) - L(Q(y_{k_1+1}) \\ &\quad - y_{k_1+1}) - BKLQ_{k_1}^{num}\frac{E_{e,k_1}}{N}| \\ &\le E_{k_1+2} + \frac{\|BKLQ_{k_1}^{num}\|}{N}\frac{1}{\theta_{na}^2}E_{e,k_1+2} \\ &=: \widetilde{E}_{e,k_1+2}. \end{aligned}$$

Iteratively, for $l \ge 3$, we have

$$\begin{aligned} &|x_{k_1+l} - \tilde{x}_{k_1+l}| \\ &\le E_{e,k_1+l} + \frac{1}{\theta_{na}^l}\frac{\left\|B_dK\bar{A}^{l-2}LQ_{k_1+1}^{num}\right\|(\theta_a - \theta_{na})E_{e,k_1}}{N} \\ &+ \frac{1}{\theta_{na}^l}\frac{\left\|B_dK\bar{A}^{l-1}LQ_{k_1}^{num}\right\|E_{e,k_1}}{N} \\ &+ \sum_{i=0}^{l-3}\frac{1}{\theta_{na}^{i+3}}\frac{\left\|B_dK\bar{A}^{i}LQ_{k_1+l-i-1}^{num}\right\|(\theta_0\theta_a - \theta_{na}^2)E_{e,k_1}}{N} \\ &=: \widetilde{E}_{e,k_1+l}. \end{aligned}$$

Since $\frac{1}{\theta_{na}} > 1$, in this case $\{\widetilde{E}_{e,k}\}$ is an increasing sequence, which contradict to the assumption that $\{|x_k - \tilde{x}_k|\}$ is a decreasing sequence. Therefore, we conclude that even if only one DoS attack happens, there exist $k > k_1$ such that $E_{e,k} < |x_k - \tilde{x}_k|$, and the state diverges eventually.

3.2 Controller under UDP-like protocol

Recall that (A,B) and (C,A) are controllable and observable, respectively, and thus (A_d,B_d) is controllable and $(C,A_d^{\eta_2})$ is observable. Let $R:=A_d^{\eta_2}(I-MC)$ and $\bar{R}:=A_d+B_dK$. Therefore, we can chose suitable $M\in R^{n_x\times n_y}$ and $K\in R^{n_u\times n_x}$ such that R is schur stable, and

$$R^{\eta_2}=0 \tag{10}$$

where η_2 is the controllability index of (A_d,B_d). The sampling period δ satisfies

$$\delta=\frac{\Delta}{\eta_2}. \tag{11}$$

Let $\{s_r\}_{r\in N_0}$ denote the sequence of successful transmission instants. The discrete-time predictor-based observer for feedback control and output quantization is described by:

$$\begin{cases} \hat{x}_{q,k+1}=A_d\hat{x}_{q,k}+B_du_{q,k}, k\delta\neq s_r \\ \hat{x}_q=\hat{x}_{q-1,\eta_2}+M(Q(y_q)-\hat{y}_{q-1,\eta_2}), k\delta=s_r \\ u_{q,k}=K\hat{x}_{q,k} \\ \hat{y}_{q,k}=C\hat{x}_{q,k} \end{cases} \tag{12}$$

where $\hat{x}_{q,k}\in R^{n_x},\hat{y}_{q,k}\in R^{n_y},Q(y_q)\in R^{n_y}$ are estimated state, estimated output and quantization of y_q, respectively. Set the initial estimated $\hat{x}_0$ to be $\hat{x}_0=0$. Through the ZOH, the control input $u(t)$ is generated as

$$u(t)=u_{q,k},\quad q\Delta+k\delta\le t<q\Delta+(k+1)\delta \tag{13}$$

for every $q\in Z_{\ge 0}$ and $k=0,\cdots,\eta_2-1$.

Remark 2. This controller structure is based on the model-based method, which was adopted in the previous study [13] on control without quantization under DoS attacks. The deadbeat gain K can be calculated by several approaches, e.g., [39], [40].

3.3 Output encoding scheme

Recall that the transmission period $\Delta=\eta_2\delta$, thus the encoder and the decoder update their quantization bounds at $q\Delta, q\in Z_{\ge 0}$. Let $E_q\ge 0$ satisfy

$$|e_{q-1,\eta_2}|\le E_q. \tag{14}$$

Since the output is continuous and the estimation error of the output satisfies

$$y_{q,k}-\hat{y}_{q,k}=Ce_{q,k}$$

then,

$$|y_{q-1,\eta_2}-\hat{y}_{q-1,\eta_2}|=|y_q-\hat{y}_{q-1,\eta_2}|\le\|C\|E_q.$$

We partition the hypercube

$$\{y\in R^{n_y}:|y_q-\hat{y}_{q-1,\eta_2}|\le\|C\|E_q\}$$

into N^{n_y} equal boxes. An index in $\{1,\cdots,N^{n_y}\}$ is assigned to each partitioned box by a certain one-to-one mapping for all $k\in Z_{\ge 1}$. The encoder sends to the decoder the number $Q^{num}(y_q)$ of the divided box containing y_k, and then the decoder recovers $Q(y_q)$ by the center of the box with the number $Q^{num}(y_q)$. If y_q lies on the boundary on several boxes, then we can choose any one of them. Propagating index implies that the encoder and the decoder have to be synchronized, only in this way can the correct state value be recovered from the decoder. The quantization error $|Q(y_q)-y_q|$ of this encoding scheme satisfies

$$|Q(y_q)-y_q|\le\frac{\|C\|}{N}E_q.$$

The sequence of error bound $\{E_q:q\in Z_{\ge 1}\}$, which satisfies (14) and exponentially decreases, will be specific in the next subsection.

Before analyzing the condition for stability, we first explain how this controller is able to maintain synchronization between the decoder and the encoder. Consider an arbitrary transmission interval $[k_r\Delta,(k_r+1)\Delta]$, and according to (12), we have

$$\hat{x}_{k_r+1}=\hat{x}_{k_r,\eta_2}=(A_d+B_dK)^{\eta_2}\hat{x}_{k_r}=0 \tag{15}$$

if a DoS attack occurs at $(k_r+1)\Delta$, and

$$\hat{x}_{k_r+1}=\hat{x}_{k_r,\eta_2}+M(Q(y_{k_r,\eta_2})-\hat{y}_{k_r,\eta_2})=\bar{R}^{\eta_2}\hat{x}_{k_r}+M(Q(y_{k_r,\eta_2})-\hat{y}_{k_r,\eta_2})=MQ(y_{k_r,\eta_2})\neq 0 \tag{16}$$

if a DoS attack does not happen at $(k_r+1)\Delta$. Because $E_q\neq 0, y_q=y_{q-1,\eta_2}\neq 0$, if we let N be an even integer, then $Q(y_q)\neq 0$. In addition, notice that $\hat{y}_{q,\eta_2}=0, q\in Z_{\ge 1}$, the center of the encoder is at the origin and thus structure (12) is not necessary in the encoder, which also means that we can encode the output with less computational resources. Therefore, the basic idea of making attacks detectable is by causing a zero input when a DoS attack occurs, i.e., $u_q=K\hat{x}_q=0$, and then the encoder can change its update scheme, which ensures that the decoder and the encoder are synchronized. Specific acknowledgement messages are thus no more needed. This allows the practical implementation by using a UDP-like communication protocol.

3.4 Result on exponential convergence

Before presenting our main result in this section, we first introduce the update rule of $\{E_q:q\in Z_{\ge 1}\}$. Notice that $A_d^2(I-MC)$ is schur stable, there exist constants M_0,M_1 such that

$$\left\|R^l\right\|\le M_0\rho^l,\left\|R^lA_d^{\eta_2}M\right\|\le M_1\rho^l. \tag{17}$$

Define constants $\theta_a,\theta_0,\theta_{na}>0$ by

$$\theta_a:=\|A_d^{\eta_2}\|,\theta_0:=M_0\rho+\frac{M_1\|C\|}{N},\theta_{na}:=\rho+\frac{M_1\|C\|}{N}.$$

Using these constants, the error bound $\{E_q:q\in Z_{\ge 1}\}$ is given by:

$$E_{q+1}: = \begin{cases} \theta_a E_q, \ \&q\Delta \neq s_r \\ \theta_0 E_q, \ \&(q-1)\Delta \neq s_r, q\Delta = s_r \\ \theta_{na} E_q, \ \&(q-1)\Delta = s_r, q\Delta = s_r \end{cases} \quad (18)$$

According to Assumption 2, the initial value E_0 is denoted by

$$|e_0| = |x_0| \leq E_{st} = :E_0. \quad (19)$$

The lemma below derives the aforementioned updating rule of the estimated error bound E_q.

Lemma 1. *Consider system (1) with controller (12), and K be chosen such that (10) holds. If the sampling period satisfies (11). Assume that $|e_{q-1,\eta_2}| \leq E_q$ and the set $\{E_q: q \in Z_{\geq 1}\}$ as in (18), then $|e_{q+l-1,\eta_2}| \leq E_{q+l}$ for all $l \in Z_{\geq 1}$.*

Proof. First, considering the system in the absence of DoS attacks, and according to (12), we deduce that

$$x_q - \hat{x}_q = (I - MC)(x_{q-1,\eta_2} - \hat{x}_{q-1,\eta_2}) - M\big(Q(y_q) - y_q\big)$$
$$x_{q,l} - \hat{x}_{q,l} = A_d^l (x_q - \hat{x}_q)$$

therefore,

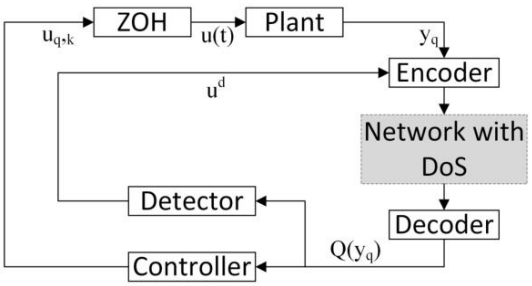


Table 2. The network structure with decoder in cascade with the controller and the detector.

$$x_{q+1} - \hat{x}_{q,\eta_2} = R(x_q - \hat{x}_{q-1,\eta_2}) - A_d^{\eta_2} M\big(Q(y_q) - y_q\big) \quad (20)$$

and iteratively,

$$\begin{aligned} \left|x_{q+l} - \hat{x}_{q+l-1,\eta_2}\right| &\leq \left|R^l (x_q - \hat{x}_{q-1,\eta_2})\right| \\ &\quad + \left|\sum_{j=0}^{l-1} R^j A_d^{\eta_2} M\big(Q(y_q) - y_q\big)\right| \\ &\leq \left\|R^l\right\| \left|e_{q-1,\eta_2}\right| + \sum_{j=0}^{l-1} \left\|R^j A_d^{\eta_2} M\right\| \frac{\|C\|}{N} E_q \\ &\leq E_{q+l}. \end{aligned}$$

The last inequality holds for $l \in Z_{\geq 1}$, if ρ, M_0, M_1 satisfy (17). Therefore, in the absence of DoS attacks, if $|e_{q-1,\eta_2}| \leq E_q$, then $|e_{q+l-1,\eta_2}| \leq E_{q+l}$ for all $l \in Z_{\geq 1}$.

Next, assume that $|e_{q-1,\eta_2}| \leq E_q$ and DoS attack occurs at $q\Delta$, then from (12) and (15), we obtain

$$|e_{q,\eta_2}| = \left|A_d^{\eta_2} e_q\right| = \left|A_d^{\eta_2} e_{q-1,\eta_2}\right|.$$

Recalling the update rule of E_q in the (18), we arrive at

$$|e_{q,\eta_2}| \leq \left\|A_d^{\eta_2}\right\| E_q \leq E_{q+1}.$$

Therefore, we conclude that $|e_{q+l-1,\eta_2}| \leqslant E_{q+l}$ holds for all $l \in Z_{\geqslant 1}$.

After deriving the update scheme for $\{E_q: q \in Z_{\geq 1}\}$, it is necessary to analysis its convergence property, which plays a crucial role in proving the stability theorem. Consider system (1) with predictor-based controller (12), and M, K be chosen such that (10) holds. If the transmission period and the sampling period satisfy (11). If Assumptions 1-4 hold.

Lemma 2. If the number of the quantization levels N is an even integer and satisfies

$$N > \frac{M_1 \|C\|}{1-\rho} \quad (21)$$

and the DoS attack satisfies

$$\frac{1}{v_d} \leq \frac{\log(1/\theta_{na})}{\log(\theta_a/\theta_{na})} - \frac{\log(\theta_0/\theta_{na})}{\log(\theta_a/\theta_{na})} \frac{1}{v_f} \quad (22)$$

then there exist $\Omega \geq 1$ and $\gamma \in (0,1)$ such that

$$E_q \leq \Omega \gamma^k E_0, \quad \forall k \in Z_{\geq 1} \quad (23)$$

Proof. The proof follows directly from that of the Lemma 3.9 in [35].

Based on Lemmas 1 and 2, we are now ready to present our stability Theorem 1, which indicates that, if the encoding scheme with the error bound $\{E_q: q \in Z_{\geq 1}\}$ updated by (18) is implemented, then the state can achieve exponential convergence under condition on DoS attacks and the quantization levels.

Theorem1. *Consider system (1) with predictor-based controller (12), and M,K be chosen such that (10) holds. If the transmission period and the sampling period satisfy (\ref{delta}). If Assumptions 1-4 hold. If the number of the quantization levels N is an even integer and satisfies (21) and the DoS attack satisfies (22) then the system achieves exponential convergence under the encoding scheme with the error bound $\{E_q: q \in Z_{\geq 1}\}$ constructed by the update rule (18).*

Remark 3. The controller discussed above can detect DoS attacks at the price of the degradation of the system robustness against the attacks. Therefore, a natural thought is generating merely switching signals for the encoder by this controller, and generating the control input by the controller same as in the TCP-like protocol. In another

Table 2. The network structure with decoder in cascade with the controller and the detector. In other words, since the general controller gain can make $\bar{R}$ schur stable and

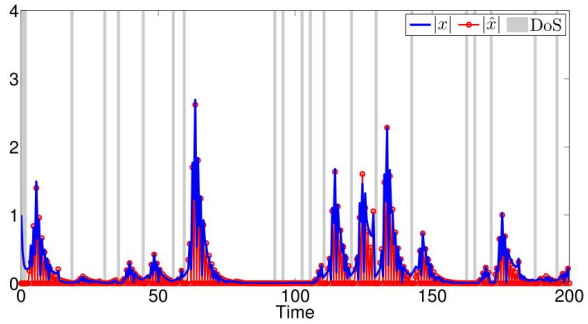


Table 3. Maximum norm of $\boldsymbol{x}$ and its estimate $\hat{\boldsymbol{x}}$ with controller (12).

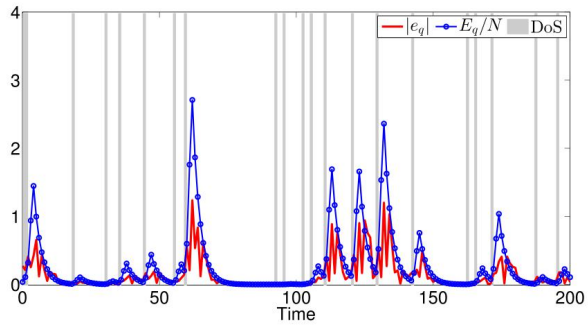


Table 4. Relationship between the $\boldsymbol{E_q/N}$ and $|\boldsymbol{e_q}|$ with controller (12).

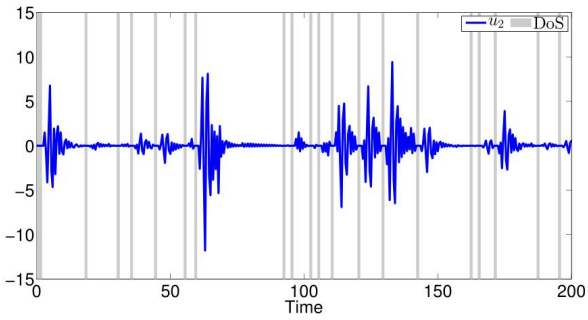


Table 5. Control input $\boldsymbol{u_{q,k}}$ with controller (12).

hence can speed up the quantization decrease process, we can divide the aforementioned controller into controller with the general K that generate $u(t)$ for controller input, and the detector with the deadbeat K that generate $u^d \in \{\text{true,false}\}$ to inform the encoder if a DoS attack happens, see Table. 2. In this way, the encoding scheme can be chosen as in [35] with simple UDP-like protocol. This structure improves the robustness of the system by scarifying computational resources.

4. NUMERICAL EXAMPLE

A linearized model of the unstable batch reactor in [35] is given by $\dot{x}(t) = Ax(t) + Bu(t)$ and $y = Cx(t)$, where

$$A = \begin{bmatrix} 1.38 & -0.2077 & 6.715 & -5.676 \\ -0.5814 & -4.29 & 0 & 0.675 \\ 1.067 & 4.273 & -6.654 & 5.893 \\ 0.048 & 4.273 & -1.343 & -2.104 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 \\ 5.679 & 0 \\ 1.136 & -3.146 \\ 1.136 & 0 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

This system (A,B,C) is observable and controllable with $\eta_1 = \eta_2 = 2$. Let the transmission period $\Delta = 0.2$, so $\delta = \Delta/\eta_2 = 0.1$. The deadbeat gain K is

$$K = \begin{bmatrix} 1.0106 & -1.5661 & 0.0385 & -4.0366 \\ 8.1074 & -0.0347 & 4.3337 & -3.6241 \end{bmatrix}.$$

M is derived by calculating the gain of the steady-state Kalman filter whose covariances of the process noise and measurement noise are I_4 and I_2, respectively,

$$M = \begin{bmatrix} 0.5534 & -0.0249 \\ -0.0287 & 0.0396 \\ 0.1489 & 0.0892 \\ 0.0810 & 0.0931 \end{bmatrix}.$$

According to Theorem 1, we discover that as the quantization level N goes to infinity, the duration and frequency bounds of DoS attacks, i.e., $\frac{1}{v_d}$ and $\frac{1}{v_f}$, get close to the line

$$\frac{1}{v_f} \approx -0.5544\frac{1}{v_f} + 0.2707.$$

From (21), the quantization level satisfies $N > 6.957$, also noticing N is an even number, so we set $N = 100$.

Over a simulation horizon of 40 s (200 time-step), according to (22), if

$$\frac{1}{v_f} < -0.5543\frac{1}{v_f} + 0.2537$$

then the closed-loop system with encoding scheme (18) can achieve stability. The DoS attacks (the gray shades) are generated randomly with $\Phi_d = 21$ and $\Phi_f = 20$. Setting $\Pi_d = 1, v_d = 10, \Pi_f = 0, v_f = 10$, and $\frac{1}{v_d} = 0.1 < 0.1983$, which satisfies the condition (22), and Tables 3 and 4 verify that the state converges to the origin in this situation. Table 3 depicts the maximum norm of the state x and the observer state $\hat{x}$. Table 4 illustrates that the error bound E_q exponentially decreases and shares the same trend with the actual error $|e_q|$. Table 5 demonstrates that when the output channel is subject to DoS attacks, then the control input is set to zero immediately, which verifies the effectiveness of our transmission strategy.

5. CONCLUSION

This paper put forth a controller with designed controller gain and specific relationship between the transmission period (i.e., time interval between two consecutive output signals) and the sampling period (i.e., time interval between two consecutive control inputs). Capitalizing on this controller, the acknowledgement messages are no more needed, thus UDP-like protocol is enough for maintaining the synchronization between the encoder and the decoder. Output encoding scheme and the corresponding condition for exponential convergence are obtained when only the output channel is quantized and is subject to DoS attacks. Modification method of this controller structure is provided to recover the robustness under TCP-like protocol at the price of the computation resources. Finally, a numerical example is given to illustrate that the proposed approaches are valid. Future developments will focus on generalizing the results by considering network phenomenon at the controller to the plant channels.

References

[1] G. Wu, G. Wang, J. Sun, and L. Xiong, "Optimal switching attacks and countermeasures in cyber-physical systems," IEEE Trans. Syst. Man Cybern. Syst., 2020, DOI: 10.1109/TSMC.2019.2945067.

[2] G. Wang, V. Kekatos, A. J. Conejo, and G. B. Giannakis, "Ergodic energy management leveraging resource variability in distribution

[3] grids," IEEE Trans. Power Syst., vol. 31, no. 6, pp. 4765–4775, Feb. 2016.

[4] A. J. Kerns, D. P. Shepard, J. A. Bhatti, and T. E. Humphreys, " Unmanned aircraft capture and control via gps spoofing," J. Field Robot., vol. 31, no. 4, pp. 617–636, July 2014.

[5] M. Lv, D. Wang, Z. Peng, L. Lu, and H. Wang, "Event-triggered neural network control of autonomous surface vehicles over wireless network," Sci. China Inf. Sci., vol. 63, Mar. 2020, DOI: 10.1007/s11432-019-2679-5.

[6] M. Zhu and S. Martłnez, "On the performance analysis of resilient networked control systems under replay attacks," IEEE Trans. Autom. Control, vol. 59, no. 3, pp. 804–808, Aug. 2014.

[7] G. Wu and J. Sun, "Optimal switching integrity attacks on sensors in industrial control systems," J. Syst. Sci. Complex, vol. 32, pp. 1290–1305, Jan. 2019.

[8] A. Cetinkaya, H. Ishii, and T. Hayakawa, "An overview on denial-of-service attacks in control systems: Attack models and security analyses," Entropy, vol. 21, no. 2, pp. 210–238, Feb. 2019.

[9] Cardenas, S. Amin, and S. Sastry, "Secure control: Towords survivable cyber-physical systems," in Proc. Int. Conf. Distrib. Comput. Syst. Workshops, July 2008, pp. 495–500.

[10] J. Qin, M. Li, L. Shi, and X. Yu, "Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks," IEEE Trans. Autom. Control, vol. 63, no. 6, pp. 1648–1663, Setp. 2018.

[11] Y. Li, L. Shi, P. Cheng, J. Chen, and D. E. Quevedo, "Jamming attacks on remote state estimation in cyber-physical systems: A game-theoretic approach," IEEE Trans. Autom. Control, vol. 60, no. 10, pp. 2831–2836, Oct. 2015.

