

Content Summary

This 2016 Summer School on Information and Communication Technology Proceedings collects research work from international cooperation between China, German, Mongolian and Russian universities. The research work was presented at the 2016 Summer School on Information and Communication Technology organized 2016 from 11th to 15th July in Harbin, China. This reading on hand contains a selection of papers presented in 2016. Key topics are information and communication Technology from different points of view. One of the main goals of this international event is to give opportunity to students and researchers of partner universities to work together during school time actively and freely in frame of science and international cooperation.

图书在版编目(CIP)数据

2016 信息与通信技术夏令营论文集 = The 2016 Summer School on Information and Communication Technology Proceedings ; 英文 / (德) 沃尔弗拉姆·哈特 (Wolfram Hardt) 等主编. -- 哈尔滨 : 哈尔滨工程大学出版社, 2016. 12

ISBN 978 - 7 - 5661 - 1408 - 2

I. ①2… II. ①沃… III. ①信息技术 - 文集 - 英文
②通信技术 - 文集 - 英文 IV. ①G202 - 53②TN91 - 53

中国版本图书馆 CIP 数据核字(2016)第 298577 号

选题策划 龚晨
责任编辑 张忠远 周一瞳
封面设计 博鑫设计

出版发行 哈尔滨工程大学出版社
社 址 哈尔滨市南岗区东大直街 124 号
邮政编码 150001
发行电话 0451 - 82519328
传 真 0451 - 82519699
经 销 新华书店
印 刷 哈尔滨圣铂印刷有限公司
开 本 787mm × 1 092mm 1/16
印 张 11
字 数 512 千字
版 次 2016 年 12 月第 1 版
印 次 2016 年 12 月第 1 次印刷
定 价 100.00 元

<http://www.hrbeupress.com>

E-mail:heupress@hrbeu.edu.cn

Preface

The International Summer School on Information and Communication Technologies look back now to 5 years' history. Initiated by foundation IBS researcher and students have joined successful the exchange and cooperation format. Beside scientific presentations the opportunity to discuss research activities partially demonstrated by case studies is an ongoing motivation for all participants. Thus foundation IBS have established very successful events for scientific dialog and cooperation. This idea is supported by Harbin University of Science and Technology (HUST) as well. Therefore, the summer school 2016 takes place at the well suited campus of HUST.

The program of this year summer school includes an educational part with lectures of invited professors, research part with presentations of the participants and work on joint projects and a culture program.

One of the main goals of this international event it to give opportunity to students and researcher of partner universities to work together during school time actively and freely in frame of science and international cooperation. By feedback of participants, this goal is reaching well and we are glad that this idea extending from year to year by different countries.

Prof. Hardt Wolfram

As a new member of this multi-national cooperation mechanism, Harbin University of Science and Technology was thrilled to extend the warm hug to scholars and students overseas to come. Our confidence and motivation to hold this session successfully stem from the time-honored friendship and cooperation with TUC, NSTU and MUST. During the 5-day academic exchanges and discussion, the young talents of respective university show their scientific strength and creativity, and more importantly, different cultures are provided the chance to inspire and understand each other, which lay a sound and pleasant tone for the future session.

We believe that through our joint effort, the mechanism will embrace a better tomorrow and add a bright hue to the national friendship and cooperation!

Prof. Fang Wenbin

CONTENTS

1. Analysis of Contactless Excitation System Resonance Compensation(Yan Meicun, Wang Xudong)	1
2. An Intelligent Hybrid PSO Algorithm for Nonsmooth Optimization in Robust Control Synthesis(Liu Jiaqi, X. Z. Gao)	7
3. A Real-time Path Planning Strategy for Autonomous Mobile Robots in 2-D Grid Maps(Li Zhi, You Bo) ...	15
4. Aspects of Medical Expert System by Pulse Diagnosis(Yumchmaa Ayush)	22
5. Case Study Method in Teaching English for Specific Purpose(Narantsetseg Ravjaa, Tatyana Borozenets, Ariunaa Gunsentsoodol)	26
6. Concept for a Multi-user Dialog Interface for Museum Guides(René Schmidt, Wolfram Hardt)	29
7. Consensus in Networks of Dynamic Agents with Non-uniform Communication Delays(ChongTan, Jinjie Huang)	36
8. Design and Simulation of MEMS Wind Speed and Direction Sensor based on Solid State Heat Transfer and Double Coordinate Model(Lan Yunpiang, Feng Qiaohua, Shi Yunbo, Yu Yang)	42
9. Developing a Collaboration Model Between Researchers Using Their Publication Bipartite Graphs (Ganbat Tsend)	47
10. Development of software and hardware complex measurement and control of devices and crystals. 2G – 401 A – V(Reva Ivan L. , Ivanov Andrey V.)	50
11. Development to Create an Electronic Tutorial on a Subject " Motor Vehicle" (Erdenetuya Amgalan, Choisuren Purevdorj)	52
12. Discrete-Time Robust LQR Control with High-gain Observer for Nano-positioning (Liu Yang, Dongjie Li, Yu Fu, Bo You)	56
13. Image process using Raspberry Pi mini computer(Tumurchudur Lkham, Uranchimeg Tudevtagva)	65
14. Image Restoration via nonlocal supervised coding(Ao Li, Deyun Chen, Guanglu Sun, Kezheng Lin)	71
15. Online Activities Club of Music Education and Music Group(Oyuntuya. G)	79
16. On Designing Reliable and Energy-Efficient MACs for Unidirectional Nodes(Philip Parsch, Alejandro Masrur)	81
17. Optimization of composite winding strategy based on robot(Xu Jiazhong, Yang Hai, Liu Meijun)	89
18. Path Planning of Mobile Robot Using Virtual Plane Approach in Dynamic Environment and Its Applications(T. Khurelbaatar, P. Enkhtsogt, B. Zorig)	105
19. Reform of Cultural studies curriculum in MUST(Narantsatsral. D)	109
20. Research on Dynamic Network Self-Configuration toward SDN(Zhong-Nan Zhao, Jian Wang)	114
21. Research on Spherical Beamforming Algorithm For Sound Source Localization(Yuechan Liu, Zhen Zhou, Chao Sun)	120
22. Resource-based Application Clustering in Wireless Sensor/Actuator Networks(Mirko Lippmann, Mirko Caspar, Wolfram Hardt)	124
23. Service Control Algorithm of Providing Efficient Video-on-Demand Service using Hybrid Mechanism (Otgonbayar Bataa, Erdenebayar Lamjav, Purevtseren Bayarsaikhan, Uuganbayar Purevdorj, Chuluunbandi	

Naimannaran, Young-il Kim, Khishigjargal Gonchigsuulaa) 126

24. Study on Process of Metamodel-based Atmosphere Pressure Modeling Method(Guobing Sun) 133

25. Towards Impact Detection and Localization on a Piezo Metal Composite(Frank Ullmann, Wolfram Hardt)
..... 138

26. The Innovation Process Making Education and Knowledge into Work Creation (M. Byambasuren, Ch.
Oyunjargal) 140

27. The Integrated Flexible Scheduling Algorithm of Complex Product based on the Shortest Processing Time
Scheduling Strategy(Xie Zhi-qiang, Xin Yu) 143

28. FPGA-based Cluster Platform for Reconfigurable Distributed Embedded Systems (Michael Nagler, Stephan
Blokzyl, Wolfram Hardt) 147

Analysis of Contactless Excitation System Resonance Compensation

Yan Meicun^{#1}, Wang Xudong^{#2}

[#] College of Electrical & Electronic Engineering, Harbin University of Science & Technology

No. 52 XueFu Road, Nangang Dist, Harbin, China

¹yanmeicun_lucky@126.com

²wxd158@163.com

Abstract— For the problem that the brush and slip ring of new energy vehicles' drive motor excitation system may bring some safe hidden troubles when the car running, a method of contactless synchronous motor rotor excitation was proposed, and the resonance compensation method was adopted to improve the contactless excitation energy transfer efficiency. The working principle of contactless excitation system was studied, established the mutual inductance model of loosely coupled transformer and analyzed the characteristics of the loosely coupled transformer, presented the design principle of contactless excitation power supply resonance compensation system and analyzed the transmission properties of the contactless energy transmission system. The simulation and experimental results show that: the series-series resonance compensation can effectively compensate the efficiency loss caused by contactless transformer's leakage inductance, increase the current and voltage's amplitude of transformer's second side. When the power supply works in all resonance condition, through the resonance compensation can significantly increase the transmission efficiency of contactless synchronous motor excitation power supply.

Keywords— Brushless DC motors, contactless excitation, loosely coupled transformers, resonance compensation, excitation power supply

I. INTRODUCTION

Due to the existence of electric brush and slip ring in traditional excited motor, it increases the contact resistance. Along with the increase of excitation current, electric brush and slip ring are often caused by poor contact fever. Seriously, it produces fire with burning electric brush and slip ring. The quality of electric brush affects the operation stability directly, as the same time with high failure rate [1–3]. In this paper, analyses the theory of contactless energy transmission and puts forward design which is a brand new way of excitation called contactless excitation system (CES) that proposes to instead the traditional excitation. It uses high frequency power supply switch has loosely coupled rotary transformer as the core. On this basis, this paper emphasizes on loose coupled rotating transformer is studied in the system. The loose coupled rotating transformer is different from the relatively static transformer in the inductive coupled power transfer (ICPT). Loose coupled rotating transformer the performance of the power transmission is the core technology of this method in the contactless excitation system. It is necessary that the whole system and operation security are due to the good operation [4–

7]. In [8] has the optimal design of magnetic circuit model transformer for ICPT. Analyses the different winding peach sticks on the effects of overall performance. In [9] analyses the adjacent windings and coaxial windings magnetic circuit model and the corresponding experimental data is presented. In [10] introduces the advantage of the S/SP compensation strategy used in ICPT system, with the voltage gain and T equivalent circuit are given at the same time.

This paper studies the working principle of contactless excitation system, the establishment of a loosely coupled transformer mutual inductance model, and uses the resonance compensation mode to improve the contactless energy transmission system transmission efficiency, we also analyse the principle of contactless excitation resonance compensation system and transmission characteristics. Finally through the simulation and prototype experiment theory are verified [11–15].

II. CONTACTLESS EXCITATION SYSTEM

The CES is designed on the basis of ICPT. It compared with traditional ICPT system the biggest characteristics are a loose coupled transformer is rotating at high speed and strict working space size. The transformer in the ICPT system is relatively static energy transmission system. A large number of literature at home and abroad show that the rotating transformer is applied to the motor magnetic force is a new field. The CES comprise DC–AC converter, loose coupled rotating transformer, AC–DC convertor and exciting windings. The switching power supply of the CES adopted isolation type DC–DC converter. Its core is loose coupled rotating transformer with the function of isolation. The transformer can be used to replace the excitation system of collector ring and brush, realize the real meaning of brushless excitation. The CES brushless excitation system is shown in Fig. 1. The whole CES is a high frequency switch power supply with special transformer. Inverter controlled switch tube by high frequency signal. At the same time, DC power supply is transformed into high frequency alternating current which is loaded into the transformer primary coil. Secondary induction of the transformer induced the high-frequency voltage, getting through a rectification filtering supply exciting windings. The secondary windings and iron core of transform-

er are coaxial arrangement with motor rotor. As a result of loose coupled rotating transformer, when secondary core rotating with the rotor, the magnetic circuit is al-

most not influenced by any factor. CES eliminate slip ring and electric brush that repeatedly wear caused by the defect completely.

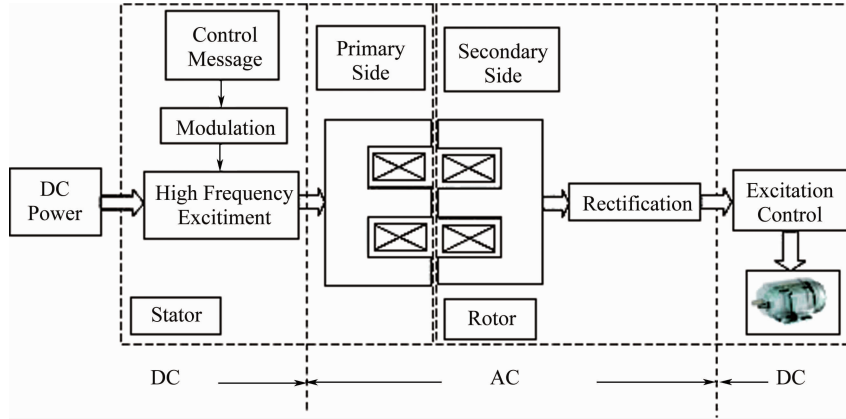


Fig. 1 Contactless excitation system

Secondary core of loose coupled rotating transformer, rectifier, motor rotor core and exciting winding are coaxial. Primary side of the Loose coupled transformer setup on the motor stator. Controller and inverter are placed in the stator shell cavity. Common DC bus is arranged on the stator casing surface. The primary core and secondary core of the loose coupled rotating transformer are relative arrangement. The primary core and secondary core of the loose coupled rotating transformer are designed between 0.5 mm to 3 mm. Assembly of loose coupled rotating transformer is shown in Fig. 2.

area to big as far as possible which causes the magnetizing inductance is increased. To reduce the leakage inductance value, it requires winding close winding on the magnetic tank center column. By physical model can be derived flux path model is shown in Fig. 3 (b). R_c is for the flux path magnetic resistance. R_{air} is for the air gap path magnetic resistance. R_{lk1} and R_{lk2} are for the coil magnetic resistance. Magnetic flux path can be concluded by physical model, which is shown in Fig. 3 (b). R_{air1} and R_{air2} are for the air gap path magnetic resistance. R_{lk} , R_{i1} and R_{i2} are for the coil magnetic resistance. Equivalent circuit is shown in Fig. 4.

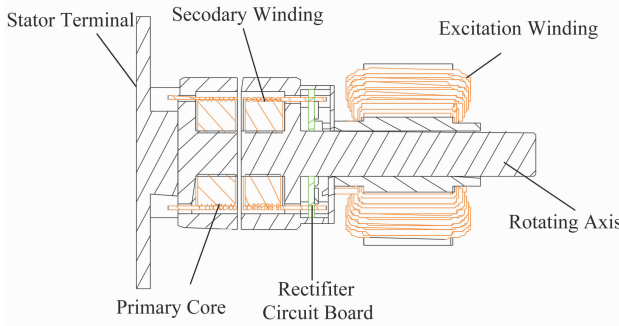


Fig. 2 Assembly drawing of the contactless excitation system

III. ADJACENT WINDINGS IN LOOSE COUPLED ROTATING TRANSFORMER WINDING

The accurate magnetic resistance model and electrical model can be obtained from the physical layout of winding topology. Adjacent windings are shown in Fig. 3. The physical model of winding topology structure of adjacent is shown in Fig. 3 (a). In order to increase the magnetizing inductance, it is necessary to let the air gap parts in mechanical performance good as far as possible under the premise of reducing, winding relative cross-sectional

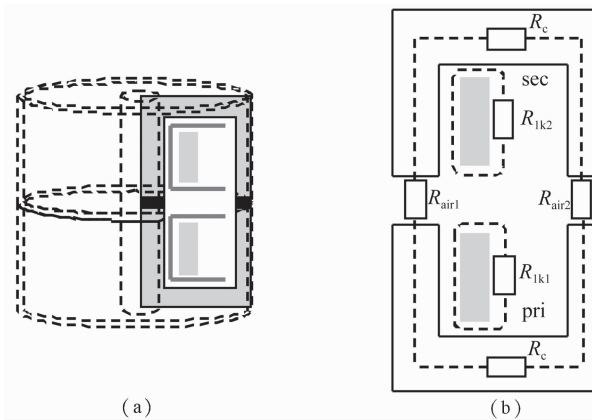


Fig. 3 Adjacent windings topology model (a) physical model; (b) flux path model

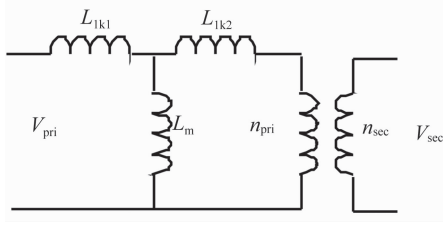


Fig. 4 Equivalent circuit

In loose coupled rotating transformer each part of the magnetic tank represents a part of inductance. The loose coupled rotating transformer magnetizing inductance and leakage inductance can be deduced by the following formula

$$L_m = N_p^2 \left(\frac{l_c}{\mu_0 \mu_c A_c} + \frac{l_g}{\mu_0 \mu_{air} A_g} \right)^{-1} \quad (1)$$

$$l_{lk} = N_p^2 \left(\frac{l_{sp}}{\mu_0 \mu_{sp} A_{sp}} \right)^{-1} \quad (2)$$

where N_p is the number of the primary windings, l_c is the length of the magnetic circuit, l_g is the length of air gap,

A_c is the cross-sectional area, A_g is the air gap cross-sectional area, l_{sp} is the length of the equivalent magnetic circuit, A_{sp} is the equivalent sectional area.

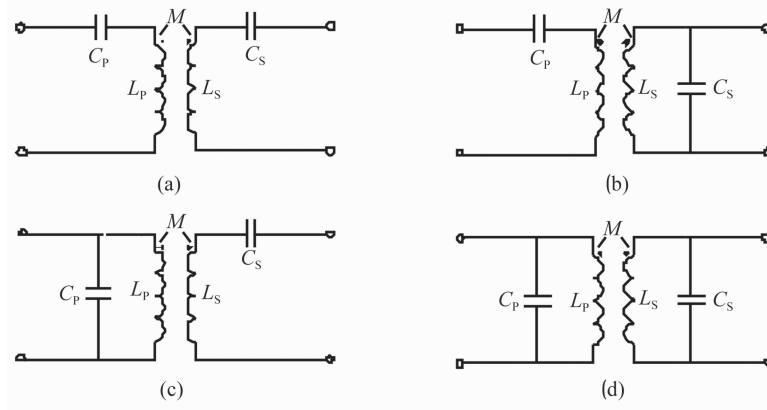
IV. COMPENSATION STRATEGY

For traditional transformer, the coupling coefficient is usually above 0.9, close to 1. The transformer in CES belongs to loose coupling system, which the coupling coefficient is below 0.9, some is even lower than 0.1, the more leakage inductance, core work in the magnetization curve linear part, due to the air gap. According to the literature above it can get the mapping impedance

$$Z_r = - \frac{j\omega M \dot{I}_s}{\dot{I}_p} = R_r + jX_r \quad (3)$$

where M is mutual inductance between primary side and secondary side, R_r is mapping resistance, X_r is mapping reactance, L_s is secondary inductance, L_p is primary inductance. R_L is load resistance.

There are 4 basic resonance compensation circuits as shown in Fig. 5.


 Fig. 5 Four basic resonance compensation circuit
(a) SS; (b) SP; (c) PS; (d) PP

When secondary side uses the series resonance compensation, the circuit is in a state of resonance.

$$\omega = \frac{1}{\sqrt{L_s C_s}} \quad (4)$$

Reflecting impedance can be simplified to

$$Z_r = \frac{\omega^2 M^2}{Z_s} \quad (5)$$

Four different resonance compensation strategies are obtained from series compensation and shunt compensation. In the CES, it determines the working frequency of the system, and then calculates the compensation capacitor vice edge. According to the resonant formula (4), we can get the capacitor in series compensation

$$C_s = \frac{1}{\omega_0^2 L_s} \quad (6)$$

where ω_0 is the frequency for secondary side resonance. Secondary side parallel resonant compensation capacitor

$$C_s = \frac{R_L^2 \pm \sqrt{R_L^2 - 4L_s^2 L_L^2 \omega^2}}{2L_s R_L^2 \omega^2} \quad (7)$$

By (6) can clearly see that, as the load capacitance value size into nonlinear change that is not fit for CES. Series compensation is superior to the parallel compensation for the secondary side in CES. At the same time mapping impedance is described by Z_r . Put (5) into (6) and (7), when secondary side uses the series compensation

$$\text{Re}Z_r = \frac{\omega^4 C_s^2 M^2 R_L}{(\omega^2 C_s L_s - 1)^2 + \omega^2 C_s^2 R_L^2} \quad (8)$$

$$\text{Im}Z_r = \frac{-\omega^3 C_s M^2 (\omega^2 C_s L_s - 1)}{(\omega^2 C_s L_s - 1)^2 + \omega^2 C_s^2 R_L^2} \quad (9)$$

When secondary side uses the shunt compensation

$$\text{Re}Z_r = \frac{\omega^4 M^2 R_L}{R^2 (\omega^2 C_s L_s - 1)^2 + \omega^2 L_s^2} \quad (10)$$

$$\text{Im}Z_r = \frac{-\omega^3 M^2 [C_s R_L^2 (\omega^2 C_s L_s - 1) + L_s]}{R_L^2 (\omega^2 C_s L_s - 1)^2 + \omega^2 L_L^2} \quad (11)$$

Re is real part. Im is imaginary part. Set the equivalent total impedance of primary Z_i , when the primary side with series compensation

$$Z_i = j\omega L_p + \frac{1}{j\omega C_p} + Z_r \quad (12)$$

When primary side the parallel compensation

$$Z_s = \frac{1}{j\omega C_p + \frac{1}{j\omega L_p} + Z_r} \quad (13)$$

The real part of Z_i is the active power. The imaginary part of Z_i is the reactive power. When $\omega = \omega_0$, the calculation results can be expressed as shown in Table I.

Table I
EXPRESSIONS OF PRIMARY SIDE CAPACITANCE

Resonant Topology	Primary Capacitance
SS	$C_p = \frac{1}{\omega_0^2 L_p}$
SP	$C_p = \frac{1}{\omega_0^2 \left(L_p - \frac{M^2}{L_s} \right)}$
PP	$C_p = \frac{L_p - \frac{M^2}{L_s}}{\left(\frac{M^2 R_L}{L_s^2} \right)^2 + \omega_0^2 \left(L_p - \frac{M^2}{L_s} \right)^2}$
PS	$C_p = \frac{L_p}{\left(\frac{\omega_0^2 M^2}{L_s^2} \right)^2 + \omega_0^2 L_p^2}$

Different compensation structures of energy transmission system have different characteristics, the transmission power with different compensation mode are gained by the following formulas

$$P_{SS} = \frac{(\omega M I_p)^2}{R_L} = \frac{U_p^2 R_L}{\omega^2 M^2} \quad (14)$$

$$P_{SP} = \frac{U_L^2}{R_L} = \frac{U_p^2 L_s}{M^2 R_L} \quad (15)$$

$$P_{PS} = \frac{U_p^2}{\frac{\omega^2 M^2}{R_L} + \frac{R_L L_p^2}{M^2}} \quad (16)$$

$$P_{PP} = \frac{U_p^2}{\frac{M^2 R_L}{L_s^2} + \frac{\omega^2 (L_p L_s - M^2)^2}{M^2 R_L}} \quad (17)$$

For analysis the power transmission characteristics of contactless resonance compensation system, respectively simulation analysed four compensation circuits under the condition of primary side input voltage is constant, the

simulation results are shown in Fig. 6.

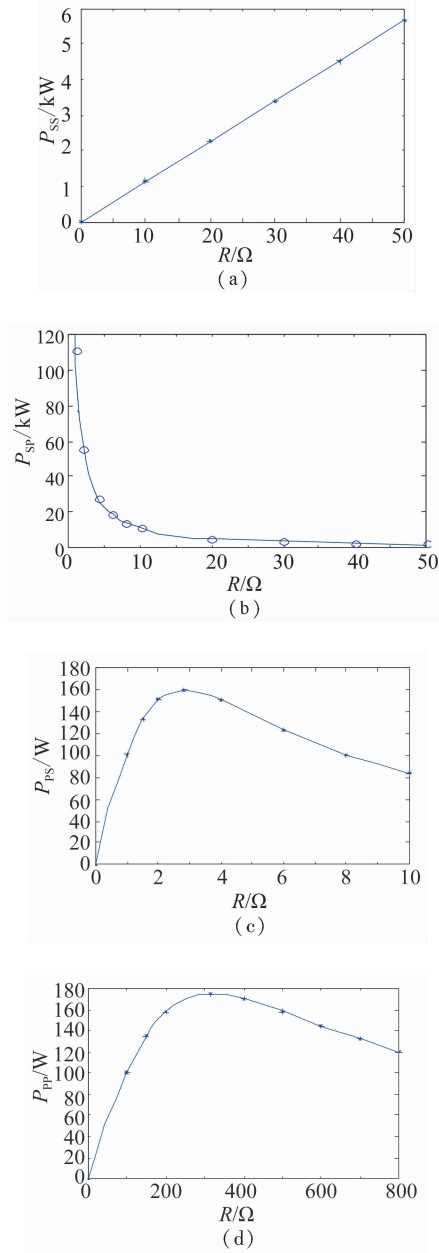


Fig. 6 Equivalent load and transmission power curve with different compensations (a) SS; (b) SP; (c) PS; (d) PP

When the primary side input voltage is a constant value, the equivalent load obtained power is proportional to the load with series-series compensation, and inverse to the load with series-parallel compensation; when the system in parallel-series and parallel-parallel compensation, the equivalent load obtained the maximum transmission power under a certain value. Because of the transmission properties of different compensation systems are different, so we must choose the most suitable compensation system according to the experimental demand.

V. EXPERIMENTAL RESULTS AND ANALYSIS

The power input of CES is 12 V and operating frequency $f=100$ kHz. LM5035 provides excitation energy through the primary side which is transferred to the secondary side by loose coupled rotating transformer. The power drives the excitation of the secondary side of loose coupled rotating transformer at high-speed rotation. Due to the transformer in CES is the loose coupled type, the rotating transformer exhibits a low-coupling coefficient. Therefore, it is necessary to add the compensation strategy which can obtain good transmission characteristics. Type in the adjacent topology winding mode with $R=1\ \Omega$ and the speed of motor is 3,000 rad/min which has the similar strain ratio characteristics of tightly coupled with the air gap is 1 mm. No compensation and add series-series compensation network secondary voltage waveform are shown in Fig. 7, current waveform are shown in Fig. 8.

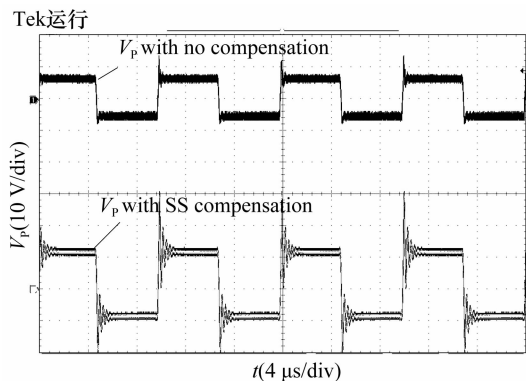


Fig. 7 The waveform of transformer primary side voltage V_p

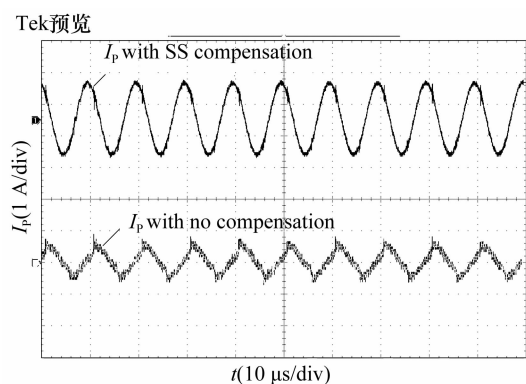


Fig. 8 The waveform of transformer primary side current I_p

From Fig. 7 and Fig. 8, we can see that no compensation's secondary side voltage is about 9 V and add series-series compensation network's secondary side is about 12 V. The current of transformer primary side approximates to sine wave. It has good transmission performance, at the same time, effectively reduces the influence of the leakage inductance. With the increase of voltage

in the secondary side it can obtain more transmission power.

VI. CONCLUSIONS

In this paper, we proposed using loosely coupled transformer as the core energy transmission part of synchronous motor excitation system, in order to realize the contactless excitation. Based on resonance compensation theory, effectively improve the energy transfer efficiency of contactless synchronous motor excitation system. We also designed and analyzed the transmission characteristics with different compensation network. By the simulation and experiment research, we verified the series-series compensation effect of loosely coupled transformer to leakage inductance, and verified that the energy transfer efficiency of contactless excitation system is improved.

ACKNOWLEDGMENT

This work was supported by a grant from the National Natural Science Foundation of China (No. 2011E070303).

REFERENCES

- [1] R. Ramakrishnan, R. Islam, M. Islam, et al, " Real time estimation of parameters for controlling and monitoring permanent magnet synchronous motors," *IEEE International Electric Machines and Drives Conference*, 2009, p. 1194.
- [2] J. N. Dong, Y. K. Huang, and L. Jin, " Review on high speed permanent magnet machines including design and analysis technologies," *Proceedings of the CSEE*, vol. 34, pp. 4640-4653, Sep. 2014.
- [3] H. Ma and X. Sun, " Design of voltage source inductively coupled power transfer system with series compensation on both sides of transformer," *Proceedings of the CSEE*, vol. 30, pp. 48-52, May. 2010.
- [4] M. Kohji, F. Hiroki, and S. Liang, " Motor Modeling for EMC simulation by 3 - D Electromagnetic Field Analysis," *IEEE International Conference on Electric Machines and Drives Conference*, 2009, p. 103.
- [5] C. Martis, H. Hedesiu, and B. Tataranu, " High-Frequency Model and Conductive Interferences of a Small Doubly Salient Permanent Magnet Machine," *IEEE International Conference on Industrial Technology*, 2004, p. 1378.
- [6] G. A. Covic, J. T. Boys, A. M. W. Tam, et al, " Self tuning pick-ups for inductive power transfer," *IEEE Power Electronics Specialists Conference*, 2008, p. 3489.
- [7] J. Y. Lee and H. Y. Shen, " Contactless inductive charging system with hysteresis loop control for small-sized household electrical appliances," *27th IEEE Applied Power Electronics Conference and Exposition*, 2012, p. 2172.
- [8] W. Zhang, Q. H. Chen, S. C. Wong, et al, " Reluc-

- tance circuit and optimization of a novel contactless transformer," *Proceedings of the CSEE*, vol. 30, pp. 108-116, Sep. 2010.
- [9] J. P. C. Smeets, D. C. J. Krop, J. W. Jansen, et al, " Optimal Design of a Pot Core Rotating Transformer," *Energy Conversion Congress and Exposition*, 2010, p. 4390.
- [10] J. Hou, Q. H. Chen, K. Q. Yan, et al, " Analysis and control of S/SP compensation contactless resonant converters," *Proceedings of the CSEE*, vol. 33, pp. 1-8, Nov. 2013.
- [11] V. Ruuskanen, M. Niemela, and J. Pyrhonen, " Modelling the brushless excitation system for a synchronous machine," *IET Electric Power Applications*, vol. 3, pp. 231-239, Mar. 2009.
- [12] Y. H. Kim and K. H. Jin, " Design and implementation of a rectangular-type contactless transformer," *IEEE Transactions on Industrial Electronics*, vol. 58, pp. 5380-5384, Nov. 2011.
- [13] K. D. Papastergiou and D. E. Macpherson, " Contact-less transfer of energy by means of a rotating transformer," *IEEE ISIE*, 2005, p. 1735.
- [14] A. J. Moradewic and M. P. Kazmierkowski, " Contactless energy transfer system with FPGA-controlled resonant converter," *IEEE Transactions on industrial electronics*, vol. 57, pp. 3181-3190, Sep. 2010.
- [15] Yanmeicun, Wang Xudong, Liu Jinfeng, et al, " Analysis of contactless excitation power supply resonance compensation," *Electric Machine and Control*, vol. 19, pp. 45-53, Mar. 2015.

An Intelligent Hybrid PSO Algorithm for Nonsmooth Optimization in Robust Control Synthesis

Liu Jiaqi^{#1}, X. Z. Gao^{*2}

[#]*College of Automation, Harbin University of Science and Technology,
No. 52 Xuefu Road, Nangang Dist, Harbin, China*

¹liujqhust@163.com

^{*}*Department of Electrical Engineering and Automation, Aalto University,
Espoo, Finland*

²xiao-zhi.gao@aalto.fi

Abstract— A new algorithm for nonsmooth optimization in fixed order robust controller design is proposed by using a sort of randomized method. This method uses hybrid Particle Swarm Optimization (PSO) to minimize the constrained objective function subject to controller structures and system stability conditions. Of particular interest is the case that fixed order robust controllers design whose optimal function is nonsmooth and even gradient is not easily computed, but of locally Lipschitz that a ε -dominance strategy can be used at locally minimizes thus a randomized optimization technique will continue without any gradient information. For this purpose, a hybrid PSO method is first introduced to minimize the objective function of controller synthesis. Next, a constraint handling method is proposed for the fixed structure robust controller with constraint is addressed based on the developed method. The proposed method is applied to some typical benchmark functions and the results of algorithm implementation are compared with a well-known method from the literature.

Keywords— Robust control, fixed order controller, nonsmooth optimization, particle swarm optimization

I. INTRODUCTION

Typical applications for robust control design method results in controllers having high requirements for robustness to parameter variations and high requirements for disturbance rejection. The controllers that result from these algorithms are typically of very high order thus unacceptably high for large system. However, if a constraint on the maximum order of the controller is set, that is lower than the order of the plant, the problem is no longer convex and is then relatively hard to solve. These problems become very complex, even when the order of the system to be controlled is low. Many researchers were now focused on the issue of finding efficient methods to solve this hard problem. In Gahinet and Apkarian, it was shown that in order to find a reduced order controller using the LMI formulation, a rank constraint had to be satisfied [1]. The fixed order robust control problem was considered in the further research, where in each iteration a convex SDP was solved in order to find a search direction [2]. Then a multi-direc-

tional search approach was considered that did not use the LMI formulation, and it seems better fitted to handle systems with large dimensions [3]. In Blanchini, the approach to overcome this problem is that directly obtain a popular controller with low order such as PID and Lag-lead [4]. In Calafiore and Campi, the fixed order robust control problem based on randomized method was presented which take inherently NP-hard into consideration [5]. In recent years, the PSO has been considerable literature on optimization problem whose objective function subject to constrains. Therefore, if the PSO can be utilized for fixed order controller design, it would be a great help for practical control engineering [6]. However, the difficult to extend these methods to a broader class of fixed structure controller design because they strongly depend on the control specific. More importantly, since many of the existing methods require in depth knowledge of intelligence computation and robust control theories, these are not easy to master for most engineers in industrial fields. Therefore, these motivates the use of an easy-to-use controller design based on efficient special purpose algorithms that can solve the more general framework of problems.