[12] H. S. Foroush and S. Martinez, "On event-triggered control of linear systems under periodic denial-of-service jamming attacks," in Proc. IEEE Conf. Decis. Control, Maui, HI, USA, Dec. 10-13 2012, pp. 2551–2556.

[13] Persis De and P. Tesi, "Input-to-state stabilizing control under denial-of-service," IEEE Trans. Autom. Control, vol. 60, no. 11, pp. 2930–2944, Nov. 2015.

[14] S. Feng and P. Tesi, "Resilient control under denial-of-service:Robust design," Automatica, vol. 79, pp. 42–51, Mar. 2017.

[15] A. Y. Lu and G.-H. Yang, "Input-to-state stabilizing control for cyber-physical systems with multiple transmission channels under denial-of-service," IEEE Trans. Autom. Control, vol. 63, no. 6, pp. 1813–1820, Jun. 2018.

[16] C. Persis and P. Tesi, "Networked control of nonlinear systems under denial-of-service," Syst. Control Lett., vol. 96, pp. 124–131, Oct. 2016.

[17] Senejohnny, "Self-triggered coordination over a shared network under denial-of-service," in Proc. IEEE Conf. Decis. Control, Osaka, Japan, Dec. 2015.

[18] A. Cetinkaya, H. Ishii, and T. Hayakawa, "Event-triggered output feedback control resilient against jamming attacks and random packet losses," in Proc. IFAC Workshop Distrib. Estimation Control Netw. Syst., vol. 48, no. 22, Sept. 2015, pp. 270–275.

[19] S. Feng and P. Tesi, "Networked control systems under denial-of-service: Co-located vs. remote architectures," Syst. Control Lett., vol. 108, pp. 40–47, Sept. 2017.

[20] V. S. Dolk, P. Tesi, C. Persis De, and W. P. M. H. Heemels, " Output-based event-triggered control systems under denial-of-service attacks," in Proc. IEEE Conf. Decis. Control, Dec. 15-18 2015, pp. 4824–4829.

[21] A. Y. Lu and G.H. Yang, "Resilient observer-based control for cyber-physical systems with multiple transmission channels under denial-of-service," IEEE Trans. Cybern., pp. 1–12, May 2019, doi: 10.1109/TCYB.2019.2915942.

[22] Cetinkaya, H. Ishii, and T. Hayakawa, "State-dependent jamming interference in networked stabilization," in Proc. IEEE Conf. Decis. Control, Miami Beach, FL, USA, Dec. 17-19 2018, pp. 7249–7254.

[23] R. W. Brockett and D. Liberzon, "Quantized

feedback stabilization of linear systems," IEEE Trans. Autom. Control, vol. 45, no. 7, pp. 1279–1289, July 2000.

[24] Y. Sharon and D. Liberzon, "Input-to-state stabilization with quantized output feedback," in Proc. Int. Conf. Hybrid Syst.: Comput. Control, St. Louis, MO, USA, Apr. 22-24, 2008, pp. 500–513.

[25] ——, "Input to state stabilizing controller for systems with coarse quantization," IEEE Trans. Autom. Control, vol. 57, no. 4, pp. 830–844, Apr. 2012.

[26] M. Wakaiki, T. Zanma, and K. Liu, "Observer-based stabilization of systems with quantized inputs and outputs," IEEE Trans. Autom. Control, vol. 64, no. 7, pp. 2929–2936, July 2019.

[27] Liberzon, "Finite data-rate feedback stabilization of switched and hybrid linear systems," Automatica, vol. 50, no. 2, pp. 409–420, Jan. 2014.

[28] ——, "On stabilization of linear systems with limited information," IEEE Trans. Autom. Control, vol. 48, no. 2, pp. 304–307, Feb. 2003.

[29] M. Wakaiki and Y. Yamamoto, "Stability analysis of sampled-data switched systems with quantization," Automatica, vol. 69, pp. 157–168, Mar. 2016.

[30] D. Liberzon, "Hybrid feedback stabilization of systems with quantized signals," Automatica, vol. 39, no. 9, pp. 1543–1554, Mar. 2003.

[31] M. Wakaiki and Y. Yamamoto, "Stabilization of switched linear systems with quantized output and switching delays," IEEE Trans. Autom. Control, vol. 62, no. 6, pp. 2958–2964, June 2017.

[32] M. Wakaiki, T. Zanma, and K. Z. Liu, "Quantized output feedback stabilization by luenberger observers," in Proc. IFAC WC, vol. 50, no. 1, Toulouse, France, July 9-14 2017, pp. 2577–2582.

[33] M. Wakaiki, A. Cetinkaya, and H. Ishii, "Quantized output feedback stabilization under dos attacks," in Proc. Amer. Control Conf., Wisconsin Center, Milwaukee, USA, June 27-29 2018, pp. 6487–6492.

[34] Yang and D. Liberzon, "Feedback stabilization of switched linear systems with unknown disturbances under data-rate constraints," IEEE Trans. Autom. Control, vol. 63, no. 7, pp. 2107–2122, July 2018.

[35] S. Feng, A. Cetinkaya, H. Ishii, P. Tesi, and C. De Persis, " Networked control under dos attacks: Trade-off between resilience and data rate," in Proc. Amer. Control Conf., Philadelphia, PA, USA, July 10-12 2019.

[36] M. Wakaiki, A. Cetinkaya, and H. Ishii, "Stabilization of networked control systems under dos attacks and output quantization," IEEE Trans. Autom. Control, pp. 2334–3303, 2019, doi:10.1109/TAC.2019.2949096.

[37] Lin, H.-Y. Su, P. Shi, Z. Shu, and Z.-G. Wu, Studies in Systems, Decision and Control, Estimation and Control for Networked Systems with Packet Losses without Acknowledgement. Springer International Publishing, 2017.

[38] W. Xu, W. Trappe, Y. Zhang, and T. Wood, "The feasibility of launching and detecting jamming attacks in wireless networks," in Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput, May 2005, pp. 46–57.

[39] P. Hespanha and A. S. Morse, "Stability of switched systems with average dewell-time," in Proc. IEEE Conf. Decis. Control, Phoenix, Arizona, USA, Dec. 1999, pp. 2655–2660.

[40] P. Van Dooren, "Deadbeat control: A special inverse eigenvalue problem," BIT, vol. 24, pp. 681–699, May 1984.

[41] M. M. Fahmy and J. O'Reilly, "Dead-beat control of linear discrete-time systems," Int. J. Control, vol. 37, no. 4, pp. 685–705, 1983.

Attack Detection against Stealthy FDI Attacks in Cyber-Physical Systems: A Stochastic Coding Detection Scheme

Haibin Guo*, Jian Sun*

*School of Automation, Beijing Institute of Technology, Beijing, 100081, P.R. China, The Key Laboratory of Intelligent Control and Decision of Complex System, Beijing Institute of Technology, Beijing, 100081, P.R. China

Abstract

This paper studies the attack detection problem against stealthy false data injection (FDI) attacks in cyber-physical systems (CPSs). Compared with the existing works of FDI detection scheme, we consider an attack scenario in which malicious attackers can obtain the exact system knowledge. A virtual system is constructed to reflect the attack process. To detect a class of stealthy FDI attacks, a stochastic coding scheme, which codes sensors measurements with a Gaussian signal, is proposed to make the CPSs generate residual differences between the normal and compromised system. By solving an optimal problem, the design of the coding covariance matrix is given. Finally, some numerical examples are provided to illustrate the effectiveness of the proposed attack detection scheme.

Keywords: Cyber-Physical Systems, Stealthy FDI attacks, Attack detection, Stochastic coding detection scheme.

1. INTRODUCTION

The cyber-physical system (CPS), which integrates computation, communication, and control [1], [2], has gained increasing attention in the recent years. Since sensors measurements and control data packets are transmitted over the wire/wireless network, malicious adversaries can damage the CPSs by intruding the communication networks and corrupting the transmitted data packets. Typical attacks include denial of service (DoS) attacks and false data injection (FDI) attacks.

DoS attacks mainly prevent the legitimate access to system components by blocking the communication network. In [3], [4], an optimal DoS attack schedule was constructed to maximize the expected average estimation error at the remote estimator. In [5], based on the signal-to-interference-plus-noise ratio-based network, the optimal DoS attack scheme with the limited power resource was proposed to damage the system performance.

FDI attacks attempt to inject false signal into the communication network to tamper the normal transmitted information, so as to degrade the system performance. Considering the limited attack power resource, in [6], [7], an optimal switching data injection attack scheme, which only corrupts partial actuators, was proposed to damage the system performance. In [8], to degrade the remote state estimation performance, a false data injection attack scheme with resource constraints was studied to only attack partial sensors. Considering the stealthy of the FDI attack, in [9], [10], an innovation-based linear attack strategy was proposed to maximize the remote estimation error covariance, meanwhile keep the attack stealthy. In [11], a stealthy two-channel FDI attack scheme was proposed for both the feedback and forward channels to disrupt the stability of the closed-loop system while avoiding the detection. In [12], based on the inaccurate model, a stealthy two-channel attack scheme, which can compromise the system without being detected, is proposed. Besides, replay attack is a special class of FDI attack scheme, which attempts to record the past data and replay the recorded data in the attack time interval [13]. In [14], in order to keep the forward channel attack stealthy, a replay attack scheme was adopted in the feedback channel.

Since the stealthy FDI attack scheme is designed to avoid the anomaly detector while maximally degrading the system performance, how to detect stealthy FDI attacks and ensure the system performance in the attack time interval becomes intractable. In [15], [16], based on the trusted sensors, a detection scheme against linear deception attacks on multi-sensor remote state estimation was studied. In [17], a Gaussian-mixture-model-based detection mechanism was proposed to detect integrity attacks. In [18], a coding matrix was adopted to code the original sensor outputs to increase the residues under FDI attacks. In [19], An active data modification detection scheme was proposed to detect the stealthy two-channel attack. In [13], an additional Gaussian signal was coded into the control input to detect the replay attack. In [20], a stochastic coding scheme, which codes sensors measurements with a Gaussian signal, was proposed to detect the replay attack.

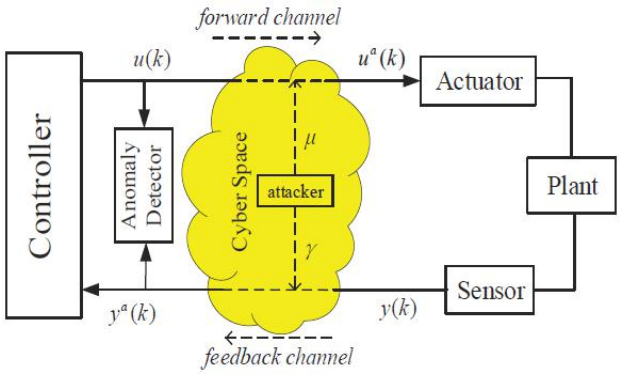


Fig.1 The closed-loop feedback control system.

The main problem of the existing works is that the FDI detection scheme [18], [19] becomes invalid when the malicious attackers obtain the exact system knowledge. This paper mainly studies a case that the malicious attackers are assumed to obtain the exact system knowledge. The main contributions of this study are that inspired by the replay attack detection scheme [20], a stochastic coding detection scheme, which codes sensors measurements with a Gaussian signal in the feedback channel, is proposed to detect a class of stealthy FDI attacks. Compared with detecting the replay attack [20], the filter gain obtained by FDI attacker holds difference with the normal filter gain, which complexes the design of the coding covariance matrix.

The reminder of this paper is organized as follows. In Section II, the problem is formulated. Section III details the FDI attack detection scheme. In Section IV, the performance of attack detection scheme is illustrated by some numerical examples. This paper is concluded in Section V.

Notations: $X \geq 0$ and $X > 0$ denote positive semi-definite matrix and positive definite matrix, respectively. $\mathcal{N}(\mu, \Sigma)$ denotes Gaussian distribution with mean μ and covariance matrix Σ. $I_m (0_m)$ denotes the $m \times m$ dimensional identity (zeros) matrix. The superscript ^{T} stands for the transposition. trX denotes the trace of matrix X.

2. PROBLEM FORMULATION

As shown in Fig. 1, a closed-loop feedback control system is illustrated. Forward channel and feedback channel make the information connection between controller and plant, which are easily corrupted by malicious attackers.

2.1 System model

Consider a linear time-invariant system

$$x(k+1) = Ax(k) + Bu(k) + \omega(k) \quad (1)$$

$$y(k) = Cx(k) + \nu(k) \quad (2)$$

where $x(k) \in \mathbb{R}^n$ denotes the system state, $y(k) \in \mathbb{R}^m$ denotes the sensor measurement, $u(k) \in \mathbb{R}^l$ denotes the system input, $\omega(k) \in \mathbb{R}^n$ and $\nu(k) \in \mathbb{R}^m$ denote the process noise and the measurement noise, respectively. $\omega(k) \sim \mathcal{N}(0, Q), Q > 0$ and $\nu(k) \sim \mathcal{N}(0, R), R \geq 0$ are zero-mean independent and identically distributed (i.i.d) Gaussian noises. The initial state $x(0) \sim \mathcal{N}(0, \Pi_0), \Pi_0 \geq 0$ is independent of $\omega(k)$ and $\nu(k)$ for all $k \geq 0$.

Assumption 1: The system is both observable and controllable under the disturbances during operation.

Assumption 2: Sensors and controller are synchronized.

Assumption 3: The system operates in the steady-state during the attack.

2.2 Control Law Based on Kalman filter

A standard Kalman filter is adopted to estimate the system state.

$$\hat{x}(k+1 \mid k) = A\hat{x}(k) + Bu(k) \quad (3)$$

$$P(k+1 \mid k) = AP(k-1)A^T + Q \quad (4)$$

$$K(k) = P(k \mid k-1)C^T (CP(k \mid k-1)C^T + R)^{-1} \quad (5)$$

$$P(k) = (I - K(k)C)P(k \mid k-1) \quad (6)$$

$$\hat{x}(k) = \hat{x}(k \mid k-1) + K(k)(y(k) - C\hat{x}(k \mid k-1)) \quad (7)$$

where $\hat{x}(k \mid k-1)$ and $\hat{x}(k)$ are the *a priori* and *a posteriori* minimum mean squared error (MMSE) estimates of the state $x(k)$, $P(k \mid k-1)$ and $P(k)$ are the corresponding error covariances. The initial conditions $\hat{x}(0 \mid -1) = 0$ and $P(0 \mid -1) = \Pi_0$.

Since the Kalman filter converges exponentially fast from any initial conditions [21], the steady-state error covariance is defined as

$$P \triangleq \lim P(k \mid k-1) \quad (8)$$

where P is the unique positive semi-definite solution of $X = AXA^T + Q - AXC^T (CXC^T + R)^{-1} CXA^T$.

The Kalman filter steady-state gain matrix

$$K \triangleq PC^T (CPC^T + R)^{-1} \quad (9)$$

Hence, (3)—(7) are reduced to

$$\hat{x}(k+1 \mid k) = A\hat{x}(k) + Bu(k) \quad (10)$$

$$\hat{x}(k) = \hat{x}(k \mid k-1) + K(y(k) - C\hat{x}(k \mid k-1)) \quad (11)$$

A state feedback control input is designed as

$$u(k) = L\hat{x}(k) \quad (12)$$

2.3 Anomaly detector

To reveal the anomalies of system, an anomaly detector is incorporated at the controller's side as shown in Fig. 1.

Define the residual

$$z(k) \triangleq y(k) - C\hat{x}(k \mid k-1) \quad (13)$$

The corresponding steady-state residual covariance is

$$\Sigma_z \triangleq \lim_{k\to+\infty} E\{z(k)z(k)^T\} = CPC^T + R \tag{14}$$

The χ^2 detector is a residue-based detector widely used to detect system anomalies [22], [23]. The detection criterion is defined as

$$g(k) = \sum_{i=k-J+1}^{k} z(i)^T \Sigma_z^{-1} z(i) \underset{H_1}{\overset{H_0}{\lessgtr}} \tau \tag{15}$$

where τ is the threshold, J is the window size of detection, $g(k)$ is χ^2 distributed with mJ degrees of freedom. The two hypotheses H_0 denotes no attack exists, H_1 denotes attack exists. If $g(k)$ is greater than τ, the detector alarm will be triggered.

2.4 Stealthy FDI attack

In this paper, we consider a class of the stealthy FDI attacks, two-channel attack [11]. Malicious attackers are assumed to obtain the exact system knowledge, i.e, A, B, C, Q, R, without the detection scheme.

The forward channel attack

$$u^a(k) = u(k) + \mu(k) \tag{16}$$
$$\mu(k+1) = F\mu(k) \tag{17}$$

where $F \in \mathbb{R}^{l\times l}$ is the attack matrix.

The feedback channel attack

$$y^a(k) = y(k) + \gamma(k) \tag{18}$$
$$\gamma(k) = -y(k) + CA\hat{x}^a(k-1) + CBu(k-1) + \xi(k) \tag{19}$$

where $\hat{x}^a(k-1)$ denotes the *a posteriori* MMSE estimate under the attack, $\xi(k) \sim \mathcal{N}(0, \Sigma_z)$.

Under this attack, the Kalman filter (10)-(11) is rewritten as

$$\hat{x}^a(k+1\,|\,k) = A\hat{x}^a(k) + Bu(k) \tag{20}$$
$$\hat{x}^a(k) = \hat{x}^a(k\,|\,k-1) + K(y^a(k) - C\hat{x}^a(k\,|\,k-1)) \tag{21}$$

Remark 1: Malicious attackers are assumed to obtain the exact system knowledge and run the corresponding Kalman filter with same as (20)—(21). This attack scheme can maximally degrade the performance of the system, while avoiding the detection of the anomaly detector. The detailed analysis is given in [11].

3. THE DESIGN OF ATTACK DETECTION SCHEME

3.1 Stochastic coding scheme

The sensors measurements are coded with a Gaussian signal.

$$y^c(k) = y(k) + y^*(k) \tag{22}$$

where $y^*(k)$ is a Gaussian white noise, $y^*(k) \sim \mathcal{N}(0, \Sigma^*)$, $\Sigma^* = diag(\sigma_1, \ldots, \sigma_m)$.

Then, the transmission data are decoded in the decoder.

$$\begin{aligned} y^d(k) &= y^c(k) - y^*(k) \\ &= y(k) \end{aligned} \tag{23}$$

Remark 2: When the Gaussian random signal generator in the decoder holds the same random seed as the coder, the coding Gaussian white noise $y^*(k)$ in the decoder is the same as the coder.

Malicious attackers can obtain the exact system knowledge by using the technology of the system identification. The virtual system is defined to describe this system.

$$x(k+1) = Ax(k) + Bu(k) + \omega(k) \tag{24}$$
$$y^c(k) = Cx(k) + \nu^c(k) \tag{25}$$

where $\nu^c(k) \triangleq \nu(k) + y^*(k)$.

Remark 3: Since the coding signal $y^*(k)$ and measurement noise $\nu(k)$ are i.i.d Gaussian noises, $\nu^c(k)$ satisfies $\nu^c(k) \sim \mathcal{N}(0, R + \Sigma^*)$.

Remark 4: Compared with original system (1)—(2), the virtual system (24)—(25) can be explained as sensors measure system output under different measurement noise.

Attackers run the corresponding Kalman filter.

$$\hat{x}^c(k+1\,|\,k) = A\hat{x}^c(k) + Bu(k) \tag{26}$$
$$\hat{x}^c(k) = \hat{x}^c(k\,|\,k-1) + K^c(y^c(k) - C\hat{x}^c(k\,|\,k-1)) \tag{27}$$

where the recursion starts from $\hat{x}^c(0\,|\,-1) = 0$, $K^c \triangleq P^cC^T(CP^cC^T + R + \Sigma^*)^{-1}$, P^c is the unique positive semi-definite solution of

$$X = AXA^T + Q - AXC^T(CXC^T + R + \Sigma^*)^{-1}CXA^T \tag{28}$$

The corresponding residual is

$$z^c(k) \triangleq y^c(k) - C\hat{x}^c(k\,|\,k-1) \tag{29}$$

The corresponding steady-state residual covariance is

$$\Sigma_{z^c} \triangleq \lim_{k\to+\infty} E\{z^c(k)z^c(k)^T\} = CP^cC^T + R + \Sigma^* \tag{30}$$

The feedback channel attack (19) is rewritten as

$$\gamma(k) = -y^c(k) + CA\hat{x}^{ca}(k-1) + CBu(k-1) + \xi^c(k) \tag{31}$$

where $\xi^c(k) \sim \mathcal{N}(\mathbf{0}, \Sigma_{z^c})$ and $\hat{x}^{ca}(k)$ satisfies

$$\hat{x}^{ca}(k+1\,|\,k) = A\hat{x}^{ca}(k) + Bu(k) \tag{32}$$
$$\hat{x}^{ca}(k) = \hat{x}^{ca}(k\,|\,k-1) + K^c\xi^c(k) \tag{33}$$

Then, with (32) and (33), we can obtain

$$\hat{x}^{ca}(k+1\,|\,k) = A\hat{x}^{ca}(k\,|\,k-1) + Bu(k) + AK^c\xi^c(k) \tag{34}$$

where the initial condition $\hat{x}^{ca}(0\,|\,-1) = 0$.

3.2 The design of stochastic coding signal

Under the feedback channel attack (31), the sensors measurements are decoded by the receiver.

$$\begin{aligned} y^{da}(k) &= y^{a}(k) - y^{*}(k) \\ &= y^{c}(k) + \gamma(k) - y^{*}(k) \end{aligned} \tag{35}$$

With (31) and (35), the residual in the compromised system is defined as

$$\begin{aligned} z^{a}(k) &\triangleq y^{da}(k) - C\hat{x}^{a}(k \mid k-1) \\ &= C(\hat{x}^{ca}(k \mid k-1) - \hat{x}^{a}(k \mid k-1)) + \xi^{c}(k) - y^{*}(k) \end{aligned} \tag{36}$$

In the compromised system, with (20) and (21), we can obtain

$$\hat{x}^{a}(k+1 \mid k) = A\hat{x}^{a}(k \mid k-1) + Bu(k) + AKz^{a}(k) \tag{37}$$

where the initial condition $\hat{x}^{a}(0 \mid -1) = 0$.