Previously, we have developed a nonsmooth optimization technique to solve H_∞ synthesis problems by using the Clarke sub-differential and the steepest descent method [6]. The method shows empirical evidence of its superiority in solving a variety of nonconvex and nonsmooth optimization problems. The purpose of this paper is to develop an intelligent computational algorithm based on randomized method for fixed/reduced structure robust controller. The main tool proposed to achieve this is a hybrid PSO, in which each particle's neighborhood is constructed with a ε -dominance strategy thus results from social learning. The PSO with embedded constraint handling method is proposed. Since that is difficult to guarantee the convergence due to the probabilistic nature of PSO, we present numerical results that demonstrate the effectiveness of the algorithm for robust controller design.

II. HYBRIDPARTICLE SWARM OPTIMIZATION AND ε -DOMINANCE STRATEGY

Constrained robust controller design is known as a complex complexity optimization. To solve this problem, it is required to not only provide effective solution algorithms aimed at nonsmooth and nonconvex functions, but also make full use of the human intelligence and knowledge-based method in order to make full use of their respective advantages. This section firstly deeply analyzes search mechanism and convergence of PSO that would be helpful for constructing extended PSO algorithms. Then, the PSO was explored by using differential analysis and Laplace transform into first order inertial element and second order oscillation element and its convergence regions are analyzed based on the spectral radius and Lyapunov stability theorem [7]. Finally, a ε -dominance Strategy is introduced into the PSO algorithm for constrained nonsmooth function which is developed to solve H_∞ synthesis problems.

A. Standard Particle Swarm Optimization

A general particle swarm model consists of a swarm of particles moving in a d -dimensional search space where the fitness f can be calculated as a certain quality measure. Each particle has a position represented by a position-vector \vec{x}_i (i is the index of the particle), and a velocity represented by a velocity-vector \vec{v}_i . The position of i -th particle and its velocity are denoted respectively, as $x_i(t) = (x_{i1}, x_{i2}, \dots, x_{id})^T \in R^d$ and $v_i(t) = (v_{i1}, v_{i2}, \dots, v_{id})^T \in R^d$, where $i \in (1, 2, \dots, d)$. Each particle remembers its own best position so far in a vector \vec{p}_i , and its j -th dimensional value is p_{ij} . The best position from the swarm thus far is then stored in a vector \vec{p}_g , and its j -th dimensional value is p_{gj} . During the iteration time t , the update of the velocity from the previous velocity is determined by Eq. (1a). Subsequently, the new position is determined by the sum of the previous position and the new velocity by Eq. (1b)

$$v_{ij}(t) = wv_{ij}(t-1) + c_1r_1(p_{ij}(t-1) - x_{ij}(t-1)) + c_2r_2(p_{gj}(t-1) - x_{ij}(t-1)) \quad (1a)$$

$$x_{ij}(t) = x_{ij}(t-1) + v_{ij}(t) \quad (1b)$$

where r_1 and r_2 are the random numbers, uniformly distributed within the interval $[0, 1]$ for the j -th dimension of i -th particle. c_1 is a positive constant termed as the coefficient of the self-recognition component, c_2 is a positive constant termed as the coefficient of the social component. The variable w is the inertia factor, for which value is typically setup to vary linearly from 1 to 0 during the iterated processing. From Eq. (1a), a particle decides where to move next, considering its own experience, which is the memory of its best past position, and the experience of its most successful particle in the swarm. The particle swarm algorithm can be de-

scribed generally as a population of vectors whose trajectories oscillate around a region which is defined by each individual's previous best success and the success of some other particle. The more detail of PSO algorithm can be found in [8]. The PSO requires only a few lines of computer code to realize algorithm. Besides, it is a simple concept which is easy to implement.

B. First Order Inertial Element

Observing Eq. (1a), the particles decide where to move next, considering its own experience, which is the memory of its best past position, and the experience of its most successful particle in the swarm. We split the three terms in Eq. (1a) into three equations

$$v_{ij}^e(t) = \chi_1 w v_{ij}(t-1) \quad (2)$$

$$v_{ij}^m(t) = \chi_2 c_1 r_1 (p_{ij}(t-1) - x_{ij}(t-1)) \quad (3)$$

$$v_{ij}^s(t) = \chi_3 c_2 r_2 (p_{gj}(t-1) - x_{ij}(t-1)) \quad (4)$$

where χ_1 , χ_2 and χ_3 are the component constants within the interval $[0, 1]$. Let $\lambda_1 = \chi_1 w$, $\lambda_2 = \chi_2 c_1 r_1$ and $\lambda_3 = \chi_3 c_2 r_2$, reduce to a single dimension, and refer to Eq. (1b), Eq. (2) to Eq. (4) can be deformed into one velocity equation and two difference equations as follows

$$v_{ij}^e(t) = \lambda_1 v_{ij}(t-1) \quad (5)$$

$$v_{ij}^m(t) = x_{ij}^m(t) - x_{ij}^m(t-1) = \lambda_2 (p_{ij}(t-1) - x_{ij}(t-1)) \quad (6)$$

$$v_{ij}^s(t) = x_{ij}^s(t) - x_{ij}^s(t-1) = \lambda_3 (p_{gj}(t-1) - x_{ij}(t-1)) \quad (7)$$

Applying the Laplace transform to Eq. (3) and Eq. (4), yield Eq. (8) and Eq. (9)

$$G_1(s) = \frac{X_i(s)}{P_i(s)} = \frac{\lambda_2}{s + \lambda_2} \quad (8)$$

$$G_2(s) = \frac{X_i(s)}{P_g(s)} = \frac{\lambda_3}{s + \lambda_3} \quad (9)$$

Taking the inverse Laplace transform, we can obtain Eq. (10) and Eq. (11)

$$g_1(t) = \mathcal{L}^{-1}\{G_1(s)\} = \lambda_2 e^{-\lambda_2 t} \quad (10)$$

$$g_2(t) = \mathcal{L}^{-1}\{G_2(s)\} = \lambda_3 e^{-\lambda_3 t} \quad (11)$$

As illustrated in Eq. (5), the velocity of the particle in the swarm model implies the inertial element. When $\lambda_1 < 1$, the velocity would be decreased gradually. The trajectories also contain inertial element in a short time-step, as described in Eq. (10) and Eq. (11). The oscillation element leads the particles to explore some new search space, which would be discussed in the next.

C. Second Order Oscillation Element

By adding a time step in Eq. (1a) and Eq. (1b). The velocity and position varying process yield an equivalent, second-order difference Eq. (12a) and Eq. (12b)

$$v_{ij}(t+1) + (c_1 r_1 + c_2 r_2 - w - 1)v_{ij}(t) + wv_{ij}(t-1) = 0 \quad (12a)$$

$$x_i(t+1) + (c_1 r_1 + c_2 r_2 - w - 1)x_i(t) + wx_i(t-1) - c_1 r_1 p_i(t) - c_2 r_2 p_g(t) = 0 \quad (12b)$$

Taking the Laplace transform of Eq. (12b), we can obtain Eq. (13) as

$$X_i(s) = \frac{\varphi_1 s P_i(s) + \varphi_2 s P_g(s)}{s^2 + (\varphi_1 + \varphi_2 - w - 1)s + w} \quad (13)$$

where $\varphi_1 = c_1 r_1, \varphi_2 = c_2 r_2$ and $\varphi = \varphi_1 + \varphi_2$. Eq. (13) can be split into two parts and rewritten as

$$G_1(s) = \frac{X_i(s)}{P_i(s)} = \frac{\tau_1 s}{s^2 + (\varphi_1 + \varphi_2 - w - 1)s + w} \quad (14a)$$

$$G_2(s) = \frac{X_i(s)}{P_i(s)} = \frac{\tau_2 s}{s^2 + (\varphi_1 + \varphi_2 - w - 1)s + w} \quad (14b)$$

where $\tau_1 = \chi_1 \varphi_1, \tau_2 = \chi_2 \varphi_2$ and χ_1, χ_2 are the component constants within the interval $[0, 1]$.

If $(\varphi - w - 1)^2 \geq 4w$, applying the inverse Laplace transform to Eq. (14a), we have

$$\begin{aligned} \tau_1 e^{-\frac{\varphi-w-1}{2}t} \cosh\left(\frac{\sqrt{(\varphi-w-1)^2 - 4w}}{2}t\right) - \\ \tau_1 \frac{\varphi - w - 1}{\sqrt{(\varphi - w - 1)^2 - 4w}} e^{-\frac{\varphi-w-1}{2}t} \\ \sinh\left(\frac{\sqrt{(\varphi - w - 1)^2 - 4w}}{2}t\right) \end{aligned} \quad (15)$$

If $(\varphi - w - 1)^2 < 4w$, applying the inverse Laplace transform to Eq. (14a), we have

$$\begin{aligned} \tau_1 e^{-\frac{\varphi-w-1}{2}t} \cos\left(\frac{\sqrt{(\varphi-w-1)^2 - 4w}}{2}t\right) - \\ \tau_1 \frac{\varphi - w - 1}{\sqrt{(\varphi - w - 1)^2 - 4w}} e^{-\frac{\varphi-w-1}{2}t} \\ \sin\left(\frac{\sqrt{(\varphi - w - 1)^2 - 4w}}{2}t\right) \end{aligned} \quad (16)$$

The first order inertial element is helpful to maintain the stability trajectory and algorithm convergence, while the second order oscillation element trends to explore some new search spaces for the better solutions. The tradeoff between exploration is the variety of algorithms game on global optimization performance, e. g., accuracy and convergence speed of optimization algorithms.

The Eq. (1a) and Eq. (1b) are rewritten in state-space form

$$\begin{aligned} \begin{bmatrix} v_{ij}(t) \\ x_{ij}(t) \end{bmatrix} &= \begin{bmatrix} w & -c_1 r_1 - c_2 r_2 \\ w & 1 - c_1 r_1 - c_2 r_2 \end{bmatrix} \cdot \\ \begin{bmatrix} v_{ij}(t-1) \\ x_{ij}(t-1) \end{bmatrix} &+ \begin{bmatrix} c_1 r_1 & c_2 r_2 \\ c_1 r_1 & c_2 r_2 \end{bmatrix} \cdot \begin{bmatrix} p_i(t-1) \\ p_g(t-1) \end{bmatrix} \end{aligned}$$

To describe in a brief way, the equation is rewritten as

$$X(t) = AX(t-1) + BU(t-1) \quad (17)$$

where

$$\begin{aligned} X(t) &= \begin{bmatrix} v_{ij}(t) \\ x_{ij}(t) \end{bmatrix}, U(t) = \begin{bmatrix} p_i(t) \\ p_g(t) \end{bmatrix} \\ A &= \begin{bmatrix} w & -c_1 r_1 - c_2 r_2 \\ w & 1 - c_1 r_1 - c_2 r_2 \end{bmatrix} \end{aligned}$$

$$B = \begin{bmatrix} c_1 r_1 & c_2 r_2 \\ c_1 r_1 & c_2 r_2 \end{bmatrix}$$

Here, $\rho(A) < 1$ represents the spectral radius of matrix A which determine convergence rate. By using the spectral radius and Lyapunov stability theorem, the necessary and sufficient condition of convergence for PSO is represented that, when w and $c_1 r_1 + c_2 r_2$ are constants, the PSO algorithm converges if and only if $0 < c_1 r_1 + c_2 r_2 < 2w + 2$ and $w < 1$ are satisfied together.

D. ε -dominance Strategy for PSO

The conception of ε -dominance was proposed by Zahara, and is defined as: if a decision vector $x_i \in \Omega$ is ε -dominated by another decision vector $x_j \in \Omega$ that must meet one of the following requirements [9]

$$\begin{cases} f_k(x_j)/(1 + \varepsilon) \leq f_k(x_i) \\ \quad \forall k = 1, 2, \dots, N \\ f_k(x_j)/(1 + \varepsilon) < f_k(x_i) \\ \quad \text{at least a } k = 1, 2, \dots, N \end{cases} \quad (18)$$

A ε -dominance strategy is introduced, thus a modified computation method of nonsmooth function and an alterable PSO are put forward, which improve the convergence speed of the algorithm and the diversity of the particles.

A general optimization problem of robust control is given by an objective function $f(\kappa)$, which is to be optimized with respect to the controller design variable κ . Besides, κ consists of m particles ($\kappa_1, \kappa_2, \dots, \kappa_m$) to search an optimal solution κ^* of $f(\kappa)$. $\kappa_i^{\text{best},k}$ denotes the best previously obtained position of the i -th particle, and $\kappa_{\text{swarm}}^{\text{best},k}$ denotes the best position in the entire swarm at the current iteration k

$$\kappa_i^{\text{best},k} = \arg \min_{\kappa_{ij}} \{f(\kappa_{ij}), 0 \leq j \leq k\} \quad (19a)$$

$$\kappa_{\text{swarm}}^{\text{best},k} = \arg \min_{\kappa_{ik}} \{f(\kappa_{ik}), \forall i\} \quad (19b)$$

Here, $f(\kappa_{ij}) = f_j(\kappa_i)/(1 + \varepsilon)$ and $f(\kappa_{ik}) = f_k(\kappa_i)/(1 + \varepsilon)$. The ε -dominance strategy can help $f(\kappa)$ to remove Pareto solutions which are close together. The goal of integrating ε -dominance strategy and PSO is to combine their advantages and avoid disadvantages. For example, ε -dominance strategy is a very efficient local search procedure but its convergence is sensitive to the ε selected value. PSO belongs to the class of global search procedures but requires much computational effort.

Consider the real robust control synthesis, it is crucial to take the given constraint conditions into account. The optimization problem $f(\kappa)$ subject to multiple constraint conditions is mathematically formulated as follows

$$f(\kappa_{ij})_{\min} = \|T_{w \rightarrow z}(K)\|_{\infty}, \quad K \in H(\kappa_{ij}) \quad (20)$$

where $K \in H(\kappa_{ij})$ represent the structural constraints on the robust controller. In next section, the original constrained optimization problem in Eq. (20) can be modified into an unconstrained problem. We extend our previous work on the PSO to develop a nonsmooth technique to compute the ε -dominance solutions.

III. NONSMOOTH OPTIMIZATION FOR ROBUST CONTROL SYNTHESIS

The main idea of constrained robust controller design is to solve H_∞ synthesis by trying to solve the optimization problem (20). As mentioned in the above section, the necessary and sufficient condition of convergence for PSO algorithm with ε -dominance strategy is already analyzed. Now, a numerical nonsmooth method is presented for solving the objective function in optimization problem (20).

A. Constraint Handling Method

A constraint handling methods is introduced, which can be embedded in hybrid PSO and ε -dominance strategy. Dong et al. proposed constraint fitness priority-based ranking method [10], and introduced a constraint fitness function $Wf(\kappa_{ij})$ for constraints, and it is computed from both inequality and equality constraints as follows.

Without loss of generality, the constrained optimization problem $f(\kappa)$ for robust controller design in Eq. (20) can be rewritten in the following canonical form

$$f(\kappa_{ij})_{\min} = \|T_{w \rightarrow z}(K)\|_\infty \quad (21)$$

subject to $h_m(\kappa_{ij}) = 0, \quad g_n(\kappa_{ij}) \leq 0$

where structural constraints $H(\kappa_{ij})$ in Eq. (20) with m equality constraints and n inequality constraints. Then, the structural constraint $H(\kappa_{ij})$ is set as

$$H(\kappa_{ij}) = \begin{bmatrix} h_m(\kappa_{ij}) = 0 \\ g_n(\kappa_{ij}) \leq 0 \\ \text{Re}[\lambda_{\max}(\sum \kappa_{ij})] \\ \text{rank } K - k \\ \|T_{w \rightarrow z}(K)\|_\infty - \gamma_\infty \end{bmatrix} \quad (22)$$

(1) For equality constraints $h_m(\kappa_{ij}) = 0$

$$W_m(\kappa_{ij}) = \begin{cases} 1, & \text{if } h_m(\kappa_{ij}) = 0 \\ 1 - \frac{|h_m(\kappa_{ij})|}{h_{\max}(\kappa_{ij})}, & \text{if } h_m(\kappa_{ij}) \neq 0 \end{cases} \quad (23a)$$

(2) For inequality constraints $g_n(\kappa_{ij}) \leq 0$

$$W_n(\kappa_{ij}) = \begin{cases} 1, & \text{if } g_m(\kappa_{ij}) \leq 0 \\ 1 - \frac{|g_m(\kappa_{ij})|}{g_{\max}(\kappa_{ij})}, & \text{if } g_m(\kappa_{ij}) > 0 \end{cases} \quad (23b)$$

The constraint fitness function evaluated at point κ_{ij} is equal to the weighted sum of the $W_m + W_n$, as defined below

$$Wf(\kappa_{ij}) = \sum_{k=1}^{m+n} w_k W_k(\kappa_{ij}), \quad \sum_{k=1}^{m+n} w_k = 1, \quad 0 \leq w_k \leq 1 \quad \forall k \quad (24)$$

where w_k is a randomly generated weight for constraint k . In that case, the original constrained optimization problem in Eq. (21) is modified as the unconstrained forms. The constraint handling method is quite easy to use and this method does not introduce any additional

decision variables such as the coefficients of the augmented Lagrangian method.

B. Nonsmooth Robust Controller Design Technique

Here we present a design procedure of fixed/reduced structure robust controllers satisfying the given performance specifications based on the hybrid PSO algorithm developed in last subsection. Note that it is straightforward to obtain the desired structured controller if we solve the optimization problem in Eq. (20) by using the PSO tool, which is the main contribution of this paper and can be easily verified in multiple robust controller design.

Given a linear time invariant (LTI) system as follows

$$\begin{bmatrix} \dot{x} \\ z \\ y \end{bmatrix} = \begin{bmatrix} A & B_1 & B_2 \\ C_1 & D_{11} & D_{12} \\ C_2 & D_{21} & 0 \end{bmatrix} \begin{bmatrix} x \\ w \\ u \end{bmatrix} \quad (25)$$

And find a static output as

$$u = K(s)y \quad (26)$$

Such that

$$\begin{aligned} K^i &= C_k^i (sI - A_k^i)^{-1} B_k^i + D_k^i \\ \alpha(K) &= \max \{ \text{Re}(\lambda) \} < 0 \\ \text{rank } K &\leq k \end{aligned} \quad (27)$$

where a dynamic output feedback $K(s)$ is stable, if and only if $\alpha(K) < 0$. $\alpha(K)$ is the spectral abscissa of a matrix, and λ is eigenvalue of matrix. Rank $K \leq k$ is fixed/reduced structured condition. The objective function of the optimization problem is difficult to solve, due to the nonsmoothness of function $f(\kappa_{ij})$ in Eq. (20), as we known that $H(\kappa_{ij})$ is the set of stabilizing $K(s)$. Finding the $K(s)$ that minimize $\|T_{w \rightarrow z}(K)\|_\infty$ is the same as finding minima of κ_{ij} in Eq. (21).

Briefly, the main optimization algorithm for problem (21) consists of following steps.

① Initialize number of particles κ_i and the algorithm parameters $w, c_1, c_2, r_1, r_2, t_0$.

② Set initial position $x_i(t_0)$ and the velocity $v_i(t_0)$.

③ Compute individual and social influence.

④ Use constraint handling method and compute position $x_i(t+1)$ and velocity $v_i(t+1)$ at next iteration.

⑤ Apply ε -dominance strategy to $\kappa_{ij}(t)$ and compare $f_i(\kappa_{ij})/(1+\varepsilon)$ with $f_{i-1}(\kappa_{ij})$.

⑥ Evaluate fitness function of Eq. (24).

If the candidate solutions are acceptable multiple minima and the robust conditions holds go to (7).

Else, update position $x_i(t_0)$ and velocity $v_i(t_0)$ and go to (3).

⑦ Print solutions.

If the particles $\kappa_{ij}(t)$ have converged, then halt.

Else, update position $x_i(t)$ and velocity $v_i(t)$ and go to (3).

Therefore, all we have to do is to solve Eq. (21), Eq. (25), Eq. (26), Eq. (27) using the above defined $f(\kappa_{ij})$, $h_m(\kappa_{ij})$ and $g_m(\kappa_{ij})$ via the hybrid PSO algorithm. To a certain extent, the necessary and sufficient condition of that $0 < c_1 r_1 + c_2 r_2 < 2w + 2$ and $w < 1$ guar-

antee the performance theoretically due to the PSO has been analyzed. Whereas, it may happen that the algorithm fail to find a feasible solution due to the probabilistic nature of PSO. In next section, the proposed hybrid PSO has been applied to solving several constrained benchmark functions and some well-known engineering design problems.

IV. NUMERICAL RESULT AND ANALYSIS

In this section, the hybrid PSO has been applied to solving several constrained benchmark functions and an engineering controller design problem [11]. The algorithm is coded in Matlab 2015b. The numerical examples of optimizing each of the benchmark functions is executed in 20 independent runs, with each run iterated 600 times, or until the convergence or a better solution is found.

During the simulations, it has been found that the population size has significant effects on the performance of the hybrid PSO. Table I shows the search results of the hybrid PSO with different population sizes for four benchmark functions G1, G2, G3 and G4 [12], and it is clearly seen that for the population sizes $15N + 1$ and $20N + 1$ better solutions are found than for $10N + 1$. Therefore, a population size of $15N + 1$ is employed in all subsequent simulations instead of $20N + 1$ to decrease the number of function evaluations.

Table II contains a summary of the execution results of the same four benchmark functions, which are compared with those of a genetic algorithm for multi-objective robust control design and a quasi-newton interior point method for reduced order H_∞ controller synthesis. For the sake of comparison, the table also gives the reference optimal values. It is seen in Table II that the hybrid PSO is competitive against MRCD and better than QN, but hybrid PSO requires less number of function evaluations than the other methods. The hybrid PSO converges in 102, 142, 58 and 72 iterations for solving G1, G2, G3 and G4.

The hybrid PSO to solve H_∞ synthesis problems will be applied for attitude control system of a satellite [13]. The disturbing torques $T_d = [T_{dx} \ T_{dy} \ T_{dz}]$, which consists of gravity gradient torque and solar light pressure torque with $\pm 20\%$ disturbance are given

$$\begin{aligned} T_{dx} = T_{dz} &= (2.5 + \sin \omega_0 t) \times 10^{-3} \text{ N} \cdot \text{m} \\ T_{dy} &= 0 \end{aligned} \quad (28)$$

The initial state of roll angle is $[\psi(0) \ \dot{\psi}(0)] = [-0.8 \ -0.35]$. For sake of briefly presentation, the system only chooses roll angle channel for robust synthesis. The change of $f(\kappa_{ij})$, which is defined as $f(\kappa_{ij}) = \|T_{w \rightarrow z}(K)\|_\infty$ for the first 400 s (which corresponds to about 600 iterations) are illustrated in The best design parameter $K = \{\kappa_{ij}\}$ obtained in 188 trials is

$$K = \begin{bmatrix} -3.673, 0 & 6.964, 4 & -1.559, 4 \\ 0.586, 3 & -0.371, 3 & -0.086 \end{bmatrix}$$

and the corresponding H_∞ performance index is 1.652, 3.

Four different plants $G_i(s)$ have been searched with

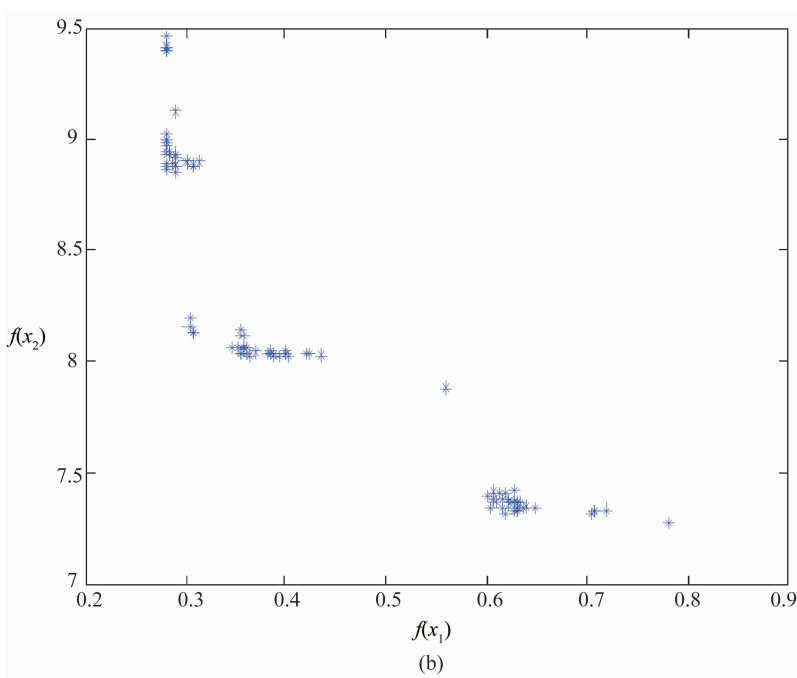
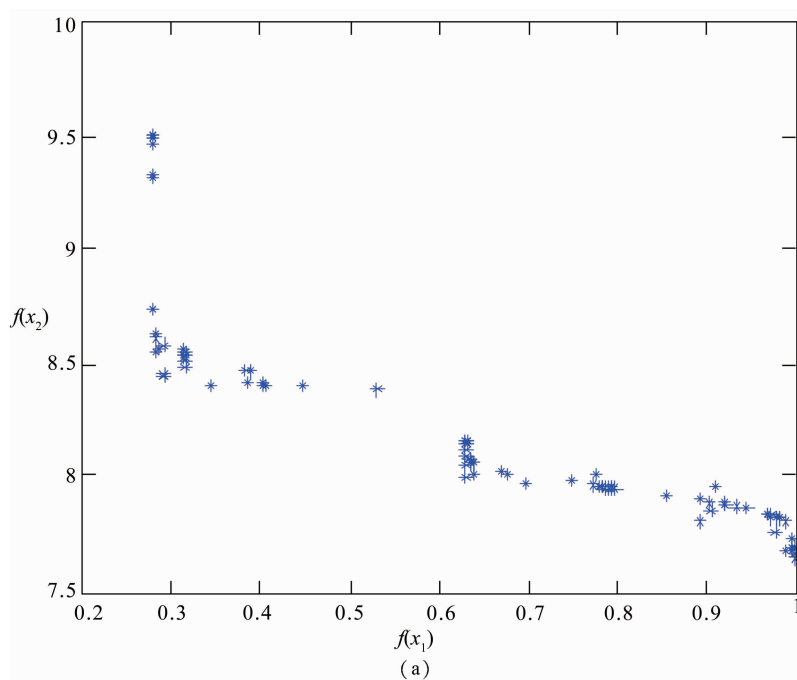
the same value of robustness parameters, and the simulation results are shown as Fig. 1. It shows that the proposed algorithm can find the satisfactory solutions within a few time with a high probability of success. From Fig. 1, the Pareto front obtained by the NSGA-II for fixed order robust control is more close to the true optimal Pareto front, and the convergence and distribution of solutions are satisfactory.

Table I
THE EFFECT OF THE POPULATION SIZE

Pop. Size	Hybrid PSO			
		10N + 1	15N + 1	20N + 1
G1	Best	0.992, 3	0.998, 3	0.998, 2
	Worst	0.989, 8	0.990, 1	0.989, 0
	Avg.	0.991, 4	0.994, 3	0.997, 8
G2	Best	-6,662.8	-6,932.11	-6,911.2
	Worst	-5,643.2	-6,343.54	-6,265.4
	Avg.	-6,254.91	-6,934.23	-6,899.11
G3	Best	24.453	24.127	24.056
	Worst	25.998	25.672	24.733
	Avg.	24.897	24.412	24.301
G4	Best	679.22	681.04	680.23
	Worst	682.50	681.88	681.34
	Avg.	680.43	681.09	680.83

Table II
TEST RESULTS OF PSO AND OTHER ALGORITHMS

Optimal value	Algorithms Compare			
		MRCD	QN	Hybrid PSO
G1 1.0	Best	0.984, 0	0.928, 7	0.998, 3
	Worst	0.952, 3	0.976, 5	0.990, 1
	Avg.	0.981, 4	0.932, 1	0.994, 3
G2 -6 961.8	Best	-6,432.1	-6,678.00	-6,932.11
	Worst	-6,001.9	-6,231.59	-6,343.54
	Avg.	-6,391.4	-6,288.86	-6,934.23
G3 24.306	Best	24.361	24.034	24.127
	Worst	25.239	25.777	25.672
	Avg.	24.410	24.519	24.412
G4 680.63	Best	677.14	692.43	681.04
	Worst	688.86	712.34	681.88
	Avg.	682.29	698.09	681.09



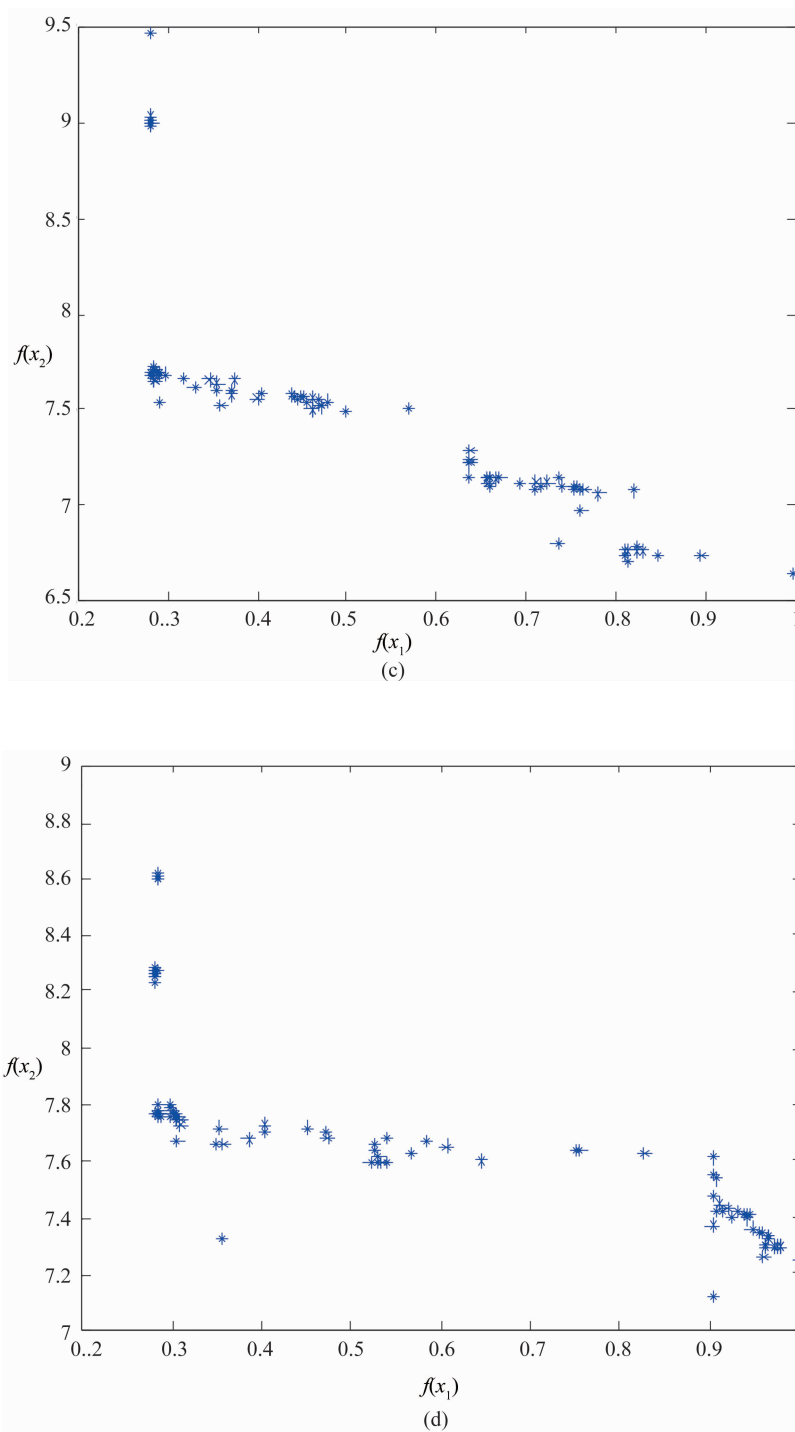


Fig. 1 Convergence property of the robust controller design

V. CONCLUSIONS

A constrained nonsmooth optimization of robust controller is designed by using a sort of hybrid PSO algorithm, which is simple and easy to implement. First, we introduce a ε -dominance strategy to improve the PSO algorithm in terms of its capacity to search for local optimal solutions, and its necessary and sufficient conditions of convergence for algorithm are analyzed. Then, the hy-

brid PSO with embedded constraint handling method is proposed for solving nonsmooth objective function. Finally, the experimental results illustrate the attractiveness of the algorithm for handling structure constraint. Practical application of the hybrid PSO method for engineering design problem is positive to promote method to obtain more accurate, reliable and efficient at locating global optima than the other alternative.

ACKNOWLEDGMENT

This work was supported in part by the Academy of Finland under Grant 135225 and Finnish Funding Agency for Technology and Innovation (TEKES).

REFERENCES

- [1] P. Gahinet and P. Apkarian, "A linear matrix inequality approach to H_∞ control," *International Journal of Robust and Nonlinear Control*, vol. 4, pp. 421-48, Apr. 1994.
- [2] P. Apkarian, D. Noll, and H. D. Tuan, "Fixed-order H_∞ control design via a partially augmented Lagrangian method," *International Journal of Robust and Nonlinear Control*, vol. 53, pp. 1137-1148, Dec. 2003.
- [3] P. Apkarian and D. Noll, "Controller design via nonsmooth multi-directional search," *SIAM Journal on Control and Optimization*, vol. 44, pp. 1923-1949, Jun. 2006.
- [4] F. Blanchini, A. Lepschy, S. Miani, and U. Viaro, "Characterization of PID and lead/lag compensators satisfying given H_∞ specifications," *IEEE Transactions on Automatic Control*, vol. 48, pp. 736-740, May. 2004.
- [5] G. C. Calafiore and M. C. Campi, "The Scenario approach to robust control design," *IEEE Transactions on Automatic Control*, vol. 51, pp. 314-331, May. 2006.
- [6] K. E. Parsopoulos and M. N. Vrahatis, "On the computation of all global minimizers through particle swarm optimization," *IEEE Transactions on Evolutionary Computation*, vol. 18, pp. 211-224, Mar. 2004.
- [7] B. X. Yue, H. B. Liu, and A. Abraham, "Dynamic trajectory and convergence analysis of swarm algorithm," *Computing and Informatics*, vol. 33, pp. 37-49, Jan. 2012.
- [8] O. Olorunda and A. Engelbrecht, "Measuring exploration in particle swarms using swarm diversity," *Proceedings of IEEE Congress on Evolutionary Computation*, Washington, pp. 1128-1134, 2008.
- [9] Zahara E and Kao Y T. "Hybrid Nelder-meade simplex search and particle swarm optimization for constrained engineering design problems," *Expert Systems with Applications*, 36 (2): 3880-3886, 2009.
- [10] Y. Dong, J. F. Tang, B. D. Xu, D. Wang, "An application of swarm optimization to nonlinear programming," *Computers and Mathematics with Applications*, vol. 49, pp. 1655-1668, Sept. 2005.
- [11] D. Ankelhed, A. Helmersson, and A. Hansson, "A quasi-newton interior point method for low order H_∞ controller synthesis," *IEEE Transactions on Automatic Control*, vol. 56, pp. 1462-1467, Jun. 2011.
- [12] S. Ramzy and A. Ali, "Design of robust mixed H_2/H_∞ PID controller using particle swarm," *International Journal of Advancements in Computing Technology*, vol. 12, pp. 53-60, May. 2010.
- [13] A. Herreros, E. Baeyens, and J. R. Peran, "MRCD: a genetic algorithm for multiobjective robust control design". *Engineering Applications of Artificial intelligence*, vol. 15, pp. 285-301, May. 2002.