Let $\mathcal{A} \triangleq A(I-KC)$, $\mathcal{B} \triangleq A(K^{c}-K)$ and $\mathcal{C} \triangleq AK$, with (34) and (37), we can obtain

$$\begin{aligned} &\hat{x}^{ca}(k+1 \mid k) - \hat{x}^{a}(k+1 \mid k) \\ &= A(\hat{x}^{ca}(k \mid k-1) - \hat{x}^{a}(k \mid k-1)) + AK^{c}\xi^{c}(k) \\ &\quad - AKz^{a}(k) \\ &= A(I-KC)(\hat{x}^{ca}(k \mid k-1) - \hat{x}^{a}(k \mid k-1)) \\ &\quad + A(K^{c}-K)\xi^{c}(k) + AKy^{*}(k) \\ &= \mathcal{A}^{k+1}(\hat{x}^{ca}(0 \mid -1) - \hat{x}^{a}(0 \mid -1)) \\ &\quad + \sum_{i=0}^{k} \mathcal{A}^{k-i}(\mathcal{B}\xi^{c}(i) - \mathcal{C}y^{*}(i)) \end{aligned} \tag{38}$$

The residual (36) in the compromised system is written as

$$z^{a}(k) = C\sum_{i=0}^{k-1} \mathcal{A}^{k-i}(\mathcal{B}\xi(i) - \mathcal{C}y^{*}(i)) + \xi^{c}(k) - y^{*}(k) \tag{39}$$

The steady-state of residual covariance is described as

$$\begin{aligned} \Sigma_{z^{a}} &\triangleq \lim_{k \to +\infty} E\{z^{a}(k)z^{a}(k)^{T}\} \\ &= \Sigma_{z^{c}} + \Sigma^{*} + \sum_{i=0}^{\infty} C\mathcal{A}^{i}(\mathcal{B}\Sigma_{z^{c}}\mathcal{B}^{T} + \mathcal{C}\Sigma^{*}\mathcal{C}^{T})\mathcal{A}^{iT}C^{T} \end{aligned} \tag{40}$$

Define $\mathcal{F} \triangleq \sum_{i=0}^{\infty} \mathcal{A}^{i}\mathcal{B}\Sigma_{z^{c}}\mathcal{B}^{T}\mathcal{A}^{iT}$ and $\mathcal{G} \triangleq \sum_{i=0}^{\infty} \mathcal{A}^{i}\mathcal{C}\Sigma^{*}\mathcal{C}^{T}\mathcal{A}^{iT}$, which can be obtained by solving the following equations.

$$\mathcal{F} - \mathcal{A}\mathcal{F}\mathcal{A}^{T} = \mathcal{B}\Sigma_{z^{c}}\mathcal{B}^{T} \tag{41}$$

$$\mathcal{G} - \mathcal{A}\mathcal{G}\mathcal{A}^{T} = \mathcal{C}\Sigma^{*}\mathcal{C}^{T} \tag{42}$$

Furthermore, (40) can be rewritten as

$$\Sigma_{z^{a}} = \Sigma_{z^{c}} + \Sigma^{*} + C(\mathcal{F} + \mathcal{G})C^{T} \tag{43}$$

Remark 5: From (28) and (30), we can see that $\Sigma_{z^{c}}$ holds a complex relation with Σ^{*}, which is different from the case of replay attack [20]. This brings difficulties to design the coding covariance matrix Σ^{*}.

Since Gaussian white noise is used to describe the system disturbance and stochastic code scheme is adopted, $g(k)$ in (15) is a stochastic variable in both normal and compromised situation, which implies anomaly detector may has false alarm rate (FAR) and attack detection rate (ADR). Therefore, we propose the design method of coding covariance matrix Σ^{*} subject to FAR and ADR in the following theorem.

Define $Z_{J} \triangleq [z(k-J+1)^{T}, \ldots, z(k)^{T}]$, $\Sigma_{J}^{*} \triangleq diag(\Sigma^{*}, \ldots, \Sigma^{*})$, $\overline{\Sigma}_{H_0} \triangleq diag(\Sigma_{z}, \ldots, \Sigma_{z})$, $\overline{\Sigma}_{H_0'} \triangleq diag(\Sigma_{z^{c}}, \ldots, \Sigma_{z^{c}})$ and $\overline{\Sigma}_{H_1} \triangleq diag(\Sigma_{z^{a}}, \ldots, \Sigma_{z^{a}})$, where H_0 and H_1 represent the system normal and compromised, respectively.

Theorem 1: Consider the system (1)—(2) equipped with the stochastic encode-decode scheme (22)—(23), and the attacker implements attack scheme (16)—(18) and (31) into two channels, then the detector (15) can detect stealthy FDI attacks.

Detector (15) is rewritten as

$$g(k) = Z_{J}^{T}\overline{\Sigma}_{H_0}^{-1}Z_{J} \underset{H_1}{\overset{H_0}{\lessgtr}} \tau \tag{44}$$

The threshold τ and the coding covariance matrix Σ^{*} can be obtained by solving the following optimization problem.

$$\begin{aligned} \min_{\Sigma^{*},\tau} \quad & tr\Sigma^{*} \\ s.t. \quad & \Sigma^{*} > 0 \\ & \tau \geq \chi_{\alpha} \\ & \lambda_{max}(\overline{\Sigma}_{H_0})\tau / \lambda_{min}(\overline{\Sigma}_{H_1}) \leq \chi_{\beta} \\ & \overline{\Sigma}_{H_1} = \overline{\Sigma}_{H_0'} + \Sigma_{J}^{*} + \overline{C}(\overline{\mathcal{F}} + \overline{\mathcal{G}})\overline{C}^{T} \\ & \overline{\mathcal{F}} - \overline{\mathcal{A}}\overline{\mathcal{F}}\overline{\mathcal{A}}^{T} = \overline{\mathcal{B}}\overline{\Sigma}_{z^{c}}\overline{\mathcal{B}}^{T} \\ & \overline{\mathcal{G}} - \overline{\mathcal{A}}\overline{\mathcal{G}}\overline{\mathcal{A}}^{T} = \overline{\mathcal{C}}\overline{\Sigma}^{*}\overline{\mathcal{C}}^{T} \end{aligned} \tag{45}$$

where α and β denote FAR and ADR, respectively.

Proof: The hypothesis testing (15) is rewritten as

$$g(k) = Z_{J}^{T}\overline{\Sigma}_{H_0}^{-1}Z_{J} \underset{H_1}{\overset{H_0}{\lessgtr}} \tau$$

Subject to FAR and ADR, the threshold τ and the coding covariance matrix Σ^* are to be designed as follows.

Firstly, FAR is guaranteed lower than a given value α.

$$\text{FAR} = \text{prob}(Z_J^T \bar{\Sigma}_{H_0}^{-1} Z_J > \tau \mid H_0) \le \alpha$$

Since $Z_J^T \bar{\Sigma}_{H_0}^{-1} Z_J$ is a standard χ^2 distribution under hypothesis H_0, τ satisfies

$$\tau \ge \chi_\alpha \tag{47}$$

Secondly, ADR is guaranteed larger than a given value β.

$$\text{ADR} = \text{prob}(Z_J^T \bar{\Sigma}_{H_0}^{-1} Z_J > \tau \mid H_1) \ge \beta \tag{48}$$

where

$$\begin{aligned} Z_J^T \bar{\Sigma}_{H_0}^{-1} Z_J &\ge \frac{1}{\lambda_{max}(\bar{\Sigma}_{H_0})} Z_J^T Z_J \\ &\ge \frac{\lambda_{min}(\bar{\Sigma}_{H_1})}{\lambda_{max}(\bar{\Sigma}_{H_0})} Z_J^T \bar{\Sigma}_{H_1}^{-1} Z_J \end{aligned} \tag{49}$$

With (48) and (49), we can obtain

$$\begin{aligned} \text{ADR} &= \text{prob}(Z_J^T \bar{\Sigma}_{H_0}^{-1} Z_J > \tau \mid H_1) \\ &\ge \text{prob}\left(\frac{\lambda_{min}(\bar{\Sigma}_{H_1})}{\lambda_{max}(\bar{\Sigma}_{H_0})} Z_J^T \bar{\Sigma}_{H_1}^{-1} Z_J > \tau \mid H_1\right) \\ &\ge \beta \end{aligned} \tag{50}$$

Therefore, (48) is satisfied when we guarantee

$$\text{prob}\left(Z_J^T \bar{\Sigma}_{H_1}^{-1} Z_J > \frac{\lambda_{max}(\bar{\Sigma}_{H_0})}{\lambda_{min}(\bar{\Sigma}_{H_1})} \tau \mid H_1\right) \ge \beta \tag{51}$$

Since $Z_J^T \bar{\Sigma}_{H_1}^{-1} Z_J$ is standard χ^2 distribution under hypothesis H_0, τ and $\bar{\Sigma}_{H_1}$ satisfy

$$\frac{\lambda_{max}(\bar{\Sigma}_{H_0})}{\lambda_{min}(\bar{\Sigma}_{H_1})} \tau \le \chi_\beta \tag{52}$$

The proof is complete.

Remark 6: Since $\lambda_{max}(\bar{\Sigma}_{H_0})\tau / \lambda_{min}(\bar{\Sigma}_{H_1})$ is a convex function, when a fixed value τ is selected to satisfy the limited second condition, the optimization problem (45) is transformed to convex optimization.

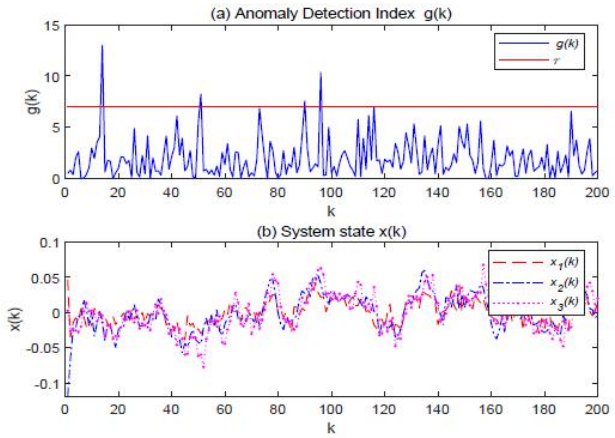


Fig.2 Response curve of normal system with the stochastic coding detection scheme: (a) Anomaly detection index $g(k)$, (b) system state $x(k)$.

Remark 7: Since Σ_{z^c} is relevant with Σ^*, how to solve the optimization problem (45) becomes intractable. Inspired by [24], Expectation-Maximization (EM) algorithm is adopted to handle this intractable problem.

4. SIMULATION RESULTS

To demonstrate the analytical results, we present some numerical simulations in this section.

Consider system

$$A = \begin{bmatrix} 0.2071 & 0.3705 & 0.0439 \\ 0.6072 & 0.5751 & 0.0272 \\ 0.6299 & 0.4514 & 0.3127 \end{bmatrix}, B = \begin{bmatrix} 0.1730 & 0.2523 \\ 0.9797 & 0.8757 \\ 0.2714 & 0.7373 \end{bmatrix},$$

$$C = \begin{bmatrix} 0.1365 & 0.8939 & 0.2987 \\ 0.0118 & 0.1991 & 0.6614 \end{bmatrix}, Q = 0.0001 I_3, R = 0.01 I_2.$$

The FAR and ADR are set as 1.0% and 95%, respectively. The threshold is set as $\tau = 7$.

The steady state Kalman filter gain

$$K = \begin{bmatrix} 0.0241 & 0.0171 \\ 0.0495 & 0.0323 \\ 0.0514 & 0.0434 \end{bmatrix}.$$

The controller gain is designed to be

$$L = \begin{bmatrix} -0.2507 & -0.2491 & -0.0132 \\ -0.4323 & -0.3704 & -0.1452 \end{bmatrix}.$$

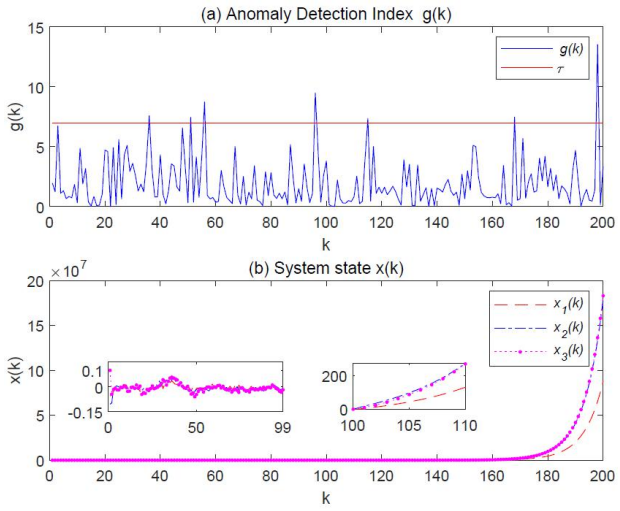


Fig.3 Response curve of the compromised system without the stochastic coding detection scheme: (a) Anomaly detection index $g(k)$, (b) system state $x(k)$.

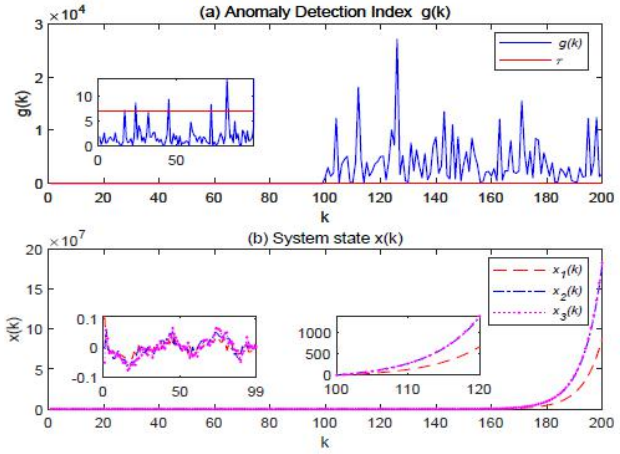


Fig.4 Response curve of the compromised system with the stochastic coding detection scheme: (a) Anomaly detection index $g(k)$, (b) system state $x(k)$.

The attack matrix is set to be

$$F = \begin{bmatrix} 0.9501 & 0.6068 \\ 0.2311 & 0.4860 \end{bmatrix}.$$

The attack time interval is set as $[100, 200]$, $\beta(100) = [10,10]^T$.

From Fig. 2, it shows that the evaluated value $g(k)$ is almost lower than the threshold τ in the normal situation with FAR 1.0% .

From Fig.3, it is clear that the evaluated value $g(k)$ in the compromised system is almost lower than the threshold τ with FAR 1.0% . This implies attack cannot be detected under the original detect scheme.

The stochastic coding scheme (22) and (23) is adopted to detect the attack. From solving the optimization problem (45), the coding covariance matrix is given as

$$\Sigma^* = \begin{bmatrix} 9.7877 & 0 \\ 0 & 9.8709 \end{bmatrix}.$$

From Fig. 4, it clearly shows that $g(k)$ is significantly larger than the threshold τ in the attack time interval $[100, 200]$, and is almost lower than the threshold τ in the normal time interval $[0, 99]$ with FAR 1.0% . This implies this detection scheme can effectively expose the stealthy FDI attacks. This proves our above analysis.

5. CONCLUSIONS

This paper has investigated the attack detection problem in cyber-physical systems against a class of stealthy FDI attacks. To detect the stealthy FDI attacks, a stochastic coding scheme, which codes sensors measurements with a Gaussian signal in the feedback channel, has been proposed in this paper. This scheme supposes attackers can obtain the exact system knowledge, except the encode-decode detection scheme, which ensures the practicality of the detection scheme. Then, the coding covariance matrix has been given by solving an optimal problem. Finally, some simulation results have illustrated the effectiveness of the proposed method.

References

[1] Y. Z. Lun, A. Dinnocenzo, F. Smarra, I. Malavolta, and M. D. D. Benedetto, “State of the art of cyber-physical systems security: An automatic control perspective,” Journal of Systems and Software, vol. 149, pp. 174–216, 2019.

[2] M. S. Mahmoud, M. M. Hamdan, and U. Baroudi, “Modeling and control of cyber-physical systems subject to cyber-attacks: A survey of recent advances and challenges,” Neurocomputing, vol. 338, pp. 101–115, 2019.

[3] H. Zhang, P. Cheng, L. Shi, and J. Chen, “Optimal denial-of-service attack scheduling with energy constraint,” IEEE Transactions on Automatic Control, vol. 60, no. 11, pp. 3023–3028, 2015.

[4] J. Qin, M. Li, L. Shi, and X. Yu, “Optimal denial-of-service attack scheduling with energy constraint over packet-dropping networks,” IEEE Transactions on Automatic Control, vol. 63, no. 6, pp. 1648– 1663, 2018.

[5] J. Qin, M. Li, J. Wang, L. Shi, Y. Kang, and W. X. Zheng, “Optimal denial-of-service attack energy management against state estimation over an sinr-based network,” Automatica, vol. 119, p. 109090, 2020.

[6] G. Wu, J. Sun, and J. Chen, “Optimal data injection

attacks in cyber-physical systems," IEEE Transactions on Systems, Man, and Cybernetics, vol. 48, no. 12, pp. 3302–3312, 2018.

[7] G. Wu, G. Wang, J. Sun, and J. Chen, "Optimal partial feedback attacks in cyber-physical power systems," IEEE Transactions on Automatic Control, pp. 1–1, 2020.

[8] F. Li and Y. Tang, "False data injection attack for cyber-physical systems with resource constraint," IEEE Transactions on Systems, Man, and Cybernetics, vol. 50, no. 2, pp. 729–738, 2020.

[9] Z. Guo, D. Shi, K. H. Johansson, and L. Shi, "Worst-case stealthy innovation-based linear attack on remote state estimation," Automatica, vol. 89, no. 89, pp. 117–124, 2018.

[10] Y. Li and G. Yang, "Optimal stealthy false data injection attacks in cyber-physical systems," Information Sciences, vol. 481, pp. 474–490, 2019.

[11] Z. Pang, G. Liu, D. Zhou, F. Hou, and D. Sun, "Two-channel false data injection attacks against output tracking control of networked systems," IEEE Transactions on Industrial Electronics, vol. 63, no. 5, pp. 3242–3251, 2016.

[12] F. Hou and J. Sun, "False data injection attacks in cyber-physical systems based on inaccurate model," pp. 5791–5796, 2017.

[13] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," IEEE Transactions on Control Systems and Technology, vol. 22, no. 4, pp. 1396–1407, 2014.

[14] Z. Li and G.-H. Yang, "A data-driven covert attack strategy in the closed-loop cyber-physical systems," Journal of the Franklin Institute, vol. 355, no. 14, pp. 6454–6468, 2018.

[15] Y. Li, L. Shi, and T. Chen, "Detection against linear deception attacks on multi-sensor remote state estimation," IEEE Transactions on Control of Network Systems, vol. 5, no. 3, pp. 846–856, 2017.

[16] A. Chattopadhyay and U. Mitra, "Attack detection and secure estimation under false data injection attack in cyber-physical systems," pp. 1–6, 2018.

[17] Z. Guo, D. Shi, D. E. Quevedo, and L. Shi, "Secure state estimation against integrity attacks: A gaussian mixture model approach," IEEE Transactions on Signal Processing, vol. 67, no. 1, pp. 194–207, 2019.

[18] F. Miao, Q. Zhu, M. Pajic, and G. J. Pappas, "Coding schemes for securing cyber-physical systems against stealthy data injection attacks," IEEE Transactions on Control of Network Systems, vol. 4, no. 1, pp. 106–117, 2017.

[19] Z.-H. Pang, L.-Z. Fan, J. Sun, K. Liu, and G.-P. Liu, "Detection of stealthy false data injection attacks against networked control systems via active data modification," Information Sciences, 2020.

[20] D. Ye, T. Zhang, and G. Guo, "Stochastic coding detection scheme in cyber-physical systems against replay attack," Information Sciences, vol. 481, pp. 432–444, 2019.

[21] R. E. Kalman, "A new approach to linear filtering and prediction problems," Journal of Basic Engineering, vol. 82, no. 1, pp. 35–45, 1960.

[22] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and diagnosis in dynamic systems," Automatica, vol. 7, no. 5, pp. 637–640, 1971.

[23] A. S. Willsky, "A survey of design methods for failure detection in dynamic systems," Automatica, vol. 12, no. 6, pp. 601–611, 1976.

[24] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," Journal of the royal statistical society series b-methodological, vol. 39, no. 1, pp. 1–22, 1977.

Analyzing Two Ways of Interdisciplinary Research; Individual Interdisciplinary Research and Collaborative Interdisciplinary Research

Masanori FUJITA*, Takato OKUDO**, Takao TERANO***, Hiromi NAGANE****
* The National Graduate Institute for Policy Studies, Tokyo, Japan
** The Graduate University for Advanced Studies, Tokyo, Japan
*** Chiba University of Commerce, Chiba, Japan
**** Chiba University, Chiba, Japan

Abstract

In this paper, we propose a method for measuring two ways in interdisciplinary research; a way for interdisciplinary research by individual researchers and another way for interdisciplinary research by collaboration of multiple researchers. In addition, by this method, a database of "KAKENHI" which is grant-in-aid for scientific research provided by the Japan Society for the Promotion of Science is used to measure interdisciplinary research from the perspective of the two research ways, and the features of interdisciplinary research in KAKENHI are analyzed for research field. As a result of the analysis, we found that (1) the number of collaborative interdisciplinary research is higher than that of individual interdisciplinary research, (2) numbers of interdisciplinary research for each field and for each combination of fields differ from field to field, (3) the relationship of numbers of interdisciplinary research of a certain field as main field and that of the same field as sub field is asymmetric. As the proposed measurement method is able to quantitatively measure interdisciplinarity among fields and their research organizations, it will be useful for decision makers of science and technology policy and strategy.

Keywords: Interdisciplinary research, collaborative research, research grant, KAKENHI, network analysis

1. INTRODUCTION

1.1 Interdisciplinary Research

Deepening and accumulating knowledge in one research field is an important issue for the progress of science and the realization of innovation. On the other hand, it is also important to search for knowledge over various fields and combine knowledge in different fields to generate new knowledge. According to March (1991), it is important to balance two types of organizational learning; “Exploitation” which deepens and utilizes existing knowledge, and “Exploration” which searches for new knowledge without being caught by existing knowledge. And organizations tend to exploit existing knowledge and neglect to explore for new knowledge [1].

1.2 Collaborative Research

As science and technology become more sophisticated and complex, the need for collaborative research is increasing. In particular, exploration of knowledge in new science and technology fields, especially in cross-cutting and fusion research fields, is an important issue. In japan, The Ministry of Education, Culture, Sports, Science and Technology's also announced, in "Implementation Policy for KAKENHI Reform", that "Challenge", "Comprehensiveness", "Fusion" and "Internationality" are required for academic research. And "Diversity" and "Overview of subdivided knowledge" are important for "Comprehensiveness", and "Collaboration and cooperation of researchers from different fields" and "Creation of new academic fields is important for "Fusion".

1.3 Two Ways of Interdisciplinary Research

Interdisciplinary research is a research that crosses over several different disciplines. There are two ways of interdisciplinary research. A way in which individual researchers conduct research across different fields, and another way in which multiple researchers specializing in different fields perform interdisciplinary activities through collaboration. In the former way, each individual researcher's load of interdisciplinary research is high, but there is no communication load with other researchers. On the other hand, in the latter way, individual researchers specialize in their specialized fields and their load of interdisciplinary research is low, but the communication load among collaborating researchers will be high. However, it has not been sufficiently clarified which research way is preferable in academia or industry and whether there are differences in research ways depending on the research field.

In this paper, we propose a method for measuring two ways in interdisciplinary research; a way for interdisciplinary research by individual researchers and another way for interdisciplinary research by collaboration of multiple researchers. In addition, by this method, a database on "KAKENHI" which is research grant provided by the Japan Society for the Promotion of Science is used to measure interdisciplinary research from the perspective of the two research ways, and the

features of interdisciplinary research in KAKENHI are analyzed for each research field. Finally, we discuss the analysis result and provide implications for the decision-makers of science and technology policy and strategy when making research policy and strategies.

In the following sections of this paper, we summarize related research, describe the analysis method, show the analysis result, and further discuss the result.

2. RELATED WORK

2.1 Interdisciplinarity

Research is an activity for creating and accumulating knowledge, and in general, there are two directions to which researchers conduct research; one is "specialty" to expertise a single research field, and the other one is "interdisciplinarity" to expand to multiple research fields. Researchers may progress toward one or both of specialty and interdisciplinarity as they gain research experience.

March, a socio-political scholar, stated that there are two types of organizational learning: "Exploitation" which deepens and utilizes existing knowledge, and "Exploration" which explores new knowledge without being caught by existing knowledge, that it is important to balance the two directions of Exploitation and Exploration, and that organizations tend to bias to Exploitation of existing knowledge and neglect Exploration for new knowledge [1]. Accordingly, Exploitation would be an approach toward specialty and Exploration would be an approach toward interdisciplinarity.

In Secientometrics and Bibliometrics which quanti-tatively study science and technology, many studies have been conducted on interdisciplinary research from the perspectives of effectiveness of interdisciplinary research, differences in interdisciplinarity depending on the criteria of researchers and fields, etc. For examples, first, when analyzing interdisciplinarity based on academic and scientific literatures, an analytical framework and indicators are required. Stirling (2007) proposed general frameworks for analyzing "Diversity" of natural sciences and social sciences [2]. Rafols, et al. (2010) extended this diversity analysis framework by adding "Coherence" analysis framework and proposed a method for assessing interdisciplinarity in Bibliometrics with a case study in the field of bionanoscience [3]. Clarifying the structure of the academic field of scientific literature is one of the central goals of scientific metrology. So Leydesdorff and Rafols et al. created science map and showed the efforts of each research institute on the science map, and stated that the ranking of journals is biased to favor research in a single field [4][5][6]. As an interdisciplinary analysis using data other than academic literature, Sun, et al. (2016) used the relationship of researchers attending different academic societies in the field of computer science and the interdisciplinary nature was visualized and analyzed based on the differences of academic societies [7]. Furthermore, as an example of analyzing the features of each researcher's criteria, Kastrin, et al. (2018) compared the principal investigators of research projects and other researchers in 19,598 Slovenian researchers, and showed that the principal investigators had superiority in productivity, joint research, internationality and interdisciplinarity [8]. As an example of analysis of interdisciplinary features in each research field, Abramo, et al. (2017) analyzed about 33,784 professors in Italy in three dimensions: range of diversity, strength, and topic relevance, and showed that interdisciplinarity was the lowest in mathematicians and the highest in chemists [9]. As an analysis of the features of interdisciplinarity from the viewpoint of time transition, Porter, et al. (2009) analyzed transition of interdisciplinarity in six research areas during the 30 years from 1975 to 2005. It showed that interdisciplinarity increased only by about 5%, though the number of citations increased by about 50% and the number of co-authors increased by about 75% [10]. Chen, et al. (2015) also analyzed the evolutionary transition between disciplines over 100 years in biochemistry and molecular biology, and showed that interdisciplinary disciplines evolved from near fields to far fields [11].

2.2 Individual Research and Collaborative Research

In science and technology innovation, the ability of collaborative research is important as well as the ability of individual researchers. In research that is an activity for creating and accumulating knowledge, the knowledge is created and accumulated in individuals and organizations to become tacit knowledge and formal knowledge, which is passed on to the next individuals and organizations. These researches include two types of research; "Individual Research" in which individual researchers refer to formal knowledge such as papers and create and accumulate knowledge, and "Collaboration Research" in which researchers can exchange tacit knowledge and creates and accumulates knowledge.

In research activities of science and technology, organizational activities and leaders within the organization are important. Crane clarified that the collaborative research groups influence the growth of scientific knowledge in related fields, and then called the network of these groups as "Invisible University." She stated that interactions among researchers and leadership in collaborative research groups have an important influence, and that ideas are propagated through the leaders [12]. In addition, Allen clarified that researchers, who have both high level of technical ability and high level of communication ability are important and called

such researchers as “Gatekeepers.” Gatekeepers contact frequently with researchers inside and outside the organization as an intermediary, translates and transmits external information inside the organization, and play a central role in communication as star researchers. He stated that Gatekeeper would improve the performance of research and development [13].

In Scientometrics and Bibliometrics, there have been many studies on organizational research. In Scientometrics, social network analysis is often applied to the studies. Typical social networks include citation networks which are knowledge networks, and co-authorship networks which are collaboration networks. The citation network is a social network that is constructed from citation relations of documents, and uses documents as nodes and citation relations as edges, and is applied to "science maps", "patent maps", etc. On the other hand, the co-authorship network is a social network that is constructed from the co-authorship relationship of documents, and use authors as nodes and co-author relation as edges, and is applied to "researcher maps", etc. Co-authorship network shows the collaboration relationship at a certain time or period, and it is possible to analyze organizational research and the roles of researchers in organizational research.

Newman (2004) conducted a structural analysis of a co-authored network of papers in the three fields of biology, physics, and mathematics. It showed that the number of authors per paper varied depending on the research style, and the number in the fields such as biology, physics where experimental research methods are often used is more than that in the fields such as mathematics where theoretical research methods were often used. Also, in physics, researchers tend to build close co-author relation on the networks, and in biology, they tend to build radial co-author relation centered on influential researchers [14]. In addition, as a study of knowledge flow based on the genealogy of researchers, Shinoda (2011) created a co-author network of papers published in Journal of Japanese Society for Artificial Intelligence, to identify the central person in the Society for Artificial Intelligence. By observing the transition of the central person, the genealogy of the AI Society was created [15].

2.3 Issues in Interdisciplinary Research

In interdisciplinary research that crosses over several different disciplines, there are two research ways; one way in which individual researchers conduct research across different fields, and another way in which multiple researchers specializing in different fields perform interdisciplinary research through collaboration.

So, which of the two research ways should researchers use to conduct interdisciplinary research? What kind of research organization should policy or strategy makers make in promoting interdisciplinary research? Is the research organization in interdisciplinary research different depending on the field? It is not clear enough to answer these questions.

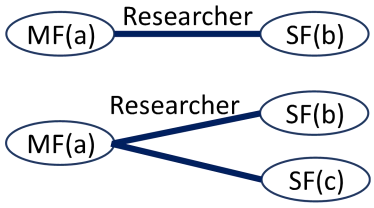
In this paper, we propose a method for measuring two ways in interdisciplinary research; a way for interdisciplinary research by individual researchers and another way for interdisciplinary research by collaboration of multiple researchers. In addition, by this method, a database on "KAKENHI" which is research grant provided by the Japan Society for the Promotion of Science is used to measure interdisciplinary research from the perspective of the two research ways, and the features of interdisciplinary research in KAKENHI are analyzed for each research field. Finally, we discuss the analysis result and provide implications for the decision-makers of science and technology policy and strategy.

3. METHOD

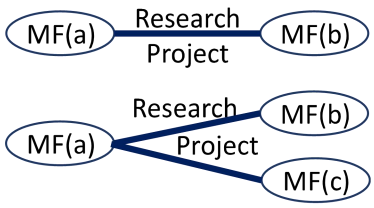
In this section, we explain the interdisciplinary networks of individual research and the interdisciplinary networks of collaborative research, then describe the database to analyze and show the analysis procedure.

3.1 Individual Interdisciplinary Network (IIDN) and Collaborative Interdisciplinary Network (CIDN)

As we previously explained, there are two ways of interdisciplinary research that crosses over several different disciplines; a way in which individual researchers conduct research across different fields, and another way in which multiple researchers specializing in



(a) Individual Interdisciplinary Network (IIDN)



(b) Collaborative Interdisciplinary Network (CIDN)

Fig. 1. Individual Interdisciplinary Network (IIDN) and Collaborative Interdisciplinary Network (CIDN)

different fields perform interdisciplinary activities through collaboration. These two ways can be represented as shown in Fig. 1 as two types of networks with research fields as nodes. In Fig. 1 (a) and (b), the nodes "MF" and "SF" mean the Main research Field and Sub research Field, and "a", "b" and "c" mean individual research fields. In addition, in Fig. 1 (a), "Researcher" at the edge means that the nodes are connected by researchers, and in Fig. 1 (b), "Research Project" at the edge means that the nodes are connected by research projects.

Fig. 1 (a), focusing on interdisciplinarity by individual researchers, shows interdisciplinary networks in which the research fields are set as the nodes of the networks and the researchers are set as the edges of the networks. we call this type of networks as "Individual Interdisciplinary Network (IIDN)". In Fig. 1 (a), MF(a) shows a main research field of a certain researcher as a node, and SF(b) shows another field (sub-field) of the researcher as a node. Furthermore, MF(a) and SF(b) are connected with the researcher as an edge.

On the other hand, Fig. 1 (b), focusing on interdisciplinarity by collaboration of multiple researchers specializing in different fields, shows interdisciplinary networks in which the research fields are set as the nodes of the networks and the research projects are set as the edges of the networks. we call this type of networks as "Collaborative Interdisciplinary network (CIDN)." In Fig. 1 (b), MF(a) shows a main research field of a certain researcher as a node, and MF(b) shows another research field of another researcher who is a collaborator of the researcher as a node. Furthermore, MF(a) and MF(b) are connected with a KAKENHI project as an edge.

3.2 Database

In analyzing the interdisciplinary research in this paper, we used the KAKEN database ("KAKEN") constructed by the National Institute of Informatics for the KAKENHI research grant program ("KAKENHI") provided by the Japan Society for the Promotion of Science.

KAKENHI includes, as research categories, Grants-in-Aid for Scientific Research (Specially Promoted Research, Scientific Research on Innovative Areas, Scientific Research (A)/(B)/(C), Challenging Research, Young Scientists, etc.), Grants-in-Aid for JSPS Fellows, Fund for the Promotion of Joint International Research, etc.

KAKEN database has information, such as research subjects, representative researchers, collaborative researchers, research budgets and expenses, research results, etc. for all of these research categories.

In this paper, we analyze “Scientific Research (A)” in KAKEN database for 11 years from 2008 to 2018 (“all period”). The research fields of Scientific Research (A) include Integrated Disciplines ("ID"), "Integrated Disciplines and Innovative Science ("II"), Humanities and Social Sciences ("HS"), Science and Engineering ("SE") and Biological Sciences ("BS") as shown in Table 1. As for research field category in analyzed period, in this paper we regard both ID and II together as same as ID, because II had been changed to ID since 2013.

3.3 Procedure of Analysis

The analysis procedure is shown as follows..

Step 1. Determination of the researcher's Main research Field (MF)

As for researchers, using the research field of each KAKENHI project in KAKEN database, the research field with the largest sum of KAKENHI projects which each researcher has obtained (in case there are multiple research fields with the largest sum of KAKENHI projects, the research field with the largest sum of KAKENHI budget in the multiple research fields) is determined as “Main research Field (MF)” of each researcher.

Table 1. Research fields of Scientific Research (A)

Field Category	Fields
Integrated Disciplines (ID)	Informatatics, Environmental science, etc.
Integrated Disciplines and Innovative Science (II)	Conprehensive Fields(Informatics, etc), New Multidisiplinary Fields (Environmental sciencs, etc.)
Humanities and Social Sciences (HS)	Humanities, Social Sciences, etc.
Science and Engineering (SE)	Mathematics and Pyhsics, Chemistry, Engineering, etc.
Biological Sciences (BS)	Biology, Acriculture, Medicine, Pharmacy, etc.

Step 2. Calculation of Field Variety (FV) for each researcher and for each MF

Calculate the sum of research fields of KAKENAI projects which each researcher has obtained each year and for the whole period, and define it as “Field Variety of Researcher (FV(researcher))” for the researcher.

In addition, calculate the average of FV(researcher) for each MF, and for each year and for whole period, and define it as “Field Variety of MF (FV(MF))”.

Step 3. Calculation of IIDN and CIDN for each MF

Calculate the number of IIDNs (individual interdisciplinary research projects) and the number of CIDNs (collaborative interdisciplinary research projects) described in the previous section for each year and for the entire period.

Step 4. Analysis of interdisciplinary research

Analyze the transition of FV, IIDN and CIDN for each MF.

4. RESULT

In this section, as a result of our analysis of researchers who won Grant-in-Aid for Scientific Research A, we show the features of FV, an index of researcher's interdisciplinarity, and then show the features of IIDN, an index of interdisciplinary projects by individual, and CIDN, an index of interdisciplinary projects by collaboration.

4.1 Features of FV of Researchers

The number of the analyzed researchers in Scientific Research A from 2008 to 2018 was 25,518. Table 2 shows the FV of the analyzed researchers. From Table 2, it is found that about 94% of the researchers in Grant-in-Aid for Scientific Research A have research in only a single field, and about 6% have research in 2 fields or more.

Fig. 2 shows the transition of FV (MF) in Scientific Research A. In Fig. 2, the numbers on the vertical axis are FV (MF) minus 1. From Fig. 2 it is found that interdisciplinarity of each field varies depending on the field. In addition, although we assumed in advance that interdisciplinarity would increase in the course of time as complex, the result of our analysis of the KAKENHI project was rather the opposite. In the case of ID, it can be seen that FV would decrease in the course of time.

Table 2. Field Variety of Researchers of Scientific Research (A)

FV	Researcher	
	Number	Ratio
1	22,518	93.7%
2	1,459	6.1%
3	44	0.2%
4	1	0.0%
Total	24,022	100.0%

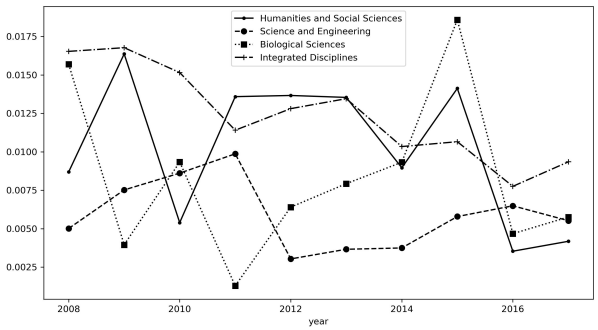


Fig. 2. Transition of Field Variety of Main Fields of Scientific Research (A)

4.2 Features of IIDN and CIDN of Projects

Table 3 shows the transition of IIDN and CIDN for each MF and SF of Scientific Research (A). Two lines of research fields are listed at the top 2 rows of each column in Table 3; the first line shows Main Field (MF), and the second line shows Sub Field (SF) which is other fields than MF. Fig. 3 shows the relationship between MFs and SFs of IIDN and CIDN in Scientific Research (A). In Fig. 3, the direction of the arrow indicates from MF to SF, and the thickness of the arrow indicates the number of IIDN and CIDN. The following can be found from Table 3 and Fig. 3.

First, as for IIDN and CIDN, the number of CIDNs is higher than that of IIDNs. The number of IIDNs is 1,739, while the number of CIDNs is 2,158. The pattern of numbers of two fields combinations in interdisciplinary research is almost similar between IIDNs and CIDNs, and combinations than those of IIDNs, though the number of IIDNs is higher than that of CIDNs only where SE is MF and ID is SF.

Second, as for research fields, the number of interdisciplinary researches for each field and for each combination of fields differ from field to field. For expla-

Table 3. Transition of IIDN and CIDN for MF and SF of Scientific Research (A)

(a) IIDN

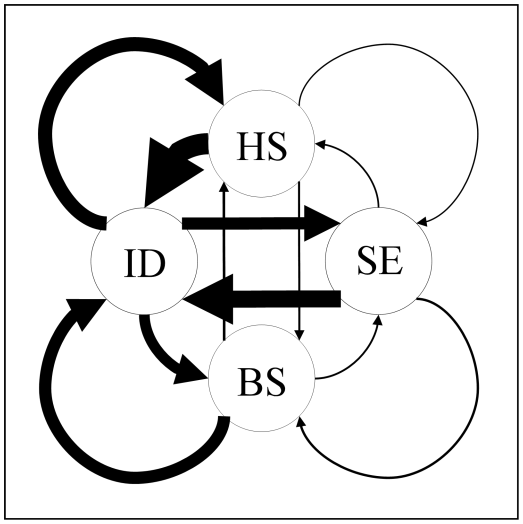
MF	ID			HS			SE			BS			Total
SF	HS	SE	BS	ID	SE	BS	ID	HS	BS	ID	HS	SE	
2008	16	18	17	30	2	4	17	2	5	21	6	5	143
2009	15	19	9	54	2	5	37	5	2	21	5	0	174
2010	19	14	15	19	1	3	25	1	5	25	2	3	132
2011	18	19	18	79	2	2	29	5	0	18	5	1	196
2012	19	22	22	37	3	5	25	2	3	20	4	2	164
2013	39	22	23	34	5	4	33	7	2	23	6	3	201
2014	29	26	22	34	3	4	35	4	3	29	3	7	199
2015	17	38	22	40	1	4	29	0	6	21	8	4	190
2016	47	16	22	23	1	1	42	1	8	19	6	6	192
2017	27	20	17	20	3	5	25	4	7	17	1	2	148
Total	246	214	187	370	23	37	297	31	41	214	46	33	1,739

(b) CIDN

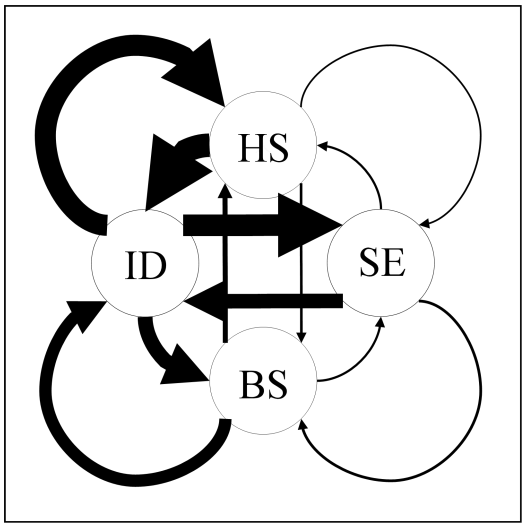
MF	ID			HS			SE			BS			Total
SF	HS	SE	BS	ID	SE	BS	ID	HS	BS	ID	HS	SE	
2008	38	18	24	27	2	0	19	2	3	13	14	6	166
2009	48	41	23	33	2	5	27	6	2	12	4	0	203
2010	31	36	29	27	1	3	19	1	4	23	2	3	179
2011	43	49	28	86	2	11	22	6	0	16	2	0	265
2012	48	41	23	22	5	3	28	2	4	24	24	3	227
2013	20	32	32	62	11	3	19	9	4	23	17	3	235
2014	48	32	47	21	3	2	37	5	7	25	9	3	239
2015	26	31	22	42	0	14	31	2	8	28	6	7	217
2016	44	50	20	60	1	0	31	1	9	26	8	9	259
2017	21	45	10	27	3	3	20	4	6	24	1	4	168
Total	367	375	258	407	30	44	253	38	47	214	87	38	2,158

ID: Integrated Disciplines HS: Humanities and Social Sciences
SE: Science and Engineering BS: Biological Sciences

ple, in both IIDN and CIDN, the number of interdisciplinary researches, where MF or SF is ID, is high, while the number of interdisciplinary researches between fields other than ID is rather low.



(a) IIDN



(b) CIDN

ID: Integrated Disciplines HS: Humanities and Social Sciences
SE: Science and Engineering BS: Biological Sciences

Fig. 3. Relationship between MFs and SFs of IIDN and CIDN in Scientific Research (A)

Finally, as for MF and SF, the relationship of numbers of interdisciplinary researches in the case a certain field is MF and the other case the field is SF is asymmetric. For examples, in both IIDN and CIDN, number of interdisciplinary research where SE is MF and the other field is SF is higher than that where the other fields are MF and SE is SF except only the case between ID and SE of CIDN. This means that SE is leading interdisciplinary research with other fields. Furthermore, each research field leads interdisciplinary research with other research field more in the order of SE, BS, HS and ID, except only the case that ID of CIDN leads interdisciplinary research with SE or BS.