A Real-time Path Planning Strategy for Autonomous Mobile Robots in 2-D Grid Maps

Li Zhi ^{#1}, You Bo ^{#2}

[#] College of Automation, Harbin University of Science and Technology

No. 52 XueFu Road, Nangang Dist, Harbin, China

¹lizhi_haligong@163.com

²youbo@hrbust.edu.cn

Abstract— The main purpose of this essay is to present the VAPF-Local A* path planning method, which combines the velocity-based artificial potential field with the improved A* algorithm. Making use of that the grid map is suitable for the real-time storage and analysis in computer, the timeliness and simplicity of VAPF is reserved. When mobile robot is stuck in local minima or oscillation using VAPF, the method switches to Local A* algorithm. The VAPF-Local A* method guarantees the timeliness, safety and stability of the locomotion of the robot.

Keywords— VAPF, A* algorithm, grid map, real-time path planning, autonomous mobile robots

I. INTRODUCTION

The 2-D path planning of mobile robots can be defined as that the robots search for a path for locomotion which fulfills certain tasks complying with certain evaluating criteria. Tasks can not only be complicated such as the coverage of the environment or searching for specific objectives, but also can be basic locomotion task from point to point. However, the complicated tasks can be decomposed into series of continuous point-to-point locomotion. Therefore, there are researchers regarding the substance of path planning as to research the point-to-point movement as in [1] and [2].

There are mainly such ways describing the working environment of mobile robots as grid map, topological map, geometric map and hybrid maps of the maps above. The grid maps generally refers to maps dividing the external environment into equal or unequal square grids, where the binary classification is done in every grid according to whether the grid area is occupied by an obstacle as in [1] to [3]. There are also researchers using special grids such as fan-shaped grids to depict the environment as in [4]. The grid map is relatively easy to be stored and analyzed by computer. Therefore, it is widely applied in map building during the real-time locomotion of mobile robots as in [5].

The path searching methods which are commonly used in grid map is A* algorithm, D* algorithm and the extended algorithms of them as in [6]. The A* algorithm belonging to HAS, heuristic search algorithm, con-

stantly approaches target location based on local information until valid path is found.

The APF model, in short of artificial potential field model, put forward by Khatib as in [7], is another commonly used path planning method applying the field theory in physics to solve the problem in robotics. There are four defects of APF summarized by Warren as in [8] as the local minima, difficult to find path among adjacent obstacles, the oscillation in narrow aisle and the oscillation caused by obstacles. There are some researchers presenting specific improvements to solve above problems as in [9] to [11], such as adding new cost functions of velocity, inertia, energy, etc.

II. ENVIRONMENT MAP REPRESENTATION

The path planning in this essay is mainly processed in 2-D plane, in which it is assumed that the binary classification of planar characteristics can be processed. To simplify the problem, it is assumed that the planar characteristics can be recognized and completely classified as passable and impassable grids. Also, obstacles are seen as immobile during current planning cycle. Generally in grid maps, the smaller the grid sizes are, the more finely the map is divided, and the smoother the path generated for mobile robots. However, if the grids are subdivided excessively, there will be a heavy burden on computer when storing and analyzing the map.

To simplify the analysis, the entire grid map is assumed as square. The grid map for the path planning of the autonomous mobile robots is presented as Fig. 1.

It is assumed the projection of robot in map is a circle with diameter of 1 unit long, while the side length of a basic square grid is set as 1. The robot can move horizontally, vertically and diagonally continuously. In order to reduce the probability that accidental collisions occurring due to be ignorant of exact information of obstacles, the dilation of obstacles is processed before path-planning. The grids of dilation, without characteristic of obstacles, can not be included in the path planned.

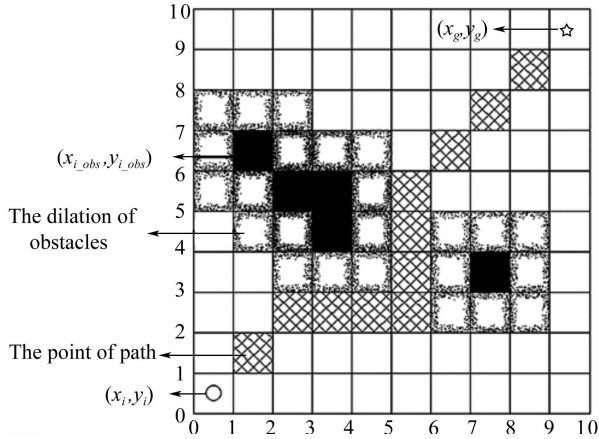


Fig. 1 The grid map for the path planning of the autonomous mobile robots with a side length of 10

In grid map, the passable region is filled with white, while the impassable region, namely obstacles, is filled with black. And the map is assumed as part of the infinite 2-D cartesian space. Before every path planning cycle, the start point is the position where the robot resides currently, which is set as (x_i, y_i) and represented by a circle. Meanwhile, the reachable goal point is selected, which is set as (x_g, y_g) , represented by a pentacle. The obstacle points is set as $(x_{i, obs}, y_{i, obs})$. The dilation of obstacles is filled with speckle. It is assumed that the goal is in the first quadrant of the robot's world coordinates.

III. PATH PLANNING ALGORITHM

A. Velocity-based Artificial Potential Field Method

The basic thought of APF method is to see the motion of the robot in certain environment as the motion in an artificial force field, where the obstacles generate repulsive forces and target point generate attractive forces on the robot. The resultant force of all forces directs the motion and affects the path of the robot generated eventually as in [7].

The attractive potential energy the robot received and the Euclidean distance from the robot to the target point is in inverse proportion. When the attractive potential energy drops to zero, the target point is reached. The attractive potential energy is described as

$$U_{att}(q) = \frac{1}{2} \eta \rho^2(q, q_g) \quad (1)$$

as in [9] where η is the gain coefficient of attractive energy, $\rho(q, q_g)$ represents a vector of which the magnitude is the Euclidean distance $|q - q_g|$ from the current position q to the target position q_g , the direction points from q to q_g .

Then the attractive force generated by the attractive potential field is described as

$$F_{att}(q) = -\nabla U_{att}(q) = \eta \rho(q, q_g) \quad (2)$$

where the direction of attractive force is the same

as $\rho(q, q_g)$.

The repulsive potential energy the robot received and the Euclidean distance from the robot to the target point is in inverse proportion as well. When the repulsive potential energy drops to zero, the robot has escaped from the region that obstacles affect. The repulsive potential energy as in [9] is described as

$$U_{rep}(q) = \begin{cases} \frac{1}{2} k \left(\frac{1}{\rho(q, q_{obs})} - \frac{1}{\rho_0} \right), & 0 \leq \rho(q, q_{obs}) \leq \rho_0 \\ 0, & \rho(q, q_{obs}) > \rho_0 \end{cases} \quad (3)$$

where the k is the gain coefficient of repulsive energy, ρ_0 being a positive constant is the maximum distance from which the obstacle affecting the robot, $\rho(q, q_{obs})$ represents a vector of which the magnitude is the Euclidean distance $|q - q_{obs}|$ from the current position q to the target position q_{obs} , the direction points from q to q_{obs} .

Then the repulsive force generated by the attractive potential field is described as

$$F_{rep}(q) = \begin{cases} k \left(\frac{1}{\rho(q, q_{obs})} - \frac{1}{\rho_0} \right) \cdot \frac{1}{\rho^2(q, q_{obs})} \cdot \nabla \rho(q, q_{obs}), & 0 \leq \rho(q, q_{obs}) \leq \rho_0 \\ 0, & \rho(q, q_{obs}) > \rho_0 \end{cases} \quad (4)$$

where the direction of attractive force is the same as $\rho(q, q_{obs})$.

Though possibly having problems of oscillation and local minima, the APF is suitable for path planning in static environment. However, in a dynamic and uncertain environment, the autonomous robot, carrying the equipment surveying and mapping the space, can not foresee the obstacles which have not been observed due to the restriction of current position. Thus the ability of avoiding obstacles is declined. Therefore, the velocity-based APF is applied to improve the robot's adaptation to real-time and dynamic environment as in [10]. The dynamic affection is mainly from the obstacles in space, so that the repulsive potential energy $U_{rep}(q)$ need to be modified.

The relative velocity potential energy as in [10] is described as

$$U_{repv}(q) = \begin{cases} \frac{1}{2} k_{repv} V_{ro}^2, & 0 \leq \rho(q, q_{obs}) \leq \rho_0 \cap \alpha \in (-\pi/2, \pi/2) \\ 0, & \text{others} \end{cases} \quad (5)$$

where the k_{repv} is the gain coefficient of velocity potential energy, V_{ro} is the relative velocity between the robot and the obstacle, α is the included angle between V_{ro} and \vec{ro} , \vec{ro} is the vector pointing from q to q_{obs} . When the robot is out of the region that obstacles affect, namely $\alpha \notin (-\pi/2, \pi/2)$, the velocity potential energy is negligible.

The velocity potential force is described as

$$F_{repv}(q) = -\nabla U_{repv}(q)$$

$$= \begin{cases} K_{repv} V_{ro}, & 0 \leq \rho(q, q_{obs}) \leq \rho_0 \cap \alpha \in (-\pi/2, \pi/2) \\ 0, & \text{others} \end{cases} \quad (6)$$

The global potential energy that the robot receives at a particular location can be concluded according to the attracting, the repulsive and the relative velocity potential energy presented above. The global potential energy is described as

$$F_{att}(q) = -\nabla U_{att}(q) = \eta \rho(q, q_g) \quad (7)$$

Considering the relative velocity, the resultant force that the robot receives is described as

$$F(q, v) = F_{att}(q) + F_{rep}(q) + F_{repv}(q)$$

$$= \begin{cases} \eta \rho(q, q_g) + k \left(\frac{1}{\rho(q, q_{obs})} - \frac{1}{\rho_0} \right) \cdot \\ \frac{1}{\rho^2(q, q_{obs})} \nabla \rho(q, q_{obs}) + K_{repv} V_{ro}, & 0 \leq \rho(q, q_{obs}) \leq \rho_0 \cap \alpha \in (-\pi/2, \pi/2) \\ \eta \rho(q, q_g) + k \left(\frac{1}{\rho(q, q_{obs})} - \frac{1}{\rho_0} \right) \cdot \\ \frac{1}{\rho^2(q, q_{obs})} \nabla \rho(q, q_{obs}), & 0 \leq \rho(q, q_{obs}) \leq \rho_0 \cap \alpha \notin (-\pi/2, \pi/2) \\ 0, & \text{others} \end{cases} \quad (8)$$

The reluctant force $F(q, v)$ is shown in the Fig. 2.

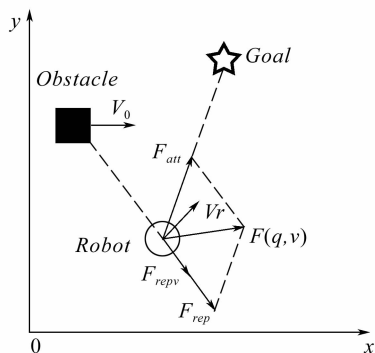


Fig. 2 The force analysis of the robot when moving in 2-D world

B. VAPF-Local A* path planning method

When the size of grid map is $N * N$, the computational complexity of the A* algorithm can reach $O(N^2)$. If the map size rose to $2N * 2N$, the computational complexity would increase to $O(N^4)$. Thus, although the A* can surely find an optimal path, it is generally inadequate for real-time motion for mobile robots only if the situation is not strict with real-time performance.

The VAPF method is more suitable for real-time path planning due to the simplicity of the algorithm, while the local minimal or oscillation possibly triggers instability of the motion. And due to the nature of the field, if the parameters of VAPF were improperly selected, the robot would be in the risk of approaching too closely to obstacle ignoring the possibility of physical contact. One kind

of local minima situation that may cause serious problem is shown the Fig. 3.

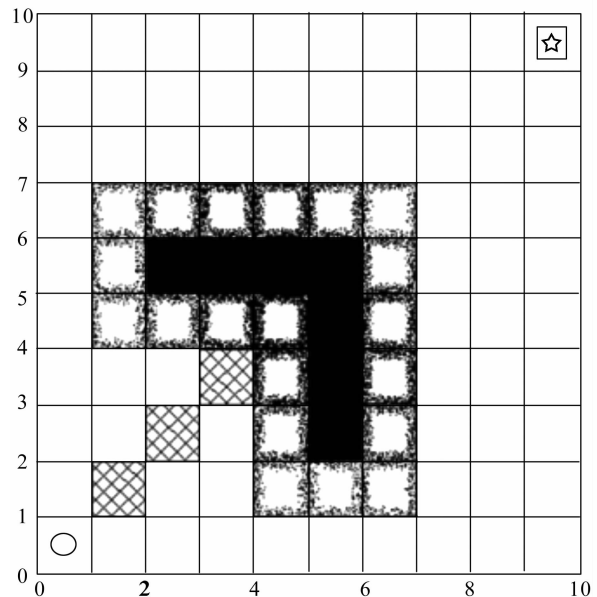


Fig. 3 One kind of local minima condition which can cause serious problems if using VAPF or APF only

Therefore, the VAPF-Local A* is represented here as a strategy for real-time path planning, which is a hybrid method with VAPF and improved A* algorithm. During every planning cycle, the start point and target point is selected after the current grid map has been updated. The real-time path planning is processed with VAPF method until the local minima or oscillation occurs. Then the path planning strategy switches to sub-process using Local A* algorithm to find a proper sub-path.

The information of the map in every planning cycle will be updated firstly. The matrix W_{list} in which storing the motion cost of each grid, where the cost of obstacles are set as 4 and the cost of the other grids as 1. After the dilation process is done, the cost of grids of dilation are set as 4. The matrix P is established, in which storing the global path generated in this planning cycle. Then the VAPF-Local A* starts.

The gains η and k , the ρ_0 and L , the magnitude of single step, are set at first, which are selected according to the situation in this planning cycle. When the robot is away from the target point (X_g, Y_g) , it is inspected whether the local minima or oscillation happens. If happens, the method switches to Local A* algorithm, which is discussed later. When robot arrives target point, the path planning process is over.

The θ_{att} and $\theta_{i_{rep}}$ are needed to calculate the component forces of $F_{att}(q)$ and $F_{rep}(q)$ projecting to x axis and y axis, which are $F_{xatt}, F_{yatt}, F_{xrep}, F_{yrep}$. The θ_{att} is the included acute angular between the X axis and the line linking q and q_g . The $\theta_{i_{rep}}$ is the included acute angular between the X axis and the line from q to q_{obs} . The

V_{ro} is needed to calculate the F_{xrepv} and F_{yrepv} , which are the projections of F_{repv} to X axis and Y axis. The θ_{sum} is the included acute angular between F_{xsum} and F_{ysum} , which are resultant forces $F(q, v)$ projecting to X axis and Y axis. The next position of robot, (x_j, y_j) , is received from L, θ_{sum} , and q . Then, the position is saved into P and it jumps back to check whether target point arrived.

Finally, the P , containing the path generated, is output. Then the next cycle of path planning is ready.

The essential part of the local A* algorithm presented here is selecting new local target point. It is assumed that, before the planning cycle starts, the robot has updated the map information by using sensors for map building such as Lidars, cameras etc., thus there is adequate information for selecting new target point.

The matrices $O_{list}, C_{list}, F_{cost}$ and P_{loc} are established as soon as the Local A* algorithm starts. The O_{list} stores the points needed to be observed. The C_{list} stores points observed. The F_{cost} stores the cost of each point arriving local target point. The P_{loc} stores the local path generated in Local A* algorithm. The initial position of robot, (X_{i_loc}, Y_{i_loc}) , is stored into O_{list} as the starting point.

Then, the local grid map is established and the local target point (X_{g_loc}, Y_{g_loc}) is selected. The local target point is on the straight line joined by (X_{i_loc}, Y_{i_loc}) and (X_g, Y_g) . The new local grid map should contain (X_{i_loc}, Y_{i_loc}) and (X_{g_loc}, Y_{g_loc}) , of which the side length, L_{loc} , is selected in advance according to the density of obstacles. The L_{loc} should fulfill the equation as

$$L_{loc} \geq \max \{ |X_{g_loc} - X_{i_loc}|, |Y_{g_loc} - Y_{i_loc}| \} \quad (9)$$

It is checked whether the (X_g, Y_g) is contained in local grid map. If contained, the global target point is set as the local target point. If the (X_{g_loc}, Y_{g_loc}) is in the same grid as an obstacle point, the local target is needed to move to the farther grid along the line discussed above. And L_{loc} should be added by 1 unit long, too. It is cyclically processed until a non-obstacle local target is found or the global target point is contained in local map.

Since the local start point and the local target point is acquired, the standard A* algorithm can be processed to get P_{loc} . It is assumed that the robot can move diagonally. Considering the robot need turning to new direction, the costs of horizontal and vertical grids are set as D , while the costs of diagonals are set as $\sqrt{2} * D$. In this essay the D is set as 4 which may vary according to the actual environment.

Finally, the local A* is finished along with that P_{loc} is pushed into P . Then, the process switches back to the main path planning process using VAPF method.

The integral process of VAPF-Local A* path planning method can refer to Algorithm 1 and Algorithm 2.

Algorithm 1: MAIN PATH PLANNING ALGORITHM WITH VAPE

```

1   $\eta, \kappa, \rho_0, L \leftarrow$  selected referring to actual condition
2  While not arrive do
3    If Local Minimum or Oscillation then
4      Switch to Local A* algorithm
5    Else
6       $\theta_{art} = \arctan | (Y_i - Y_g) / (X_i - X_g) |$ 
7       $F_{xart}, F_{yart} \leftarrow$  using  $\theta_{art}$  and formula(2)
8       $\theta_{i\_rep} = \arctan | (Y_{i\_obs} - Y_i) / (X_{i\_obs} - X_i) |$ 
9       $F_{xrep}, F_{yrep} \leftarrow$  using  $\theta_{i\_rep}$  and formula(4)
10      $F_{xrepv}, F_{yrepv} \leftarrow$  using  $V_{ro}$  and formula(6)
11      $F_{xsum}, F_{ysum} \leftarrow F_{xart}, F_{yart}, F_{xrep}, F_{yrep}, F_{xrepv}, F_{yrepv}$ 
12      $\theta_{stum} = \arctan | F_{ysum} / F_{xsum} |$ 
13      $(X_j, Y_j) \leftarrow L, \theta_{stum}, (X_i, Y_i)$ 
14     Push  $(X_j, Y_j)$  into  $P$ 
15 End if
16 End while
17 Output  $P$ 

```

Algorithm 2: LOCAL A* Algorithm

```

1   $O_{list}, C_{list}, F_{cost}, P_{loc} \leftarrow$  sizes vary
2  Push  $(X_{i\_loc}, Y_{i\_loc})$  into  $O_{list}$ 
3  While  $(X_{g\_loc}, Y_{g\_loc})$  not find do
4     $L_{loc} = \max \{ |X_{g\_loc} - X_{i\_loc}|, |Y_{g\_loc} - Y_{i\_loc}| \}$ 
5    If  $L_{loc} > L_{thr}$  then
6      Approach to  $(X'_{i\_loc}, Y'_{i\_loc})$ 
7      Push nodes from  $(X_{i\_loc}, Y_{i\_loc})$  to  $(X'_{i\_loc}, Y'_{i\_loc})$  into
8       $P_{loc}$ 
9      Switch back to Algorithm 1
10   End if
11   End while
12   Using stand ar d A* Algorithm to get  $P_{loc}$ 
13   Output  $P_{loc}$  and push  $P_{loc}$  into  $P$ 

```

IV. SIMULATION

The grid map model is established in MATLAB 2014b. The computer carries Intel (R) Core (TM) i5-3210M and 8 GB RAM running Windows 7 64 bit.

Under the condition that there is no obstacle and the program is optimized, the simulation of the standard A* algorithm and the VAPF-Local A* method in this essay is done. Actually, only the attractive potential field exists and functions now. And is set as 1.5. The simulations of the two methods are performed 10 times to get the average time that the MATLAB spends to process path planning. The result can refer to Fig. 4.

The Fig. 4 shows that the path-planning time of both methods increases as the magnitude of the grid map grows, while the time VAPF-Local A* spends is always more than A* algorithm.

Under the condition that the obstacles are sparse, which means there are 5 obstacles that can affect the ro-

bot moving, the A* algorithm and the VPAF-Local A* are simulated in maps of different sizes. The average

time of path planning is shown in Fig. 5.

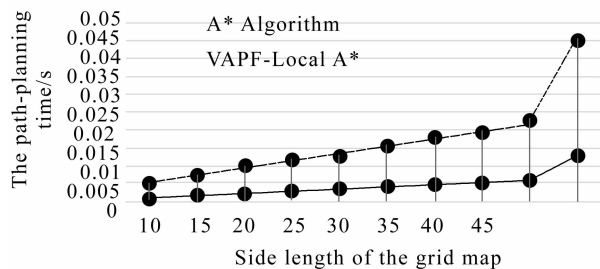


Fig. 4 The path-planning time of A* Algorithm and VAPF-Local A* in map without obstacles. the VAPF-Local A* is roughly 4 times faster than A* Algorithm under this condition

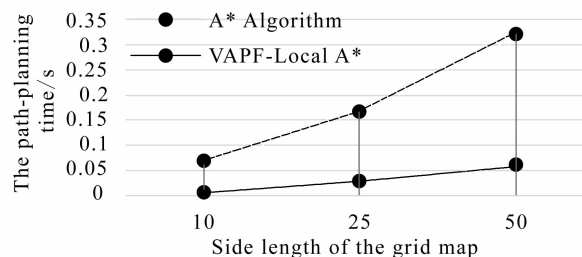
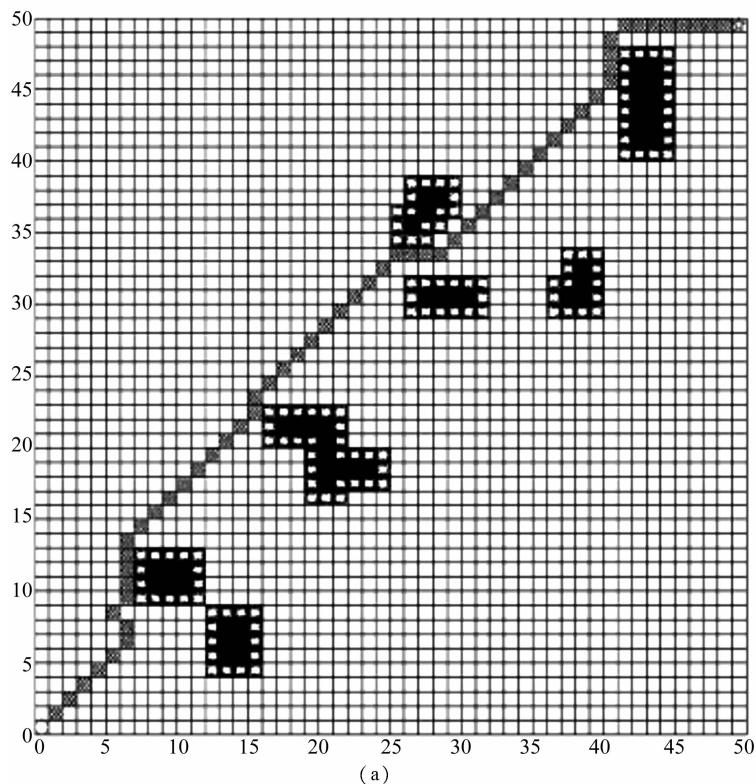


Fig. 5 The path-planning time of A* Algorithm and VAPF-Local A* in grid maps containing sparse obstacles, which means that only 5 obstacles affect the path, the VAPF-Local A* is roughly 6 times faster than A* Algorithm under this condition

The Fig. 5 shows that the path-planning time of VAPF-Local A* is even much more less than the A* algorithm when obstacles are sparse.

There are 41 obstacles in the map and there is a high risk of local minima and oscillation. And the final path is shown in Fig. 6.



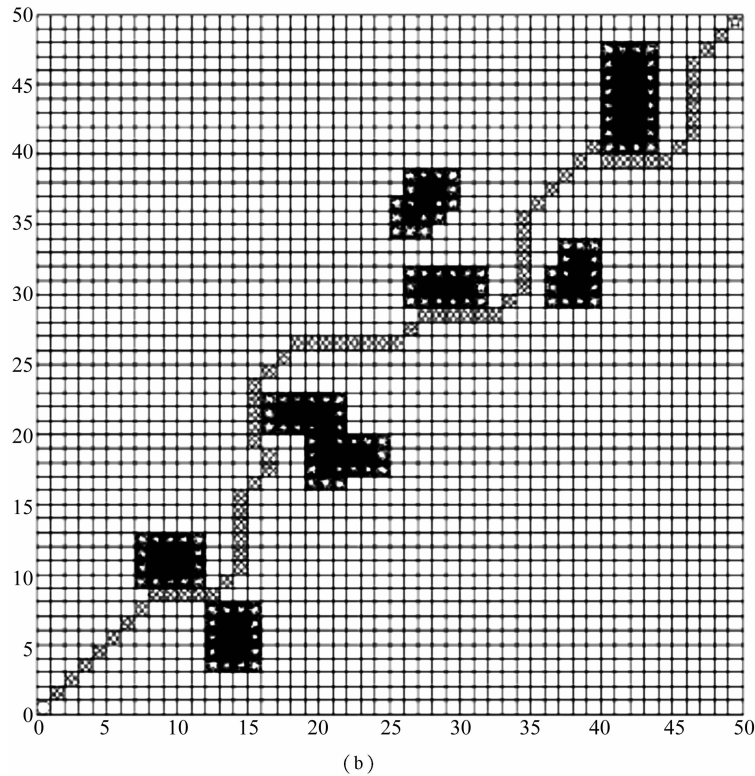


Fig. 6 The simulation of two methods in the map with the side length of 50, the A* algorithm takes 0.374 seconds to get a path of 59 steps, while VAPF-Local A* takes 0.505 s to get a path of 69 steps. The attractive force is rather higher near the start point in large maps, which make the parameters are hard to be selected
(a) path generated by A* algorithm; (b) path generated by VAPF-Local A*

Fig. 6 shows that during the process of VAPF-Local A* planning the path, local minima occurred 4 times. It takes plenty of planning time to get out of the local minima, which prolongs the path planning time of VAPF-Local A*. However, if using the traditional APF or VAPF independently, the robot will stuck in local minima quite early, or the target is not reachable since there are obstacles too close to target.

With the VAPF-Local A* path-planning method, it is relatively much more efficient for the autonomous mobile robot to generated a path if the environment has a moderate intensity of obstacles. Even when the obstacles are intensive, it is feasible to get out of local miniature or oscillation with acceptable path-planning time and total steps for most situations.

V. CONCLUSION

In huge maps where the obstacles spread intensively, traditional AFP or VAPF is possible to encounter local minima or oscillation so that hardly to implement the path planning. When the map is huge enough, A* algorithm can take too much time in allocating resources of computer that debases the timeliness and consistency of mobile robot's motion.

The VAPF-Local A* method is proposed here to insure efficient and real-time path planning when the ob-

stacles are sparse in the map. Still it functions properly encountering more obstacles and larger map. The method processed in grid maps is suitable for the navigation and SLAM (simultaneous localization and mapping) of autonomous mobile robots embedded with sensors providing information for map building, such as Lidar, camera, etc.

The path planned by VAPF-Local A* is not smooth enough. It is needed to be post-processed to get an adaptive path for real-world mobile robots. The research team will focus on the application of this method to the wheeled autonomous mobile robot in further works.

ACKNOWLEDGMENT

The authors would like to thank professor Ding Liang and professor Gao Haibo from Harbin Institute of Technology for providing the research environment for this study.

REFERENCES

- [1] D. Zhu and M. Yan, "Survey on technology of mobile robot path planning. Control and Decision," 2010, 25(7):961-967.
- [2] S. Alejandro, M. Rafael and H. Seth, "An efficient motion strategy to compute expected-time locally optimal continuous search paths in known environ-

- ments," *Journal of Advanced Robotics*. 2009, 23 (12-13) :1533-1560.
- [3] E. Alberto, Using occupancy grids for mobile robot perception and navigation. *Computer*, 1989, 22 (6) :46-57.
- [4] T. Li, S. Sun, Y. Gao, Fan-shaped grid based global path planning for mobile Robot. *Robot*, 2010, 32 (4) :547-552.
- [5] D. Cole, P. Newman, Using laser range data for 3D SLAM in outdoor environments. *IEEE International Conference on Robotics and Automation*, 2006, 1556-1563.
- [6] E. Fernandes, P. Costa, J. Lima, Towards an orientation enhanced astar algorithm for robotic navigation. *IEEE International Conference on Industrial Technology*, 2015, June :3320-3325.
- [7] O. Khatib, Real-time obstacle avoidance for manipulators and mobile robots. *IEEE International Journal of Robotics Research*. 1985, 5(1) :90-98.
- [8] O. Warren, Global path planning using artificial potential fields. *IEEE International Conference on Robotics and Automation*. 1989, 316-321.
- [9] H. Dong, Parameter Selecting in Artificial Potential Functions for Local Path Planning. *IEEE International Journal of Fuzzy Logic and Intelligent Systems*, 2005, 5(4) :339-346.
- [10] Z. Yu, J. Yan, J. Zhao, Mobile robot path planning based on improved artificial potential field method. *Journal of Harbin Institute of Technology*, 2011, 43(1) :50-55.
- [11] C. Cheng, D. Zhu, B. Sun, Path planning for autonomous underwater vehicle based on artificial potential field and velocity synthesis, *IEEE Canadian Conference on Electrical and Computer Engineering*, 2015, June :717-721.

Aspects of Medical Expert System by Pulse Diagnosis

Yumchmaa Ayush^{#1}

[#]Faculty of Automation and Computer Engineering

Novosibirsk State Technical University

Novosibirsk, Russia

¹yumchmaa@ must.edu.mn

Abstract— Nowadays, medical application has been required fast diagnosis. Thus, development trend of computational technology attempts to build a new clinical decision-making support system at the expert-level. One of the fast diagnosis methods in medical practice is pulse diagnosis. This method can be used both of Traditional and European medicine. Especially, pulse diagnosis in the traditional medicine is based on the experience of the physician and he can say to patient about his health condition directly. Therefore, computational multi-diagnostic system to accurate diagnosis pulse palpation is important. The paper discusses about an opportunity to build a medical expert system by pulse diagnosis.

Keywords— diagnosis, computational method, traditional pulse diagnosis, European pulse diagnosis, expert system.

I. INTRODUCTION

Development of computational technology improves the quality of medical diagnosis. Especially, medical application has been required fast and correct diagnosis. Since 1970, scientists have been developed many computational methods in field of medicine. One of these methods is known as medical expert system. This system attempts to make medical decision at the expert-level.

II. MEDICAL DIAGNOSTIC EXPERT SYSTEM BY PULSE DIAGNOSIS

A. Structure of Medical expert system by pulse diagnosis

Any expert system consists of basic parts. There are: knowledge-base, database and user interface.

From this Fig. 1, medical diagnosis expert system by pulse diagnosis may consist of double database.

For example; nowadays, researchers have been built web-based system for field of medical diagnosis. One of these systems is Aimedica system from Russia. This system has built in 2013 and now it has test version. This medical diagnostic system was modeled as an advisory system for doctors, medical specialties and medical students. This system gives responses based on the more than 22 million medical scientific research papers and results of medical practices, which from Pub Med library. In this system have information about 255 different diseases, 2,270 symptoms of diseases, short conclusions about disease, list of the medical specialties correspond-

ing to the diagnosed disease, list of the laboratory analyses necessary for specification of diagnosis, and function of verification of diagnosis [1].

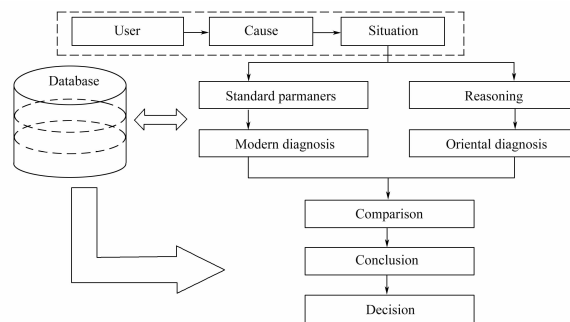


Fig. 1 Structure of medical diagnostic expert system by pulse diagnosis

Second, pulse diagnosis is applied in the traditional and European medicine. In traditional medicine, pulse diagnosis bases on the experience of the physician and he knows all human body state on the pulse wave characteristics. Rather, in the European medicine, necessarily require sophisticated equipment for inspection about human cardiovascular system. The reasons of change blood pressure depend on many factors, and also indicators of blood pressure give important information about a human body state. Researchers build system which used traditional pulse diagnosis and as known as the automated system of pulse diagnostics. This system is registering pulse wave from 6 different points in human arm and simultaneously diagnosing functional condition of 12 internal organs. This investigation proceeds in the field of mathematical processing of pulse signals, creation of environment of expert systems [2].

If such these two databases are will be in one system, it means application of information technologies for medicine and multi-diagnosing process.

B. Mathematical method for Data analysis of Medical expert system by pulse diagnosis

Recent time, mathematicians and physicians are developing several methods of computational diagnosis. It helps to simplify data analysis. All medical diagnosis is based on the pathophysiological problem and they can

be able to question of probability. And medical diagnosis is concepts question of human opinion, too. Because, human makes decisions based on the experience gained something you have learned from his life. In this case, under the concept of probability, we consider it a subjective measure. It helps to doctor to express confidential degree of given sign, a symptom or the diagnosis.

According to holistic viewpoint of all this scientific-technical developments, mathematical methods have recognized as a popular complementary and develop continuously. One of the most used mathematical methods in medical diagnosis is Bayesian approach. Bayesian approach is based on a statistical conclusion and applied in medical diagnosis for many years.