5. DISCUSSION

5.1 Interdisciplinary Research and Collaborative Research

To progress science and realize innovation, not only deepening and accumulating knowledge in one field, but also seeking knowledge in various fields and combining knowledge in different fields to obtain new knowledge is important.

Regarding the progress of science and technology, Newton, the great scientist, stated that the new discoveries are accumulated based on the previous discoveries, using the metaphor of "Standing on the Shoulder of Giants". That is, scientific progress has been accumulated and built on the research activities of many previous researchers [16]. The accumulation of new knowledge on top of past knowledge is called "knowledge accumulation" in this paper, as an approach toward specialization.

On the other hand, regarding innovation, Schumpeter, the great economist, defined five types of “Neue Kombination" (= New Combination); (1) production of new things, (2) introduction of new production methods, (3) development of new sales channels, (4) acquisition of new sources of raw materials or semi-finished products. , (5) Realization of a new organization, and stated that these new combinations are necessary for economic development [17]. Creation of new knowledge by combining various knowledge is called "knowledge combination" in this paper, as an approach toward interdisciplinarity.

As previously mentioned, March, an sociopolitical scholar, stated that there are two types of organizational learning; “exploitation” which means utilization of existing knowledge and “exploration” which means search for new knowledge without being caught by existing knowledge, and he added that it is important to balance the two types but showed that organizations tend to bias the utilization of existing knowledge and neglect to search for new knowledge [1]. This may suggest that the balance between specialization and interdisciplinarity is important but that specialization tends to be dominant to interdisciplinarity.

Fig. 4 shows the research activities to progress science and realize innovation from the above-mentioned two types of viewpoints; specialty/exploitation/knowledge

accumulation, and interdisciplinarity/exploration/ knowledge search.

Furthermore, in Fig. 5, based on the results of our analysis in this paper, we positioned the four of research fields of Scientific Research (A) on the square that has one axes of interdisciplinarity, i.e. specialty or interdisciplinarity, and the other axes of research organization, i.e. individual or collaboration. In Fig. 5, the interdisciplinarity axis positions each research field based on the sum of IIDN and CIDN. And the research organization axis positions each research field based on difference of CIDN and IIDN (CIDN minus IIDN).

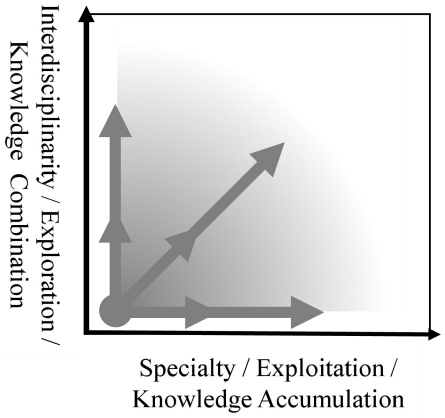
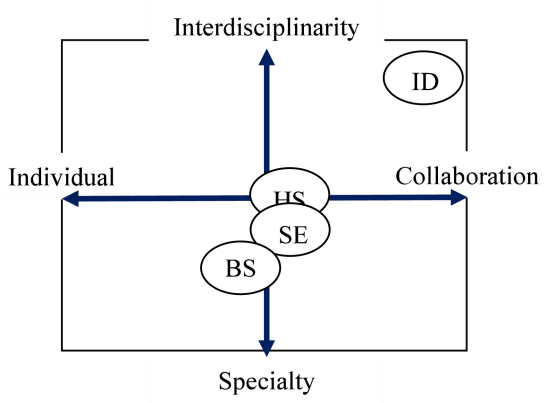


Fig. 4 Research activities from two types of viewpoints



ID: Integrated Disciplines HS: Humanities and Social Sciences
SE: Science and Engineering BS: Biological Sciences

Fig. 5. Positioning of four research fields of Scientific Research (A) from viewpoints of interdisciplinarity/ specialty and collaboration/individual

From Fig. 5, as for the four analyzed research fields of KAKENHI, ID field has a particularly high level of both interdisciplinary research and collaborative research, while BS field has a low level of both interdisciplinary research and collaborative research. HS field and SE field are located between ID field and BS field. As the importance of Liberal Arts, which conducts research and education across social science and natural sciences, is increasing, further interdisciplinary research across these fields is desired.

As the creation of knowledge in new areas of science and technology, especially in the area of cross-cutting and fusion research fields becomes important, and science and technology become more sophisticated and complex, the Ministry of Education, Culture, Sports, Science and Technology of Japan (“MEXT”) announced KAKENHI reform policy. In this policy, “Integration” and “Fusion” are mentioned as important issues. In addition, it mentions that as for "Integration", "Diversity" and "Bird's-eye view of subdivided knowledge" are important, and that as for "Fusion", "Creation of new academic fields" through "Cooperation and collaboration of researchers in different fields" is necessary. In promoting this policy, the proposed method in this paper will be useful as a method for quantitatively measuring the interdisciplinarity and research organizations.

5.2 Detecting Emerging Fields

In Section 4.1, we analyzed the transition of the interdisciplinarity. The purpose was to try to detect the emerging areas by analyzing the fields of interdisciplinary growth, because we assumed in advance that interdisciplinarity of the emerging areas would increase in the course of time, as science and technology become more sophisticated and complex. In the above-mentioned MEXT's KAKENHI reform policy, “Challenging” in addition to “Integration” and “Fusion” is also mentioned as the first requirements for academic research. As a result of our analysis, it was not possible to identify such emerging areas where interdisciplinarity is growing, though it was found that research has different interdisciplinary features in each research field. Our analysis result of the KAKENHI project was rather the opposite, and interdisciplinarity tends to decline in ID field.

In science technology and innovation policies and strategies, it is desired to make the policies and strategies based on the evidence [18] [19]. In this research, KAKEN projects were analyzed by the proposed method in the four major categorized fields shown in Table 1, i.e. Integrated Disciplines, Humanities and Social Sciences, Science and Engineering, and Biological Sciences. In the future, more detailed analysis results will be able to obtained by analyzing fields that are further divided, because MEXT 's KAKENHI reform policy announced to review and reconstruction of field categories. We plan to challenge for detection of emerging research fields in academia and business by measuring the transition of the interdisciplinary field based on the reconstructed field categories.

6. CONCLUSION

In this paper, we proposed a method for measuring two ways in interdisciplinary research; a way for

interdisciplinary research by individual researchers and another way for interdisciplinary research by collaboration of multiple researchers.

In addition, by this method, a database of "KAKENHI" which was grant-in-aid for scientific research provided by the Japan Society for the Promotion of Science is used to measure interdisciplinary research from the perspective of the two research ways, and the features of interdisciplinary research activities in KAKENHI were analyzed for each research field.

As a result of the analysis, we found that (1) the number of collaborative interdisciplinary research is higher than that of individual interdisciplinary research, (2) numbers of interdisciplinary research for each field and for each combination of fields differ from field to field, (3) the relationship of numbers of interdisciplinary research of a certain field as main field and that of the same field as sub field is asymmetric.

Recently, science and technology become more sophisticated and complex, and the need for interdisciplinarity and collaborative research increases. As the proposed measurement method is able to quantitatively measure interdisciplinarity among fields and their research organizations, it will be useful for decision makers of science and technology policy and strategy. In the future, we intend to detect the newly emerging academic and business fields by measuring and analyzing the transition of the interdisciplinary field between the more subdivided fields.

Acknowledgement

This study was supported by JSPS KAKEN 18H00840.
We are very grateful to Prof. Koichi Sumikura of The National Graduate Institute for Policy Studies for his cooperation.

References

[1] James G. March, Exploration and Exploitation in Organizational Learning, Organization Science, 2 (1), 71–87, 1991.
[2] Stirling, A., A general framework for analysing diversity in science, technology and society, Journal of the Royal Society Interface, 4-15(22), 707-719, 2007.
[3] Rafols, I., Meyer, M., Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience, Scientometrics, 82(2), 263–287, 2010.
[4] Leydesdorff, L., Rafols, I., A Global Map of Science Based on the ISI Subject Categories, Journal of the American Society for Information Science and Technology, 60(2), 348-362, 2009.
[5] Rafols, I., Porter, A. L., Leydesdorff, L., Science Overlay Maps: A New Tool for Research Policy and Library Management, Journal of the American Society for Information Science and Technology, 61(9), 1871-1887, 2010.
[6] Rafols, I., Leydesdorff, L., O'Hare, A., et al., How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management, Research Policy, 41(7), SI, 1262-1282, 2012.
[7] Sun, X., Ding, K., Lin, Y., Mapping the evolution of scientific fields based on cross-field authors, Journal of Informetrics, 10(3), 750-761, 2016.
[8] Kastrin, A., Klisara, J., Luzar, B., et al., Is science driven by principal investigators?, Scientometrics, 117(2) 1157-1182, 2018.
[9] Abramo, G., D'Angelo, C. A. Di Costa, F., Specialization versus diversification in research activities: the extent, intensity and relatedness of field diversification by individual scientists, Scientometrics, 112(3), 1403-1418, 2017.
[10] Porter A. L., Rafols, I., Is science becoming more interdisciplinary? Measuring and mapping six research fields over time, Scientometrics, 81(3), 719-745, 2009.
[11] Chen, S., Arsenault, C., Gingras, Y., et al., Exploring the interdisciplinary evolution of a discipline: the case of Biochemistry and Molecular Biology, Scientometrics, 102(2), 1307-1323, 2015.
[12] D. Crane, Invisible Colleges, The University of Chicago, 1972.
[13] Thomas J. Allen, Managing the Flow of Technology, MIT Press, 1979.
[14] Mark E. J. Newman, Coauthorship networks and patterns of scientific collaboration, Proceedings of the National Academy of Sciences, 101(suppl. 1), 5200-5205, 2004.
[15] Shinoda, K., A Family Tree of Artificial Intelligence Research in Japan, Journal of Japanese Society for Artificial Intelligence, Vol.26, No. 6, pp. 584-589, 2011.
[16] H. W. Turnbull, ed., The Correspondence of Isaac Newton: 1661-1675, 1, Published for the Royal Society at the University Press, 1959.
[17] J. A. Schumpeter, Theorie der Wirtschaftlichen Entwicklung, 1926.
[18] Macilwain,Colin, What science is really worth, Nature, Vol.465, pp.682-684, 2010.
[19] Lane, J., Bertuzzi, S., Measuring the Results of Science Investments, Science, Vol. 331, pp.678-680, 2011.

Development Direction of Human Coexistence Robot Partner Based on Smart Device

Jinseok WOO*, Yasuhiro OHYAMA*, Naoyuki KUBOTA**
* Department of Mechanical Engineering
Tokyo University of Technology
Hachioji, Tokyo, Japan
** Graduate School of Systems Design
Tokyo Metropolitan University
Hino, Tokyo, Japan

Abstract

Social isolation can cause problems for human both mentally and physically. In particular, the isolation of the elderly causes serious problems such as lonely death. However, social participation can lead to healthier lives by reducing isolation. Many technologies are being developed to reduce social isolation and loneliness. We considered using a robot partner in order to prevent social isolation. In this paper, we explain the current state about our development of robot partners, and we discuss the development direction of the robot. First, we describe our robot partner system. Next, we explain the modular structured systems for the development of the robot partner system. Finally, we present a few examples of the robot system and address the applicability of the proposed system.

Keywords: Robot partner system, Robot design development, System integration, Social implementation

1. INTRODUCTION

Recently, it is expected that many problems will arise due to global aging. A typical problem is a decrease in the working population, and there is also a problem with elderly people's care. In addition, the problem of social isolation may arise due to the increase in single-person households. Social isolation can cause problems for human both mentally and physically [1]. In particular, the isolation of the elderly causes serious problems such as lonely death [2]. However, social participation can lead to healthier lives by reducing isolation [3].

Recently, we have been trying to keep a social distance to reduce the chances of touch infection due to the COVID-19 epidemic [4]. However, the relationship between people is significant, and if there is a social distance, there will be a number of problems [5]. This condition would intensify for elderly people who need a caregiver or have a social networking meeting with their friends. Humans should be assisted by human beings. However, when humans can not go, we suggest that a robot partner is required for the next solution. The implementation of non-face-to-face applications and services using robots, AI, and IoT will become more involved in the future.

Previously, we had many experiments to understand the architecture and requirements of the robot system [6]. After merging our thoughts, we were able to find out the following: First, elderly people have a preconceived notion of robots based on TV media details. Second, the first experience of a real robot is unpleasant, since there is no previous knowledge of robots. Thirdly, in order to commercialize a robot partner, an additional modification is required to make a fair price which is important for the creation of marketing strategies. Forth, their family wants to use the robot device to ensure the safety of elderly people. These issues have become important information for our robot development direction. One of our previous studies was about how the robot boosts motivation by doing gymnastics together [7]. As a result, consumers have been shown to be inspired to better their own wellbeing by working with robots. We, therefore, continued with the creation of a robot system capable of performing such robotic services. On the basis of these results, we are introducing the process of developing the robot partner system and finding the future image of the robot partner.

This paper is organized as follows. Section II describes the specification of the robot partner system according to hardware and software system structures considering the development process so far. Section III presents the proposed Human-friendly robot partner system for social implementation. Section IV explains that shows the ability of the robot partner using the proposed method. Finally, in section V, we summarize this paper, and discuss the effectiveness of this research.

2. LITERATURE REVIEW

The origin of the word ‘robot’ can be found in the “R.U.R.” by Novelist Karel Capek [9]. Based on the

"R.U.R.", the word 'robot' has frequently been used to indicate automatic machines in the world. Etymology has become a 'robota', which means 'labor' in Czech. Afterward, robots have been used for various purposes as human assistants in dirty, dangerous, and difficult (3D) works. Nowadays, robots are approaching us in a familiar form as well as 3D work [10]. For example, "Pepper" is developed by SoftBank Robotics (formerly Aldebaran Robotics), and, this robot is utilized for elder care and child education [11, 12]. "Paro" is a baby seal-type robot for the mental health care of elderly people [13, 14]. This robot is being used in numerous elderly facilities and is showing its effectiveness. There is also research on developing a walking partner robot that can interact by sharing information on the visual scene of the environment in which humans belong [15]. And, as a study from the perspective of physical interaction with humans, there is a study on robots that lead to physical human-robot interaction during dance training [16]. Such activities of robot partners are expected and used in various places where people live. Previously, people use mobile communication devices to send a phone call or text message [17]. But now, with the advent of smart devices, the paradigm of the communication environment is shifting. Smart devices are the central elements of the IoT (Internet of Things) environment [18]. As technology evolves, many smart devices are being developed with low price and high specifications. As a result, various applications have been developed based on the high specifications of smart devices equipped with various sensors such as gyro sensors, motion sensors, proximity sensors. There are various types of robots to which smart device technology is applied. For example, a robot has been developed that can conduct meetings remotely from a distance. The telepresence robot is called "Double" [19]. It supports human-to-human communication while remotely controlled by smart devices. "RoBoHoN" is Android OS based tiny human-type robot which can interact with human [20]. It could connect 3G/4G and Wi-Fi networks in order to phone and internet use.

Since 2009, we have started developing robots based on smart devices in our study. Recently, due to the advancement of 3D printer technology, more complex hardware designs can be produced. The description of the development of our robot partner system is covered in the next section.

3. ROBOT PARTNER SYSTEM

In recent years, robots have been built to provide a range of supports, led by robots for cleaning. In addition, social robots are being actively built to provide services in individual homes. However, while numerous social robots are being built and commercialized, not many robots will be used. There may be a variety of reasons why social robots have not been able to make widespread use of the general home. Among them, we have found two situations based on previous experiments [6]. The first is that it is difficult to consider the desires of users. Secondly, there is also the fact that it is not easy to set up a service system for people to meet the needs of customers. Therefore, in order to build a robot partner that can function as a social robot, we have defined needs and developed a platform for a long time. The following describes the scope of our study.

3.1 Robot Development using Design Thinking

Previously, we visited and demonstrated several facilities to evaluate the needs of robots [6 - 8]. There were various opinions, so we synthesized the opinions and considered the system necessary for the robot. Therefore, we introduce the development direction of robot partners based on the smart device using design thinking [21]. The flow of design thinking can be expressed as shown in Fig.1.

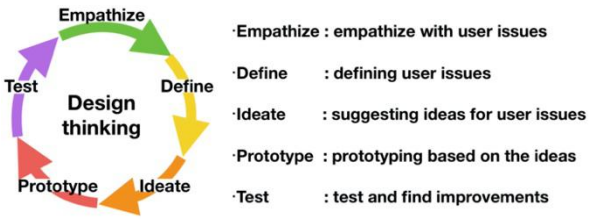


Fig.1 The example of design thinking process

Firstly, we performed robot demos to empathize with users, visited facilities for the elderly and private homes, and assessed needs using robots. There were different opinions, so we summarized the opinions and suggested the method suitable for the robot. Next, on the basis of opinions, as a result of the creation and operation of a robotic exercise support system for the health management of the elderly, it was found that the following points are essential for the development of social robots. The question is whether the content can be freely adapted to the user's needs, whether the device can be used in a comfortable manner without becoming cumbersome to the user, and whether the user expectations will choose the robot's architecture. Thus, a social robot must be built by empathizing with the situation of these users. The development path for the robot partner's design is determined, and ideas are produced accordingly, according to the above information (Fig. 2).

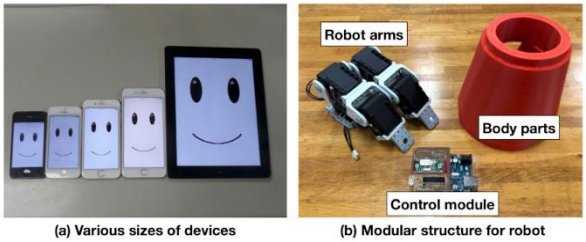


Fig.2 Robot Partner Development.

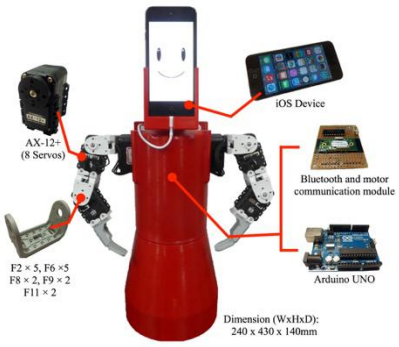


Fig.3 Robot Partner: iPhonoid-C

As a result, we are developing a robotic system using smart devices to develop a robot that can be used without reticence to users, as shown in Fig.3. In order to help the design according to the preference of each user, we are creating a 3D printer robot. The robots that have been built are getting ideas for the next creation by actually using them and conducting the tests as shown in Fig.4.

(a) Interaction with the robot (b) Information support by robot partners

Fig.4 Robot partner with modular structure.

3.2 Human Robot Interaction System

The design of a robot is important for the realization of natural communication with humans. It is important to have a design that lets you know its purpose and know how to use it by looking at the exterior of the robot. So far, we have developed different types of robot partners. In order to interact with humans, our robot partner system uses various sensors and communication systems built into the smart device, and the robot body is designed using only simple communication between the drive system, including the actuator and the smart device. In addition, the smart device can be moved around by separating from the robot's body. In this way, knowledge of different circumstances can be accessed. In addition,

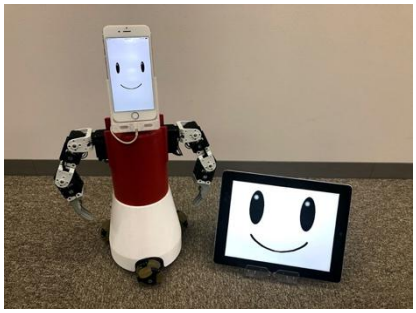


Fig.5 Robot partner equipped with moving mechanism

by separating the robot body, the smart device can be relocated around and, thus, knowledge of different circumstances can be collected. In addition, if the internal details of the smart device can be accessed, it would be possible to provide adequate help for the situation of the user. The robot companion is also in a position to provide knowledge support, such as a personal assistant or care help for the elderly. Also, if a smart device is installed on a mobile robot, it can be used for guidance in a public place (Fig.5).

3.3 Emotional Empathy

In human speech, facial and gesture expressions contain elements of emotion, and the type of communication is tailored to their emotional state. In other words, it is difficult to control the flow of speech and enjoy contact when the facial expressions of the other person are unknown. It is also important to convey this emotional state to the robot partner as well. Since it is difficult to identify these emotional states to solve this issue, non-verbal communication research uses an emotional model to understand the emotional state.

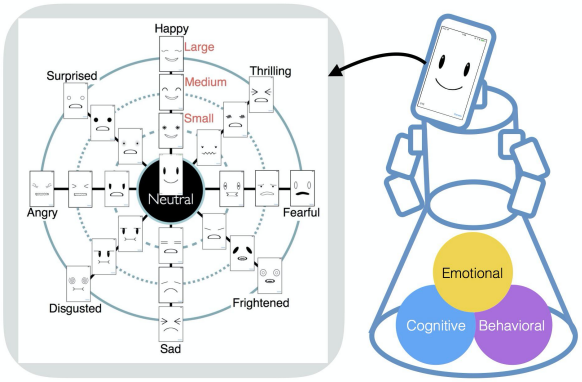


Fig.6 Facial expression design based on robot emotional model

In our study, an emotional model consisting of emotion, feeling, and mood is used, and it has an effect on the selection of phrases and the use of gestures [22, 23]. We modeled the robot partner's facial expressions using a smartphone computer. We are trying to alter facial expressions using 25 styles of facial patterns using emotion model values as shown in Fig.6.

4. DEVELOPMENT OF HUMAN-FRIENDLY ROBOT PARTNERS

4.1 Development of Software system

In order to build a robot, elements with different functions, including AI (Artificial Intelligence) elements, are needed. We consider a system architecture that integrates modules that are appropriate for the construction of robot partner systems that meet the

diverse needs of users. As a result, in our study, we are developing a robot partner with the structure shown in Fig.7. By modularizing the robot partner framework and integrating the components, it is simple to build a system that meets the needs of the user. Each robot system module consists of four layers, as shown in Fig.7 in order to incorporate the basic functions of the robot. Consists of the first hardware layer, the second library layer, the third part layer, and the fourth device layer. Based on the configuration of this module, a robot system that meets the needs of users can be built and configured.

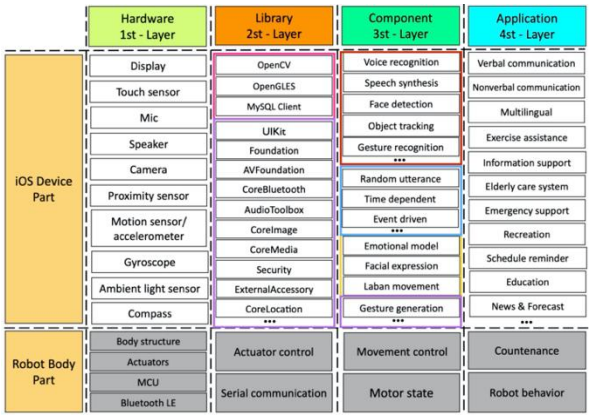


Fig.7 Modularization of Robot Element Technologies

4.2 Development of Hardware system

In order to minimize the cost of implementing robot partners, we have established a robot partner that primarily uses smart devices, as shown in Fig.8. We are constructing a robotic partner system based on iOS and Android devices (Fig.8(b)). This robot partner system has a simple module consisting of a motor, a microcomputer control and a smart device.

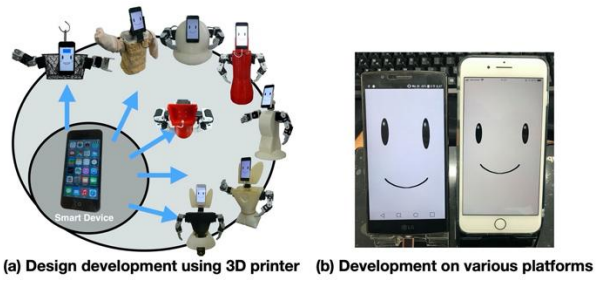


Fig.8 Various design and platforms for robot partner developments

Fig.9 Example of a panoramic photo using a robot's waist degree of freedom

The robot design shown in Fig.8(a) makes it possible to develop robots with various physical properties. For example, using the smart device and a robot body structure, it could confirm the environment through panoramic images of the surrounding environment as shown in Fig.9.

In addition to the development of 3D printers in recent years, personal and small-scale manufacturing can be achieved by uploading concept drawings of robot designs. Thus, by forming a community that can share 3D drawings of robot partners, robot designs can be shared and selected (Fig.10). It is, therefore, possible to design a robot partner that suits the user's preferences.

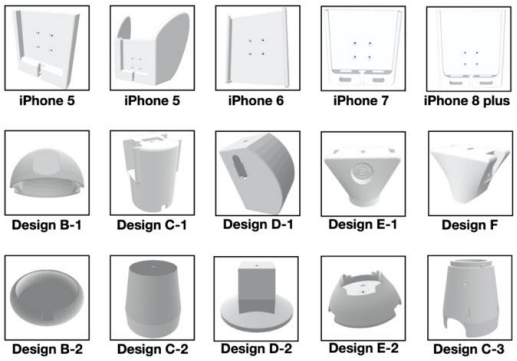


Fig.10 Examples of design drawings for 3D printing

5. SOCIAL IMPLEMENTATION OF ROBOT PARTNERS

5.1 The Design Selection of Robot Partner

A design support framework to improve the efficiency of robot creation has been developed in this paper. Here, we are presenting an example of a method for making robot parts using a 3D printer in a general home. Using a general design application needs some basic drawing skills. We have therefore built a framework that can indirectly pick a design based on their context information (Fig.11). We could pick and print each component using the smart device AR technology as shown in Fig.11(c).

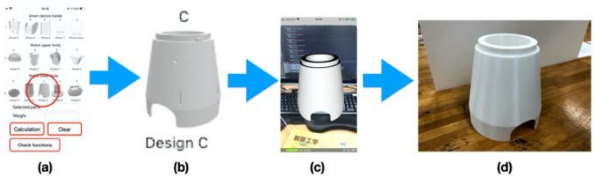


Fig.11 The process of making robot parts

5.2 Information Supporting System using Robot Partners

We present examples of the social implementation of human-robot interactions for education and reception. For users who do not have advanced programming skills, we

have built a multi-robot interaction framework, such as the development of voice content for robot partners. Users who build robot content can demonstrate interactions from multiple robots by changing the parameters shown in Table 1. By editing the event guide details in the content design prototype, the event guide can be assisted as shown in Fig.12.

Table 1. The example of interaction design template

robot_id	scenario_id	emotion_id	…	language	sentences
A_robot	0	1	…	en-US	Are you sure you want to use English?
B_robot	0	1	…	ja-JP	日本語でよろしいでしょうか。
C_robot	0	1	…	ko-KR	한국어를 사용하시겠습니까?
…	…	…	…	…	…

Fig.12 Conversation system of multi robot

In order to help elderly people, we should include resources such as the recognition of medical problems and emergency signaling. As seen in the figure. 13, when a robot has been able to detect a fall due to sensory input, the robot ensures that the user is safe. If there is no response from the user, it is possible to connect with the outside through the smart device network.

Fig.13 Example of elderly care system

6. CONCLUSION

In this paper, we have come up with new purposes for smart devices. Also, we have been evaluating a range of services using robot partners. Users may verify the safety of their family and contact them via a robot partner. However, also from the point of view of robot science, we believe that human services should be delivered by human hands. However, we should also consider the situation where a human being is unable to provide services due to a labor shortage. We, therefore, intend to continually build robot partner systems using smart devices so that there is no void in human support. Lastly, we presented innovations and examples related to the creation of the social implementation of robotic partners

As the future works, for the ease of system access to create a robot, we are considering designing a web page-based interface selection method using both a computer and a smart device.

Acknowledgment

The authors would like to thank Wei Hong Chin for his kind proofreading.

References

[1] Cacioppo, John T., and Louise C. Hawkley. "Perceived social isolation and cognition." Trends in cognitive sciences 13, No. 10, 2009, pp. 447-454.

[2] Tomaka, Joe, Sharon Thompson, and Rebecca Palacios. "The relation of social isolation, loneliness, and social support to disease outcomes among the elderly." Journal of aging and health 18, No. 3, 2006, pp. 359-384.

[3] Ejiri, Manami, Hisashi Kawai, Yoshinori Fujiwara, Kazushige Ihara, Yutaka Watanabe, Hirohiko Hirano, Hun Kyung Kim, Kaori Ishii, Koichiro Oka, and Shuichi Obuchi. "Social participation reduces isolation among Japanese older people in urban area: A 3-year longitudinal study." PloS one 14, No. 9, 2019, 11pages.

[4] Sun, Chanjuan, and John Z. Zhai. "The Efficacy of Social Distance and Ventilation Effectiveness in Preventing COVID-19 Transmission." *Sustainable Cities and Society*, 2020, 10pages.

[5] Werner, Perla. "Social distance towards a person with Alzheimer's disease." International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences 20, No. 2, 2005, pp. 182-188.

[6] Woo, Jinseok, Kazuyoshi Wada, and Naoyuki Kubota. "Robot partner system for elderly people care by using sensor network." In *2012 4th IEEE RAS & EMBS international conference on biomedical robotics and biomechatronics (BioRob)*, 2012, pp. 1329-1334.

[7] Ono, S., J. Woo, Y. Matsuo, J. Kusaka, K. Wada, and N. Kubota. "A Health Promotion Support

System for Increasing Motivation Using a Robot Partner." *Transactions of the Institute of Systems. Control and Information Engineers* 284, 2015, pp. 161-171.

[8] Yamamoto, Shion, Jinseok Woo, Wei Hong Chin, Keiichi Matsumura, and Naoyuki Kubota. "Interactive Information Support by Robot Partners Based on Informationally Structured Space." *Journal of Robotics and Mechatronics* 32, No. 1, 2020, pp. 236-243.

[9] Čapek, K. (1920). *R.U.R(Rossum's universal robots)*. Aventinum.

[10] Hayashi, Kotaro, Masahiro Shiomi, Takayuki Kanda, and Norihiro Hagita. "Are robots appropriate for troublesome and communicative tasks in a city environment?." IEEE Transactions on Autonomous Mental Development 4, no. 2 (2011): 150-160.

[11] Van der Putte, Daisy, Roel Boumans, Mark Neerincx, Marcel Olde Rikkert, and Marleen de Mul. "A social robot for autonomous health data acquisition among hospitalized patients: an exploratory field study." In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 658-659. IEEE, 2019.

[12] Pandey, Amit Kumar, and Rodolphe Gelin. "A mass-produced sociable humanoid robot: Pepper: The first machine of its kind." IEEE Robotics & Automation Magazine 25, no. 3 (2018): 40-48.

[13] Shibata, Takanori, and Kazuyoshi Wada. "Robot therapy: a new approach for mental healthcare of the elderly–a mini-review." Gerontology 57, no. 4 (2011): 378-386.

[14] Shibata, Takanori, Yukitaka Kawaguchi, and Kazuyoshi Wada. "Investigation on people living with seal robot at home." International journal of social robotics 4, no. 1 (2012): 53-63.

[15] Totsuka, Ryusuke, Satoru Satake, Takayuki Kanda, and Michita Imai. "Is a robot a better walking partner if it associates utterances with visual scenes?." In *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 313-322. 2017.

[16] Granados, Diego Felipe Paez, Jun Kinugawa, Yasuhisa Hirata, and Kazuhiro Kosuge. "Guiding human motions in physical human-robot interaction through COM motion control of a dance teaching robot." In 2016 IEEE-RAS 16th International conference on Humanoid Robots (Humanoids), pp. 279-285. IEEE, 2016.

[17] Campbell, Scott W., and Yong Jin Park. "Social implications of mobile telephony: The rise of personal communication society." Sociology compass 2, no. 2 (2008): 371-387.

[18] Stojkoska, Biljana L. Risteska, and Kire V. Trivodaliev. "A review of Internet of Things for smart home: Challenges and solutions." Journal of Cleaner Production 140 (2017): 1454-1464.

[19] Niemelä, Marketta, Lina Van Aerschot, Antti Tammela, Iina Aaltonen, and Hanna Lammi. "Towards ethical guidelines of using telepresence robots in residential care." International Journal of Social Robotics (2019): 1-9.

[20] Kobayashi, Toru, Koushin Kuriyama, and Kenichi Arai. "SNS Agency Robot for Elderly People Realizing Rich Media Communication." In 2018 6th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud), pp. 109-112. IEEE, 2018.

[21] H. Plattner, “An Introduction to Design Thinking Process Guide.” The Institute of Design at Stanford, 2010.

[22] Jinseok Woo, Janos Botzheim, and Naoyuki Kubota. "Facial and gestural expression generation for robot partners." In *2014 international symposium on micro-nanomechatronics and human science (MHS)*, 2014, pp. 1-6.

[23] Jinseok Woo, János Botzheim, and Naoyuki Kubota. "Emotional empathy model for robot partners using recurrent spiking neural network model with Hebbian-LMS learning." Malaysian Journal of Computer Science 30, No. 4, 2017, pp. 258-285.

Accelerated First-Order Distributed Method for Nash Equilibria of Convex-Concave Two-Network Bilinear Zero-Sum Games

Xianlin Zeng*, Xia Jiang*, Jian Sun*, Jie Chen*,**
* Key Laboratory of Intelligent Control and Decision of Complex Systems, School of Automation, Beijing Institute of Technology, 100081, Beijing, China
** School of Electronic and Information Engineering, Tongji University, 200092, Shanghai, China

Abstract

This paper proposes an accelerated first-order continuous-time method for the distributed Nash equilibrium seeking of a class of two-network zero-sum problems, which are convex-concave bilinear saddle point problems. First-order methods in optimization, which only use subgradients of functions, are frequently used in distributed/parallel algorithms for solving large-scale and big data problems due to their simple structures. However, in the worst cases, first-order methods for convex-concave bilinear saddle point problems often converge at a rate of $O(1/t)$. In contrast to existing time-invariant first-order methods, this paper designs the accelerated algorithm by modifying the saddle point dynamics using the derivative information of variables. If parameters of the proposed algorithm are proper, the algorithm owns $O(1/t^2)$ convergence without any strict or strong convexity requirement. The proposed distributed algorithm is used to solve a Nash equilibrium of the two-network zero-sum game with a $O(1/t^2)$ convergence rate.

Keywords: Distributed first-order method, two-network zero-sum games, accelerated convergence, convex-concave bilinear saddle point problems.

1. Introduction

Convex-concave bilinear saddle point problems (BSPP) are an important model in optimization and game theory. From application perspective, many real-life problems in various fields in the science and engineering, e.g. signal/image processing (see [1]), machine learning (see [2]), and power allocation of smart grids [3] are convex-concave bilinear saddle point problems. In addition, using the popular (augmented) Lagrangian method [4], both linearly constrained optimization problems and zero-sum games with linear couplings can be cast as convex-concave bilinear saddle point problems and solved in the primal-dual framework. Hence, lots of efforts have been dedicated to designing efficient algorithms for BSPP.

Due to the big data/scale of many modern applications and rise of distributed optimization/game, first-order methods, which only access cost functions and their subgradient information, have been particularly revisited because first-order methods may drastically simplify the computation complexity of problems and adjust to distributed deigns easily. For BSPP, existing first-order methods are mostly (time-invariant) first-order primal-dual algorithms that have $O(1/t)$ or asymptotic convergence rates under the worst choice of convex/concave cost functions. When considering unconstrained convex optimization problems which are a special case of BSPP, the best rate of convergence for first-order primal-based algorithms under the worst case has been proved to be $O(1/t^2)$ (see [5]). Specifically, the Nesterov accelerated method developed in [5] by using a vanishing damping coefficient has $O(1/t^2)$ convergence rate and has been proved to be optimal in some sense (see [6]). Following Nesterov's theme, accelerated method has been extended to nonsmooth minimization problems [7] and minimizing composite functions [8]. Recently, ordinary differential equations for modeling discrete-time accelerated methods have been revisited and analyzed to get a better understanding of design intuitions of the accelerated optimization methods [9]-[11]).

However, most existing literature on accelerated optimization methods mainly focus on primal-based methods. The design of accelerated first-order methods with $O(1/t^2)$ convergence rate for BSPP is still a challenging problem due to the involvement of dual variables and cost functions. In [12], authors proposed an accelerated primal-dual method for solving deterministic and stochastic BSPP by assuming cost functions are general smooth convex or simple functions that can be solved efficiently. [13] has proposed an accelerated linearized augmented Lagrangian method and an accelerated alternating direction method of multipliers for solving structured linearly constrained convex programming. Results in [13] are based on the assumption that cost functions have the composite convexity structure with easy minimization operations or proximal mappings. From the practical point of view, cost

functions in BSPP are often general nonsmooth convex functions that do not have good structures in these previous works. Hence, it is of great importance to study accelerated primal-dual methods for nonsmooth BSPP.

In addition, much attention has been recently paid to distributed algorithms for optimization and games over multi-agent systems. For distributed optimization problems, first-order methods have been applied to optimal consensus problems (see [14][15]), resource allocation (see [16]), and extended monotropic optimization problems (see [17]). Furthermore, by considering the presence of adversaries in optimization, a substantial body of research on distributed algorithms for Nash equilibrium of noncooperative games [18]-[21] and generalized Nash equilibria seeking problems [22]-[25] recently. When the cost functions are strongly convex, distributed algorithms for optimization and games often have an exponential convergence rate under mild conditions. Despite the recent advances on distributed optimization and games, distributed algorithms for constrained optimization problems and games without strong convexities are mostly asymptotically convergent.

Two-network zero-sum games are an important class of noncooperative games, which can also be viewed as a distributed optimization problem in presence adversaries. Considering strictly concave-convex and locally Lipschitz objective functions, [26] proposed distributed algorithms using time-invariant saddle-point dynamics with an asymptotic convergence rate. [27] proposed a distributed algorithm for nonsmooth minmax saddle point problems subject to bounded constrains.

The main shortcoming of existing works for BSPP and distributed two-network zero-sum games is that they only have $O(1/t)$ or asymptotic convergence rates under the worst choice of convex/concave cost functions. This shortcoming motivates this paper, where we aim to propose an accelerated first-order method for distributed two-network zero-sum games with a faster convergence rate than $O(1/t)$.

The main contributions of this paper can be stated as follows:

• This paper proposes a distributed accelerated algorithm, which extends the Nesterov accelerated method [5], for a class of two-network zero-sum games. To our best knowledge, this is the first accelerated first-order continuous-time method for nonsmooth convex-concave two-network bilinear zero-sum games. Compared with existing results in [26], the proposed method shows a faster convergence rate under some conditions.

• By using the the Lyapunov approach, we give rigorous proofs that the proposed algorithm can converge at a rate of $O(1/t^2)$ by choosing proper parameters.

The paper is organized as follows. Section 2 introduces mathematical preliminaries. Section 3 gives the problem formulation and proposes an accelerated primal-dual continuous-time algorithm. Section 4 proves the convergence of the proposed algorithm and shows the $O(1/t^2)$ rate of convergence under some conditions. Then Section 5 shows numerical examples to verify the efficacy of the proposed algorithm. Finally, Section 6 gives concluding remarks.

2. Mathematical Preliminary

2.1 Notation

The notation used in this paper is defined in Table 1.

Table 1 Notation

$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the set of n -dimensional real column vectors
$\mathbb{R}^{n\times m}$	the set of n -by- m real matrices
I_n	the $n\times n$ identity matrix
$(\cdot)^{\mathrm{T}}$	transpose
$\mathrm{rank}(A)$	the rank of the matrix A
$\mathrm{range}(A)$	the range of the matrix A
$\ker(A)$	the kernel of the matrix A
$\mathbf{1}_n$	the $n\times 1$ ones vector
$\mathbf{0}_n$	the $n\times 1$ zeros vector
$A\otimes B$	the Kronecker product of matrices A and B
$\|\cdot\|$	the Euclidean norm
$A>0$ $(A\geq 0)$	matrix $A\in\mathbb{R}^{n\times n}$ is positive definite (positive semi-definite)
$\overline{\mathcal{S}}$	the closure of the subset $\mathcal{S}\subset\mathbb{R}^n$
$\mathrm{int}(\mathcal{S})$	the interior of the subset $\mathcal{S}$
$\dim(\mathcal{S})$	the dimension of the vector space $\mathcal{S}$
$\mathcal{B}_\epsilon(\alpha),\alpha\in\mathbb{R}^n,\epsilon>0,$	the open ball centered at α with radius ϵ
$\mathrm{dist}(p,\mathcal{M})$	the distance from a point p to the set $\mathcal{M}$
$x(t)\to\mathcal{M}$ as $t\to\infty$	$x(t)$ approaches the set $\mathcal{M}$

Let $f:\overline{\mathbb{R}}_+\to\overline{\mathbb{R}}_+$ be a continuous-time function. $f(t)=O(1/t^n)$ denotes that there exists a constant $C>0$ such that $f(t)\leq Ct^{-n}$ for all $t\geq 0$. An undirected graph $\mathcal{G}$ is denoted by $\mathcal{G}(\mathcal{V},\mathcal{E},A)$, where $\mathcal{V}=\{1,\ldots,n\}$ is a set of nodes, $\mathcal{E}\subset\mathcal{V}\times\mathcal{V}$ is a set of edges, and $A=[a_{i.j}]\in\mathbb{R}^{n\times n}$ is an adjacency matrix such that $a_{i.j}=a_{j.i}>0$ if $(j,i)\in\mathcal{E}$ and $a_{i.j}=0$ otherwise. The Laplacian matrix is $L_n=D-A$, where $D\in\mathbb{R}^{n\times n}$

is diagonal with $D_{i,i} = \sum_{j=1}^{n} a_{i,j}$, $i \in \{1,\dots,n\}$. If the graph $\mathcal{G}$ is undirected and connected, then $L_n = L_n^\top \geq 0$, $\operatorname{rank}(L_n) = n-1$ and $\ker(L_n) = \{k\mathbf{1}_n : k \in \mathbb{R}\}$.

2.2 Differential Inclusion

Consider a differential inclusion [28] in the form of

$$\dot{x}(t) \in \mathcal{H}(t, x(t)), \quad x(0) = x_0, \quad t \geq 0, \qquad (1)$$

where the set-valued map $\mathcal{H} : \mathbb{R}_+ \times \mathbb{R}^q \to \mathfrak{B}(\mathbb{R}^q)$ is upper semi-continuous, compact, and the images of $\mathcal{H}(t,x)$ are convex.

Let $\tau > 0$. A solution of (1) defined on $[t_0, \tau] \subset [t_0, \infty)$ is an absolutely continuous function $x : [t_0, \tau] \to \mathbb{R}^q$ such that holds for almost all $t \in [t_0, \tau]$ (in the sense of Lebesgue measure).

Recall that the solution $t \mapsto x(t)$ to is a right maximal solution if it cannot be extended forward in time. We assume that all right maximal solutions to exist on $[t_0, \infty)$.

Let $V : [t_0, \infty) \times \mathbb{R}^q \to \mathbb{R}$ be a locally Lipschitz continuous function such that $V(t,\cdot)$ is convex for any $t \geq t_0$ and $V(\cdot, x)$ is differentiable for any $x \in \mathbb{R}^q$. Let $\partial_x V(t,x)$ be the subdifferential of $V(t,x)$ with respect to x. The set-valued derivative $\mathcal{L}_{\mathcal{H}} V : \mathbb{R} \times \mathbb{R}^q \to \mathcal{P}(\mathbb{R})$ of V with respect to is defined as

$$\mathcal{L}_{\mathcal{H}} V(t,x) = \Big\{ a \in \mathbb{R} : \text{there exists } v \in \mathcal{H}(t,x) \text{ such that } \nabla_t V(t,x) + p^{\mathrm{T}} v = a \text{ for all } p \in \partial_x V(t,x) \Big\}.$$

Recall that convex functions are regular. We have $\dot{V}(t,x) \in \mathcal{L}_{\mathcal{H}} V(t,x)$ almost surely.

Consider a dynamical system given by

$$\ddot{u}(t) + \partial F(u(t)) \ni h(t, u(t), \dot{u}(t)), \quad u(t_0) = u_0, \quad \dot{u}(t_0) = \dot{u}_0, \quad (2)$$

where $t \geq t_0$, $u \in \mathbb{R}^d$ is the state of the system, function $h : \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz in its last two arguments with respect to the first one.