Application of a probabilistic method for diagnosis requires preliminary statistical processing of existing clinical material in the form a large number of case histories. From Bayesian research can be assume, that the patient can have simultaneously only one disease from some group of i of the diseases, denoted a set of diseases. $B \in \{\delta_1, \delta_2, \dots, \delta_3\}$. Thus, diagnosis problem is to choose one of the most probable diseases from many groups. For recognition of disease use set of symptoms $C \in \{c_1, c_2, \dots, c_j\}$, the characterizing any properties of the investigated patients. In traditional medicine, certain value of a symptom is denoted as $\Pi \in \{\Pi_1, \Pi_2, \dots, \Pi_j\}$, and called as an indicator. Also, it is necessary to note on the movement of a pulse wave. The experimental physicians can distinguish by pulse approximately 360 indicators [3]. See in Fig. 2.

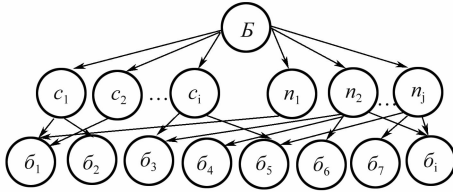


Fig. 2 Conditional probability of disease

For example: From Fig. 1 can denote conditional probabilities of disease $P(\delta_1)$. It can be $P(\delta_1) = P(c_1, \delta_1) + P(\Pi_1, \delta_1)$ or $P(\delta_1) = P(c_2, \delta_1) + P(\Pi_1, \delta_1)$.

All probabilities are conditional. Mathematical definitions of conditional probability of a symptom C at this disease of B according to Bayesian theorem

$$P(C | B) = \frac{P(C \cap B)}{P(B)} \quad (1)$$

Algebra transformations of the equation lead to a basic Bayesian formula

$$P(B | C) = \frac{P(C | B)P(B)}{P(C)} \quad (2)$$

For application of diagnosis probabilistic algorithm must to find previously probabilities of symptoms according to statistical data, for which approximately accept their relative frequencies.

$$P(\delta_i) = P(C | \delta_i) + P(\Pi | \delta_i) \quad (3)$$

δ_i - is one of only set of possible and incompatibility

diseases. Then

$$P(C) = \sum_i P(C | \delta_i)P(\Pi | \delta_i)P(\delta_i) \quad (4)$$

In medical practice, physicians often meet the undefined state. In this case need to find an informal assessment of that corresponds to the observed facts of any diseases [4].

In that case, for calculation $P(C)$ must previously to calculate work of the form $P(c_1 | c_2 \wedge \dots \wedge c_j)P(c_2 | c_3 \wedge \dots \wedge c_j) \dots P(c_j)$. Independence symptoms and indicators $c_j \Pi_j$ follows $P(c_j | \Pi_j)$, from what follows relation $P(c_j \wedge \Pi_j) = P(c_j)P(\Pi_j)$. In the European medicine, if all symptoms are independent, at the diagnosis accept only symptom. And in traditional medicine, physicians consider all indicators of disease depend on pulse wave. Bayesian approach allows to calculate all values $P(c_j | \delta_i)$ and to make reasoning of mutual independence symptoms. In medical diagnostic expert system, if the patient has indications and symptoms $P(c_1 \wedge \dots \wedge c_j)$, and there are certain conditions $P(\Pi_1, \Pi_2, \dots, \Pi_j)$, Then can to conclude confidence, that the patient has a disease δ_i [4, 5, 6]. Bayesian formula is accepted for a case of all symptoms.

III. ADVANTAGES DIAGNOSTIC MEDICAL EXPERT SYSTEM BY PULSE DIAGNOSIS

Development of computational technologies gives the chance to reach new level at present visually the course of a disease according to mathematical models. Any attempt of developing a medical diagnostic expert system dealing with human disease diagnosis has to overcome various difficulties. Expert systems for a medicine developed usually for physician, nurse or for procedure of healthcare. Doctors have own features of the experience and feelings. Therefore, they must estimate management of patient treatment and patient's situation of diseases. Most of medical expert systems have the content of database mainly contains the current situation of patients. For example: name, age, sex, symptoms [7]. Most medical decision-making programs have based their advice on the data available at one particular time. In actual practice, the physician receives additional information from tests and observations over time and reevaluates the diagnosis and prognosis of the patient. Both the progression of the disease and the response to previous therapy are important for assessing the patient's situation [8].

Complete medical diagnostic expert systems can create new medical knowledge learning is seen to the effective characteristic of an intelligent being. They contain medical knowledge about a very specially defined task, and are able to reason with data from individual patients to come up with reasoned conclusions. Consequently, one of the major ambitions of artificial intelligence has been to develop computational technologies that can learn from experience of experts. The resulting developments in the artificial intelligence field of medical diagnostic expert systems have resulted in a set of tech-

niques which have the potential to describe the way in which knowledge is created. Although there are many variations, the knowledge within an experts system is typically represented in the form of a set of rules [9].

Expert systems are the commonest type of artificial intelligence systems in routine clinical application. Now this goal has been moving to more high-level. Nowadays, there are online versions of medical expert systems, where any user can establish the diagnosis with this, or that share of probability. The old system can be easily incorporated into the new system with the reused modules. The theoretical mathematical framework of new expert system has a key component. And also, the automated diagnostic systems can be equipped with the sophisticated equipment for inspection of pulse, which is based on "smart" technology [10].

While working on the research work is to get the following scientific results:

- ① To develop models of interrelations of traditional and European medicine;
- ② To develop fuzzy models for "symptoms-diseases";
- ③ To develop the algorithms of consecutive diagnostics directed to allocation of the most probable disease;
- ④ To develop the human-machine interfaces for interaction of the doctor and user;
- ⑤ To develop the diagnostic medical expert system interacting with the diagnostic equipment (pulse control and etc.).

Recent time, modern medicine is focusing big conception that to teach early prevent disease and probable diseases identify source properly. It helps to early diagnosis any diseases. This means to use development of computational technology for a medicine and increase opportunities multi-diagnosing process. Major goal of this research work is that to increase the probability assignment to the diagnostic possibilities in European and traditional medicines.

Advantages are:

- ① To have integrated medical diagnostic expert system;
- ② Help saving time of the physician for diagnosis of any diseases;
- ③ To improve quality of medical diagnosis;
- ④ Creating integrated database of traditional and European medical diagnosis for medical specialties and student's training;
- ⑤ Giving some opportunities for early diagnosis any diseases before will be chronic;
- ⑥ Giving explanation about cause of diseases based on the traditional and European medicine.

IV. CONCLUSION

Nowadays, expert system in field of medicine has been developed. It helps to doctors to express confidential degree of given sign, a symptom or the diagnosis. Use the traditional method has the advantage of fast diagnosing the disease. This traditional diagnosis to char-

acterize pulse wave, which based on the experience of traditional physician, introduce to this modern web-based expert system will be create newly quality and fast medical multi-diagnostic expert system. And this system will help to medical specialties to make correct diagnosis of disease and will be highly level training tools for medical students and clinical youngest physicians. For best result, it consider on the traditional pulse diagnosis. It research work is devoted to development of citizen's intellectual primary medical diagnostics methods and it's an actual task. Modern approach to such kind of expert systems offers their content on the basis of million sources. There is also big gap in terms of traditional and European medicine.

Whole of pulse diagnosis in traditional medicine bases on the experience of physician that he collected during his whole life practice. Therefore, in the future this multi-diagnostic medical expert system will be called experience-based system.

ACKNOWLEDGMENT

Thank you to Scholarship of the Ministry of Education and Science of the Russian Federation, that for gave me chance to do research work about Medical Diagnostic Expert System.

REFERENCES

- [1] Aimedia; web-site. Available at: <http://www.aimedia.ru/info.jsp>.
- [2] FGDN, IFM SO RAN, project [Project of The Buryat Scientific Center of the Siberian Branch of the Russian Academy of Sciences]. Automatizirovannii pulisodiagnosticheskii kompleks tibetskoi meditsini [The automated pulse-diagnostics complex in Tibetan medicine]. Available at: <http://ipms.bsnet.ru/labs/lvdjs/resultats.html>.
- [3] Tsidipov Ts. Pulisovaya diagnostika Tibetskoi meditsini [Pulse diagnosis in Tibetan Medicine]. Novosibirsk, Nauka Publ., 1988. 133 p.
- [4] Lektsii. Net; web-site [Lectures. Net; website]. Available at: <http://lektsii.net/1-46254.html>.
- [5] Grif M. G, Yumcmaa A., Primenenie ekspertnykh system pulisovoy diagnostiky [Application of expert systems by pulse diagnostics] Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta-Transaction of scientific papers of the Novosibirsk state technical university, 2015, no. 3(81), pp. 114-133.
- [6] Lusted L. Vvedenie problemi prinyatiya reshenii v meditsine. [Introduction to medical decision making], Moscow, Mir Publ., 1971, 235 p.
- [7] Xiaoping Li, Junhui Chen, Qiong Xu, Hujian Dong. Research of the medical expert system under a new architecture. 2nd International Conference on Industrial Mechatronics and Automation, ICIMA 2010, Vol. 2, 30-31 May 2010, China, 2010, pp. 244-247.

- [8] Shortliffe E. H. , Fagan L. M. Expert systems research: Modeling the medical decision making process. Workshop on Integrated Approached to Patient Monitoring, 5-6 March 1982, Stanford, USA, 1982, 23 p. Available at: <http://i.stanford.edu/pub/cstr/reports/cs/tr/82/932/CS-TR-82-932>.
- [9] Prasadl, Kristina Prasad, Sagar Y. An approach to develop expert systems in medical diagnosis using machine learning algorithms (asthma) and a performance study. International journal on soft computing(IJSC) , Vol. 2 , № (1) , 2011 , pp. 26-33. Availale at: <http://airccse.org/journal/ijsc/papers/2111ijsc03>.
- [10] Uranchimeg T. , Uyanga S. , Yumchmaa A. A Prototype of Expert System for Rural Medical Centers. 7th International Conference on Frontiers of Information Technology, Applications and Tools, and the 4th PT-ERC International Symposium on Personalized Medicine, FITAT/ISPM 2014, 29 July-01 August 2014, Thailand, 2014, pp. 23-26.

Case Study Method in Teaching English for Specific Purpose

Narantsetseg Ravjaa^{#1}, Tatyana Boroznets^{#2}, Ariunaa Guntsentsoodol^{#3}

¹rna9mn1@yahoo.com

²talebor@mail.ru

³mesforlanar@yahoo.com

[#]Dept. of English study, IFL, MUST
Ulaanbaatar, Mongolia

Abstract— As the country's demand, Mongolian educational organizations must prepare well trained and educated specialists that are able to work in any fields of their professions corresponding to the international standards. Teaching ESP (English for Specific Purpose) is much different from teaching general English with its purposes. Through ESP, English teachers give their students not only language knowledge and skills, but also give them employable skills. In other words, students can acquire the skills to use their English language knowledge in their work places, so the methods used to teach ESP have to be different. This paper aims to introducing one of the active methods of teaching, which is called "case study". Case study involves real stories of real business organizations including their success and failure, presents complex situations that are contextually rich on dilemmas, conflicts, and problems. The method of case study gives students possibilities to improve their language skills to solve problems, negotiate, etc. . Thus, this method develops students' critical thinking, consolidates already acquired knowledge, and work place skill-managerial skill, as well.

Keywords— Workplace skill, methodology, confidence, education, real situation

I . INTRODUCTION

Nowadays, the main role of teachers is not limited by transferring the information directly to students. Also, they should be creative users of technology, leaders among the public advisers for students, experts for training process and educational servants in society.

The result of any learning process depends on many factors like individuals' interest and their psychological readiness to raise their qualification and their ability to acquire modern learning techniques.

Today's widely used methodology is " student centered teaching". It is a teaching technology based on competitiveness (ability) and at giving independent studying skills for learners.

The training that is conducted in globalized environment based on advanced methodology has spread in many countries. In other words, it is the combination of training the teacher and learner do research together and in turn, teachers encourage learners. These days, advanced training methods have been widely used in many colleges and universities in the world.

II . WHAT IS CASE STUDY?

The important way to make a decision is a case study which aims to providing the ability and skills to perform the main function of science creatively. It is a practical studying method that requires intellectual skills, general mental process as well as all kinds of scientific knowledge. Case means occurrence, situation and extra example. Also, in the study of education, it is described as a research of situation. Since ancient times, foundation of case study has been laid as a performance of debate and discussion. The case study is as follows:

- ① the most practical and fastest research method to diagnose and analyze the situation and find the appropriate ways;
- ② exercises to describe the situation and raise a question using the information from the specific organization;
- ③ exercises to learn from the historical events analyze and study experience [1].

The purpose of the case study is as follows:

- ① to select the necessary information;
- ② to elaborate the information gathered;
- ③ to learn how to analyze;
- ④ to suggest many versions and ideas;
- ⑤ to develop logical skills;
- ⑥ to work in team;
- ⑦ use theoretical knowledge.

The case study is considered as a main method of training and is used in the leading and the most reputable universities of the world.

The " case study" for students introduces the real situations, gives them the possibility to test themselves, increases the confidence to defend their ideas and get along with members of team, and teaches how to use knowledge and skills the students acquired during the class. The authors have noticed that the students who use the " case study" in class tend to be motivated, get involved in lessons, brainstorm quickly, raise interesting ideas, accumulate experience, and have more confidence as well.

The " case study" provides teachers possibilities to use theoretical knowledge to be understood by students in the fastest way and the goal to be achieved. It makes

teaching and learning process more interesting and helpful.

Case is occurrence, situation and extra example of a certain company or business organizations as described above. It could be a company's privacy or secret. So in order to use the "case study" method, a teacher should get a permission from an organization to gather the information, define the situation precisely while doing a research based on a real situation or data. So the real names and other relevant information of the company and employees should be changed.

The sequence of class activities for case study includes reading the prompts by students and imagining it in their minds, selecting the useful information which is needed for the suggestion, diagnosing, raising and comparing many versions, analyzing and approving.

We must mention that there are no absolutely true answers, the main goal of using this method is that a teacher must support and encourage students to express their own ideas, direct and demonstrate students the ways definite problems were solved in the history.

III. CASE STUDY AND ESP

Case studies in English for Specific Purpose may vary in performance and can be used differently, depending on the case itself and on the goals of Engineering, Business and Social Science English courses. The process of creation of sample case is complex and is carried out in a number of stages, which are:

- ①determination of didactic aims that should be clearly and laconically formulated;
- ②choosing a case layout;
- ③gathering of relevant facts;
- ④creating a model of the situation;
- ⑤selecting a type of case;
- ⑥writing a case in words;
- ⑦analysis of precision and efficiency of case;
- ⑧using of a case in the process of teaching.

An effective case study is one that:

- ①tells a "real" and engaging story;
- ②raises a thought-provoking issue;
- ③has elements of conflict;
- ④promotes empathy with the central characters;
- ⑤lacks an obvious or clear-cut right answers;
- ⑥encourages students to think and take a position;
- ⑦portrays actors in moments of decision;
- ⑧provides plenty of data about character, location, context, actions;
- ⑨is relatively concise [2].

According to [2], six steps provide a common basis of how to conduct a decision, shown in Table I.

Table I
SIX STEPS PROVIDE A COMMON BASIS
OF HOW TO CONDUCT A DECISION

	Content	Sample tasks to resolve
1	Give the group enough time to read and think about the case, even to prepare a company advertisement as home task in form of a film or slide-show. *)	What is the main sense of the case? What are the alternative actions? What are the main challenges the central character is facing?
2	Introduce the case and give some instruction for how to approach it, how you want the students to consider this problem.	" You are a member of an advertising campaign for one of the products or services. Present your campaign to the management concerned. "
	Break down the stages you want participants to take in developing the case.	" Firstly..., secondly..., finally.... "
	Specify certain information you want your students to concentrate on.	" I want you to ignore the political views of the countries "
3	Form groups and monitor them to be convinced that all students are involved in discussion. Formulate the task for each group precisely.	Each group will be preparing a mini-presentation on one of the companies in their own words based on the information given in a form of analyst's report
	Make sure that even very timid students could participate in the process by distributing the roles inside each small group.	" The student playing role of an executive director of a Mongolian company would chair the meeting. " Another student should take notes about the discussion, and there may be role-cards for every student.
4	Have students present their solutions, write their conclusions on the board, then return to the students involving them into oral discussion, thus encouraging them to be sober.	

Table I Continued

	Content	Sample tasks to resolve
5	Ask clarifying questions to move discussion to another level and probe for deeper analysis without over-directing.	Examine the students' own assumptions, substance of their claims and illustrations.
6	Synthesize issues raised to bring different strands of the discussion back together at the end of lesson.	Show the students what they have learnt and make feedback from their work. One or more students can implement the task, not necessarily the instructor.

*) An original student's work is available

Some variations on this general method include having students do outside research individually or in groups to bring to bear on the case in question, and comparing the actual outcome of a real-life dilemma to the solutions generated in class [3].

IV. CONCLUSION

In summary we should note that we have used the above-described case-study method in our English language teaching process at MUST. Using the case study in class is beneficial, because it gives students a perfect opportunity to use creatively the active vocabulary gained on the lessons of ESP, to show their professional knowledge and skills, and to adapt to the real and potential situations. Then it helps students to learn how to work out the information, to work in team, and develops logical mental process. If teachers use more methodologies like a case study in their teaching process, students will be able to deal with all the problems they will face

in their future work.

Also, it improves quality of teaching. The method provides the complete theoretical knowledge, develops creativeness of both teachers and students. Also, it helps to achieve the main goal of educational process, i. e. preparation of highly qualified specialists in all fields of society.

ACKNOWLEDGMENT

The authors would like to express our thanks to T. Hutchinson and A. Waters, Sunjidmaa A. and Bathurel G. for their contribution with great ideas to support our paper. These people's works motivate the authors to develop our teaching experiences and practical knowledge in a more theoretical way.

We are also grateful to the students who helped the authors with their hardworking to be born this paper.

REFERENCES

- [1] T. Hutchinson and A. Waters, *English for Specific Purposes: A Learning-Centered Approach*, Cambridge, Britain: Cambridge University Press, 1997.
- [2] Davis B. G. *Tools for teaching*. San Francisco, V. S. : Jossey-Bass, 1993.
- [3] G. Tatyana, Beckisheva, et. al. , *Case study as active method of teaching Business English* (www. sciencedirect. com).
- [4] T. D. Evans and M. J. S. John, *Development in English for Specific Purposes: A multi-disciplinary approach*, Cambridge, Britain: Cambridge University Press, 2000.
- [5] A. Sunjidmaa, G. Bathurel, *Case study will be one of the contemporary methods for training*. 2013. Proceedings, IFOST, Vol. 2 Ulaanbaatar.
- [6] J. Munby, *Communicative Syllabus Design*, Cambridge, Britain: Cambridge University Press, 1998.

Concept for a Multi-User Dialog Interface for Museum Guides

René Schmidt^{#1}, Wolfram Hardt^{#2}

[#]Computer Science Department, Technische Universität Chemnitz
D-09107 Chemnitz, Germany

¹rene.schmidt@informatik.tu-chemnitz.de

²wolfram.hardt@informatik.tu-chemnitz.de

Abstract— Interaction with speech is the most natural way of communication. Since Google and Apple have demonstrated how speech recognition engines can simplify the interaction between humans and computers, speech recognition becomes increasingly important in different research areas. In the field of museum guides there are already numerous robots using speech as communication interface, e. g. Robotinho or Fritz. This paper describes a concept for a spoken dialog systems implemented in a museum guide, directs to complete computability in order to realize a more intuitive dialog behavior. We identify state of the art dialog system properties and introduce a concept addressing hugely reverberant recognition environments. Our proposal provides speaker separation and identification which is a fundamental feature in multi-human communication scenarios. Additionally, the concept provides a calculable processing time for adapting the response times in respect to content and question type.

Keywords— Multi-User Dialog Interface, Robot, Speech.

I. INTRODUCTION

Museum guides like Robotinho [1], Fritz [2] or Minerva [3] have been established in the museum area. Most recently used as communication interface is the touch interface. However, the most intuitive way to interact with a machine is the human speech. Some robots like Fritz use spoken dialog systems to communicate with the visitors [2], but mostly this dialog interfaces are just rudimentary systems which try to realize intuitive dialogs by using filler words if the calculation time takes to long [4]. On the contrary, psychological research shows that a suitable amount of time between the moment of questioning and the moment of answering depends on the context and question type [5] and has to be adapted. In order to archive this goal, a complete predictable system has to be realized with the aim to reduce the processing time to a minimum and a huge amount of computability. For the use on a mobile robot a specialized hardware is necessary. Current solution would offer a digital preprocessing in an FPGA or ASIC and subsequent further processing in a processor. The disadvantage of this approach is, that the possibility's for

a suitable hardware/software partitioning is very restricted, because of high transfer times between FPGA and processor. At this point, the new SoC techniques from Xilinx and Altera provide completely new possibility's for a more efficient hardware/software partitioning, due to the common use of AXI-Interfaces. On this account, a efficient usage of the parallelism characteristics of the FPGA can be drastically increased. Therefore, the computability of a system can be enhanced, by using the pipeline calculation of the FPGA. Due to the usage of standalone applications, there is no additional delay based on interrupt from operation systems. In this manner, a deterministic process is ensured. Another challenge in museum context is the necessity to interact with multiple users. On this account, a usage of multi-user applications is inevitable. For this purpose we introduce a multi-user dialog system directed to the use on a mobile robot. In the following, first we give some general information about Spoken Dialog Systems, before we describe our Top-Level concept, including the definition of the main features. Afterwards we outline the main units and discuss different options for each algorithm. Finally we draw our conclusion.

II. SPOKEN DIALOG SYSTEMS

Spoken dialog systems is a fast growing field of research. In the last years, numerous commercial applications have been developed, like "Siri" by Apple or "OK Google" by Google. Older applications were deployed usually in call centers, e. g. HMIHY [6] or information systems like JUPITER [7] or the DARPA communicator [8]. Nowadays, Spoken Dialog Systems have been deployed in smart home environments, e. g. SCARS [9] or intelligent robots [1], [10], [2].

An example for a conventional Spoken Dialog Systems is shown in Fig. 1. Its main components are the Speech Recognition Engine (SR), the Natural Language Understanding (NLU), the Dialog Manager and the Text to Speech Engine (TTS). The Speech Recognition Engine converts the analog signal of a spoken language to a sequence of parameter vectors, which is transformed into a sequence of words. Afterwards, the Natural Language Understanding extracts the semantic of utterances and passes it to the Dialog Manager. The Dialog Manager analyses each utterances and based on the respective his-

¹The author gratefully acknowledges funding by the DFG (GRK 1780/1).

²We gratefully acknowledge the cooperation of our project partners and the financial support of the DFG (Deutsche Forschungsgemeinschaft) within the Federal Cluster of Excellence EXC 1075 "MERGE".

tory and the actual utterance, makes a decision of the next output. This output is transformed by the Text to Speech Engine into speech [11], [12], [13].

These days commercial approaches apply a Voice Browser based approach. It is fed with documents described in speech specific languages like Voice XML. The Voice Browser communicates with a web server, which can activate the speech resources e. g. Speech Recognition Engine, Text to Speech Engine. The respective result is sent to the web server, which proceeds on the received data and send the final result back to the Voice Browser. On this account, the latter is able to perform the deficed actions [13].

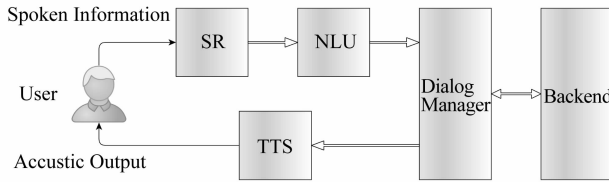


Fig. 1 Traditional Dialog System

However, this concept focus on a multi-user dialog system optimized for the usage on SoC Architectures, with the focus of reducing the processing time as good as possible such as a calculable response time. For this purpose a web solution is not possible, because the delay for the web communication always depends on too many parameters. In summery the traditional spoken dialog approach provide a good basis, but has to be adapted in order to work with multiple users and deal with the rough museum environment.

III. ARCHITECTURE

A. Overview

The development concept provides a modular approach, which is easy to maintain and a highly fiexible research platform. Fig. 2 composes an overview of the concept archi-ecture. As input source is a microphone array(MA) of M microphones and the produced output $S_m(t)$ defined, where $1 \leq m \leq M$. $S_m(t)$ describes the digital output from microphone m at time t . The microphone array feeds the Preprocessing Unit(PPU), which full fills the major task to reduce reverberation in the room. The filtered signal is represented by $S_{P_m}(t)$. Since the system shall be implemented in a museum guide, different speakers have to be separated for the increased possibility that more than just one person or group talking to each other. For this purpose, the Speaker Identification Unit(SIU) separates different speakers and proceeds a speaker localization. The result is represented by S_{B_n} with $0 \leq n \leq N$ while N displays the amount of speakers. Based on this data, the Speech Recognition Engine(SRE)process transforms the given digital waveform signal into a sequence of para meter vectors for each speaker. Afterwards these vectors are converted in-

to a word sequence $S_{w_n}(t)$. The Dialog Manager(DM) collects the generated data and decides about the relevant speaker, based on direction, distance and level of loudness. Afterwards, the Dialog Manager proceeds the utterance analysis and decides, based on his policy, about the output $O(t)$. In the end, the speech is synthesized by the Text to Speech Engine(TTS).

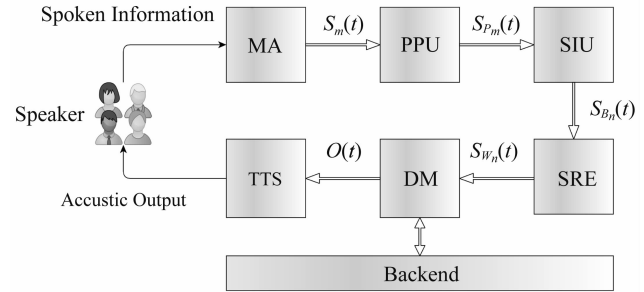


Fig. 2 Concept Overview

B. Preprocessing

The Preprocessing Unit is fed by the microphone array output $S_m(t)$. The major task of this unit consists of preparing the signal for the whole speech recognition process. Since this concept is developed for evaluating communication in a museum, the most important task the Preprocessing Unit faces consists of removing reverberation in big halls. Actual dereverberation techniques can be divided in two big categories, Front-End-Based Approaches and the Back-End-Based Approaches. Front-End-Based approaches aim to dereverberate corrupted feature vectors, while Back-End-Based Approach aim to adapt the acoustic model or tailor the decoder to the corrupted model. Since we also need the dereverberation for the localization algorithms(see chapter III-C3) and not just for the speech recognition, a Front-End-Approach has to be chosen. The Front-end-approach has the advantage that no changes in the Back-End processing steps are required and they are completely independent of the following calculations. Additionally, most of the Front-End-Approaches can be implemented with less processing costs and combined with noise reduction methods [14]. In Order to find a well suited algorithm, several algorithms has to be evaluated. Evaluation parameters should be the increasing accuracy of the localization algorithm, possibility for parallelism on the FPGA, resource usage and calculation time in Software rather throughput time on the FPGA.

C. Speaker Identification

The Speaker Identification represents one of the major parts of the proposed concept. It consists of four additional components, the Speaker Separation Unit(SSU), the Speaker Estimation Unit(SEU), Speaker Localization Unit(SLU) and the Beamforming Unit(BFU). The Speaker Separation is needed for the decomposing of different sound sources and different speakers. The re-

sult of the Speaker Separation is a record $S_{m,n}(t)$ which represents every detected speaker nper microphone m at the time t . This results in a mapping problem, since the result of each speaker has to be mapped to the same speaker on an other microphone. On this account the Speaker Estimation generates a hypothesis about the correlated speaker and is represented by $S_{E_{m,n}}(t) \equiv (S_{m,n}(t), L_{m,n}(t))$ where $L_{m,n}(t)$ represents an estimation parameter for speaker n on microphone m . $S_{E_{m,n}}(t)$ is fed to the Speaker Localization, which calculates the position of the speaker in 3D room, in order to track the speaker. The result can be displayed as $S_{L_{m,n}}(t) \equiv (S_{m,n}(t), pos_{m,n}(t))$. The Beamforming performs the fusion of the different microphone signals with the result $S_{B_n}(t) \equiv (S_n(t), pos_n(t))$, with $S_n(t)$ represents the combined signal for speaker nover all Mmicrophones, mapped by the corresponding position of the speaker. $S_{B_n}(t)$ represents the final result of the component Speaker Identification and is passed to the Speech Recognition Engine.

(1) Speaker Separation

Speaker Separation or Blind Source Separation is a topical research area, where numerous methods and solutions have been developed [15], [16], [17]. The basic idea is to separate the speakers voice from undesired background noise. Most of the approaches based on the Independent Component Analysis(ICA) with a Principal Component Analysis(PCA) as preprocessing technique [15], [16], [17]. The ICA assumes statistical independence of the original signals. However, the ICA can separate the maximum number of sound sources as microphones are available. Zhe Wang et al. [18] overcomes this problem. They proposed a blind source separation algorithm, that does not depend on the amount of microphones. This algorithm is an optimized Voice Activity Detection(VAD) algorithm combined with a noise removing technique and rests upon spectral subtraction, Zero-Crossing-Energy VAD and Entropy-Based methods. Next to the independence of the available number of microphones, the described algorithm is independent from prior clean speech variance estimations and does not require additional models or trainings. In order to rind the best fitting approach, both techniques has to be evaluated regarding the possibility of parallelism, processing time, resource usage and throughput time.

(2) Speaker Estimation

During the process of Speaker Separation different waveform for each speaker are generated and certain noise suppression are performed. The results can be displayed as $S_{m,n}(t)$. On this base, the component Speaker Estimation calculates identification features in order to map a speaker to its corresponding parts on the other microphones. This aspects is addressed within the Speaker Recognition or Speaker Diarization, which describes common fields of research [19], [20], [21], [22]. Since we have multiple microphones recording at the same time, we can assume that their features are

quite equal. [23] shows that the LPCC feature gives a solid first estimation on the speakers identity, with an error rate less than 22%. If additional pitch features and maximum autocorrelation values (MACV) are used, the identification error can be reduced to less than 16%. But it is also shown that, considering pitch and MACV feature on their own, provides bad estimations with error rates higher than 60%. Nevertheless, developing our concept we have to evaluate whether MACV feature provide a suitable first estimation or if LPCC feature has to be used. Also for this algorithm processing time, resource usage and parallelism capability has to be evaluated. As result of this component a record, represented by the current separated speaker waveform for each microphone and the added feature, is obtained.

(3) Speaker Localization

The Speaker Localization calcu-lates a location estimation of the speaker, in order to get information about his or her position in the 3D space as well as the distance from the system. This enables the Dialog Manger to estimate, if the detected speaker is talking to the system or if the detected data are irrelevant. Additionally, such information can be used to provide an eye-to-eye communication while used on a robot.

Within the field of Speaker Localization or Sound Source Localization various methods have been developed to calculate the position of a sound source. They can be separated into three categories, beamforming methods [24], [25], time difference of arrival (TDOA) methods [26], [27] and techniques which adopt the measured Head Related Transfer Function (HRTF) [28]. [29] presents an approach based on the summed GCC, in order to calculate the elevation angle and azimuth angle for speaker localization in 3D space. However, most of the algorithms just provide angels but does not calculate the distance of the speaker. Valin et al. [27] proposes an TDOA based approach with distance measurement in a three meter range. According to the authors, the low range is based on the noise and reverberation of their laboratory. Though beamforming methods need a high number of microphones and heavy data processing to achieve valid results, whereas HRTF methods need big databases of the specific robot platform. On the contrast, TDOA needs a moderate amount of microphones and has quite low computational coasts [29]. Based on this comparison and the usage of the calculated time difference between the microphones which can be used for beamforming (see chapter III -C4) an TDOA based approach is most suitable. Since there are just localization approaches which calculate angle and distances and define the position of a speaker relative to the system, a new algorithm for calculating the position in 3D space based on TDOA has to be developed.

(4) Beamforming

The Beamforming component has to merge different corresponding speaker signals. There are numerous methods for this microphone array processing problem. The classic solution [30] is based on the calculation of $y[n]$ with

$$y[n] = \sum_{m=0}^{M-1} \alpha_m x_m[n - \tau_m] \quad (1)$$

where α_m symbolize a weight for microphone m (standard $\alpha_m = \frac{1}{M}$) and τ_m defined in (2) by using speed of sound v and base distance between the microphones b

$$\tau_m = \frac{b \cos \theta}{v} \quad (2)$$

Another approach comprises the adaptive array processing, using an Generalized Sidelobe Canceller (GSC), as an alternative to the Frost beamformer [30]. The GSC calculates a fix beamform and uses a blocking matrix to block signals from the gaze direction, in order to remove unwanted signals. [31] propose special beamforming algorithms for speech recognition. This

approaches bases on trained filter-and-sum array processing, optimized for speech recognition with improvements of 36% compared to classical beamforming methods. However, in our concept we calculate angel θ in the Speaker Localization component and b is constant over the whole system, providing perfect requirements to calculate the beamforming with the classic approach, since τ_m can be directly calculated and used in equation (1) for the final merge process. However, it has to be evaluated whether the beamforming algorithm provide the necessary accuracy for the following speech recognition. Also resource management and timing analysis has to be done, in order to find the best hardware/software partitioning.

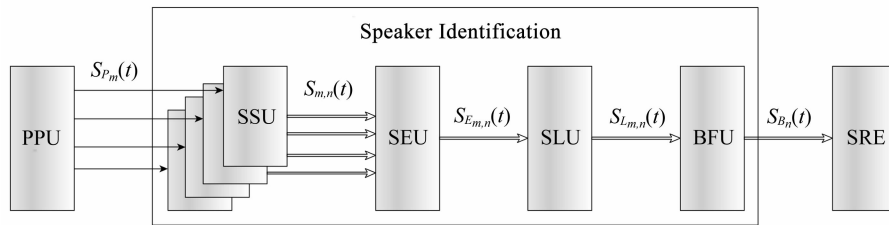


Fig. 3 Speaker Identification

D. Speech Recognition

In this section, the conversion from digital waveform into word sequence is described. On this account, we have to define the following requirements. The main focus is a constant processing time. However, the highest processing time in a dialog system has the speech recognition engine. Standard speech recognition engines wait with the processing until the speaker stops speaking, afterwards the processing starts. This behavior leads to a big incalculable delay, which is a big problem for our concept. However, this concept still rely on standard ASR Technologies, but not on the usually behavior. In order to get a fixed processing time, the system will split the utterances in constant windows. This windows will be proceed separately. The described processing leads to another problem. The benefit of the NLU is getting less effective which leads to higher word error rates. In order to face this problem, this concept focus on keyword spotting approaches, just as shown in [32]. This approach has another big advantage related to the German language. In Germany, a huge amount of various dialects exists, which has to be understood by the system. These dialects can differ substantially and result in bad ASR outcomes with word error rates of approximately 30% challenging the system rather hard [33]. In future work, it has to be evaluated whether the usage of windowed utterances leads to a constant processing time and the reliability of the keyword detection has to be determined.