Definition 2.1:

A function $u : [t_0, T] \to \mathbb{R}^d$ is a solution of with initial condition if

- u is Lipschitz continuous, with values in $\operatorname{dom}(F)$,
- $\dot{u}$ is an absolutely continuous function,
- function holds almost everywhere in $[t_0, T]$.

Here we provide a result (a special case of Theorem 3.1 of [29]) of the existence of solutions to system (2).

Assumption 2.1:

(H1) function h is a continuous function from $[t_0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ to $\mathbb{R}^d$ and is Lipschitz continuous in its last two arguments uniformly with respect to the first one;

(H2) function F is a convex function from $\mathbb{R}^d$ to $\mathbb{R} \cup \{+\infty\}$, lower bounded, non-identically equal to $+\infty$ and lower semicontinuous.

Lemma 2.1: Let Assumption 2.1 hold. For any initial condition $(u_0, \dot{u}_0)$, system has a solution in the sense of Definition 2.1.

3. Two-network Bilinear Zero-sum Game

3.1 Problem Description

The two-network bilinear zero-sum game is conducted by two networks $\mathcal{G}_1$ and $\mathcal{G}_2$, which are composed of n_1 and n_2 agents, respectively. In this game, strategy variables, feasibility set, and payoff functions are defined as follows.

Strategy variables of $\mathcal{G}_1$ ($\mathcal{G}_2$) are defined as $\mathbf{x} = [x_1^\top, \cdots, (x_{n_1})^\top]^\top \in \mathbb{R}^{pn_1}$ ($\mathbf{y} = [y_1^\top, \cdots, (y_{n_2})^\top]^\top \in \mathbb{R}^{qn_2}$). The local variable of agent i in $\mathcal{G}_1$ ($\mathcal{G}_2$) is $x_i \in \mathbb{R}^p$ ($y_i \in \mathbb{R}^q$).

Feasibility sets of $\mathcal{G}_1$ and $\mathcal{G}_2$ are given by

$$\Omega_1 = \{\mathbf{x} \in \mathbb{R}^{pn_1} : x_i = x_j, \forall i, j \in \{1,\dots,n_1\}\}$$
$$\Omega_2 = \{\mathbf{y} \in \mathbb{R}^{qn_2} : y_i = y_j, \forall i, j \in \{1,\dots,n_2\}\}.$$

The **payoff function** of the two-network zero-sum game is $U : \mathbb{R}^{pn_1} \times \mathbb{R}^{qn_2} \to \mathbb{R}$ given by

$$U(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}) + \mathbf{y}^\top H(\mathbf{x} - a) - g(\mathbf{y}), \qquad (3)$$

with

$$\begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{n_1} f_i(x_i), \\ g(\mathbf{y}) &= \sum_{i=1}^{n_2} g_i(y_i), \qquad (4) \\ \mathbf{y}^\top H(\mathbf{x} - a) &= \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} y_j^\top H_{i,j}(x_i - a_i), \end{aligned}$$

where $H_{i,j} \in \mathbb{R}^{q \times p}$ is a zero matrix if agent i of $\mathcal{G}_1$ and agent j of $\mathcal{G}_2$ cannot observe each other's choice; agent $i \in \{1, \cdots, n_1\}$ of $\mathcal{G}_1$ knows the information of $f_i(\cdot)$, x_i, $H_{i,j}$, a_i, and y_j; agent $j \in \{1, \cdots, n_2\}$ knows the information of $g_j(\cdot)$, $H_{i,j}$, and y_j; define $\mathcal{N}_0$ as the set of connectivity between two graphs such that $(i,j) \in \mathcal{N}_0$ if agent i of $\mathcal{G}_1$ and agent j of $\mathcal{G}_2$ can observe each other's choice; define L_1 as the Laplacian matrix of $\mathcal{G}_1$ and $\mathcal{G}_2$

The two-network bilinear zero-sum game can be reformulated as

$$\begin{aligned} &\min_{\mathbf{x} \in \mathbb{R}^{pn_1}} \max_{\mathbf{y} \in \mathbb{R}^{qn_2}} \quad U(\mathbf{x}, \mathbf{y}) \\ &\text{s.t.} \qquad\qquad \mathbf{L}_1 \mathbf{x} = 0_{pn_1}, \quad \mathbf{L}_2 \mathbf{y} = 0_{qn_2}, \end{aligned} \qquad (5)$$

where $\mathbf{L}_1 = L_1 \otimes I_p$, $\mathbf{L}_2 = L_2 \otimes I_a$, and L_1 and L_2 are the Laplacian matrices of $\mathcal{G}_1$ and $\mathcal{G}_2$.

Assumption 3.1: Graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ are connected and undirected.

Lemma 3.1: Point $(\mathbf{x}^*, \mathbf{y}^*) \in \mathbb{R}^{pn_1} \times \mathbb{R}^{qn_2}$ is a Nash equilibrium of problem (5) if and only if there exists $\lambda^* \in \mathbb{R}^{pn_1}$ and $\mu^* \in \mathbb{R}^{qn_2}$ such that

$$\begin{aligned} &\mathbf{L}_1\mathbf{x}^* = && 0_{pn_1}, \\ &\mathbf{L}_2\mathbf{y}^* = && 0_{qn_2}, \\ &\partial f(\mathbf{x}^*) + H^\top \mathbf{y}^* + \mathbf{L}_1\lambda^* \ni && 0_{pn_1}, \\ &\partial g(\mathbf{y}^*) - H(\mathbf{x}^* - a) + \mathbf{L}_2\mu^* \ni && 0_{qn_2}. \end{aligned} \tag{6}$$

Define the Lagrangian function as

$$S(\mathbf{x}, \mu, \mathbf{y}, \lambda) = U(\mathbf{x}, \mathbf{y}) + \lambda^\top \mathbf{L}_1\mathbf{x} - \mu^\top \mathbf{L}_2\mathbf{y} + \frac{1}{2}\mathbf{x}^\top \mathbf{L}_1\mathbf{x} - \frac{1}{2}\mathbf{y}^\top \mathbf{L}_2\mathbf{y}.$$

Lemma 3.2: Point $(\mathbf{x}^*, \mu^*, \mathbf{y}^*, \lambda^*) \in \mathbb{R}^{pn_1} \times \mathbb{R}^{qn_2} \times \mathbb{R}^{qn_2} \times \mathbb{R}^{pn_1}$ satisfies if and only if

$$S(\mathbf{x}, \mu, \mathbf{y}^*, \lambda^*) \geq S(\mathbf{x}^*, \mu^*, \mathbf{y}^*, \lambda^*) \geq S(\mathbf{x}^*, \mu^*, \mathbf{y}, \lambda). \tag{7}$$

3.2 Distributed Algorithm and Convergence

For ease of notion, we define $\mathbf{z} = [\mathbf{x}^\mathrm{T}, \mu^\mathrm{T}, \mathbf{y}^\mathrm{T}, \lambda^\mathrm{T}]^\mathrm{T} \in \mathbb{R}^{2pn_1+2qn_2}$. Let $D_1 = \mathrm{diag}[\alpha_{1,1}, \ldots, \alpha_{1,n_1}] \in \mathbb{R}^{n_1 \times n_1}$, $D_2 = \mathrm{diag}[\alpha_{2,1}, \ldots, \alpha_{2,n_2}] \in \mathbb{R}^{n_2 \times n_2}$, and define $\mathbf{D}_1 = D_1 \otimes I_p$, $\mathbf{D}_2 = D_2 \otimes I_a$.

Then we propose a dynamics which solves . In particular, the algorithm for $\mathcal{G}_1$ and $\mathcal{G}_2$ are proposed as

$$\begin{aligned} \ddot{x}_i(t) \in\ & -\frac{\alpha_{1,i}}{t}\dot{x}_i(t) - \partial f_i(x_i(t)) - \sum_{k=1}^{n_1} a_{i\,k}(x_i(t) - x_k(t)) \\ & - \sum_{k=1}^{n_2} H_{i,k}^\top \left(y_k(t) + \frac{t}{2}\dot{y}_k(t)\right) \\ & - \sum_{k=1}^{n_1} a_{i,k}\left(\lambda_i(t) + \frac{t}{2}\dot{\lambda}_i(t) - \lambda_j(t) - \frac{t}{2}\dot{\lambda}_j(t)\right), \\ \ddot{\lambda}_i(t) =\ & -\frac{\alpha_{1,i}}{t}\dot{\lambda}_i(t) + \sum_{k=1}^{n_1} a_{i,k}\left(x_i(t) + \frac{t}{2}\dot{x}_i(t) - x_k(t) - \frac{t}{2}\dot{x}_k(t)\right), \\ \ddot{y}_j(t) \in\ & -\frac{\alpha_{2,j}}{t}\dot{y}_j(t) - \partial g_i(y_i(t)) - \sum_{k=1}^{n_2}(y_j(t) - y_k(t)) \\ & + \sum_{k=1}^{n_1} H_{k,j}(x_k(t) + \frac{t}{2}\dot{x}_k(t) - a_k) \\ & - \sum_{k=1}^{n_2} a_{j,k}\left(\mu_j(t) + \frac{t}{2}\dot{\mu}_j(t) - \mu_k(t) + \frac{t}{2}\dot{\mu}_k(t)\right), \end{aligned}$$

where $t \geq t_0$, $i \in \{1, \ldots, n_1\}$ and $j \in \{1, \ldots, n_2\}$.

The compact form of the algorithm is

$$\begin{aligned} \ddot{\mathbf{x}}(t) \in\ & -\frac{1}{t}\mathbf{D}_1\dot{\mathbf{x}}(t) - \partial f(\mathbf{x}(t)) - \mathbf{L}_1\mathbf{x}(t) - H^\top\left(\mathbf{y}(t) + \frac{t}{2}\dot{\mathbf{y}}(t)\right) \\ & - \mathbf{L}_1(\lambda(t) + \frac{t}{2}\dot{\lambda}(t)), \\ \ddot{\lambda}(t) =\ & -\frac{1}{t}\mathbf{D}_1\dot{\lambda}(t) + \mathbf{L}_1\left(\mathbf{x}(t) + \frac{t}{2}\dot{\mathbf{x}}(t)\right), \end{aligned} \tag{8}$$

$$\begin{aligned} \ddot{\mathbf{y}}(t) \in\ & -\frac{1}{t}\mathbf{D}_2\dot{\mathbf{y}}(t) - \mathbf{L}_2\mathbf{y}(t) - \partial g(\mathbf{y}(t)) + H(\mathbf{x}(t) + \frac{t}{2}\dot{\mathbf{x}}(t) - a) \\ & - \mathbf{L}_2\left(\mu(t) + \frac{t}{2}\dot{\mu}(t)\right), \\ \ddot{\mu}(t) =\ & -\frac{1}{t}\mathbf{D}_2\dot{\mu}(t) + \mathbf{L}_2\left(\mathbf{y}(t) + \frac{t}{2}\dot{\mathbf{y}}(t)\right), \end{aligned}$$

Define the gap function as

$$\begin{aligned} G(\mathbf{x}, \mathbf{y}) =\ & S(\mathbf{x}, \mu, \mathbf{y}^*, \lambda^*) - S(\mathbf{x}^*, \mu^*, \mathbf{y}, \lambda) \\ =\ & f(\mathbf{x}) - f(\mathbf{x}^*) + g(\mathbf{y}) - g(\mathbf{y}^*) \\ & + \mathbf{y}^{*\top} H(\mathbf{x} - \mathbf{x}^*) - (\mathbf{y} - \mathbf{y}^*)^\top H(\mathbf{x}^* - a) \\ & + \lambda^{*\top}\mathbf{L}_1\mathbf{x} + \mu^{*\top}\mathbf{L}_2 y + \frac{1}{2}\mathbf{x}^\top\mathbf{L}_1\mathbf{x} + \frac{1}{2}\mathbf{y}^\top\mathbf{L}_2\mathbf{y} \end{aligned}$$

Define function

$$\begin{aligned} V(t, \mathbf{z}, \dot{\mathbf{z}}) =\ & V_1(t, \mathbf{z}, \dot{\mathbf{z}}) + V_2(t, \mathbf{x}, \dot{\mathbf{x}}) + V_3(t, \mu, \dot{\mu}) \\ & + V_4(t, \mathbf{y}, \dot{\mathbf{y}}) + V_5(t, \lambda, \dot{\lambda}) \end{aligned} \tag{9}$$

such that

$$\begin{aligned} V_1(t, \mathbf{x}, \mathbf{y}) =\ & \frac{1}{2}t^2 G(\mathbf{x}, \mathbf{y}), \\ V_2(t, \mathbf{x}, \dot{\mathbf{x}}) =\ & \| \mathbf{x} + \frac{t}{2}\dot{\mathbf{x}} - \mathbf{x}^*\|^2 + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top(\mathbf{D}_1 - 3I_{pn_1})(\mathbf{x} - \mathbf{x}^*), \\ V_3(t, \mu, \dot{\mu}) =\ & \| \mu + \frac{t}{2}\dot{\mu} - \mu^*\|^2 + \frac{1}{2}(\mu - \mu^*)^\top(\mathbf{D}_2 - 3I_{qn_2})(\mu - \mu^*) \\ V_4(t, \mathbf{y}, \dot{\mathbf{y}}) =\ & \| \mathbf{y} + \frac{t}{2}\dot{\mathbf{y}} - \mathbf{y}^*\|^2 + \frac{1}{2}(\mathbf{y} - \mathbf{y}^*)^\top(\mathbf{D}_2 - 3I_{qn_2})(\mathbf{y} - \mathbf{y}^*), \\ V_5(t, \lambda, \dot{\lambda}) =\ & \| \lambda + \frac{t}{2}\dot{\lambda} - \lambda^*\|^2 + \frac{1}{2}(\lambda - \lambda^*)^\top(\mathbf{D}_1 - 3I_{pn_1})(\lambda - \lambda^*) \end{aligned}$$

where $L(\cdot)$ is the Lagrangian function and (x^*, λ^*) is a saddle point of $L(\cdot)$. Clear, function V is positive definite with respect to $(x, \lambda, t\dot{x}, t\dot{\lambda})$ for all $t \geq 0$.

Theorem 3.1:

Let $\mathbf{z}(t)$ be a trajectory of algorithm .

(i) The trajectory of $(\mathbf{z}(t), t\dot{\mathbf{z}}(t))$ is bounded for $t \geq t_0$.

(ii) The trajectory $(\mathbf{x}(t), \mathbf{y}(t))$ satisfies the convergence properties $G(\mathbf{x}(t), \mathbf{y}(t)) = O(1/t^2)$.

4. Theoretical Analysis

4.1 Convergence Proof

The proof of the convergence performance in Theorem 3.1 is given as following.

(i) It follows from algorithm (8) that there exist $f_{\mathbf{x}} \in \partial f(\mathrm{x})$ and $g_{\mathbf{v}} \in \partial g(\mathrm{y})$ such that

$$\begin{aligned} \ddot{\mathbf{x}} =\ & -\frac{1}{t}\mathbf{D}_1\dot{\mathbf{x}} - f_{\mathbf{x}} - H^\top(\mathbf{y} + \frac{t}{2}\dot{\mathbf{y}} - b) - \mathbf{L}_1\left(\lambda + \frac{t}{2}\dot{\lambda}\right) - \mathbf{L}_1\mathbf{x}, \\ \ddot{\mathbf{y}} =\ & -\frac{1}{t}\mathbf{D}_2\dot{\mathbf{y}} - g_{\mathrm{y}} + H(\mathbf{x} + \frac{t}{2}\dot{\mathbf{x}} - a) - \mathbf{L}_2\left(\mu + \frac{t}{2}\dot{\mu}\right) - \mathbf{L}_2\mathbf{y}. \end{aligned}$$

Then the derivative of V_i 's along the trajectory of algorithm **(8)** satisfies that

$$\begin{aligned}\dot{V}_1(t,\mathbf{x},\mathbf{y}) &= t[S(\mathbf{x},\mu,\mathbf{y}^*,\lambda^*)-S(\mathbf{x}^*,\mu^*,\mathbf{y},\lambda)]\\ &+\frac{1}{2}t^2[f_{\mathbf{x}}+H^\top\mathbf{y}^*+L_1\lambda^*+L_1x]^\top\dot{\mathbf{x}}\\ &+\frac{1}{2}t^2[g_{\mathbf{y}}-H(\mathbf{x}^*-a)+\mathbf{L}_2\mu^*+L_2\mathbf{y}]^\top\dot{\mathbf{y}}.\end{aligned}$$

We have

$$\begin{aligned}\dot{V}_1(t,\mathbf{x},\mathbf{y}) &= t[S(\mathbf{x},\mu,\mathbf{y}^*,\lambda^*)-S(\mathbf{x}^*,\mu^*,\mathbf{y},\lambda)]\\ &+\frac{1}{2}t^2[f_{\mathbf{x}}-f_{\mathbf{x}^*}+\mathbf{L}_1\mathbf{x}]^\top\dot{\mathbf{x}}+\frac{1}{2}t^2[g_{\mathbf{y}}-g_{\mathbf{y}^*}+\mathbf{L}_2\mathbf{y}]^\top\dot{\mathbf{y}}.\end{aligned}\quad(10)$$

Similarly,

$$\begin{aligned}\dot{V}_2(t,\mathbf{x},\dot{\mathbf{x}}) &= (\mathbf{x}+0.5t\dot{\mathbf{x}}-\mathbf{x}^*)^\top(3\dot{\mathbf{x}}+t\ddot{\mathbf{x}})+(\mathbf{x}-\mathbf{x}^*)^\top(\mathbf{D}_1-3I_{pn_1})\dot{\mathbf{x}}\\ &= (\mathbf{x}-\mathbf{x}^*)^\top(\mathbf{D}_1\dot{\mathbf{x}}+t\ddot{\mathbf{x}})+1.5t\|\dot{\mathbf{x}}\|^2+0.5t^2\dot{\mathbf{x}}^\top\ddot{\mathbf{x}}\\ &= -t(\mathbf{x}-\mathbf{x}^*)^\top\left(f_{\mathbf{x}}+H^\top\mathbf{y}+\mathbf{L}_1\lambda+\mathbf{L}_1\mathbf{x}\right)-\frac{t^2}{2}(\mathbf{x}-\mathbf{x}^*)^\top(H^\top\dot{\mathbf{y}}+\mathbf{L}_1\dot{\lambda})\\ &\quad -0.5t\dot{\mathbf{x}}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\mathbf{x}}-\frac{1}{2}t^2\dot{\mathbf{x}}^\top\left(f_{\mathbf{x}}+H^\top\mathbf{y}+\mathbf{L}_1\lambda+\mathbf{L}_1\mathbf{x}\right)\\ &\quad -0.25t^3\dot{\mathbf{x}}^\top(H^\top\dot{\mathbf{y}}+\mathbf{L}_1\dot{\lambda})\\ &= -t(\mathbf{x}-\mathbf{x}^*)^\top(f_{\mathbf{x}}-f_{\mathbf{x}^*})-t(\mathbf{x}-\mathbf{x}^*)^\top H^\top(\mathbf{y}-\mathbf{y}^*)-t\mathbf{x}^\top\mathbf{L}_1(\lambda-\lambda^*)\\ &\quad -t\mathbf{x}^\top\mathbf{L}_1\mathbf{x}-\frac{t^2}{2}(\mathbf{x}-\mathbf{x}^*)^\top(H^\top\dot{\mathbf{y}}+\mathbf{L}_1\dot{\lambda})\\ &\quad -0.5t\dot{\mathbf{x}}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\mathbf{x}}-0.25t^3\dot{\mathbf{x}}^\top(H^\top\dot{\mathbf{y}}+\mathbf{L}_1\dot{\lambda})-\frac{1}{2}t^2\dot{\mathbf{x}}^\top\mathbf{L}_1\mathbf{x}\\ &\quad -\frac{1}{2}t^2\dot{\mathbf{x}}^\top(f_{\mathbf{x}}-f_{\mathbf{x}^*})-\frac{1}{2}t^2\dot{\mathbf{x}}^\top H^\top(\mathbf{y}-\mathbf{y}^*)-\frac{1}{2}t^2\dot{\mathbf{x}}^\top\mathbf{L}_1(\lambda-\lambda^*),\end{aligned}\quad(11)$$

$$\begin{aligned}\dot{V}_3(t,\mu,\dot{\mu}) &= (\mu+0.5t\dot{\mu}-\mu^*)^\top(3\dot{\mu}+t\ddot{\mu})+(\mu-\mu^*)^\top(D_2-3I)\dot{\mu}\\ &= t(\mu-\mu^*)^\top\mathbf{L}_2\mathbf{y}+\frac{1}{2}t^2(\mu-\mu^*)^\top\mathbf{L}_2\dot{\mathbf{y}}\\ &\quad +0.5t\dot{\mu}^\top(3I_{qn_2}-\mathbf{D}_2)\dot{\mu}+0.5t^2\dot{\mu}^\top\mathbf{L}_2\mathbf{y}+\frac{1}{4}t^3\dot{\mu}^\top\mathbf{L}_2\dot{\mathbf{y}},\end{aligned}\quad(12)$$

$$\begin{aligned}\dot{V}_4(t,\mathbf{y},\dot{\mathbf{y}}) &= (\mathbf{y}+0.5t\dot{\mathbf{y}}-\mathbf{y}^*)^\top(3\dot{\mathbf{y}}+t\ddot{\mathbf{y}})+(\mathbf{y}-\mathbf{y}^*)^\top(\mathbf{D}_1-3I_{pn_1})\dot{\mathbf{y}}\\ &= (\mathbf{y}-\mathbf{y}^*)^\top\left(-tg_{\mathbf{y}}+tH(\mathbf{x}-a)-t\mathbf{L}_2\mu-t\mathbf{L}_2\mathbf{y}\right)+\frac{t^2}{2}(\mathbf{y}-\mathbf{y}^*)^\top(H\dot{\mathbf{x}}-\mathbf{L}_2\dot{\mu})\\ &\quad -0.5t\dot{\mathbf{y}}^\top(\mathbf{D}_2-3I_{qn_2})\dot{\mathbf{y}}+0.25t^3\dot{\mathbf{y}}^\top(H\dot{\mathbf{x}}-\mathbf{L}_2\dot{\mu})\\ &\quad +0.5t^2\dot{\mathbf{y}}^\top\left(-g_{\mathbf{y}}+H(\mathbf{x}-a)-\mathbf{L}_2\mu-\mathbf{L}_2\mathbf{y}\right)\\ &= -t(\mathbf{y}-\mathbf{y}^*)^\top(g_{\mathbf{y}}-g_{\mathbf{y}^*})+t(\mathbf{y}-\mathbf{y}^*)^\top H(\mathbf{x}-\mathbf{x}^*)-t\mathbf{y}^\top\mathbf{L}_2(\mu-\mu^*)-t\mathbf{y}^\top\mathbf{L}_2\mathbf{y}\\ &\quad +\frac{t^2}{2}(\mathbf{y}-\mathbf{y}^*)^\top(H\dot{\mathbf{x}}-\mathbf{L}_2\dot{\mu})-0.5t\dot{\mathbf{y}}^\top(\mathbf{D}_2-3I)\dot{\mathbf{y}}\\ &\quad -0.5t^2\dot{\mathbf{y}}^\top(g_{\mathbf{y}}-g_{\mathbf{y}^*})+0.5t^2\dot{\mathbf{y}}^\top H(\mathbf{x}-\mathbf{x}^*)-0.5t^2\dot{\mathbf{y}}^\top\mathbf{L}_2(\mu-\mu^*)\\ &\quad -0.5t^2\dot{\mathbf{y}}^\top\mathbf{L}_2\mathbf{y}+0.25t^3\dot{\mathbf{y}}^\top(H\dot{\mathbf{x}}-\mathbf{L}_2\dot{\mu}),\end{aligned}\quad(13)$$

$$\begin{aligned}\dot{V}_5(t,\lambda,\dot{\lambda}) &= (\lambda+0.5t\dot{\lambda}-\lambda^*)^\top(3\dot{\lambda}+t\ddot{\lambda})+(\lambda-\lambda^*)^\top(\mathbf{D}_1-3I_{pn_1})\dot{\lambda}\\ &= t(\lambda-\lambda^*)^\top\mathbf{L}_1x+\frac{1}{2}t^2(\lambda-\lambda^*)^\top\mathbf{L}_1\dot{\mathbf{x}}\\ &\quad -0.5t\dot{\lambda}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\lambda}+0.5t^2\dot{\lambda}^\top\mathbf{L}_1\mathbf{x}+\frac{1}{4}t^3\dot{\lambda}^\top\mathbf{L}_1\dot{\mathbf{x}}.\end{aligned}\quad(14)$$