E. Dialog Management

The Dialog Manager is fed by the keyword sequence $S_w(t)$ for each speaker n . This leads to the task of to identifying the relevant speaker, based on the position in 3D space and loud-ness level mainly, because of the high probability of a speaker in the near environment in front of the robot talking to the system rather than a quiet voice from the background. The non sort out speaker are analyzed by the system, and the content retrieval takes part. Classic dialog management systems would use finite state machines for modeling the flow of the dialog [13]. However, this approach has the huge disadvantage, that the complexity of the created graph spreads massively with growing knowledge database resulting in heavy maintenance coasts. Currently, two major approaches are suited to the described concept, the ontology based approach [34], [35] and the POMDP [12] based approach. The POMDP approach rests upon a observable Markov Model, whose transitions and observations are modeled as set of probabilities defining the dialog model. A second statistical model describes the next action and therefor the policy of the system [12]. A core limitation of this approach comprises the fact, that it requires a lot of training. On the other hand, the ontology based approach provides a more flexibile and easier reconfigurable system, which provides easy access to databases while using SPARQL [36], [37]. The exhibits can be organized in a hierarchical ontology, improving scaleability. Bearing in mind, that the concept need a constant and low processing time, the best per-

formance is expected by the ontology approach since the SPARQL database is provided by the embedded device itself. In order to evaluate the dialog manager, the processing time has to be evaluated in respect to the database size and the amount of possible keywords. Based on that it has to be analyzed, whether a processing time can be calculated based on the amount of keywords and the database size.

F. Natural Language Generation

Finally the system has to generate the output speech. On this account we analyzed different text to speech (TTS) synthesizers, Mary TTS V 5.1.2 [38], Festival TTS V 2.4 [39] and MBROLA V3.02b [40], regarding to their nature appearance, changeability and amount of prepared German voices. German voices are provided by all synthesizers except Festival TTS, which just comprises English and Spanish voices. With relation to changeability MBROLA provides the biggest amount of freedom by enabling the user to define every phone by his own. Mary TTS offers a Voice Import Tool, giving a user the opportunity to generate own voices. The natural appearance of the voice was best related for Mary TTS, whereas MBROLA was not able to convince in this respect. In summary Mary TTS provides a very naturally prepared voice and offers a tool for generating own voices, which enables us to analyze different voices, genders and intonations.

IV. CONCLUSION

The presented concept proposes a modular and flexible concept for an analysis platform for a multi-user content retrieval based spoken dialog system for implementation on a museum guide. In this paper the current state of the art approaches are discussed and a new concept is deduced. The described concept offers a Preprocessing Unit to provide the required robustness in reverberant environments. In order to deal with multiple speakers, a unlimited speaker separation approach is introduced. The Dialog Management provides a natural dialog flow and identifies the communication partner, based on the calculated speaker position in 3D space. Additionally the concepts, based on a keyword spotting approach in combination with a ontology based Dialog Management, which enables the system to deal with various dialects and accents of the German language. Finally the system provides a natural voice, which can easily be adapt, in order to analyses effects on humans during a guidance.

REFERENCES

- [1] F. Faber, M. Bennewitz, and et. al. , "The humanoid museum tour guide robotinho," in Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on. IEEE, 2009, pp. 891-896.
- [2] M. Bennewitz, F. Faber, D. Joho, and S. Behnke, "Fritz-a humanoid communication robot," in Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on. IEEE, 2007, pp. 1072-1077.
- [3] S. T. et al. , " Probabilistic algorithms and the interactive museum tour-guide robot minerva," The International Journal of Robotics Research, vol. 19, no. 11, pp. 972-999, 2000.
- [4] T. Shiwa and et al. , " How quickly should communication robots respond?" in Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on. IEEE, 2008, pp. 153-160.
- [5] S. Strömbergsson and et al. , " Timing responses to questions in dialogue. " in INTERSPEECH, 2013, pp. 2584-2588.
- [6] A. L. Gorin, G. Riccardi, and J. H. Wright, " How may I help you?" Speech communication, vol. 23, no. 1, pp. 113-127, 1997.
- [7] V. Zue, S. Seneff, J. R. Glass, J. Polifroni, C. Pao, T. J. Hazen, and L. Hetherington, " Juplter: a telephone-based conversational interface for weather information," Speech and Audio Processing, IEEE Transactions on, vol. 8, no. 1, pp. 85-96, 2000.
- [8] M. A. Walker, J. S. Aberdeen, J. E. Boland, E. O. Bratt, J. S. Garofolo, L. Hirschman, A. N. Le, S. Lee, S. Narayanan, K. Papineni et al. , " Darpa communicator dialog travel planning systems; the june 2000 data collection. " in INTERSPEECH. Citeseer, 2001, pp. 1371-1374.
- [9] H. Hannes, " Einsatzvon sprach-und gerauschsensitiven systemen in smart home umgebungen," HASE 10 HCI Aspects of Smart Environments, pp. 19-28, 2010.
- [10] M. Bennewitz, F. Faber, D. Joho, M. Schreiber, and S. Behnke, " Towards a humanoid museum guide robot that interacts with multiple persons," in Humanoid Robots, 2005 5th IEEE-RAS International Conference on. IEEE, 2005, pp. 418-423.
- [11] C. Lee, S. Jung, K. Kim, D. Lee, and G. G. Lee, " Recent approaches to dialog management for spoken dialog systems," Journal of Computing Science and Engineering, vol. 4, no. 1, pp. 1-22, 2010.
- [12] S. Young, M. Gasic, B. Thomson, and J. D. Williams, " Pomdp-based statistical spoken dialog systems: A review," Proceedings of the IEEE, vol. 101, no. 5, pp. 1160-1179, 2013.
- [13] R. Pieraccini and J. Huerta, " Where do we go from here? research and commercial spoken dialog systems," in 6th SIGdial Workshop on Discourse and Dialogue, 2005.
- [14] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, " Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," Signal Processing Magazine, IEEE, vol. 29, no. 6, pp. 114-126, 2012.

- [15] H. Saruwatari and et. al. , "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 2, pp. 666-678, 2006.
- [16] H. Saruwatari, S. KURITA, and T. Kazuuya, "Blind Source Separation combining frequency-domain ICA and beamforming," *Science And Technology*, no. 1, pp. 2733-2736, 2001.
- [17] W. Guo and Q. Zong, "A blind separation method of instantaneous speech signal via independent components analysis," in *Consumer Electronics, Communications and Networks (CECNet)*, 2012 2nd International Conference on. IEEE, 2012, pp. 3001-3004.
- [18] Z. Wang and et. al. , "Automatic Multi-Speaker Speech Recognition System Based on Time-Frequency Blind Source Separation under Ubiquitous Environment," in *9th Conference on Industrial Electronics and Applications (ICIEA)*, 2014, pp. 101-106.
- [19] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, vol. 52, no. 1, pp. 12-40, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.specom.2009.08.009>
- [20] S. E. Tranter and D. a. Reynolds, "An overview of automatic speaker diarization systems," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, pp. 1557-1565, 2006.
- [21] K. H. Yuo and et. al. , "Combination of autocorrelation-based features and projection measure technique for speaker identification," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 4, pp. 565-574, 2005.
- [22] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 1, pp. 24-33, 1977.
- [23] B. Wildermoth and K. K. Paliwal, "Use of voicing and pitch information for speaker recognition," *SST 2000. Proceedings of the 8th Australian International Conference on Speech Science and Technology*, no. 1, pp. 324-328, 2000.
- [24] S. Woo and et. al. , "Combined architecture of adaptive beamforming and blind source separation for speech recognition of intelligent service robots," in *Intelligent Pervasive Computing, 2007. IPC. The 2007 International Conference on. IEEE*, 2007, pp. 214-219.
- [25] S. Argentieri, P. Danes, and P. Soueres, "Modal analysis based beam-forming for nearfield or farfield speaker localization in robotics," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on. IEEE*, 2006, pp. 866-871.
- [26] I. Markovi'c and et. al. , "Speaker localization and tracking with a microphone array on a mobile robot using von Mises distribution and particle filtering," *Robotics and Autonomous Systems*, vol. 58, no. 11, pp. 1185-1196, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.robot.2010.08.001>
- [27] J. -m. Valin and et. al. , "Robust Sound Source Localization Using a Microphone Array on a Mobile Robot," *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pp. 1128-1233. , 2003.
- [28] M. Togami and Y. Kawaguchi, "Head orientation estimation of a speaker by utilizing kurtosis of a DOA histogram with restoration of distance effect," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings*, pp. 133-136, 2010.
- [29] B. Kwon and et. al. , "Sound source localization for robot auditory system using the summed GCC method," *2008 International Conference on Control, Automation and Systems, ICCAS 2008*, vol. 1, no. 1, pp. 241-245, 2008.
- [30] M. L. Seltzer, "Microphone array processing for robust speech recognition," Ph. D. dissertation, Carnegie Mellon University Pittsburgh, PA, 2003.
- [31] M. L. Seltzer and B. Raj, "Speech-recognizer-based filter optimization for microphone array processing," *IEEE Signal Processing Letters*, vol. 10, no. 3, pp. 69-71, 2003.
- [32] W. Li and Q. Liao, "Keyword-speci? normalization based keyword spotting for spontaneous speech," in *Chinese Spoken Language Processing (ISCSLP)*, 2012 8th International Symposium on. IEEE, 2012, pp. 233-237.
- [33] N. Beringer and et. al. , "German Regional Variants—A Problem for Automatic Speech Recognition?" *Processing of the 5th International Conference on Spoken Language ICSLP-1998*, pp. 85-88, 1998.
- [34] D. Milward and M. Beveridge, "Ontology-based Dialogue Systems," *Proceedings of the 3rd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, Acapulco, Mexico*, pp. 9-18, 2003. [Online]. Available: <http://www.ida.liu.se/labs/nlplab/ijcai-ws-03/papers/milward.pdf>
- [35] Y. Vanrompay and et. al. , "Ontology-Based User Preferences and Social Search for Spoken Dialogue Systems," *2012 Seventh International Workshop on Semantic and Social Media Adaptation and Personalization*, pp. 113-118, 2012. [Online]. Available: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6406827>
- [36] J. P'erez and et. al. , "Semantics and complexity of sparql," *ACM Transactions on Database Systems (TODS)*, vol. 34, no. 3, p. 16, 2009.
- [37] E. Kaufmann and et. al. , "Querix: A Natural Language Interface to Query Ontologies Based on Clarification Dialogs," In: *5th ISWC*, no. November, pp. 980-981, 2006. [Online]. Available: <http://citeseerx.ist.psu.edu/viewdoc/summary?>

- doi = 10.1.1.94.6633
- [38] M. Schröder and J. Trouvain, "The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching," *International Journal of Speech Technology*, vol. 6, pp. 365-377, 2003.
- [39] P. Taylor, A. W. Black, and R. Caley, "The architecture of the festival speech synthesis system," 1998.
- [40] T. Dutoit and et. al. , "The MBROLA project: towards a set of high quality speech synthesizers free of use for non commercial purposes," *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP '96*, vol. 3, pp. 1393-1396, 1996.

Consensus in Networks of Dynamic Agents with non-uniform communication delays

ChongTan^{#1}, Jinjie Huang^{*2}

[#]*School of Computer Science and Technology, Harbin University of Science and Technology Harbin, China*

¹tc20021671@126.com

^{*}*School of Automation, Harbin University of Science and Technology Harbin, China*

²huangjinjie163@163.com

Abstract— This study addresses the problem of distributed consensus for discrete-time networked multi-agent systems (NMASs) with communication delays in the transmission network. Different from existing techniques, the networked predictive control scheme is introduced to compensate for communication delays actively and overcome the difficulties induced by the delays. A distributed consensus protocol based on the predictions of states at current time is designed. For discrete-time NMASs with a directed topology and non-uniform constant delays, necessary and sufficient conditions for the consensus are given. Numerical examples are presented to demonstrate the effectiveness of theoretical results.

Keywords— Consensus, multi-agent systems, delays, networked predictive control, discrete-time systems.

I. INTRODUCTION

Consensus problem has received extensive attention in the distributed coordination and cooperative control of networks of dynamic agents [1] to [3], which roughly speaking means to design a distributed protocol such that as time goes on, all agents can asymptotically reach an agreement on certain quantities of interest based on the local information.

To achieve consensus and cooperative behaviors for networked multi-agent systems (NMASs), it is important that the agents in the group are capable of exchanging information through the communication networks, which implies that the network-induced delay can occur inevitably, due to the limited bandwidth of the communication channels and the finite transmission speed. Time delays can degrade the performance of NMASs and even cause the divergence. Hence, it is essential reducing or eliminating the negative effect of delays on NMASs. Recently, many research works on the consensus problem of NMASs with communication delays have emerged. For example, Sun et. al. [4] studied the average consensus problem of the NMASs with undirected switching topology and multiple time-varying communication delays, and the maximal allowable upper bound was determined by a linear matrix inequality method. Based on a tree-type transformation method, results of [4] were extended to the directed network case in [5]. A leader-following consensus problem of second-order multi-agent systems with non-uniform time-varying delays was further investigated and a new neighbor-based protocol was designed

in [6].

However, when there exist communication delays in NMASs, the vast majority of existing literature takes a passive acceptance approach, the prediction intelligence of each individual is neglected. Inspired by numerous results on the predictive intelligence of natural bio-groups, Zhang et. al. [7], [8] designed a small-world predictive protocol for the A/R and Vicsek models, and proposed centralized and decentralized model predictive control protocols for linear dynamic networks without leaders, which shows that the predictive protocols can accelerate consensus speeds and reduce sampling frequencies. Farrai-Trecate et. al. [9] considered the input saturation constraints case and proposed a decentralized predictive mechanism and predictive pinning control to achieve the consensus and improve consensus performance. However, the communication delays are not considered in [7] to [9]. For continuous-time first-order and second-order multi-agent systems with a uniform constant communication delay, Fang et. al. [10], [11] introduced a weighted average prediction into the existing consensus protocol to simultaneously improve the maximum tolerant delay and consensus convergence speed. Therefore, it is a promising topic how to improve the performance of NMASs by fully utilizing the prediction intelligence of each individual.

This paper investigates the consensus problem of networked multi-agent systems with a directed topology and diverse communication delays, where the dynamics of all agents are described by discrete-time linear time-invariant systems. By introducing the networked predictive control scheme (NPCS) [12], [13] to compensate for communication delays actively, a new distributed consensus protocol is put forward. The predictions of states at current time, instead of available delayed states, are exploited to design the consensus protocol. For NMASs with identical linear agents and bounded non-uniform communication delays, necessary and sufficient conditions of the consensus are provided. Finally, a numerical simulation is further presented to demonstrate that performance of NMASs with communication delays, based on networked predictive control method, is very close to that of systems without a network delay.

The rest of the paper is organized as follows. Section II gives the problem formulation and some preliminaries. Based on the NPCS, the design of distributed pro-

tocol and analysis of consensus are discussed in Sections II-A and III, respectively. To illustrate theoretical results, a numerical simulation is provided in Section IV. Finally, some concluding remarks are drawn in Section V.

Some remarks on notation are given as follows. Let $M_{m,n}(\mathbb{F})$ be the set of all m -by- n matrices over a field \mathbb{F} , and $M_{n,n}(\mathbb{F})$ is abbreviated to $M_n(\mathbb{F})$, where \mathbb{F} is real number field \mathbb{R} or complex number field \mathbb{C} . The set of nonnegative integers is denoted by \mathbb{Z}^+ . \otimes stands for the Kronecker product of matrices. $\|\cdot\|$ represents l^2 norm on vectors or its induced norm on matrices. $\mathbf{1}_N$ denotes a N -dimension column vector with all entries equal to one. $\mathbf{0}$ and \mathbf{I} represent zero matrix and identity matrix with an appropriate dimension respectively. For a matrix $A \in M_n(\mathbb{C})$, define $A^0 = I_n$. A matrix $A \in M_n(\mathbb{C})$ is said to be Schur if all eigenvalues of A locate in the open unit disk centered at the origin. $e_i(N, n) = [0_n \cdots I_n \cdots 0_n]^T \in M_{Nn,n}(\mathbb{R})$ is block entry matrix, i. e. $e_i(N, n)$ is composed of N matrices belonging to $M_n(\mathbb{R})$, where the i -th block is I_n , others are 0_n , $I_n \in M_n(\mathbb{R})$ and $0_n \in M_n(\mathbb{R})$, $i = 1, 2, \dots, N$.

II. PRELIMINARIES AND PROBLEM FORMULATION

In this section, some basics of graph theory are introduced to formulate the consensus problem. Let $\mathcal{V} = \{1, 2, \dots, N\}$ be an index set of N agents, where i represents the i -th agent. A weighted digraph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ can be used to conveniently represent the communication relationship among N agents in NMASs, where $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges and $A = [a_{ij}] \in M_N(\mathbb{R})$ is a nonnegative weighted adjacency matrix. An edge from i to j is denoted by $e_{ij} = (i, j)$ with the adjacency element a_{ji} , i. e. $e_{ij} \in \mathcal{E} \Leftrightarrow a_{ji} > 0$. The directed edge $e_{ij} \in \mathcal{E}$ means that agent j can receive the information from agent i . Self-loop (i, i) is not allowed, i. e. $a_{ii} = 0$ for all $i \in \mathcal{V}$. The set of neighborhood of agent i is denoted by $N_i = \{j \in \mathcal{V}; (j, i) \in \mathcal{E}\}$. A sequence of edges $(i_1, i_2), (i_2, i_3), \dots, (i_{f-1}, i_f)$ with $(i_j, i_k) \in \mathcal{E}$ and $i_j \in \mathcal{V}$ for all $j = 1, 2, \dots, f \in \mathbb{Z}^+$, is called a directed path from agent i_1 to agent i_f . If there exists a directed path from agent i to agent j , then agent j is said to be reachable from agent i , or agent i is said to be reachable to agent j . The set of all reachable agents to agent i is denoted by N_i^* . The Laplacian matrix $\mathcal{L} = [l_{ij}]_{N \times N}$ of the weighted digraph \mathcal{G} is defined as $l_{ii} = \sum_{j=1, j \neq i}^N a_{ij}$ and $l_{ij} = -a_{ij} \forall i \neq j$. Obviously, all the row-sums of \mathcal{L} are zero, which implies that \mathcal{L} has always an eigenvalue zero corresponding the right eigenvector $\mathbf{1}_N$. For convenience, \mathcal{L} is partitioned as $\mathcal{L} = [l_1^T \ l_2^T \ \cdots \ l_N^T]^T$ with $l_i \in M_{1,N}(\mathbb{R})$, $i = 1, 2, \dots, N$.

Consider an NMAS composed of N discrete-time linear agents, where the dynamics of agent i are described by

$$\begin{aligned}
 x_i(t+1) &= Ax_i(t) + Bu_i(t), t \in \mathbb{Z}^+ \\
 x_i(t) &= \varphi_i(t), -\tau_M \leq t \leq 0, i \in \mathcal{V}
 \end{aligned} \quad (1)$$

where x_i and u_i are the state and control input of agent i , respectively, $A \in M_n(\mathbb{R})$ and $B \in M_{n,m}(\mathbb{R})$ are constant matrices.

In NMAS(1), the information exchanged among different agents is achieved by a network, which implies that communication time-delays often occur due to physical characteristics of medium transmitting the information (e. g. acoustic wave communication between underwater vehicles), diversity of signals (boolean codes, images, videos, etc.), as well as bandwidth of communication channels. So it is assumed that agent i receives information from agent j ($j \in N_i^*$) with a constant communication delay τ_{ij} at time t , where τ_{ij} is a known and positive integer. It yields that agent i is compelled to receive τ_{ij} -step lag data from agent j due to network delays. And it is assumed that agent i receives information from itself without a delay, i. e. $\tau_{ii} = 0$, $\forall i \in \mathcal{V}$. Let $\tau_M = \max_{i,j \in \mathcal{V}} \tau_{ij}$. And $\varphi_i(\cdot)$ represents a given initial state, $i \in \mathcal{V}$.

The main objective of this paper is to this paper, the following assumption can reasonably be made.

Assumption 1: The states of all agents are available. And each agent i can receive information from agent j , $\forall j \in \{i\} \cup N_i^*$.

A. Design of Protocol Based on the NPCS

Due to network delays, agent i ($i \in \mathcal{V}$) can not get the current information from agent j ($j \in N_i^*$) at time t , but only obtain information at time $t - \tau_{ij}$. The most of existing consensus protocols are designed by exploiting delayed states. for example, $u_i(t) = K \sum_{j \in N_i} a_{ij} (x_j(t - \tau_{ij}) - x_i(t - \tau_{ij}))$, $i \in \mathcal{V}$. Because limited and outdated data can not reflect current situations and dynamics of the systems completely, dynamic responses and control performances can not be ideal based on traditional consensus protocols. Therefore, the NPCS is employed to predict current states of agents and compensate for communication delays actively in this paper. Prediction states, instead of received outdated states, at current time are exploited to design the consensus protocol.

Let $\tau_i = \max_{j \in N_i^*} \tau_{ij}$, $i \in \mathcal{V}$. For agent i , one possible way to construct the state predictions of agent j ($j \in \{i\} \cup N_i^*$) from time $t - \tau_i + 1$ to t and control inputs predictions from time $t - \tau_i$ to t are presented as follows.

Step 1: The control input of agent j ($j \in \{i\} \cup N_i^*$) at time $t - \tau_i$, based on receiving the available state data at time t , can be constructed as

$$\hat{u}_j(t - \tau_i | t - \tau_i) = K \sum_{p \in N_j} a_{jp} \Delta x_{j,p}(t - \tau_i)$$

where $\Delta x_{j,p}(t - \tau_i) = x_p(t - \tau_i) - x_j(t - \tau_i)$ is the state difference between agent j and agent p at time $t - \tau_i$, and $K \in M_{m,n}(\mathbb{R})$ is a feedback gain matrix to be designed.

Step 2: On the basis of linear system model (1) and

control input $\hat{u}_j(t - \tau_i | t - \tau_i)$ obtained in the first step, the predictions of state and control input for agent $j (j \in \{i\} \cup N_i^*)$ at time $t - \tau_i + 1$ can be constructed as

$$\begin{aligned} \hat{x}_j(t - \tau_i + 1 | t - \tau_i) &= Ax_j(t - \tau_i) + \\ &B \hat{u}_j(t - \tau_i | t - \tau_i)(t - \tau_i | t - \tau_i) \\ \hat{u}_j(t - \tau_i | t - \tau_i)(t - \tau_i + 1 | t - \tau_i) \\ &= K \sum_{p \in N_j} a_{jp} \Delta \hat{x}_{j,p}(t - \tau_i + 1 | t - \tau_i) \end{aligned}$$

where $\hat{x}_j(t - p | t - q)$ ($p < q$) denotes the state prediction for time $t - p$ on the basis of the states up to time $t - q$, and $\Delta \hat{x}_{j,p}(t - p | t - q) = \hat{x}_p(t - p | t - q) - \hat{x}_j(t - p | t - q)$ is the state predictions difference between agent j and agent p at time $t - p$.

Step $k + 1$: By the way of iteration, the predictions of state and input for agent $j (j \in \{i\} \cup N_i^*)$ at time $t - \tau_i + k$ can be constructed as

$$\begin{aligned} \hat{x}_j(t - \tau_i + k | t - \tau_i) &= A \hat{x}_j(t - \tau_i + k - 1 | t - \tau_i) \\ &+ B \hat{u}_j(t - \tau_i + k - 1 | t - \tau_i), \\ \hat{u}_j(t - \tau_i + k | t - \tau_i) &= \\ K \sum_{p \in N_j} a_{jp} \Delta \hat{x}_{j,p}(t - \tau_i + k | t - \tau_i) \end{aligned}$$

Step τ_i : The predictions of state and control input for agent $j (j \in \{i\} \cup N_i^*)$ at time $t - 1$ can be constructed as

$$\begin{aligned} \hat{x}_j(t - 1 | t - \tau_i) &= A \hat{x}_j(t - 2 | t - \tau_i) + \\ &B \hat{u}_j(t - 2 | t - \tau_i) \\ \hat{u}_j(t - 1 | t - \tau_i) &= K \sum_{p \in N_j} a_{jp} \Delta \hat{x}_{j,p}(t - 1 | t - \tau_i) \end{aligned}$$

Step $\tau_i + 1$: Finally, the prediction of state for agent $j (j \in \{i\} \cup N_i)$ at time t can be constructed as

$$\begin{aligned} \hat{x}_j(t | t - \tau_i) &= \\ A \hat{x}_j(t - 1 | t - \tau_i) + B \hat{u}_j(t - 1 | t - \tau_i) \end{aligned}$$

Hence, for NMAS (1) with constant communication delays, the protocol of agent i is designed as

$$\begin{aligned} u_i(t) &= u_i(t | t - \tau_i) \\ &= K \sum_{j \in N_i} a_{ij} \Delta \hat{x}_{i,j}(t | t - \tau_i) \quad (2) \end{aligned}$$

where $K \in M_{m,n}(\mathbb{R})$ is a feedback gain matrix to be designed.

In the proposed consensus protocol, state predictions at current time are exploited instead of received outdated states.

Remark 1: By using the prediction method, the state prediction of agent $j (j \in N_i^*)$ at time t can be obtained, based on information up to time $t - \tau_i$. When $\tau_{ij} < \tau_i$, the states and inputs from time $t - \tau_i$ to time $t - \tau_{ij}$ are available at time t , it is unnecessary to predict states and inputs within the finite time points $\{t - \tau_i, t - \tau_i + 1, \dots, t - \tau_{ij}\}$. Besides, agent i receives data from itself without a delay. So it is not necessary to predict the states and inputs of agent i .

Therefore, on some level, the proposed predictive method in this paper increases the computational burden and is slightly conservative. However, the proposed method follows the uniform rule for all agents, overcomes

design obstacle from different time delays τ_{ij} , simplifies the prediction procedure and reduces difficulties of the theoretical analysis and derivation. In particular, when there are not external disturbances, predictive values will be equal to actual values. The accuracy and precision of the propose method would not be affected though the additional computation load of algorithm is increased. Hence, this paper provides a method of dealing with non-uniform communication delays, though it is mildly conservative.

The major work of this paper aims to design consensus protocols such that all the states of the multi-agent systems achieve consensus. The definition of the consensus is presented as follows.

Definition 1: For NMAS (1), protocol (2) is said to (asymptotically) solve consensus problem if the states of system (1) satisfy $\lim_{t \rightarrow \infty} \|x_i(t) - x_j(t)\| = 0, \forall i, j \in \mathcal{I}$.

III. CONSENSUS OF DISCRETE-TIME NMAS WITH DELAYS

In this section, the consensus analysis of NMAS(1) with a directed topology and constant communication delays is provided.

Note that

$$\begin{aligned} &((l_i \mathcal{L}^k) \otimes I_n) A_c = \\ &A((l_i \mathcal{L}^k) \otimes I_n) - BK((l_i \mathcal{L}^{k+1}) \otimes I_n) \end{aligned}$$

where $A_c = I_N \otimes A - \mathcal{L} \otimes (BK)$, and l_i is the i -th row of the Laplacian matrix $\mathcal{L}, i = 1, 2, \dots, N$. By the way of iteration, the state prediction of agent $j (j \in \{i\} \cup N_i)$ at time t on the basis of data up to time $t - \tau_i$ can be expressed by

$$\begin{aligned} \hat{x}_j(t | t - \tau_i) &= \\ A \hat{x}_j(t - 1 | t - \tau_i) + B \hat{u}_j(t - 1 | t - \tau_i) \\ &= A^{\tau_i} x_j(t - \tau_i) + \Omega_j(\tau_i) x(t - \tau_i) \quad (3) \end{aligned}$$

where $\Omega_j(\tau_i) = \sum_{q=1}^{\tau_i} (l_j \mathcal{L}^{q-1}) \otimes f(A^{\tau_i - q})$ and $f(A^h)$ is all the items that degree of A is h in $(A - BK)^{\tau_i}$ expansion, $h = 0, 1, \dots, \tau_i - 1$.

For simplicity, denote

$$\begin{aligned} x(t) &= [x_1^T(t) \quad x_2^T(t) \quad \dots \quad x_N^T(t)]^T, \\ u(t) &= [u_1^T(t) \quad u_2^T(t) \quad \dots \quad u_N^T(t)]^T, \\ \delta_i(t) &= x_i(t) - x_1(t), i = 1, 2, \dots, N, \\ \delta(t) &= [\delta_2^T(t) \quad \delta_3^T(t) \quad \dots \quad \delta_N^T(t)]^T. \end{aligned}$$

The following theorem provides a necessary and sufficient condition of protocol (2) solving the consensus problem of NMAS(1).

Theorem 1: Consider NMAS(1) with a directed topology $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ and non-uniform constant communication delays. Protocol (2) solves the consensus problem if and only if the following linear discrete-time system with multiple delays is asymptotically stable.

$$\delta(t + 1) = H\delta(t) + \sum_{i=1}^N H_{\tau_i} \delta(t - \tau_i), t \in \mathbb{Z}^+$$

where

$$\begin{aligned}
 H &= I_{N-1} \otimes A \\
 \hat{R} &= R_b^T (R_b R_b^T)^{-1} \\
 R_b &= \begin{bmatrix} -1_{N-1} & I_{N-1} \end{bmatrix} \\
 H_{\tau_1} &= \sum_{q=0}^{\tau_1-1} (1_{N-1} l_1 \mathcal{L}^q \hat{R}) \otimes (BKf(A^{\tau_1-q})) \\
 H_{\tau_j} &= -e_{j-1}(N-1, n) \sum_{q=0}^{\tau_j-1} (l_j \mathcal{L}^q \hat{R}) \otimes \\
 &\quad (BKf(A^{\tau_j-q})) \quad j=2, 3, \dots, N
 \end{aligned}$$

and $f(A^h)$ is all the items that the degree of A is h in $(A - BK)^{\tau_i}$ expansion, $h=0, 1, \dots, \tau_i, i=1, 2, \dots, N$.

Proof:

Substituting (3) into (2) derives

$$\begin{aligned}
 u_i(t) &= K \sum_{j \in N_i} a_{ij} \Delta \hat{x}_{i,j}(t | t - \tau_i) \\
 &= -K \sum_{j=1}^N l_{ij} \hat{x}_j(t | t - \tau_i) \\
 &= \Theta_i(\tau_i) x(t - \tau_i), \quad i \in \mathcal{V} \quad (4)
 \end{aligned}$$

where

$$\Theta_i(\tau_i) = - \sum_{q=0}^{\tau_i} L_j \mathcal{L}^q \otimes (Kf(A^{\tau_i-q}))$$

Introduce some auxiliary matrices

$$R = R_b \otimes I_n, R_e = \begin{bmatrix} R_b^T & 1_N \end{bmatrix}^T, R = R_e \otimes I_n.$$

It is easily verified that R_e and R are inverse matrices and

$$\begin{aligned}
 I_{Nn} &= R^T (RR^T)^{-1} R \\
 &= [R_e^T (R_e R_e^T)^{-1} R_e] \otimes I_n \\
 &= [R_b^T (R_b R_b^T)^{-1} R_b + \frac{1}{N} 1_N 1_N^T] \otimes I_n \\
 &= R^T (RR^T)^{-1} R + \frac{1}{N} (1_N 1_N^T) \otimes I_n \quad (5)
 \end{aligned}$$

It follows from $\mathcal{L} 1_N = 0$ and $l_i 1_N = 0, i=1, 2, \dots, N$, that

$$\begin{aligned}
 & \left(\sum_{q=0}^{\tau_i} (l_i \mathcal{L}^q) \otimes (BKf(A^{\tau_i-q})) \right) \left((1_N 1_N^T) \otimes I_n \right) \\
 &= \sum_{q=0}^{\tau_i} (l_i \mathcal{L}^q 1_N 1_N^T) \otimes (BKf(A^{\tau_i-q})) \\
 &= 0 \quad (6)
 \end{aligned}$$

So, it follows from (5) and (6) that

$$\begin{aligned}
 & \sum_{q=0}^{\tau_i} (l_i \mathcal{L}^q) \otimes (BKf(A^{\tau_i-q})) x(t - \tau_i) \\
 &= \left(\sum_{q=0}^{\tau_i} (l_i \mathcal{L}^q) \otimes (BKf(A^{\tau_i-q})) \right) R^T (RR^T)^{-1} \delta(t - \tau_i) \\
 &= \Xi_i(\tau_i) \delta(t - \tau_i)
 \end{aligned}$$

where

$$\Xi_i(\tau_i) = \sum_{q=0}^{\tau_i} (l_i \mathcal{L}^q \hat{R}) \otimes (BKf(A^{\tau_i-q})), \quad i=1, 2, \dots, N$$

Hence, the relative state error system can be presented as

$$\begin{aligned}
 \delta(t+1) &= R x(t+1) \\
 &= R [(I_N \otimes A) x(t) + (I_N \otimes B) u(t)] \\
 &= H_0 \delta(t) - R \begin{bmatrix} \Xi_1(\tau_1) \delta(t - \tau_1) \\ \Xi_2(\tau_2) \delta(t - \tau_2) \\ \vdots \\ \Xi_N(\tau_N) \delta(t - \tau_N) \end{bmatrix} \quad (7)
 \end{aligned}$$

$$= H_0 \delta(t) + \sum_{i=1}^N H_{\tau_i} \delta(t - \tau_i)$$

It can be seen that system (7) is a linear discrete-time system with multiple delays. Therefore, that protocol (2) solves the consensus problem of NMAS (1) is transformed into the asymptotical stability problem of system (7) with multiple delays. So protocol (2) solves the consensus problem if and only if system (7) is asymptotical stable.

Lemma 1: Let the Laplacian matrix \mathcal{L} of digraph \mathcal{G} be partitioned as $\begin{bmatrix} \mathcal{L}_{11} & \mathcal{L}_{12} \\ \mathcal{L}_{21} & \mathcal{L}_{22} \end{bmatrix}$ with $\mathcal{L}_{11} \in \mathbb{R}$ and $\mathcal{L}_{22} \in M_{N-1}(\mathbb{R})$. For any positive integer n .

$$R_b \mathcal{L}^n = (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12})^n R_b$$

where $R_b = \begin{bmatrix} -1_{N-1} & I_{N-1} \end{bmatrix}$.