By summing up - and simplifying items, we have

$$\begin{aligned}\dot{V}(t,\mathbf{z},\dot{\mathbf{z}}) &= t[S(\mathbf{x},\mu,\mathbf{y}^*,\lambda^*)-S(\mathbf{x}^*,\mu^*,\mathbf{y},\lambda)]\\ &\quad -t(\mathbf{x}-\mathbf{x}^*)^\top(f_{\mathbf{x}}-f_{\mathbf{x}^*})-t\mathbf{x}^\top\mathbf{L}_1\mathbf{x}-0.5t\dot{\mathbf{x}}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\mathbf{x}}\\ &\quad -t(\mathbf{y}-\mathbf{y}^*)^\top(g_{\mathbf{y}}-g_{\mathbf{y}^*})-t\mathbf{y}^\top\mathbf{L}_2\mathbf{y}-0.5t\dot{\mathbf{y}}^\top(\mathbf{D}_2-3I_{qn_2})\dot{\mathbf{y}}\\ &\quad -0.5t\dot{\lambda}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\lambda}-0.5t\dot{\mu}^\top(\mathbf{D}_2-3I_{qn_2})\dot{\mu}\\ &= t\mathbf{A}-0.5t\mathbf{B}\end{aligned}\quad(15)$$

where

$$\begin{aligned}\mathbf{A} &= \big(f(\mathbf{x})-f(\mathbf{x}^*)+g(\mathbf{y})-g(\mathbf{y}^*)+\mathbf{y}^{*\top}H(\mathbf{x}-a)\\ &\quad -\mathbf{y}^\top H(\mathbf{x}^*-a)+\lambda^{*\top}\mathbf{L}_1\mathbf{x}+\mu^{*\top}\mathbf{L}_2\mathbf{y}-\frac{1}{2}\mathbf{x}^\top\mathbf{L}_1\mathbf{x}-\frac{1}{2}\mathbf{y}^\top\mathbf{L}_2\mathbf{y}\\ &\quad -(\mathbf{x}-\mathbf{x}^*)^\top(f_{\mathbf{x}}-f_{\mathbf{x}^*})-(\mathbf{y}-\mathbf{y}^*)^\top(g_{\mathbf{y}}-g_{\mathbf{y}^*})\big)\end{aligned}$$

$$\begin{aligned}\mathbf{B} &= \dot{\mathbf{x}}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\mathbf{x}}+\dot{\mathbf{y}}^\top(\mathbf{D}_2-3I_{qn_2})\dot{\mathbf{y}}\\ &\quad +\dot{\lambda}^\top(\mathbf{D}_1-3I_{pn_1})\dot{\lambda}+\dot{\mu}^\top(\mathbf{D}_2-3I_{qn_2})\dot{\mu}\le 0\end{aligned}$$

By the third and fourth equations in , it follows that

$$\begin{aligned}\mathbf{A} &= f(\mathbf{x})-f(\mathbf{x}^*)-(\mathbf{x}-\mathbf{x}^*)^\top f_{\mathbf{x}}-\frac{1}{2}\mathbf{x}^\top\mathbf{L}_1\mathbf{x}\\ &\quad +g(\mathbf{y})-g(\mathbf{y}^*)-(\mathbf{y}-\mathbf{y}^*)^\top g_{\mathbf{y}}-\frac{1}{2}\mathbf{y}^\top\mathbf{L}_2\mathbf{y}\\ &\le 0\end{aligned}$$

Because f and g are convex, it is clear that $f(\mathbf{x})-f(\mathbf{x}^*)-(\mathbf{x}-\mathbf{x}^*)^\top f_{\mathbf{x}}\le 0$ and $g(\mathbf{y})-g(\mathbf{y}^*)-(\mathbf{y}-\mathbf{y}^*)^\top g_{\mathbf{y}}\le 0$. Note that $\mathbf{L}_1\ge 0$ and $\mathbf{L}_2\ge 0$. It follows from (15) that

$$\dot{V}(t,\mathbf{z},\dot{\mathbf{z}})\le 0.$$

Recall that function V is positive definite with respect to $(\mathbf{z},t\dot{\mathbf{z}})$ for all $t\ge 0$. The trajectory of $(\mathbf{z},t\dot{\mathbf{z}})$ is bounded for $t\ge 0$.

(ii) Since $\dot{V}(t,\mathbf{z}(t),\dot{\mathbf{z}}(t))\le 0$ then

$$V(t,\mathbf{z}(t),\dot{\mathbf{z}}(t))\le m_0\triangleq V(0,\mathbf{z}(0),\dot{\mathbf{z}}(0)).$$

It follows from that

$$\frac{1}{2}G(\mathbf{x}(t),\mathbf{y}(t))\le\frac{1}{t^2}m_0.$$

Hence, $G(\mathbf{x}(t),\mathbf{y}(t))=O(\frac{1}{t^2})$, $\mathbf{x}^\top\mathbf{L}_1\mathbf{x}=O(\frac{1}{t^2})$, and $\mathbf{y}^\top\mathbf{L}_2\mathbf{y}=O(\frac{1}{t^2})$. One can prove $\|\dot{\mathbf{x}}(t)\|=O(\frac{1}{t})$ and $\|\dot{\lambda}(t)\|=O(\frac{1}{t})$ using similar arguments. ■

4.2 Comparison with Existing Results

In this subsection, we compare the rate of convergence with the algorithm proposed in [26]. Specifically, the design in [26] for this problem is

$$
\begin{aligned}
\dot{\mathbf{x}} \in & \quad -\partial_{\mathbf{x}} S(\mathbf{x},\mu,\mathbf{y},\lambda), \quad \mathbf{x}(0)=\mathbf{x}_0, \\
\dot{\lambda} = & \quad \nabla_{\lambda} S(\mathbf{x},\mu,\mathbf{y},\lambda), \quad \lambda(0)=\lambda_0, \\
\dot{\mathbf{y}} \in & \quad -\partial_{\mathbf{y}}\left[-S(\mathbf{x},\mu,\mathbf{y},\lambda)\right], \quad \mathbf{y}(0)=\mathbf{y}_0, \\
\dot{\mu}(t) = & \quad -\nabla_{\mu} S(\mathbf{x},\mu,\mathbf{y},\lambda), \quad \mu(0)=\mu_0,
\end{aligned} \tag{16}
$$

The convergence and boundedness of algorithm are proved in [26]. We further show the rate of convergence of algorithm , which is not obtained in [26]. Define the ergodic trajectory as $\hat{\mathbf{x}}(t)=\frac{1}{t}\int_0^t \mathbf{x}(s)\mathrm{d}s$, $\hat{\mathbf{y}}(t)=\frac{1}{t}\int_0^t \mathbf{y}(s)\mathrm{d}s$, $\hat{\lambda}(t)=\frac{1}{t}\int_0^t \lambda(s)\mathrm{d}s$, and $\hat{\mu}(t)=\frac{1}{t}\int_0^t \mu(s)\mathrm{d}s$.

Lemma 4.1: Let $(\mathbf{x}(t),\mathbf{y}(t),\lambda(t),\mu(t))$ be a trajectory of algorithm (16). The ergodic trajectory $(\hat{\mathbf{x}}(t),\hat{\mathbf{y}}(t),\hat{\lambda}(t),\hat{\mu}(t))$ satisfies the convergence properties $G(\hat{\mathbf{x}}(t),\hat{\mathbf{y}}(t))=O(1/t)$.

PROOF: Define function

$$V(\mathbf{x}(t),\mathbf{y}(t),\lambda(t),\mu(t))=\frac{1}{2}\|\mathbf{x}-\mathbf{x}^*\|^2+\frac{1}{2}\|\mathbf{y}-\mathbf{y}^*\|^2+\frac{1}{2}\|\lambda-\lambda^*\|^2+\frac{1}{2}\|\mu-\mu^*\|^2$$

where $(\mathbf{x}^*,\mathbf{y}^*,\lambda^*,\mu^*)$ is an equilibrium of algorithm (16). The derivative of V is

$$\dot{V}(\mathbf{x},\mathbf{y},\lambda,\mu)=(\mathbf{x}-\mathbf{x}^*)^\top\dot{\mathbf{x}}+(\mathbf{y}-\mathbf{y}^*)^\top\dot{\mathbf{y}}+(\lambda-\lambda^*)^\top\dot{\lambda}+(\mu-\mu^*)^\top\dot{\mu}.$$

Note that S is convex (concave) with respect $\mathbf{x}$ and μ ($\mathbf{y}$ and λ). It follows from algorithm that

$$(\mathbf{x}-\mathbf{x}^*)^\top\dot{\mathbf{x}}+(\mu-\mu^*)^\top\dot{\mu}\le S(\mathbf{x}^*,\mu^*,\mathbf{y},\lambda)-S(\mathbf{x},\mu,\mathbf{y},\lambda)$$

$$(\mathbf{y}-\mathbf{y}^*)^\top\dot{\mathbf{y}}+(\lambda-\lambda^*)^\top\dot{\lambda}\le S(\mathbf{x},\mu,\mathbf{y},\lambda)-S(\mathbf{x},\mu,\mathbf{y}^*,\lambda^*)$$

Hence,

$$\dot{V}(\mathbf{x},\mathbf{y},\lambda,\mu)\le S(\mathbf{x}^*,\mu^*,\mathbf{y},\lambda)-S(\mathbf{x},\mu,\mathbf{y}^*,\lambda^*)=-G(\mathbf{x},\mathbf{y})\le 0$$

Hence

$$V(\mathbf{x}(t),\mathbf{y}(t),\lambda(t),\mu(t))-V(\mathbf{x}_0,\mathbf{y}_0,\lambda_0,\mu_0)\le\int_0^t G(\mathbf{x}(s),\mathbf{y}(s))\mathrm{d}s\le($$

It follows from Jensen's inequality for the convex-concave function S that

$$S(\mathbf{x}^*,\mu^*,\hat{\mathbf{y}}(t),\hat{\lambda}(t))\ge\frac{1}{t}\int_0^t S(\mathbf{x}^*,\mu^*,\mathbf{y}(s),\lambda(s))\mathrm{d}s$$

$$S(\hat{\mathbf{x}}(t),\hat{\mu}(t),\mathbf{y}^*,\lambda^*)\le\frac{1}{t}\int_0^t S(\mathbf{x}(s),\mu(s),\mathbf{y}^*,\lambda^*)\mathrm{d}s$$

Hence,

$$G(\hat{\mathbf{x}}(t),\hat{\mathbf{y}}(t))\le\frac{1}{t}\int_0^t G(\mathbf{x}(s),\mathbf{y}(s))\mathrm{d}s\le\frac{1}{t}V(\mathbf{x}_0,\mathbf{y}_0,\lambda_0,\mu_0).$$ ■

5. Numerical Simulation

Consider the problem of two-network zero-sum game . We take $p=q=2$, $n_1=n_2=25$, $H=I_{100}$, and $a=0_{100}$.

(1) Convex function

$$f_i(x_i)=\frac{2}{3}\log(1+(x_{i,1}+i)^2)+\frac{1}{3}\log(1+(x_{i,2}-i)^2)$$

$$g_i(y_i)=\log((y_{i,1}-iy_{i,2})^2+1/2).$$

(2) Quadratic convex functions $f_i(x_i)=(x_{i,1}-x_{i,2}+i)^2$ and $g_i(y_i)=(y_{i,1}+y_{i,2}+i)^2$.

For case (1), Fig. 1 gives trajectories of $S(\cdot)$ (defined in (7)) along algorithm the algorithm with $\alpha=2,4$, and shows that $\alpha=4$ gives a better performance. The theoretical proof of step size α to the convergence result is not discussed due to space limitation. Fig. 2 compares algorithm and primal-dual algorithm in [26] for case 1. Fig. 2 indicates that the proposed algorithm has a better performance than the primal-dual method in [26] for case 1. Fig. 3 shows that, for case 2 whose cost functions are quadratic, the primal-dual method in [26] has a better convergence performance than the proposed algorithm . The reason for this is that the primal-dual method in [26] is an affine algorithm whose convergence rate is linear.

6. Conclusion

This paper has focused on designing an accelerated first-order algorithm for the distributed Nash equilibrium seeking of a class of two-network zero-sum problems, a class of convex-concave bilinear saddle point problems. By using the derivative information, this paper has developed a continuous-time algorithm having $O(1/t^2)$ convergence by choosing proper parameters. The paper has proved the correctness and convergence of the algorithm based on a Lyapunov approach.

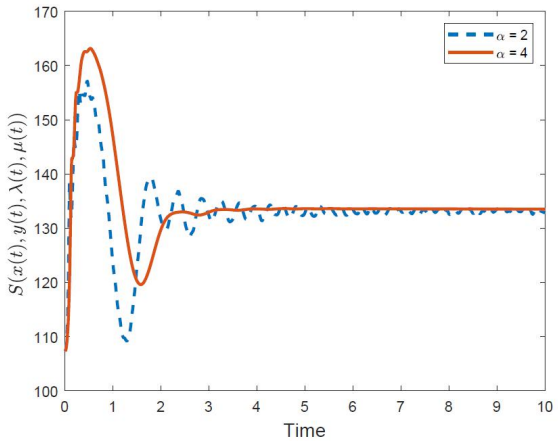


Fig. 1. Trajectories of S(·) along algorithm **(8)** with α = 2; 4 for case 1

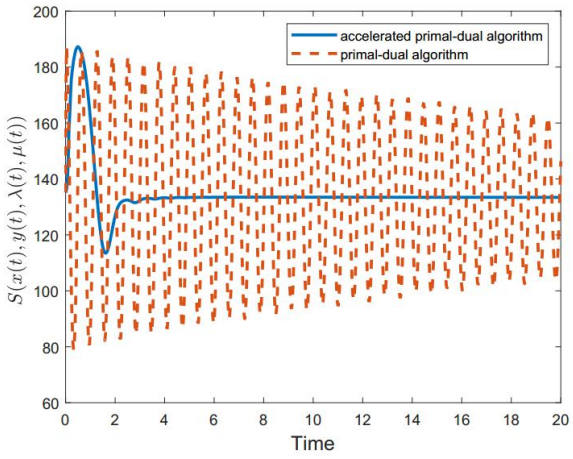


Fig. 2. Trajectories of S(·) along algorithm and primal-dual algorithm in [9] for case 1

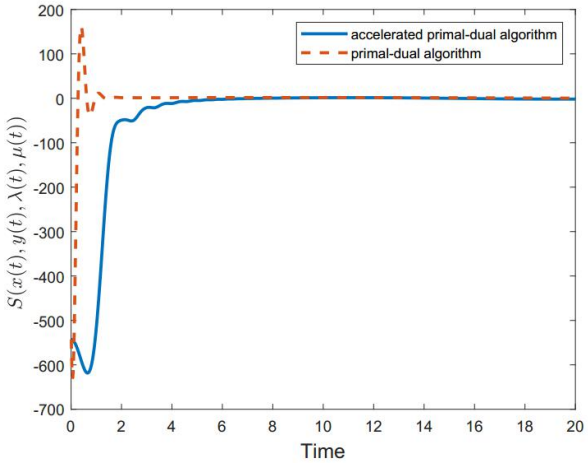


Fig. 3. Trajectories of S(·) along algorithm and primal-dual algorithm in [26] for case 2

References

[1] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 21(1): 183-202, 2009.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3: 1-122, 2011.

[3] T. Ibaraki and N. Katoh. *Resource allocation problems: algorithmic approaches*. MIT Press, Cambridge, MA, United States, 1988.

[4] D. P. Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, New York, United States, 2014.

[5] Y. Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady*, 27: 372-376, 1983.

[6] A. Nemirovskii and D. Yudin. *Problem Complexity and Method Efficiency in Optimization*. John Wiley & Sons, 1983.

[7] Y. Nesterov. Smooth minimization of non-smooth functions. *Mathematical Programming*, 103(1): 127-152, 2005.

[8] Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1): 125-161, 2013.

[9] H. Attouch, Z. Chbani, and H. Riahi. Rate of convergence of the nesterov accelerated gradient method in the subcritical case $\alpha \leq 3$. *ESAIM: Control, Optimisation and Calculus of Variations*, 25, 2015.

[10] W. Su, S. Boyd, and E. J. Candes. A differential equation for modeling Nesterov's accelerated gradient method: Theory and insights. *Advances in Neural Information Processing Systems*, 3(1): 2510-2518, 2015.

[11] A. Wibisono, A. C. Wilson, and M. I. Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47): 7351-7358, 2016.

[12] Y. Chen, G. Lan, and Y. Ouyang. Optimal primal-dual methods for a class of saddle point problems. *SIAM Journal on Optimization*, 24(4): 1779-1814, 2014.

[13] Y. Xu. Accelerated first-order primal-dual proximal methods for linearly constrained composite convex programming. *SIAM Journal on Optimization*, 27(3): 1459-1484, 2017.

[14] S. S. Kia, J. Cortes, and S. Martinez. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. *Automatica*, 55: 254-264, 2015.

[15] P. Yi, Y. Hong, and F. Liu. Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. *Systems & Control Letters*, 83: 45-52, 2015.

[16] P. Yi, Y. Hong, and F. Liu. Initialization-free distributed algorithms for optimal resource allocation with feasibility constraints and application to economic dispatch of power systems. *Automatica*, 74: 259-269, 2016.

[17] X. Zeng, P. Yi, Y. Hong, and L. Xie. Distributed continuous-time algorithms for nonsmooth extended monotropic optimization problems. *SIAM J. Control and Optimization*, 56(6): 3973-3993, 2018.

[18] S. Liang, P. Yi, and Y. Hong. Distributed Nash equilibrium seeking for aggregative games with coupled constraints. *Automatica*, 85: 179-185, 2017.

[19] J. S. Shamma and G. Arslan. Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria. *IEEE Transactions on Automatic Control*, 50(3): 312-327, 2005.

[20] M. Ye and G. Hu. Distributed Nash equilibrium seeking by a consensus based approach. *IEEE Transactions on Control of Network Systems*, 62(9): 4811-4818, 2017.

[21] M. Ye, G. Hu, F. Lewis, and L. Xie. Simultaneous Nash equilibrium seeking and social cost minimization in graphical *N*-coalition non-cooperative games. arXiv:1708.02394.

[22] J. Ghaderi and R. Srikant. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica*, 50(2): 3209-3215, 2014.

[23] J. Pang, G. Scutari, F. Facchinei, and C. Wang. Distributed power allocation with rate constraints in Gaussian parallel interference channels. *IEEE Transactions on Information Theory*, 54(8): 3471-3489, 2008.

[24] P. Yi and L. Pavel. Distributed generalized Nash equilibria computation of monotone games via a preconditioned proximal point algorithm. arXiv:1705.01624v2.

[25] M. Zhu and E. Frazzoli. Distributed robust adaptive equilibrium computation for generalized convex games. *Automatica*, 63: 82-91, 2016.

[26] B. Gharesifard and J. Cort´es. Distributed convergence to Nash equilibria in two-network zero-sum games. *Automatica*, 49(6): 1683-1692, 2013.

[27] S. Yang, J. Wang, and Q. Liu. Cooperative-competitive multiagent systems for distributed minimax optimization subject to bounded constraints. *IEEE Trans. Automati. Contr.*, 64(4): 1358-1372, 2019.

[28] J. P. Aubin and A. Cellina. *Differential Inclusions*. Springer-Verlag, Berlin, Germany, 1984.

[29] L. A. Paoli. An existence result for vibrations with unilateral constraints: Case of a nonsmooth set of constraints. *Mathematical Models and Methods in Applied Sciences*, 10(6): 815-831, 2000.