Proof: Mathematical induction is used to prove the conclusion. Because $\mathcal{L} 1_N = 0$, which implies $\mathcal{L}_{11} + \mathcal{L}_{12} 1_{N-1} = 0$ and $\mathcal{L}_{21} + \mathcal{L}_{22} 1_{N-1} = 0$.

When $n=1$,

$$\begin{aligned}
 R_b \mathcal{L} &= \begin{bmatrix} \mathcal{L}_{21} - 1_{N-1} \mathcal{L}_{11} & \mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12} \\ 1_{N-1} \mathcal{L}_{12} 1_{N-1} - \mathcal{L}_{22} 1_{N-1} & \mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12} \end{bmatrix} \\
 &= \begin{bmatrix} -(\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12}) 1_{N-1} & \mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12} \end{bmatrix} \\
 &= (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12}) R_b
 \end{aligned}$$

Assume that the equality is true for $n=k$: $R_b \mathcal{L}^k = (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12})^k R_b$. From this, it need to be shown that the equality continues to hold for $n=k+1$.

$$\begin{aligned}
 R_b \mathcal{L}^{k+1} &= R_b \mathcal{L}^k \mathcal{L} = (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12})^k R_b \mathcal{L} \\
 &= (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12})^{k+1} R_b
 \end{aligned}$$

The proof is completed.

Especially, when $\tau_1 = \tau_2 = \dots = \tau_N \equiv \tau$, the following result can be obtained easily.

Corollary 1: Consider NMAS(1) with a directed topology $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$ and non-uniform constant communication delays. when $\tau_1 = \tau_2 = \dots = \tau_N \equiv \tau$, protocol (2) solves the consensus problem if and only if the following linear discrete-time system with a constant delay τ is asymptotically stable.

$$\delta(t+1) = H \delta(t) + H_\tau \delta(t - \tau), \quad t \in \mathbb{Z}^+ \quad (8)$$

where $H_\tau = - \sum_{q=0}^{\tau} (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12})^{q+1} \otimes (BKf(A^{\tau-q}))$, and H is defined as Theorem 1.

Proof: when $\tau_1 = \tau_2 = \dots = \tau_N \equiv \tau$, it follows from Lemma 1 that

$$\begin{aligned}
 & R \begin{bmatrix} \Xi_1(\tau_1) \delta(t - \tau_1) \\ \Xi_2(\tau_2) \delta(t - \tau_2) \\ \vdots \\ \Xi_N(\tau_N) \delta(t - \tau_N) \end{bmatrix} \\
 &= R \sum_{q=0}^{\tau} \begin{bmatrix} (l_1 \mathcal{L}^q \hat{R}) \otimes (BKf(A^{\tau-q})) \\ (l_2 \mathcal{L}^q \hat{R}) \otimes (BKf(A^{\tau-q})) \\ \vdots \\ (l_N \mathcal{L}^q \hat{R}) \otimes (BKf(A^{\tau-q})) \end{bmatrix} \delta(t - \tau) \\
 &= R \sum_{q=0}^{\tau} (\mathcal{L}^{q+1} \hat{R}) \otimes (BKf(A^{\tau-q}))
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{q=0}^{\tau} (R_b \mathcal{L}^{q+1} R_b^T) (R_b R_b^T)^{-1} \otimes (BKf(A^{\tau-q})) \\
 &= \sum_{q=0}^{\tau} (\mathcal{L}_{22} - 1_{N-1} \mathcal{L}_{12})^{q+1} \otimes (BKf(A^{\tau-q}))
 \end{aligned}$$

Hence, system (7) can be reduced to system (8). From Theorem 1, the result holds.

IV. SIMULATION

In this section, a numerical simulation is presented to demonstrate the applicability of the obtained theoretical results.

Example 1: Consider a NMAS with four agents with a directed topology $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A})$, where $\mathcal{V} = \{1, 2, 3, 4\}$, $\mathcal{E} = \{(1, 2), (1, 4), (2, 3), (3, 1), (4, 1)\}$ and corresponding adjacency elements are 1 (see Fig. 1). The dynamics of agent $i (i = 1, \dots, 4)$ is described by the following linear discrete time system.

$$x_i(t+1) = \begin{bmatrix} 1.5 & -0.5 \\ 0.2 & 0.5 \end{bmatrix} x_i(t) + \begin{bmatrix} 0.5 \\ -1 \end{bmatrix} u_i(t) \quad (9)$$

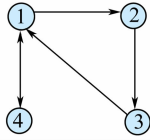


Fig. 1 Fixed topology among four agents

Assuming there exist communication delays τ_{ij} vary randomly in a finite set $\{1, 2, 3\}$. By using employing a cone complementary linearization algorithm and [14, Theorem 3], a feedback gain is obtained as $K = [-0.154, 7 \quad 0.331, 3]$, which guarantee that switched system(7) with delays is asymptotically stable. Hence, from Theorem 1, the protocol (2) solves the consensus problem.

Set initial conditions of the system states to be $\varphi_i(t) = 0, t = -3, -2, -1, i = 1, 2, 3, 4$. $\varphi_1(0) = [1 \quad 0]^T$, $\varphi_2(0) = [2 \quad -1]^T$, $\varphi_3(0) = [1 \quad -1]^T$ and $\varphi_4(0) = [-1 \quad -1]^T$. The state trajectories of the closed-loop NMASs with communication delay are shown in Fig. 2. Solid lines represent the case of no communication delays, and dashed lines represent the case of communication delays. It illustrates that performance of NMASs with communication delays, based on networked predictive control method, is very close to that of systems without a network delay. The simulation result further validates that the NPCS can compensate for communication delays actively and improves the performance.

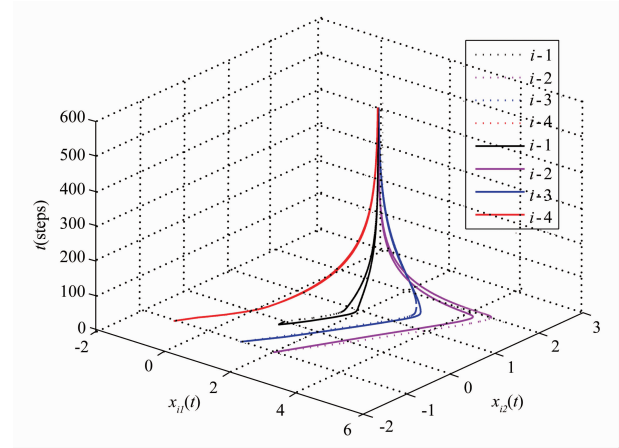


Fig. 2. State trajectories of agents

V. CONCLUSIONS

The consensus problem of linear discrete-time NMASs with a directed topology and diverse communication delays has been studied. To overcome possible negative effects of communication delays, the networked predictive control method is introduced to compensate for time delays actively. The predictions of states at current time, instead of available delayed states, are exploited to design the distributed consensus protocol. For NMASs with identical linear agents and bounded communication delays, the consensus problem of NMASs is transformed into asymptotical stability problem of linear discrete-time switched system with delays, and then necessary and sufficient conditions of the consensus have been provided. A numerical simulation has been provided to further validate the effectiveness of the proposed theoretical results.

In this paper, an approach has been provided to deal with bounded and diverse communication delays, based on the common upper bound function. However, the proposed predictive method increases the computation burden and is slightly conservative. So some issues will deserve further investigation in our future work, such as how to reduce the design conservatism, and the analysis of the time-varying topology and data dropout cases.

ACKNOWLEDGMENT

This work is supported by China Postdoctoral Science Foundation Funded Project under Grants 2015M581463 and Natural Science Foundation of Heilongjiang under Grant F2016025.

REFERENCES

- [1] Z. Y. Fei, H. J. Gao, and W. X. Zheng, "New synchronization stability of complex networks with an interval time-varying coupling delay," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 56, no. 6, pp. 499-503, 2009.

- [2] W. L. Lu and T. P. Chen, "Global synchronization of discrete-time dynamical network with a directed graph," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 2, pp. 136-140, 2007.
- [3] M. Y. Chen, "Synchronization in complex dynamical networks with random sensor delay," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 1, pp. 46-50, 2010.
- [4] Y. G. Sun, L. Wang, and G. M. Xie, "Average consensus in networks of dynamic agents with switching topologies and multiple time-varying delays," *Syst. Control Lett.*, vol. 57, no. 2, pp. 175-183, 2008.
- [5] Y. G. Sun and L. Wang, "Consensus of multi-agent systems in directed networks with nonuniform time-varying delays," *IEEE Trans. Automat. Control*, vol. 54, no. 7, pp. 1607-1613, 2009.
- [6] W. Zhu and D. Z. Cheng, "Leader-following consensus of second-order agents with multiple time-varying delays," *Automatica*, vol. 46, no. 12, pp. 1994-1999, 2010.
- [7] H. T. Zhang, M. Z. Q. Chen, G. B. Stan, T. Zhou, and J. M. Maciejowski, "Collective behavior coordination with predictive mechanisms," *IEEE Circ. Syst. Mag.*, vol. 8, no. 3, pp. 67-85, 2008.
- [8] H. T. Zhang, M. Z. Q. Chen, and G. B. Stan, "Fast consensus via predictive pinning control," *IEEE Trans. Circuits Syst. I, Reg. Pap.*, vol. 58, no. 9, pp. 2247-2258, 2011.
- [9] G. Ferrari-Trecate, L. Galbusera, M. P. E. Marciandi, and R. Scattolini, "Model predictive control schemes for consensus in multi-agent systems with single-and double-integrator dynamics," *IEEE Trans. Automat. Control*, vol. 54, no. 11, pp. 2560-2572, 2009.
- [10] H. J. Fang, Z. H. Wu, and J. Wei, "Improvement for consensus performance of multi-agent systems based on weighted average prediction," *IEEE Trans. Automat. Control*, vol. 57, no. 1, pp. 249-254, 2012.
- [11] Z. H. Wu, H. J. Fang, and Y. Y. She, "Weighted average prediction for improving consensus performance of second-order delayed multi-agent systems," *IEEE Trans. Syst. Man Cybern. Part B—Cybern.*, vol. 42, no. 5, pp. 1501-1508, 2012.
- [12] G-P. Liu, "Predictive controller design of networked systems with communication delays and data loss," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 6, pp. 481-485, 2010.
- [13] G. P. Liu, Y. Q. Xia, J. Chen, D. Rees, and W. S. Hu, "Networked predictive control of systems with random network delays in both forward and feedback channels," *IEEE Trans. Ind. Electron.*, vol. 54, no. 3, pp. 136-140, 2007.
- [14] L. Zhang, P. Shi, and M. Basin, "Robust stability and stabilisation of uncertain switched linear discrete time-delay systems," *IET Control Theory Appl.*, vol. 2, no. 7, pp. 606-614, 2008.

Design and Simulation of MEMS Wind Speed and Direction Sensor based on Solid State Heat Transfer and Double Coordinate Model

Lan Yunpiang^{#1}, Feng Qiaohua^{#2}, Shi Yunbo^{#3}, Yu Yang^{#4}

[#]The Higher Educational Key Laboratory for Measurement & Control Technology and Instrumentation of Heilongjiang Province, Institute of Measurement-Control Technology & Communications Engineering Harbin University of Science and Technology
No. 52 Xuefu Road, Nangang Dist, Harbin, China

¹lan_yun_ping@163.com

²fengqiaohua80@126.com

³shiyunbo@126.com

⁴mail_yyang@163.com

Abstract— In order to solve the problems of low conductivity, narrow range, and low accuracy, a novel MEMS wind speed and direction sensor based on solid state heat transfer and double coordinate model is reported in this paper. The sensor has the following prominent characteristics, using silicon as structure, using Ni-Cr alloy film as heater and using Al₂O₃ film as solid state medium heat conduction layer, eventually an eight array wind speed and direction sensor model based on solid media heat transfer and double coordinate model can be structured, which can improve the accuracy and range of the sensor. Furthermore, the rationality and feasibility of the design have been verified through simulation experiment in this paper.

Keywords— Solid state heat transfer, double coordinate model, wind speed and direction sensor, eight array sensing element

I. INTRODUCTION

Wind is a nature phenomenon due to the flow of air existing in everywhere of human life. Meanwhile, it is an important renewable energy and has great influence on agriculture and transportation. How to measure wind speed and direction accurately is the key problem of how to use it efficiently. So the study and development on wind speed and direction are very important subjects. Nowadays, wind speed and direction sensors are becoming more and more important in many applications on, such as aircrafts, meteorology, ocean, environment, industry, transportation and so on [1].

Various detection requirements on wind speed and direction result in various kinds of it. The cup shaped and propeller shaped wind gauge are two typical applications among mechanical wind speed meter, which have low cost, simple measuring method, better measuring results, but they also have many drawbacks including slower reaction rate, bigger volume, moving parts easy to

wear, frequent maintenance, higher later using cost and so forth [2]. With the rapid development of laser and ultrasonic technology, wind meter based on the two technologies has become possible. However, the high price limits its wide applications to a great extent. Wind meter sensor based on MEMS process has got more and more widely applications for its smaller dimension, easy integration, mass production, high precision and hard to wear. What is worth to say, low heat transfer efficiency, narrow range, vulnerable to environmental interference and other shortcomings have existed in present report of wind speed and direction sensor. So the main problem is how to improve its sensitivity and accuracy [3].

In this paper, a wind speed and direction integrated sensor is designed based on MEMS technology. The heat can be transferred to the sensitive unit, through the Al₂O₃ solid medium heat conduction layer, which can increase the temperature of the sensing unit, expand the range of the sensor. Improve the accuracy of the sensor by a double coordinate calculation model.

II. MODEL DESIGN AND MANUFACTURING METHOD

Fig. 1 presents a schematic illustration of the proposed wind meter sensor. The device is consisted of eight parts, which are Si substrate, SiO₂ insulation layer, Ni-Cr alloy heating film, Al₂O₃ solid medium heat conduction layer, eight array sensitive cell, pad, connecting line and groove. Ni-Cr alloy heating film located in the middle of SiO₂ insulation layer and Al₂O₃ solid medium heat conduction layer.

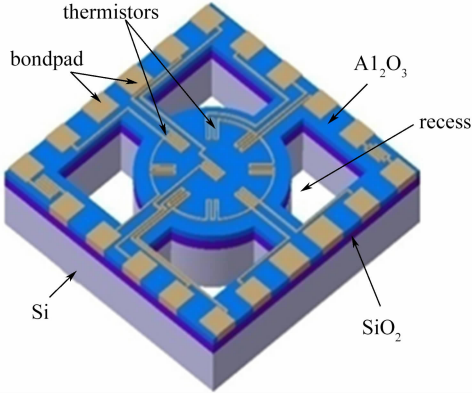


Fig. 1 Whole structure

The fabrication process is as seen in Fig. 2. Firstly, prepare crystal(100)Si as the substrate with the size of $8,000 \times 8,000 \times (200 \sim 300) \mu\text{m}^3$. Then form SiO_2 insulation layer within $0.4 \sim 0.5 \mu\text{m}$ thickness through the dry-wet-dry oxygen alternating oxygen process. After that, sputter a film of Ni-Cr alloy heating film on the SiO_2 insulation layer. It follows with the Al_2O_3 solid medium heat conduction layer of $0.6 \mu\text{m}$ except the pads areas. Then generate the sensitive elements and connecting lines on Al_2O_3 layer. Meanwhile, there will be growth on the corresponding thickness of the pads. In order to form the micro-cantilever beams, firstly, we use Hydrochloric acid, HF acid and EPW corrosion agent to etch Al_2O_3 solid medium heat conduction layer, Al_2O_3 solid medium heat conduction layer and Si in order from the upside and then reuse the EPW corrosion agent to etch the Si from the downside. After all the procedures mentioned above, the sensor we proposed can be formed entirely.

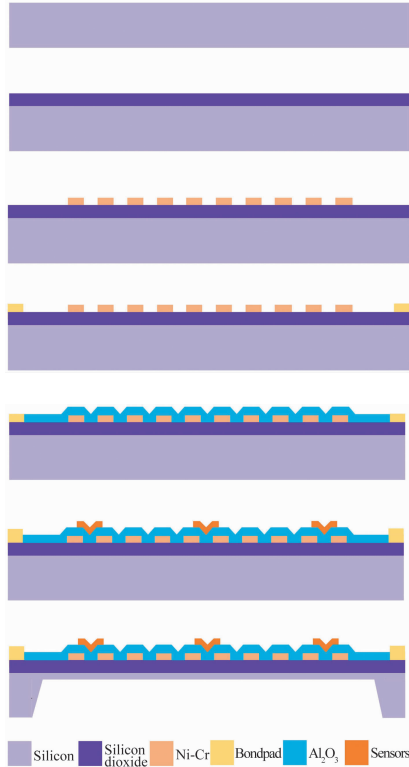


Fig. 2 Fabrication process of the wind speed and direction sensor

III. DOUBLE COORDINATE MODEL AND SIMULATION

A. Building and Calculation of Double Coordinate Model

Eight sensitive elements are distributed on the surface of the sensor to improve the measuring accuracy as we can see from Fig. 3. Using the double coordinate model to calculate wind speed and direction was the first time. There are two kinds of common working methods of hot film wind speed and direction sensor, namely, constant power(CP) and constant temperature difference(CTD). For two dimensional thermal temperature difference type wind speed meter chip[4], the temperature gradient induced by fluid can be decomposed on the XOY coordinate as follows.

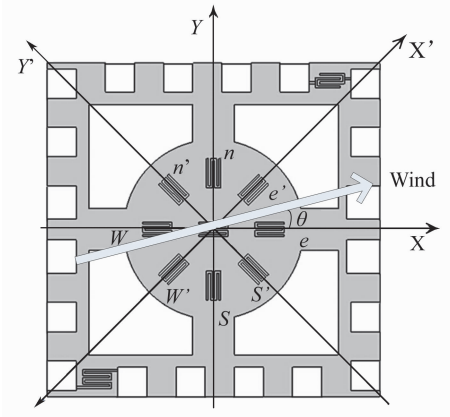


Fig. 3 Vertical view of sensor element

ΔT_0 is the temperature difference between the heating electrode and fluid. The value of the temperature difference ΔT is not only related to ΔT_0 but also related to the gas flow v , $F(v)$ is the function of v , s is a constant determined by fluid characteristics and sensor's structure. According to boundary layer theory and King function, when the wind blows over the surface of the sensor, we can determined the temperature difference as follows[5]

$$\Delta T = s \cdot \Delta T_0 \cdot F(v) \quad (1)$$

θ is the angle between the wind direction and the positive X axis. According to the trigonometric function, temperature difference ΔT can be divided into horizontal difference ΔT_{ew} and vertical difference ΔT_{ns}

$$\begin{aligned} \Delta T_{ew} &= \Delta T \cdot \cos \theta = s \cdot \Delta T_0 \cdot F(v) \cdot \cos \theta \\ \Delta T_{ns} &= \Delta T \cdot \sin \theta = s \cdot \Delta T_0 \cdot F(v) \cdot \sin \theta \end{aligned} \quad (2)$$

Assuming that the value of the resistor is R and the temperature coefficient of the sensitive is α , according to Wheatstone Bridge, the output voltage can be calculated as

$$\Delta U_{ew} = \frac{\Delta R_{ew}}{R} U = \frac{\alpha \cdot \Delta T_{ew}}{R} U$$

$$\begin{aligned}
 &= \frac{\alpha \cdot s \cdot \Delta T_0 \cdot F(v) \cdot \cos \theta}{R} \\
 \Delta U_{ns} &= \frac{\Delta R_{ns} U}{R} = \frac{\alpha \cdot \Delta T_{ns} U}{R} \\
 &= \frac{\alpha \cdot s \cdot \Delta T_0 \cdot F(v) \cdot \cos \theta}{R} \quad (3)
 \end{aligned}$$

From formula(2) and(3) ,we can know that the wind velocity satisfying the following formula

$$\begin{aligned}
 F(v) &= \frac{\sqrt{\Delta T_{ew}^2 + \Delta T_{ns}^2}}{s \cdot \Delta T_0} \\
 &= \frac{R}{\alpha \cdot s \cdot \Delta T_0 \cdot U} \sqrt{\Delta T_{ew}^2 + \Delta T_{ns}^2} \quad (4)
 \end{aligned}$$

Calculate the inverse function in formula (4) to get the wind speed v

$$v = F^{-1} \left(\frac{R}{\alpha \cdot s \cdot \Delta T_0 \cdot U} \sqrt{\Delta T_{ew}^2 + \Delta T_{ns}^2} \right) \quad (5)$$

Refer to formula (2) and (3) , wind direction can be defined as

$$\theta = \arctan \frac{\Delta T_{ns}}{\Delta T_{ew}} = \arctan \frac{\Delta U_{ns}}{\Delta U_{ew}} \quad (6)$$

Similarly, the temperature gradient induced by the fluid decomposed by the XOY' axis can be expressed as

$$v' = F^{-1} \left(\frac{R}{\alpha \cdot s \cdot \Delta T_0 \cdot U} \sqrt{\Delta U'_{ew}^2 + \Delta U'_{ns}^2} \right) \quad (7)$$

$$\theta' = \arctan \frac{\Delta T'_{ns}}{\Delta T'_{ew}} = \arctan \frac{\Delta U'_{ns}}{\Delta U'_{ew}} \quad (8)$$

Through formula (5) and (7) , the final wind speed can be obtained as

$$\begin{aligned}
 W_s &= \frac{v + v'}{2} \\
 &= \frac{1}{2} \left(F^{-1} \left(\frac{R}{\alpha \cdot s \cdot \Delta T_0 \cdot U} \sqrt{\Delta U_{ew}^2 + \Delta U_{ns}^2} \right) + \right. \\
 &\quad \left. F^{-1} \left(\frac{R}{\alpha \cdot s \cdot \Delta T_0 \cdot U} \sqrt{\Delta U'_{ew}^2 + \Delta U'_{ns}^2} \right) \right) \quad (9)
 \end{aligned}$$

Through formula(6) and(8) ,the final wind direction can be obtained as

$$\begin{aligned}
 W_d &= \frac{\theta + \theta' - 45^\circ}{2} \\
 &= \frac{1}{2} \arctan \frac{\Delta U_{ns}}{\Delta U_{ew}} + \frac{1}{2} \left(\arctan \frac{\Delta U'_{ns}}{\Delta U'_{ew}} - 45^\circ \right) \quad (10)
 \end{aligned}$$

Through the double coordinate model,we can achieve two measurements of the same wind speed and direction,and two calculations,so as to improve the accuracy of the measurement.

B. Simulation

In order to save time,reduce the cost of sensor manufacturing,shorten the period of sensor development,optimize the performance of the device,the simulation of the sensor is essential. The simulation is a multi domain coupling simulation,which involves thermal simulation domain,fluid simulation domain and electric energy simulation domain. Using the constant power mode,the simulation of the multi coupling field can be simplified

to the temperature simulation in the fluid solid coupling field(FSI)by calculating the value of the heating resistors and the voltage at both ends.

The prominent advantage of thermal type anemometer is that,it can measure the wind direction and has greater sensitivity in smaller wind speed. The disadvantage is ΔT will be saturated when the wind speed is large, because the upstream temperature can not be lower than the environment temperature and the downstream temperature can not be higher than heating temperature, so there will be range saturation in the wind speed measurement[6]. Using the simplified model for simulation, the hierarchical structures are shown in Fig. 4 (a) and Fig. 4(b). Fig. 4(a) exhibits the model with Al_2O_3 solid heat conducting medium layer. Fig. 4 (b) exhibits the model without Al_2O_3 solid heat conducting medium layer,and the Ni-Cr alloy heating film and eight array elements are in the same plane.

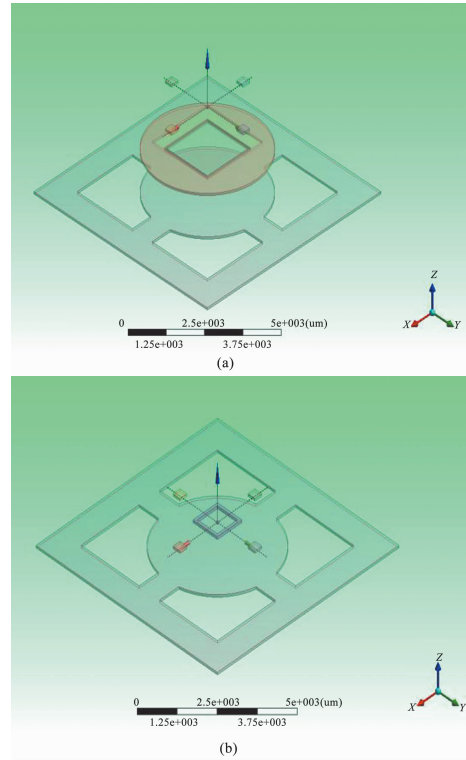


Fig. 4 Simplified hierarchical structure model
 (a)hierarchical structure with Al_2O_3 heat conduction layer;
 (b)hierarchical structure without Al_2O_3 heat conduction layer

Applied the same heat generation rate on the heating body of the same volume and the temperature distribution field is shown in Fig. 5. As we can see from the figures, the temperature of the sensitive element in Fig. 5(a) was significantly higher than that in Fig. 5(b). Heat transferred to sensitive unit is more by Al_2O_3 solid heat conducting medium layer. Therefore, there is a higher static temperature (i. e. no wind blowing through the sensor unit) ,or to achieve the same temperature with low power consumption. Therefore, the method of heat conduction of the solid medium can not only improve the range of the sensor, but also can im-

prove its sensitivity.

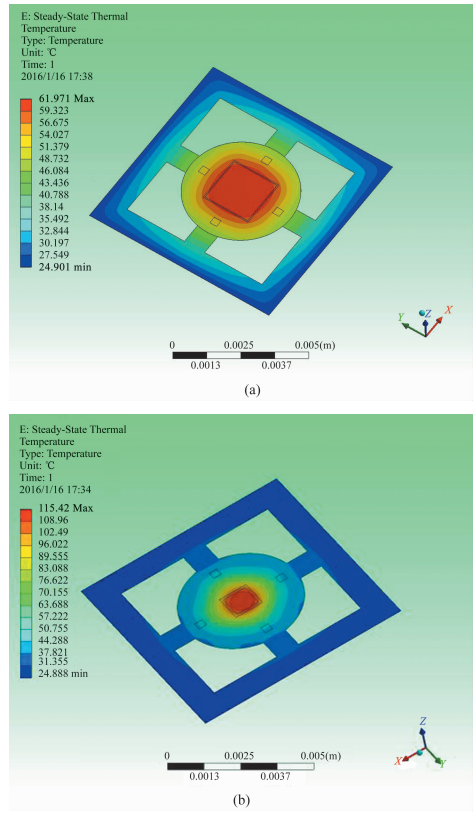


Fig. 5 Temperature distribution field
 (a) temperature distribution field with Al_2O_3 heat conduction layer;
 (b) temperature distribution field without Al_2O_3 heat conduction layer

To achieve the temperature distribution field of the design model, we applied $4.9 \times 10^{12} \text{ W/m}^3$ heat generation rate on the Ni-Cr alloy film. Fig. 6 shows the simulation results when the wind speed is 4 m/s and the wind direction with the positive direction of X-axis is zero. Fig. 6(a) is the whole temperature distribution field and Fig. 6(b) is the refinement enlarge the Al_2O_3 insulating layer temperature distribution field. The upstream temperature distribution was significantly lower than that of the downstream temperature, that means, the model has better perception to the wind speed.

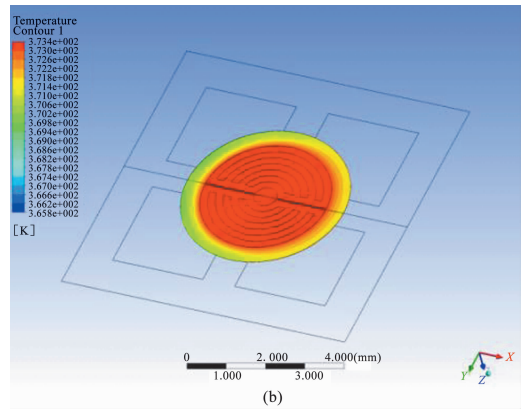
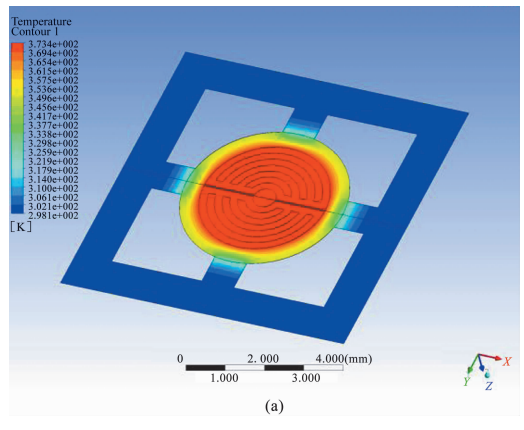
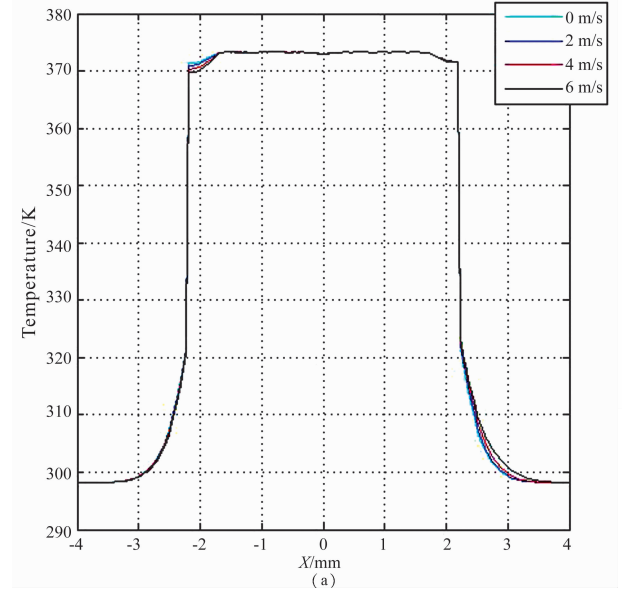


Fig. 6 Temperature distribution field
 (a) the whole temperature distribution field;
 (b) refinement enlarge of certain section

Fig. 7 (a) exhibits the whole temperature distribution curve when the wind speed is 0 m/s, 2 m/s, 4 m/s and 6 m/s. Fig. 7(b) is the amplifier of Fig. 7(a) at certain section, what can be seen clearly in Fig. 7(b) is the temperature gradually decreased as the increase of the wind speed, and the value of the sensitive cell ΔT can be affected by the distance from the sensor to the center. In later works, we can expand the simulation range to determine the optimal distance from sensitive element to center in measurement range [7] [8].



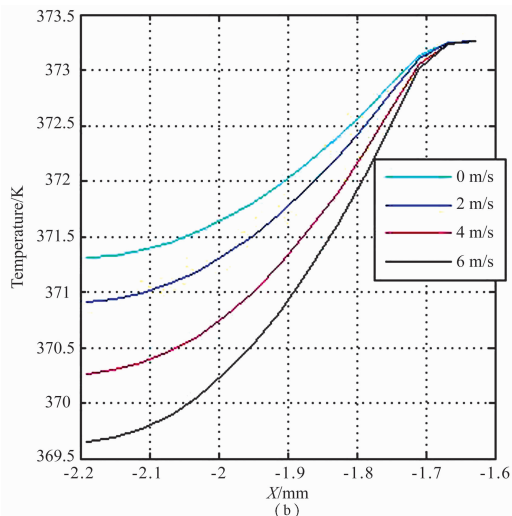


Fig. 7 Temperature distribution curve
(a) whole temperature distribution curve;
(b) amplifier of certain section

IV. CONCLUSIONS

In summary, using the double coordinate model calculation method can improve the sensor accuracy, using Al_2O_3 solid heat conducting medium layer to transfer heat can obtain the higher temperature distribution field, improving the sensitivity and range of the sensor. The methods presented in the paper provide theoretical basis and simulation verification for improving the accuracy and the range of the sensor, and lay the theoretical basis for the manufacture, widen the application field and application prospect of the wind speed and direction sensor.

ACKNOWLEDGMENT

Thanks for the key projects of Heilongjiang province natural science foundation(No. ZD201217) and the funding of Harbin application technology and the development project(No. 2013DB1BG013).

REFERENCES

- [1] R. Y. Que, R. Zhu. A Compact Flexible Thermal Flow Sensor for Detecting Two-Dimensional Flow Vector [J]. March 2015, pp. 1931-1932.
- [2] Guo Yuhong. Research and implementation of Wind speed and direction Measurement Based on Ultrasonic[D]. May 2014, pp. 2-3.
- [3] Djuzhev Nikolay, Novikov Dmitry, Ryabov Vladimir. Application of the streamlined body for properties amplification of thermal resistive MEMS gas flow sensor[J]. Solid State Phenomena 2016, pp. 247-249.
- [4] K. A. A. Makinwa, J. H. Huijsing. Constant Power Operation of a Two-Dimensional Flow Sensor [J]. IEEE Transactions on Instrumentation and Meas-

urement, August 2002, pp. 840-843

- [5] Zhao Wenjie, Shi Yunbo, Luo Yi, Li Guolong, Zhou Zhen. Design of the wind velocity sensor based on AIN thermal isolation MEMS array [J]. November 2012, pp. 2820-2821.
- [6] Gao Donghui, Qin Ming, Huang Qing'an. Thermal simulate of a silicon gas flow sensor and it's packaging [J]. Chinese Journal of Semiconductors, 2005, pp. 26:368.
- [7] Ali Sukru Cubukcu, Eugen Zernickela, Uwe Buerklina, Gerald Anton Urbana A. 2D thermal flow sensor with sub-mW power consumption [J]. Sensors and Actuators A: Physical, May 2010, pp. 452-454.
- [8] Ali Sukru Cubukcu, Gerald A. Urban. Sensitivity-Maximizing and Error-Reducing Design of a Flow and Thermal Property Sensor [C]. 9th. Int. Conf. on Thermal, Mechanical and Multiphysics Simulation and Experiments in Micro-Electronics and Micro-Systems, Euro SimE 2008, pp. 3-5.

Developing a Collaboration Model Between Researchers Using Their Publication Bipartite Graphs

Ganbat Tsend^{#1}

[#] *Information Technology Department, Mongolian University of Science and Technology
Ulaanbaatar, Mongolia*

¹ganbat@must.edu.mn

Abstract— We can construct bipartite graphs of scientist, researcher's publication list. When we build bipartite graphs, we will describe collaboration models for some interesting applications can be used to solve problems such as finding particular researcher, it's collaboration, some statistics and relevant operations such as add paper, classification, integration of research areas.

To solve the wide variety of problems that can be studied using graphs, we will introduce several graph algorithms using formal methods

Keywords— algorithms, software, graphs, formal method

I. INTRODUCTION

Graphs are discrete structures consisting of vertices and edges that connect these vertices. Graphs are used as models in a variety of area. For instance, graphs are used to represent who influences whom in an organization, acquaintance ships between people in social networks such as Facebook, Youtube, LinkedIn, telephone calls between telephone numbers, road networks in transportation, collaboration. In such models, vertices represent person and edges present their connections. For scientist and researcher's publication list, vertex set can be divided into two disjoint subsets to present authors and their research papers.

II. PUBLICATION BIPARTITE GRAPHS

We consider following publication list. Scientist, researcher's publication list generally can be with following structure, as shown in Table I.

Table I
PUBLICATION LIST

	Authors	Paper	Other info
1	Khuder A, Ganbat Ts	N-gram analysis of a Mongolian Text	2011, MUST
2	Dolgorsuren B, Ganbat Ts	Development of Com- puter law syllabus	2011, ШУТИС

Table I Continued

	Authors	Paper	Other info
3	Munkhbuyan B, Ganbat Ts, Batzolboo B	Enterprise Archite- cture Development strategy in Higher Ed- ucation	2011, EICTHE
4	Zolboo D, Ganbat Ts	The survey on online service at metropolis State	2012, FITAT
5	Zolboo D, Ganbat Ts	Survey of comparison of the governmental online service	2012, MUST
6	Dolgorsuren B, Ganbat Ts	Some translation prob- lems on software tech- nology term	2012, MUST
7	Khuder A, Ganbat Ts	Part of Speech Tagging Experiments on Mon- golian Language	2012, MUST
8	Dolgorsuren B, Ganbat Ts	Development of com- puter model to calcu- late statistics about traffic jam	2013, MMT

We can denote above publication list by $G(A, P, N)$ bipartite graphs where A set is consist of six authors $A = \{a1, a2, a3, a4, a5, a6\}$, P set is consist of eight papers $P = \{p1, p2, p3, p4, p5, p6, p7, p8\}$ and N set is their connections $N = \{(a1, p1), (a1, p2), (a1, p3), (a1, p4), (a1, p5), (a1, p6), (a1, p7), (a1, p8), (a2, p1), (a2, p7), (a3, p2), (a3, p6), (a3, p8), (a4, p3), (a5, p3), (a6, p4), (a6, p5)\}$ where $A \cap P = \emptyset$.

Each author and paper are represented by vertex and connection is represented by an edge, as shown in Fig. 1.

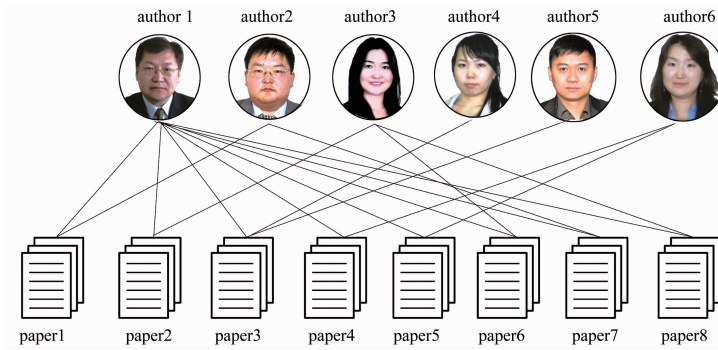


Fig. 1 Publication bipartite graphs

There are many useful ways to represent graphs. Here we represented $G = (A, P, N)$ graph using adjacency matrices

$$m(i,j) = \begin{cases} 1, & (a,p) \text{ is edge of } G \\ 0, & \text{none} \end{cases}$$

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

We can easily obtain some interested information from M matrix such as calculate total paper's number for particular author, author's number for particular paper, total collaborated author's number for particular author, find a paper title with maximum collaborated authors, author name with maximum collaborated authors, who has maximum papers.

Following we represented author's collaboration graph and its matrix, as shown in Fig. 2 and Fig. 3.

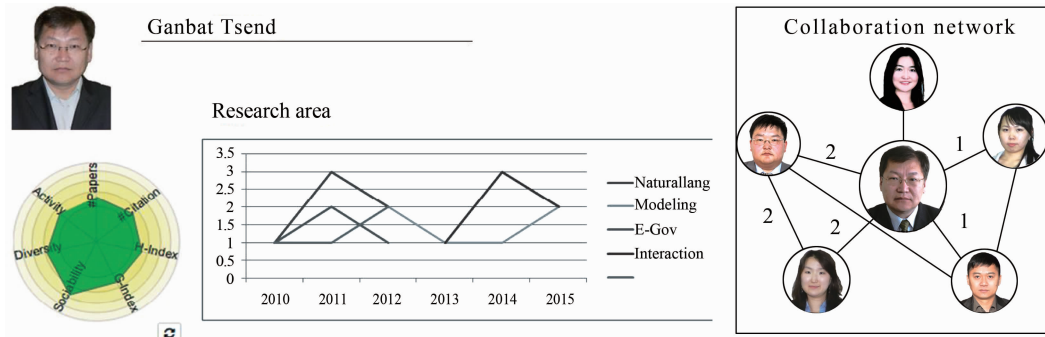


Fig. 2 Collaboration graph for author1


 <p>Ganbat Tsend MUST Ganbat_tsend@ yahoo.com</p>	Statistics	
	Publication years	1992-2016
	Publications(Book/Paper)	70 (13/57)
	Cited	5
	vAvailable download	20
	Downloaded(6 weeks)	5
	Downloaded(Year)	10
	Total Downloaded	30
	Average download per paper	2.1
	Average cited per paper	0.75
	Collaborators	20
	Published Journals/proceedings	40

Fig. 3 Researcher's statistics

For M matrix, given row's values sum is author's total paper's number, sum of column is collaborated author's total number. All row number is author's total number, all column number is paper's total number.

$$C = \begin{vmatrix} 0 & 2 & 3 & 1 & 1 & 2 \\ 2 & 0 & 0 & 0 & 1 & 0 \\ 3 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 3 & 0 \\ 1 & 1 & 0 & 3 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 \end{vmatrix}$$

adjacency matrix for author1

III. ALGORITHM DESIGN

We can develop algorithms to check publication list whether it is publication list or not, whether paper is registered into publication list, add paper to publication list, remove paper from publication list. Here we defined one algorithm to add paper to publication list using formal method

Type A, P, N
 G = A × P × N

value add: A × P × G → G
 add(a, p, g) ≡
 mk_G({a} Ug. A, {P} Ug. p, {a, p} Ug. N)
 pre(not a ∈ A) and(not p ∈ P)

IV. CONCLUSIONS

① We illustrated publication list by set, graphs, matri-

ces. At result, we developed some algorithms to obtain interested statistics.

② As the list, we defined collaboration graphs between researchers.

③ We defined some relevant operations on the list and collaboration graphs. And we developed some algorithms.

④ Using RSL, we defined formal specification of publication list and collaboration graphs.

REFERENCES

- [1] K. H. Rosen. *Discrete mathematics and its applications*. 7th edition. Mc Graw-Hill. 2012.
- [2] Deo Narsingh. *System simulation with digital computer*, ISBN 978-81-203-0028-6.
- [3] Deo Narsingh. *Graph theory with applications to engineering and computer science*. ISBN:978-81-203-0145-0.
- [4] Ganbat Tsend. *System simulation using computer*. MUST edition. Ulaanbaatar. 2004. 2nd edition. In mongolian.
- [5] Ganbat Tsend, Baatarkhuu Enkhtuya. *Develepment of portal for Research publication. Project report*. 2015. Ulaanbaatar. MEDS. Mongolian Science Technology Foundation.

Development of software and hardware complex measurement and control of devices and crystals. 2G-401 A-V

Reva Ivan L. ^{#1}, Ivanov Andrey V. ^{#2},
Trushin Viktor A. ^{#3}, Aravenkov Alexandr. A. ^{#4}

[#] Automation and Computer Engineering, Novosibirsk State Technical University
av. Karl Marks 20, Novosibirsk, Russia, 630073-All authors

¹reva@corp.nstu.ru

²andrej.ivanov@corp.nstu.ru

³trushin@corp.nstu.ru

⁴aravenkov@corp.nstu.ru

Abstract— Today, more and more often it is necessary to upgrade equipment, carrying out measurement and control of production parameters in production. This need arises from the need to increase the volume of production, automation of the measurement process, grading, design of reporting protocols. Thus arises the need for automated measurement of parameters of the complex, which controls by the computer, performing processing of measurement results already in digital form, as well as manage the process of grading the finished product.

Keywords— Instrument making, upgrade equipment, automated measurement, modules power supplies, information-measuring systems

I. INTRODUCTION

Today, more and more often it is necessary to upgrade equipment, carrying out measurement and control of production parameters in production. This need arises from the need to increase the volume of production, automation of the measurement process, grading, design of reporting protocols. Existing systems are fully analog and require significant operator involvement in the process of measurement. Thus arises the need for automated measurement of parameters of the complex, which controls by the computer, performing processing of measurement results already in digital form, as well as manage the process of grading the finished product. With regard to thezener-noise generators (devices/crystals) 2G-401 A-V(vary the noise power spectral density and spectral width) , where control parameters are: a constant voltage

in the reverse current of 100 mA, the difference is in constant tension with the reverse currents of 100 mA and 1 mA, DC voltage at a forward current of 10 mA and spectral characteristics (noise spectral density [1] , boarded frequency).

Measurements should be made in 3 variants: on crystals (using a probe) in the cassette 36 ap paratus (thermal chambers) , ongardingment finished devices, which requires contacting the respective control devices. The complex has been implemented on the modules produced by " L Card" [2] For the measurement of the modules have been selected LTR210 (two channels of ADC with a sampling frequency of 10 MHz each, beyond the measurements ± 10 V) , thus, the highest measured signal frequency 4.5 MHz. So the contact devices TTL implemented logic (5 B) for the control channel modules LTR43 were selected (up to 32 I / O signals TTL).

Optional modules power supplies have been designed and manufactured to set the mode power devices / crystals. Due to the fact that there are two modules LTR210 full measurement channels and measurements are made in the cartridge 36 to devices, it has been realized 4 parallel measuring channels for modules 2 LTR210. It helps speed up the measurement process in the cassette (9 measurement cycles instead of 36). For switching devices in the cassette and the measuring current source input module has been implemented multichannel analog switch that connects one of 9 to the output devices.

Block diagram of the three uses of the complex are shown in Fig. 1 and Fig. 2.

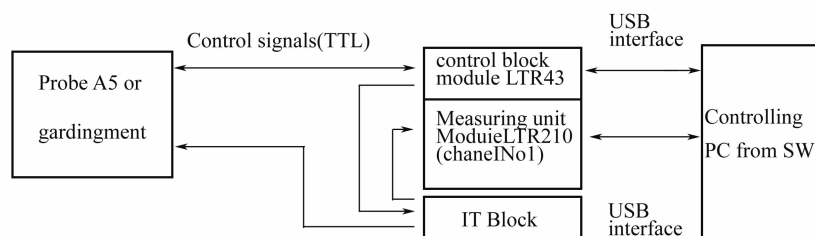


Fig. 1 Flowchart complex crystals as measured with a probe

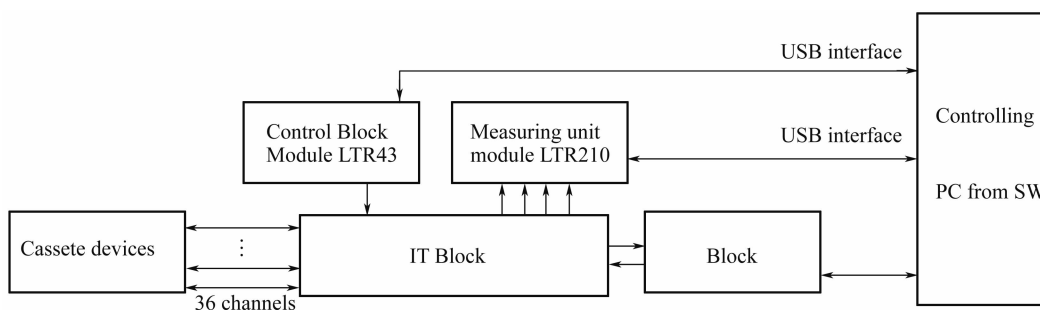


Fig. 2 A block diagram of the complex in measuring instruments in the cassette

Implementation of the software is executed in an environment LabView, modules "A Card" is delivered in a set with the libraries for programming in LabView, C ++ ,etc.

II. CONCLUSIONS

The result was a universal set that can be used in any of the above. Successfully conducted production tests, which showed that the complex complies with the technical.

III. ACKNOWLEDGMENT

Thank administration Harbin University of Science

and Technology and Novosibirsk State Technical University.

REFERENCES

- [1] GOST 18986. 23-80 Zener semiconductor. Metro-hole measuring noise spectral density. -Moscow, 2001. -12 p.
- [2] Kreytovay LTR system. Manual [Elec-throne resource]. URL; <http://www.lcard.ru/download/ltr.pdf>(Obra-tion date;13.06.2016).

Development to Create an Electronic Tutorial on a Subject "Motor Vehicle"

Erdenetuya Amgalan^{#1}, Choisureen Purevdorj^{#2}

[#]*Department of Transportation*

School of Mechanical Engineering and Transportation at MUST,

Ulaanbaatar, Mongolia

¹erdenetuya@must.edu.mn

²choisureen@must.edu.mn

Abstract— Education is one of the fastest-growing economic and social sectors in the world, and the use of new technologies is an integral and driving component of that growth. The problems of creating an electronic tutorial for the discipline "Motor Vehicle", developed by the Department of Transportation. Presented the scientific and methodical approach to creating a poster based on the general principles of development of multimedia training materials and considers specific discipline.

Keywords— Course, e-learning, multimedia, engineering education

I. INTRODUCTION

Multimedia instruction is one of the current examples of a new area of instructional research and practice that has generated a considerable amount of excitement. Like other new areas, its early advocates begin with a set of assumptions about the learning and access problems it will solve and the opportunities it affords [1].

So, many different definitions of multimedia have been offered [2]. "Multimedia" usually refers to the capacity of computers to provide real-time representations of nearly all existing media and sensory modes of instruction. Sensory modes are distinguished from media because they relate to the sensory format of information so that it is compatible with one of the five senses. Visual and aural forms of information can be provided by a variety of media whereas taste, smell, and texture representations in media are very limited.

A study subject "Motor Vehicle" is compulsory for students of three professions. Students from the profession "Management for Transportation" consider that the subject is the hardest. Discipline "Motor Vehicle" gives students a basic understanding of the relationship of systems and mechanisms and the study of unit and operating principle of the vehicle components and assemblies.

Currently, there is a tendency for the development and wide application of multimedia teaching aids in teaching various technical disciplines. Multimedia technology represents an opportunity to create educational materials that contain images, text, accompanied by sound, video, animation and other visual effects.

II. PURPOSE AND PROBLEM STATEMENT

On the urgent need for the use of multimedia educational materials for studying technical subjects seen in a number of facts that in recent years increasingly observed in the learning process.

First, the students, most of which have very advanced skills in using computer technology, readily accept submissions in electronic form. Secondly, there is a disturbing trend: students—experienced computer users—found it as they think the easiest way to "assimilation" of the course—search the web material on the program of discipline, which are not always of good quality, methodologically aligned, and sometimes openly illiterate.

That is why the goal is to create an electronic material, obviously, will be demanded of students, they have to offer professionals.

Distinguish two main tasks of multimedia tools used in the learning process: improving students' independent work and improving the classroom teacher of students, primarily through the most intuitive visualization of lecture materials.

The last factor is extremely relevant to the discipline of "Motor Vehicle", which is associated with specific features of this course. During the lecture the teacher often spends much time on explaining the rather complicated for the perception of the material. This applies, for example, sections of the "principle of transmission" "steering" as well as sections, which are accompanied by an analysis of electrical circuits. There is a need to replace the unproductive part of the teacher with the board to represent the visual materials in the computer as follows: in the form of "live" illustrations, animation, video images. This saves time and enhances the teacher's positive reaction of students at the stated material.

III. PRINCIPLES FOR THE DEVELOPMENT OF ELECTRONIC LEARNING MATERIALS

The most important principle on which to base the development of electronic teaching materials,—the princi-

ple of interaction—is the effective joint work of the two main actors in this process: a practitioner—the teacher disciplines and the developer—a specialist in multimedia technologies.

The problem is that the teacher, not a computer specialist, does not know how to implement existing ideas in his presentation. This is the key role of specialist media or developer. This specialist has the necessary skills to use software tools to implement those ideas that have defined the teacher to visualize the material.

The developer has methodological experience of forming a multimedia educational material, operates ways of presenting information, visual or acoustic, and uses different channels for information. Integration of all these mediums into a single product of complex structure and the task of the developer. Thus, selection of objects visualization and creative work to create a picture of the project a task the teacher-practitioner. The task of the developer to implement this project available to it software tools with the above rules. Becomes clear that the effective interaction of the teacher and the developer, their combined creativity are key to generating high-quality multimedia materials [5] and [6].

IV. THE STRUCTURE OF THE ELECTRONIC TUTORIAL

The main components of electronic textbooks are structured lecture materials, multimedia presentation materials, exercises and tests for knowledge control, vir-

tual labs.

Multimedia presentations of lecture material with application of visualization tools are most appropriate for studying objects and processes of the following nature:

- ① Objects, processes and phenomena that are inaccessible to direct observation;
- ② Processes and phenomena occurring in the movement and development;
- ③ Very slow or very rapid processes and phenomena.

For the maintenance of discipline "Motor Vehicle" is characterized by all kinds of these objects and phenomena. For example: the electrochemical processes in the battery, automatic transmission, electrical equipments of automobile that are extremely difficult to understand for students. The study of such objects requires the development of video or animation of options. See Fig. 1.

Visual range, designed for the module is logically divided into separate slides, the sequence of which, according to the principle of integrity, coincides with the order of presentation during the lecture. Teacher can use such presentation materials for a visual accompaniment lecture. In addition, these materials can be used by students and for independent work, as presented on the slides concepts and definitions contain references to the relevant text fragments.

Text of reference for monitoring students' knowledge formed for self-knowledge as a student and teacher to assess the assimilation of the section as a form of analysis of current performance.

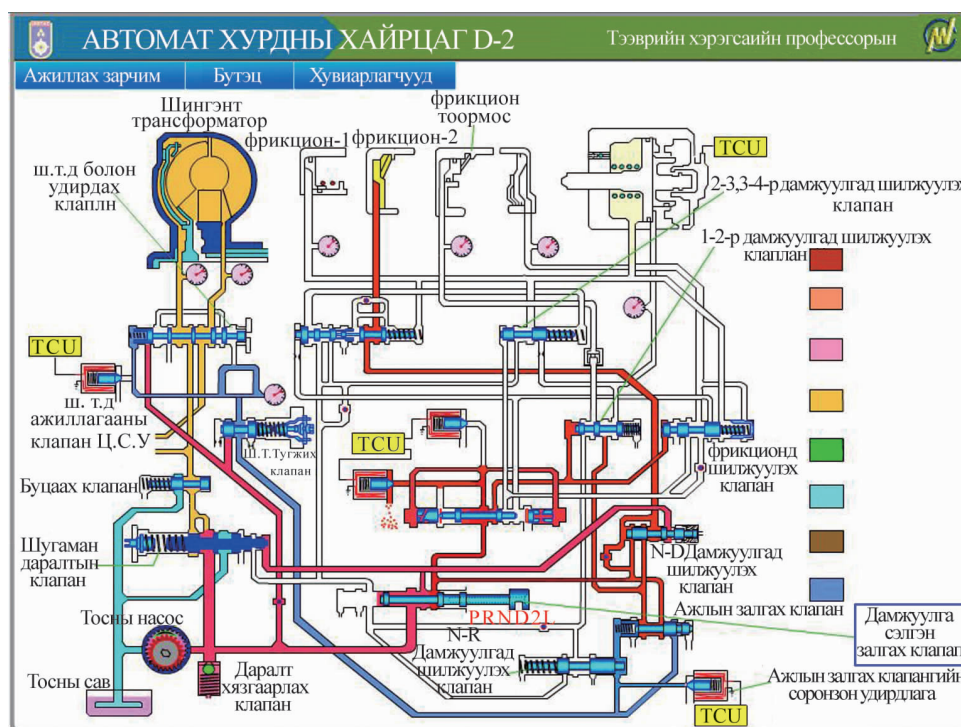


Fig. 1 Automatic transmission working principle

Technology is required to produce and deliver e-learning. Different tools can be used to produce content, depending on which file format will be used and the na-

ture of the desired final product.

Microsoft PowerPoint or even Microsoft Word can be sufficient to create simple learning resources like a pres-

entation or a tutorial. However, more sophisticated tools are required if you want to create interactive content.

Courseware authoring tools are special-purpose tools that create interactive content. They add text, graphics and other media, but also provide a framework to organize pages and lessons for reliable navigation. While most of these tools are stand-alone package that incorporate assessment and quiz capabilities, some integrate those functions from other programs.

To create media components, authoring tools need auxiliary software (e. g. Adobe Photoshop for bitmap graphics, Adobe Illustrator for vector images or Adobe Flash for animations) and other tools for video and sound creation and compression [3].

Generally, programming tools (particularly those that are sophisticated and complex) require professional expertise and considerable development time, while authoring tools can be used by people without programming skills. The main advantage of authoring tools is that they are easier and faster to use, and they therefore shorten development time [4].

Many authoring tools were simple PowerPoint "add-ons", able to convert a set of slides directly from PowerPoint. For example, iSprint Presenter or Articulate transform standard PowerPoint presentations into Flash.

Pressing on Preview > Preview slides or Publish, a PPT presentation is automatically converted into Flash. See Fig. 2.

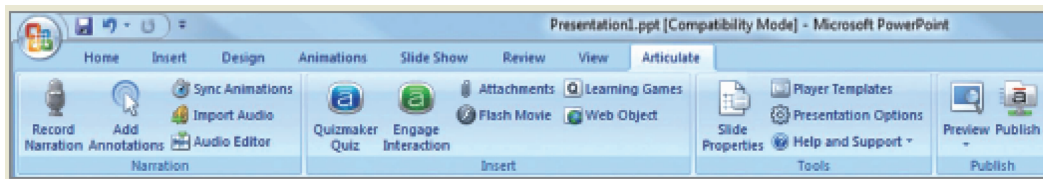


Fig. 2 The articulate tool is nested inside PowerPoint

As authoring tools evolved, they integrated many other useful features and new easy-to-use templates to accommodate rich media interactions, quiz makers, video converter, etc. for more engaging and complete learning experiences.

Authoring tools can be grouped under three main categories according to the architecture they use for authoring:

- ① Template-based tools;
- ② Timeline-based tools;

③ Object-based tools.

Timeline-based tools, such as Adobe Flash, are widely used to create animations and robust interactive applications with their own scripting languages and timeline that organizes and controls content over time.

Currently, our young teachers who hold to use different programs for, namely Macromedia Flash animation and develop animated posters and uses in her/his classes. See Fig. 3.

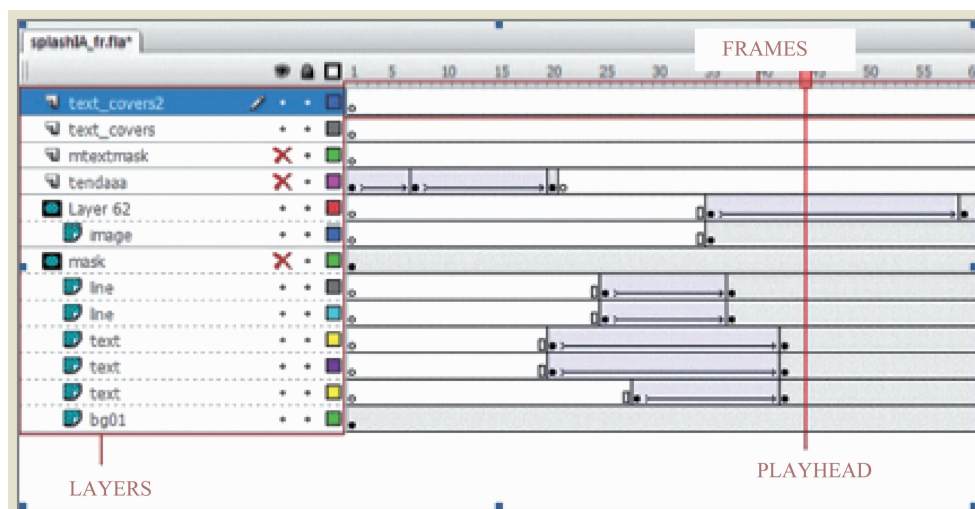


Fig. 3 Adobe flash used for an animated splash in courses development

V. CONCLUSIONS

Interactive e-lessons are created by the development

team. A number of authoring tools exist for producing courseware. Authoring tools are specifically designed for producing e-learning content without needing programming skills. Compared with template-based tools, object-

based tools offer more flexibility for content developers but require more development time.

To date, the developed parts of electronic lesson in several sections of discipline "Motor Vehicle". The testing of these electronic modules in the student groups showed the effectiveness of their use to improve student learning of educational material.

VI. ACKNOWLEDGMENT

This work was supported by project of Transportation Department of School of Mechanical Engineering and Transportation, Mongolian University of Science and Technology.

REFERENCES

- [1] American society for Training and Development. E-Learning: if you build it, will they come? Alexandria VA; ASTD, 2001.
- [2] Clark, R. E. , *Learning from Media: Arguments, Analysis and Evidence*. Greenwich, CM; Information Age Publishers. 2001.
- [3] Clark R. C. , Mayer R. E. , *e-Learning and the Science of Instruction-Proven Guidelines for Consumers and Designers of Multimedia Learning*, Second Edition, Pfeiffer 2005.
- [4] E-Learning methodologies. A guide for designing and developing e-learning courses, FAO, 2011.
- [5] Petrova LG, Ostroukh AV, Barinov, KA *Principles for the development of electronic educational resources for higher education; the Security Council. "Engineering Pedagogy"*. -M. , MGADA, 2008. pp. 87-102.
- [6] Prikhodko W. , Melezinek A. , Petrova L. *Development of New Multimedia Course on Engineering Pedagogic*. In: "Joint Forces in Engineering Education Towards Excellence", SEFI-IGIP Joint Annual Conference, 1-4 July 2007, Miskolc, Hungary.

Discrete-Time Robust LQR Control with High-gain Observer for Nano-positioning

Liu Yang^{#1}, Dongjie Li^{#2}, Yu Fu^{#3}, Bo You^{#4}

[#]College of Automation, Harbin University of Science and Technology

No. 52 Xuefu Street, Nangang District, Harbin, China 150001

¹yangliuheu@gmail.com

²dongjieli2013@163.com

³173325586@qq.com

⁴youbo@hrbust.edu.cn

Abstract— This paper proposed a robust linear quadratic regulation (LQR) controller based on a discrete-time estimation. The controller is designed for nano-positioning of piezoelectric actuators (PEAs). PEAs have been widely used in precision positioning systems because of the advantages of its inherently small displacement resolution. Due to its nonlinear effects, mainly hysteresis, can drastically degrade the positioning and tracking control accuracy. Therefore, it is desirable to develop advanced controllers to compensate hysteresis effect for improving the trajectory tracking performance. This controller consists of three parts which are a nominal feedforward control input, a LQR control input and a control input based on system uncertainties compensation. Further, as the only measurable information is the position, a high-gain observer is adopted to estimate the states. The robust stability of the designed controller is proved through a Lyapunov stability analysis. According to simulation results, the proposed controller is effective for both positioning and tracking applications. Moreover, it can provide a high resolution of the system, which is less than 1 nm and robustness of external disturbances.

Keywords— Piezoelectric actuators (PEAs); discrete linear quadratic regulation (LQR) control; hysteresis; nonlinear system.

I. INTRODUCTION

Piezoelectric actuators and PEA-driven positioning systems have been widely employed in diverse applications of micro and nan positioning such as atomic force microscopes [1], [2], adaptive optics [3], micromanipulators [4], and data storage [5] due to their high displacement resolution and large actuating force. Recently, more applications of PEAs come out in biomedical engineering. A 2-D imaging spectrometer is developed, which provides 2-D multispectral mappings for biomedical detection, diagnosis of intractable diseases and imaging technology. This spectrometer employs multiple Piezoelectric actuators to provide spectral tuning of the desired optical signal transmittance by selecting the gap spacing of a tuneable optical filter. For this application, the requirements on the positioning of PEAs are extremely high, i. e. in nanometer level. Furthermore, owing to the noise of the sensor that is used for PEAs position measurement, it is desirable to obtain a high resolution of PEAs.

Nevertheless, the main challenge of using piezoactuated systems arises from the nonlinearities attributed to hysteresis and creep when the PEAs are driven by a voltage amplifier. In the inverse-based feedforward compensation control, hysteresis is compensated using an inverse of the hysteresis model. In all positioning control applications, the hysteresis and creep effects of PEAs still have shown to be able to significantly degrade the system performance and even system stability, although some feedback control methods have been applied [6]. Hysteresis is the nonlinear dependence of a system not only on its current input but also on its past input. Under an open-loop voltage-drive approach, the hysteresis can induce a positioning error as high as 10% ~ 15% of the PEAs travel range. Fig. 1 shows an simulation observed hysteresis of the PEA model used in this research. Alternatively, hysteresis can be significantly suppressed by operating the PEAs using a charge amplifier [6] to [9]. Nonetheless, because of its complex implementation and high cost [10] the charge amplifier has not been widely adopted. Creep is the slow variation in the PEA displacement that occurs without any accompanying change in the input voltage [11]. It is caused by the same piezoelectric material properties as PEA hysteresis. Being a slow and a small effect which is on the order of 1% of the last displacement per time decade, creep sometimes can be mitigated in closed loop and high frequency operations [12], [13]. However, when the PEAs are applied in slow or static applications, creep must be considered to avoid large positioning error [14]. Therefore, the development of advanced controllers in order to suppress the effect of hysteresis in PEAs has drawn more attention.

Various control strategies have been reported for positioning and tracking control of PEAs, of which three types of control approaches are typically used in the control of PEAs. First one is open-loop control schemes which are usually employed in applications in which position feedback are difficult to implement due to mechanical constraints, e. g. atomic force microscopes [1], [2], [15]. In such control schemes, inverse model of the PEA to be controlled is found and then cascaded to the PEA. However, the major disadvantage of the open-loop control schemes is that their positioning per-

performances are highly sensitive to unknown effects such as model errors, external disturbances, and changes in the dynamics of the PEA. Considering that hysteresis modeling is a sophisticated procedure, feedback control techniques without taking hysteresis into account have been developed as the second type of control strategies, such as PID (proportional-integral-derivative) control that is widely used because of its simplicity and capability of eliminating steady state errors [16], robust control which tries to find the control law via optimizing an objective function that incorporates the robustness objective [17] to [21], sliding mode control which can completely reject the effects of model imperfection and system uncertainties resulting in strong robustness [22] to [24], repetitive control [25] and other control methods [26] to [30]. The last one is feedback with feedforward control method. Feedforward is sometimes used to augment feedback controllers for nonlinearity compensation. However, a precise inverse model of the PEA is necessary in this control scheme, which is sometimes difficult to be obtained.

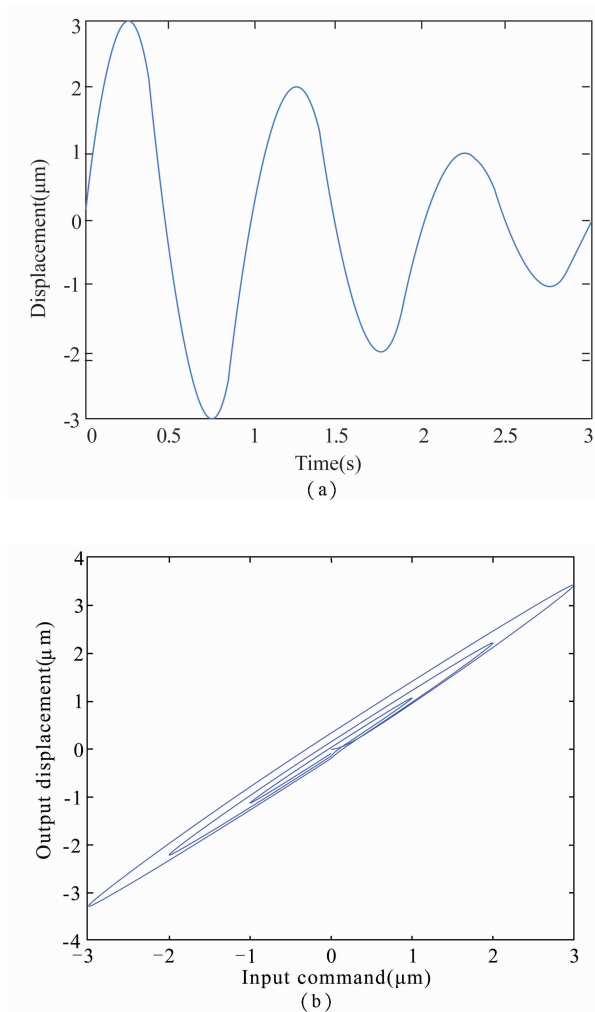


Fig. 1 Simulation observed hysteresis of the PEA model used in this research
(a) a 1-Hz input displacement signal applied to the PEA model;
(b) hysteresis loop obtained by simulation

In the literature, it has been known that optimal control is applied as an approach to attain the expected system dynamic and steady-state performances. However, it is essential to design an accurate model of the controlled plant for such controllers. Currently, the simplified model of the piezoelectric actuator which is considered as a second-order system only roughly approximate the reality in spite of the hysteresis effect. Douglas proposed a type of robust optimal controller [31], which improved the LQR control technique performances to parametric uncertainties but limited robustness against unstructured uncertainties. An optimal LQR method was discussed by Shieh and Chiu [32] for a piezoelectric microspitoner control. This optimal controller developed based on the error state-space dynamic model but without considering the system uncertainties compensation. In [33], a LQR controller is employed for piezoelectric actuated nanopositioner. This control strategy consists of solving the feedback gain so as to minimize the objective function. However, it does not include the part used to compensate the system nonlinearities and disturbances, which is not able to be robust to the system. In this paper, a robust LQR controller which consists of a nominal feedforward control input, a LQR control input and a compensator for system uncertainties is proposed. Moreover, a high gain observer is applied to estimate the system's full states. For the Robust LQR controller design, the PEA system is considered as a second-order system, and the hysteresis loop is modeled included as a nonlinear system for accurate simulation. The stability of the designed controller is proved by using Lyapunov stability theory, and the positioning and tracking performances of the resulting control system illustrate that the proposed controller can provide both high displacement resolution and precise tracking performance.

This paper is organized as follows. In Section 2, the problem formulation is presented. In Section 3, the proposed robust LQR controller is designed and its stability is proved through a Lyapunov stability analysis. A high gain observer is designed and simulations demonstration of the proposed controller is shown in Section 4. Section 5 concludes this paper.

II. PROBLEM STATEMENT

A class of single input nonlinear systems with dynamic processes can be defined as

$$x^{(n)} + F(x) + \Delta F(x) = [b(x) + \Delta b(x)]u(t) + p(t) \quad (1)$$

where $x^{(n)} = [x, \dot{x}, \dots, x^{(n-1)}]^T$ is the state vector, $u(t)$ is the control input, $f(X, t)$ and $b(X, t)$ are in general nonlinear and possibly time-varying, and $p(t)$ denotes the disturbances of the system. The superscript n on $x(t)$ signifies the order of differentiation.

All the uncertainties are bounded and can be combined together as

$$P(t) = \Delta F + \Delta b u(t) + f(t) = x^{(n)} - F - b u(t) \quad (2)$$

The discrete-time estimation of the uncertainties $P(t)$ is approximated as

$$\bar{P}(t) = \bar{x}^{(n)} - F - bu(t - T) \quad (3)$$

where $\bar{P}(t)$ represents a calculated state since the measurement of higher order states of the system (e. g. , velocity) cannot always be realized, T is the sampling time interval, and $u(t - T)$ denotes the control input in the previous time-step.

In practice, the sampling frequency is selected high enough to ensure that $u(t) \approx u(t - T)$.

According to [34], the state discretization is computed based on a backward difference equation

$$\bar{x}^{(n)}(t) = \bar{x}^{(n-1)}(t) - \bar{x}^{(n-1)}(t - T) \quad (4)$$

Traditionally, consider the piezoelectric actuator as a second-order system, which can be written as

$$\ddot{x}(t) + 2\xi\omega_n \dot{x}(t) + \omega_n^2 x(t) = k\omega_n^2 u(t) + f(t) \quad (5)$$

where $x(t)$ and $\dot{x}(t)$ are the system state variables, ξ , ω_n , and k are the damping ratio, the natural frequency, and the gain of the second-order system respectively. $f(t)$ represents the bounded external disturbances of the system.

In order to take the uncertainties of the system into account, the system dynamics can be rewritten as

$$\ddot{x} = -2\xi^N \omega_n^N \dot{x} - (\omega_n^N)^2 x + k^N (\omega_n^N)^2 u + f_d \quad (6)$$

where the superscript N denotes the nominal value of the parameter. ξ^N , ω_n^N , and k^N take the forms

$$\xi^N \omega_n^N = \xi \omega_n - \Delta\xi \omega_n \quad (7)$$

$$(\omega_n^N)^2 = \omega_n^2 - \Delta\omega_n^2 \quad (8)$$

$$k^N (\omega_n^N)^2 = k\omega_n^2 - \Delta k\omega_n^2 \quad (9)$$

where $\Delta\xi\omega_n$, $\Delta\omega_n^2$, and $\Delta k\omega_n^2$ are the parametric uncertainties, and $f_d = -2\Delta\xi\omega_n \dot{x} - \Delta\omega_n^2 x + \Delta k\omega_n^2 u + f$ denotes the named equivalent disturbances, which contains all the uncertainties, such as unknown constant parameters, nonlinear terms, and bounded external disturbances.

Based on the discrete-time system uncertainties' estimation method above, given the PEA model 6, the estimation of system uncertainties \bar{f}_d can be given by

$$\bar{f}_d(t) = \ddot{x}(t) + 2\xi^N \omega_n^N \dot{x}(t) + (\omega_n^N)^2 x(t) - k^N (\omega_n^N)^2 u(t - T) \quad (10)$$

Then the discrete-time dynamics system becomes

$$\ddot{x}(t) + 2\xi^N \omega_n^N \dot{x}(t) + (\omega_n^N)^2 x(t) = k^N (\omega_n^N)^2 u(t - T) + \bar{f}_d(t) + \check{f}_d(t) \quad (11)$$

where $\check{f}_d(t) = f_d(t) - \bar{f}_d(t)$ is a estimation error of real uncertainties.

Assumption 1: *The lumped system equivalent disturbances f_d is bounded by $|f_d| \leq D$. D is a given positive constant.*

Our goal is to design a robust optimal controller to achieve following of the prescribed reference signals, denoted by x_d .

III. CONTROLLER DESIGN

In this section, a systematic design procedure for a robust LQR controller is shown in three steps.

Firstly define an error system $E(t)$ as

$$E(t) = [e_1(t) e_2(t) e_3(t)]^T \quad (12)$$

where $e_1 = x_d - x$, $e_2 = \dot{e}_1$, and $e_3 = \int_0^t e_1 d\tau \cdot e_3$ is introduced in order to design a static feedback controller to improve steady-state performances for the nominal closed-loop system. Then, the error system can be described in a state-space form as

$$\begin{aligned} \dot{E}(t) &= AE(t) + B[-k^N (\omega_n^N)^2 u(t) + \\ \ddot{x}_d(t) + 2\xi^N \omega_n^N \dot{x}_d(t) + (\omega_n^N)^2 x_d(t) - f_d(t)] \end{aligned} \quad (13)$$

where

$$A = \begin{bmatrix} 0 & 1 & 0 \\ -(\omega_n^N)^2 & -2\xi^N \omega_n^N & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad (14)$$

The controller is constructed by three parts: a nominal feedforward controller, a nominal LQR controller, and a compensator which is applied to compensate all the system uncertainties. Thus, the control input $u(t)$ has the following form:

$$u(t) = u^{FF}(t) + u^{LQR}(t) + u^{CU}(t) \quad (15)$$

where u^{FF} is a nominal feedforward control input based on a discrete-time estimation, u^{LQR} is a control input designed by LQR method, and u^{CU} is a control input based on system uncertainties compensation for the discrete-time estimation error.

The nominal feedforward controller is given by

$$u^{FF}(t) = \frac{1}{k^N (\omega_n^N)^2} [\ddot{x}_d(t) + 2\xi^N \omega_n^N \dot{x}_d(t) + (\omega_n^N)^2 x_d(t) - \bar{f}_d(t)] \quad (16)$$

This control input is employed to obtain a nominal linear error system in order to design the LQR controller and robust controller. Now the error system can be described as

$$\dot{E}(t) = AE(t) + B[-k^N (\omega_n^N)^2 (u^{LQR}(t) + u^{CU}(t)) - \check{f}_d(t)] \quad (17)$$

Then, the nominal LQR control input u^{LQR} is designed for the following nominal error system

$$\dot{E}(t) = AE(t) + B[-k^N (\omega_n^N)^2 u^{LQR}(t)] \quad (18)$$

Consider the cost functions of the forms

$$J = \int_0^\infty [E^T(t)QE(t) + r(-k^N (\omega_n^N)^2 u^{LQR}(t))^2] dt \quad (19)$$

where Q is a symmetric and positive definite matrix and r is a positive constant. The LQR controller is given by

$$u^{LQR}(t) = \frac{1}{K^N (\omega_n^N)^2} KE(t) \quad (20)$$

where the state-feedback gain K can be given as

$$K = r^{-1} B^T P \quad (21)$$

where P is the positive definite solution of the associated Riccati equation

$$A^T P + PA - r^{-1} P B B^T P + Q = 0 \quad (22)$$

The compensator is designed to produce a compensating signal to restrain the influences of the equivalent uncertainties $f_d(t)$ in Eq. (13). The compensator is constructed with a signum function as follows

$$u^{cu} = \frac{1}{k^N (\omega_n^N)^2} \lambda \operatorname{sgn}(E^T P B) \quad (23)$$

where λ is a given constant that is $\lambda > D > 0$.

Theorem 1: For a single-input second-order nonlinear system given by Eq. (5) with the proposed error system defined in Eq. (12) both the system stability and tracking convergence are guaranteed if the control law is given by Eq. (15).

Proof: To have a concise manner of representation, in the rest of this paper the time variable t will be omitted. Considering the positive definite Lyapunov function

$$V = E^T P E \quad (24)$$

By differentiating V with respect to time and substituting Eqs. (21) and (22) into it, one have

$$\begin{aligned} \dot{V} &= E^T P [A E + B(-k^N (\omega_n^N)^2 (u^{LQR} + u^{cu}) - \tilde{f}_d)] + \\ &\quad [A E + B(-k^N (\omega_n^N)^2 (u^{LQR} + u^{cu}) - \tilde{f}_d)] P E \\ &= E^T P A E + 2E^T P B [-k^N (\omega_n^N)^2 (u^{LQR} + u^{cu}) - \\ &\quad \tilde{f}_d] + E^T A^T P E \\ &= E^T (P A + A^T P) E - 2k^N (\omega_n^N)^2 E^T P B u^{LQR} - \\ &\quad 2k^N (\omega_n^N)^2 E^T P B u^{cu} - 2E^T P B \tilde{f}_d \\ &= E^T (P A + A^T P) E - 2r^{-1} B^T P E^T P B E - \\ &\quad 2k^N (\omega_n^N)^2 E^T P B u^{cu} - 2E^T P B \tilde{f}_d \\ &= E^T (A^T P + P A - P B r^{-1} B^T P) E - \\ &\quad E^T P B (r^{-1}) B^T P E - 2\lambda E^T P B \operatorname{sgn}(E^T P B) - \\ &\quad 2E^T P B \tilde{f}_d \\ &< -E^T Q E - E^T P B (r^{-1}) B^T P E - (2\lambda | E^T P B | \\ &\quad + 2\tilde{f}_d E^T P B) \\ &< 0 \end{aligned} \quad (25)$$

This shows that the controller satisfies the condition to enable the system stable in finite time. According to the definition of Eq. (12), if $e \rightarrow 0$ and $\dot{e} F \rightarrow 0$, then $x \rightarrow x_d$ and $\dot{x} F \rightarrow \dot{x} F_d$ as $t \rightarrow \infty$. Therefore, the control law ensures both the stability of the system and the convergence of the motion tracking.

IV. SIMULATION RESULTS

In this section, the proposed Robust LQR controller is validated through simulations with a sample rate of 10 kHz. The results are shown and discussed in this section.

A. Controller Implementation Issues

Insight into Eq. (15) to Eq. (23) reveals that both full-state feedback and the full-state trajectory are required to implement the proposed controller. Although the full-state trajectory can be generated by differentia-

ting the desired position trajectory in advance, the velocity feedback must be estimated since only the position can be measured by the available capacitive sensor.

The full state can be estimated by resorting to the measured position using a backward difference equation [34]. However, due to the limitations of the accuracy and quantization noise, the achievable bandwidth of the feedback controller is restricted as a result. Alternatively, a closed-loop high-gain observer can be employed without the above limitations [35]. Therefore, a one-input two-output high-gain observer is developed to estimate the full state as follows

$$\begin{aligned} \dot{\hat{X}} F &= A \hat{X} + B x \\ Y &= C \hat{X} \end{aligned} \quad (26)$$

with

$$A = \begin{bmatrix} -\beta_1/\tau & 1 \\ -\beta_2/\tau^2 & 0 \end{bmatrix}, B = \begin{bmatrix} \beta_1/\tau \\ -\beta_2/\tau^2 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad (27)$$

where the input to the observer is the measured position x , and the output of the observer is the full-state feedback, i. e., $y = [\hat{x}, \dot{\hat{x}}]^T$. The bandwidth of the observer is determined by the design of the gains β_1 and β_2 , and the accuracy of the estimated state relies on the design parameter τ , as τ approaches zero, the estimation becomes exact.

Thus, the controller consists of the above proposed three control inputs and the high-gain observer. The inputs of this controller are the desired position trajectory and measured position, while the output is the voltage that will be applied to PZT actuator. Specifically, once the desired position trajectory x_d is given, the velocity and acceleration trajectories (\dot{x}_d and \ddot{x}_d) can be obtained.

B. PEA Model

For the purpose of simulation, one applies a Bouc-Wen model for hysteresis in this paper. In view of the fact that the hysteresis is the main nonlinearity that can be regarded as the uncertainty of the PEAs system, the hysteresis is modeled integrated into the second-order PEA model for accurate simulation. The Bouc-Wen model has already been verified that the Bouc-Wen model is applicable to describe the hysteresis loop of PEAs [36]. Thus, The piezoelectric actuator model with nonlinear hysteresis for simulation can be written as

$$\ddot{x} + 2\xi^N \omega_n^N \dot{x} + (\omega_n^N)^2 x = (\omega_n^N)^2 (k^N u - h) \quad (28)$$

$\dot{h} = \alpha \dot{u} - \beta | \dot{u} | h | h |^{n-1} - \gamma \dot{u} | h |^n$ (29) where h is the nonlinear hysteresis that represents the hysteretic loop in terms of displacement whose magnitude and shape are determined by parameters α, β, γ, d is the piezoelectric coefficient, u denotes the input voltage, and the order n governs the smoothness of the transition from elastic to plastic response. For the elastic structure and material, $n = 1$ is assigned in Eq. (29) as usual.

Based on the Bouc-Wen PEA model for simulation,

according to the uncertainties estimation strategy 10, the discrete-time estimation of system uncertainties \bar{f}_d is rewritten as

$$\bar{f}_d(t) = \ddot{x}(t) + 2\xi^N \omega_n^N \dot{x}(t) + (\omega_n^N)^2 x(t) - (\omega_n^N)^2 [k^N u(t - T) - h(t - T)] \quad (30)$$

The modified system uncertainties \bar{f}_d is applied in control law for simulation. The above parameters chosen in this paper are calculated through simulations from [34], [37] and the value of these parameters are shown in Table I. The sampling time interval T is 0.000,1 s.

C. Step Response

First, the set-point capability of the LQR controller is examined. The parameters of this controller for step response has been shown in Table II, and the results for 1.5 μm amplitude step signal is illustrated in Fig. 2 for a clear explaining.

Table I
PARAMETERS OF THE PEA WITH BOUC-WEN MODEL

Parameter	Value
n	1
ξ^N	$1.231,5 \times 10^4$
ω_n^N	$1.222,5 \times 10^6$
k^N	$1.733,9 \times 10^{-6}$
α	0.357,5
β	0.036,4
γ	0.027,2

Table II
PARAMETERS OF THE IMPLEMENTED CONTROLLER

Parameter	Value
λ	8.5×10^6
τ	0.2
β_1	2.5×10^3
β_2	8.0×10^4
r	1
Q	$\text{diag}(150, 1.0 \times 10^{-8}, 8.0 \times 10^5)$

The step response shown in [38], Fig. 4 indicated that a settling time approximating to 100 ms was achieved by employing model predictive output integral discrete sliding mode control. Compared with this result, the response of proposed controller observed from Fig. 2 provides a fast convergence and smooth control in approximately 49 ms (1% settling time) with a 5.93% overshoot which is relative small.

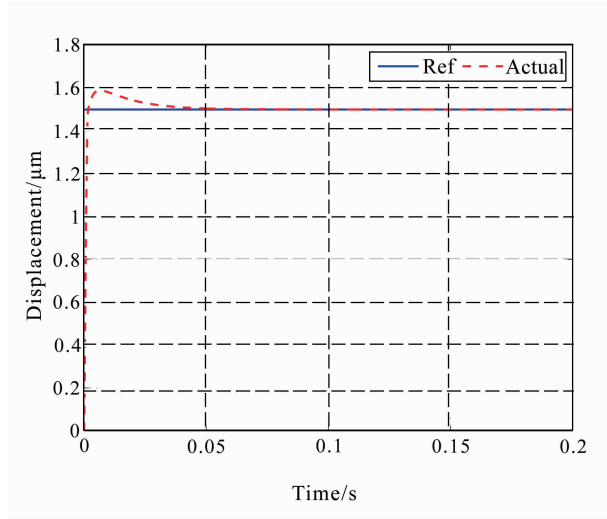


Fig. 2 Simulation responses to step signals with amplitudes of 1.5 μm .

D. Sinusoidal Tracking

Fig. 3 and Fig. 4 have shown the performances of tracking a waveform of 5 μm peak-to-peak (p-p) amplitude by employing the robust LQR controller. It can be clearly observed from the trajectories that the proposed controller can track the desired signal precisely and chattering free. Additionally, this controller is able to suppress the error within $\pm 0.020,2 \mu\text{m}$ and $\pm 0.065,6 \mu\text{m}$ which are 0.81% and 2.62% of the tracking range with the frequency of 10 Hz and 20 Hz, respectively.

E. Responses to staircase signal

A staircase signal is applied to the proposed controller for tracking control. Fig. 5 shows that a step of the staircase signal covering the range of 1 μm by 1,000 steps with each step lasting for 0.1 s. From the figure, one can see that the controller can guarantee the steady-state error of 0 nm for approximating 88% duration of the step. Thus, the steps can be identified indicates that the positioning resolution of the proposed controller is less than 1 nm.

F. Responses to triangular waveform

The performances for triangular waveform tracking are shown in Fig. 6 and Fig. 7. The maximum tracking errors are respective $\pm 0.004,5 \mu\text{m}$ and $\pm 0.045,5 \mu\text{m}$ when the fundamental frequency of the triangular waveform is 1 Hz and 10 Hz. Moreover, it can be seen from the figures the duration of the transient phase for both of 1 Hz and 10 Hz are zero, and the steady errors are also zero. Thus, the proposed controller can track the triangular waveform precisely and chattering free.

G. Robustness Analysis

The robustness of the proposed controller is examined by applying an external disturbance to the control input of the PEA system. This unknown disturbance input is

set as $\epsilon = [1 \sin(10t + 2) + 1]$. The desired displacement is the same as the 10 Hz one used in sinusoidal tracking. The results are shown in Fig. 8 and the maximum tracking error remains below 1% of the tracking

range which is $\pm 0.023,4 \mu\text{m}$. This suggests the robustness of the proposed LQR controller despite the presence of an external disturbance adding to the control input u .

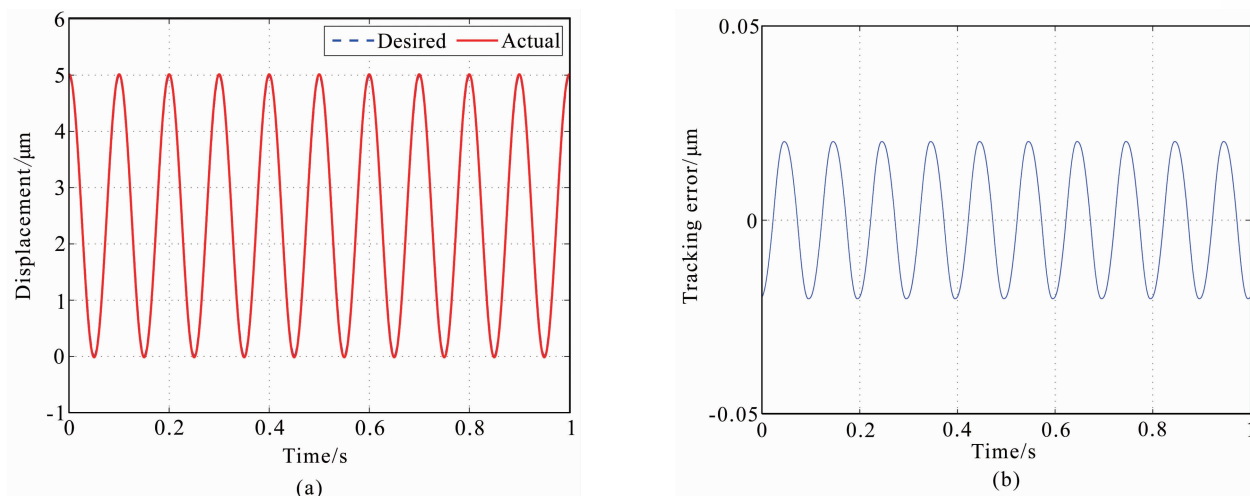


Fig. 3 Simulation results of response to a 10 Hz sinusoidal signal

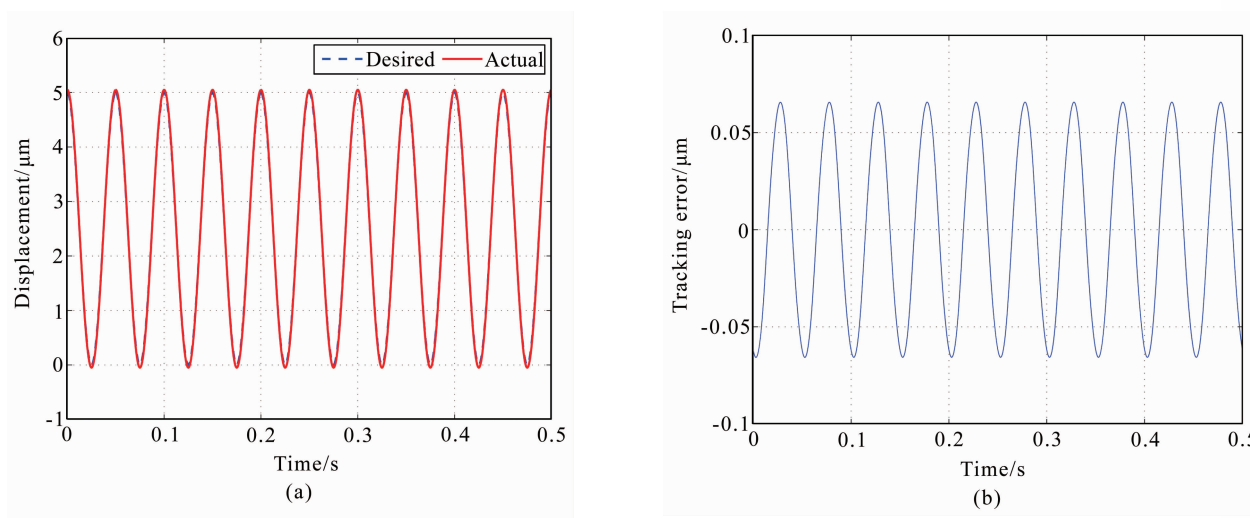


Fig. 4 Simulation results of response to a 20 Hz sinusoidal signal

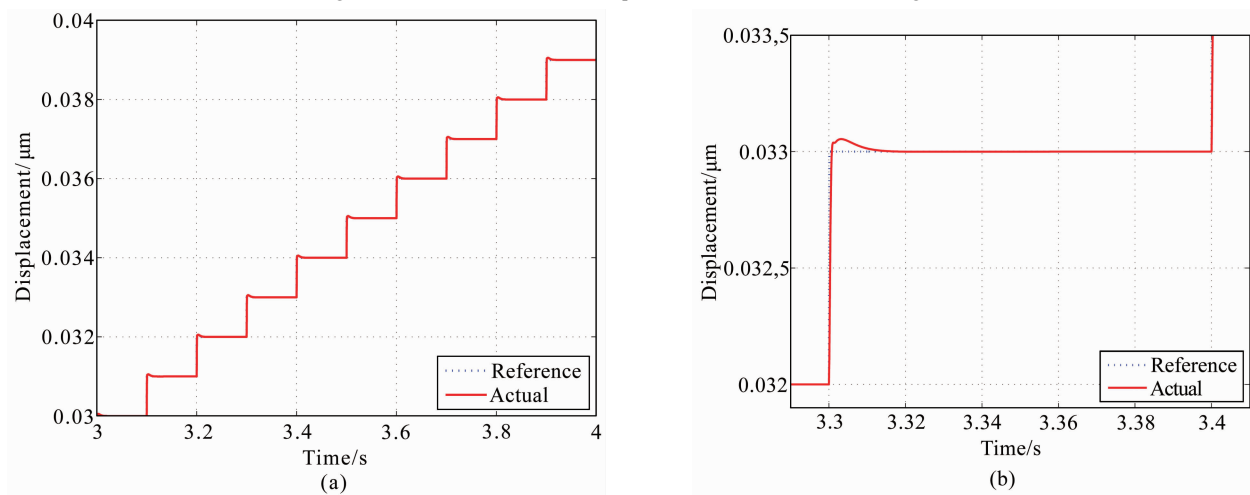


Fig. 5 Simulation responses to staircase signals covering the range 1 μm with 1,000 steps

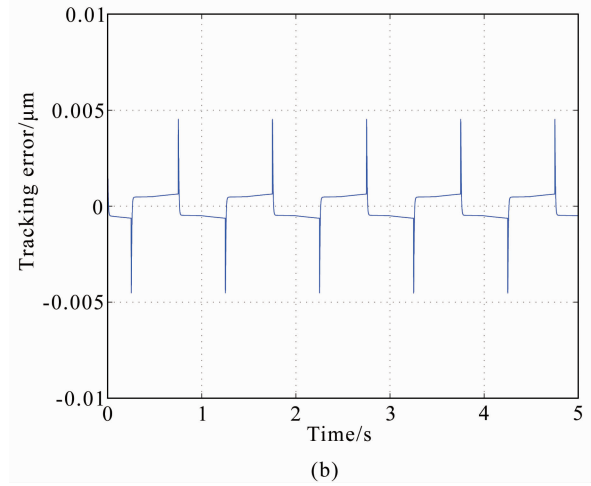
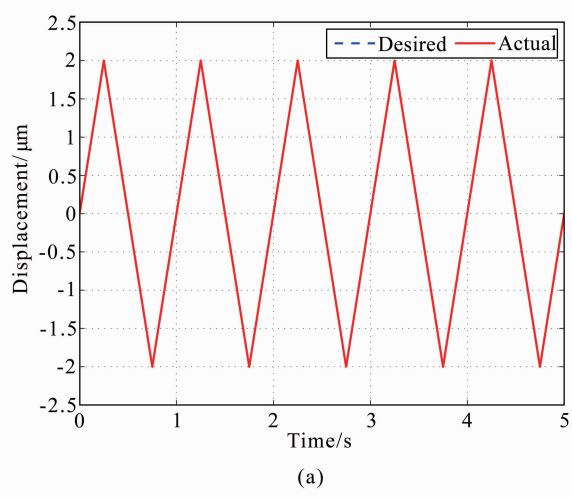


Fig. 6 Simulation results of response to a 1 Hz triangular waveform

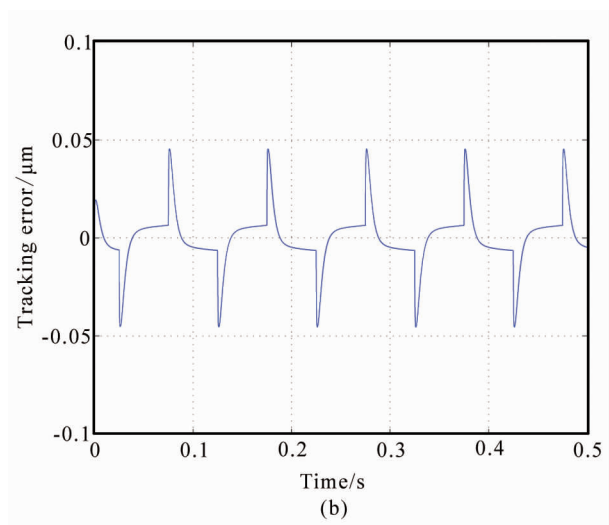
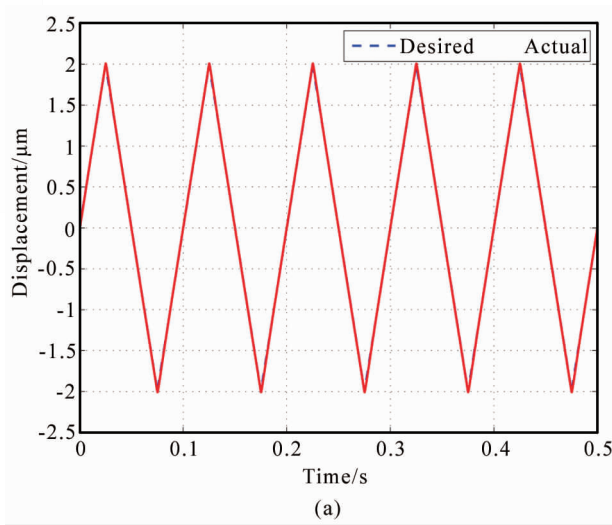


Fig. 7 Simulation results of response to a 10 Hz triangular waveform

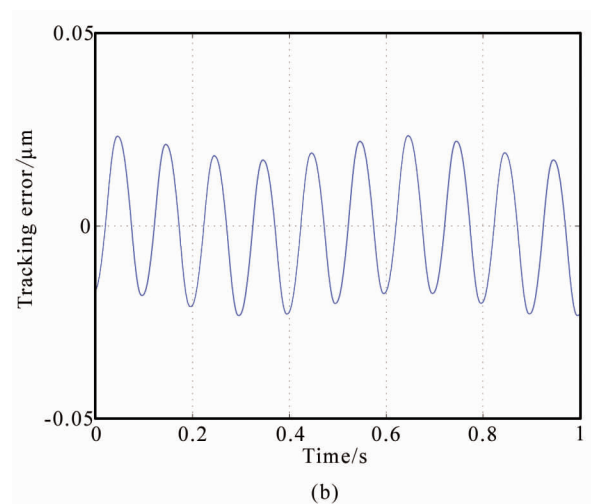
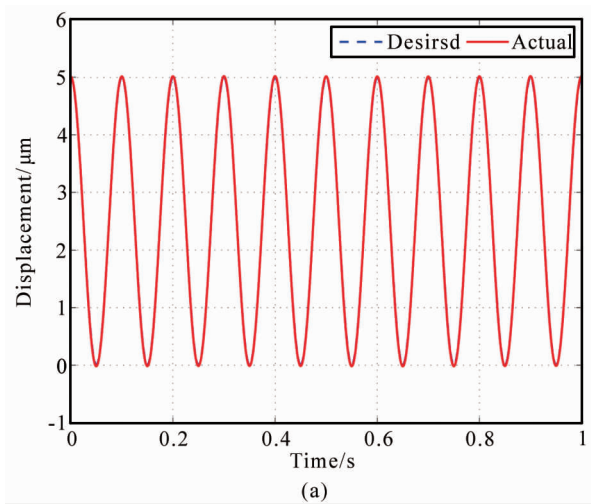


Fig. 8 Robustness analysis of the proposed controller

V. CONCLUSIONS

The main contribution of this paper lies in the proposal of a discrete-time robust LQR controller for the piezoelectric actuator nanopositioning and tracking without the accurate hysteresis model. The controller was designed based on Lyapunov stability analysis and its performance was verified by a series of simulation studies. According to the response simulation results, the proposed controller can obtain a fast transient response with a low overshoot. Results of sinusoidal motion and triangular waveform tracking illustrate that the designed controller has precise tracking performances with small peak-to-peak tracking errors. With the proposed controller, the resolution which is less than 1 nm of the system has been achieved, which validates the effectiveness of the designed controller and the hysteresis has been significantly compensated to a low level of magnitude as well. In the future, a discrete-time observer used to estimate all the uncertainties of the system should be considered to be designed.

REFERENCES

- [1] Y. Yong, S. Moheimani, B. J. Kenton, and K. Leang, "Invited review article: High-speed flexure-guided nanopositioning: Mechanical design and control issues," *Review of scientific instruments*, vol. 83, no. 12, p. 121101, 2012.
- [2] Y. Suzuki, N. Sakai, A. Yoshida, Y. Uekusa, A. Yagi, Y. Imaoka, S. Ito, K. Karaki, and K. Takeyasu, "High-speed atomic force microscopy combined with inverted optical microscopy for studying cellular events," *Scientific reports*, vol. 3, 2013.
- [3] H. Song, G. Vdovin, R. Fraanje, G. Schitter, and M. Verhaegen, "Extracting hysteresis from nonlinear measurement of wavefront-sensorless adaptive optics system," *Optics letters*, vol. 34, no. 1, pp. 61-63, 2009.
- [4] J. -j. Wei, Z. -c. Qiu, Y. -c. Wang *et al.*, "Experimental comparison research on active vibration control for flexible piezoelectric manipulator using fuzzy controller," *Journal of Intelligent and Robotic Systems*, vol. 59, no. 1, pp. 31-56, 2010.
- [5] K. W. Chan and W. -H. Liao, "Precision positioning of hard disk drives using piezoelectric actuators with passive damping," in *Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference on*. IEEE, 2006, pp. 1269-1274.
- [6] S. Devasia, E. Eleftheriou, and S. R. Moheimani, "A survey of control issues in nanopositioning," *Control Systems Technology, IEEE Transactions on*, vol. 15, no. 5, pp. 802-823, 2007.
- [7] G. Clayton, S. Tien, A. Fleming, S. Moheimani, and S. Devasia, "Inverse-feedforward of charge-controlled piezopositioners," *Mechatronics*, vol. 18, no. 5, pp. 273-281, 2008.
- [8] J. Minase, T. -F. Lu, B. Cazzolato, and S. Grainger, "A review, supported by experimental results, of voltage, charge and capacitor insertion method for driving piezoelectric actuators," *Precision Engineering*, vol. 34, no. 4, pp. 692-700, 2010.
- [9] Y. T. Ma, L. Huang, Y. B. Liu, and Z. H. Feng, "Note: Creep character of piezoelectric actuator under switched capacitor charge pump control," *Review of Scientific Instruments*, vol. 82, no. 4, pp. 046 106-046 106, 2011.
- [10] G. -Y. Gu, M. -J. Yang, and L. -M. Zhu, "Real-time inverse hysteresis compensation of piezoelectric actuators with a modified prandtl-ishlinskii model," *Review of Scientific Instruments*, vol. 83, no. 6, pp. 065 106-065 106, 2012.
- [11] T. Fett and G. Thun, "Determination of room-temperature tensile creep of pzt," *Journal of materials science letters*, vol. 17, no. 22, pp. 1929-1931, 1998.
- [12] L. Deng and Y. Tan, "Modeling of rate-dependent hysteresis in piezo-electric actuators," in *Control Applications, 2008. CCA 2008. IEEE International Conference on*. IEEE, 2008, pp. 978-982.
- [13] W. T. Ang, P. K. Khosla, and C. N. Riviere, "Feedforward controller with inverse rate-dependent model for piezoelectric actuators in trajectory-tracking applications," *Mechatronics, IEEE/ASME Transactions on*, vol. 12, no. 2, pp. 134-142, 2007.
- [14] R. Robinson, "Interactive computer correction of piezoelectric creep in scanning tunneling microscopy images," *Journal of Computer-Assisted Microscopy*, vol. 2, no. 1, pp. 53-58, 1996.
- [15] S. R. Cohen and A. Bitler, "Use of afm in bio-related systems," *Current Opinion in Colloid & Interface Science*, vol. 13, no. 5, pp. 316-325, 2008.
- [16] H. Xu, T. Ono, and M. Esashi, "Precise motion control of a nanopositioning pzt microstage using integrated capacitive displacement sensors," *Journal of Micromechanics and Microengineering*, vol. 16, no. 12, p. 2747, 2006.
- [17] A. Sebastian and S. M. Salapaka, "Design methodologies for robust nano-positioning," *Control Systems Technology, IEEE Transactions on*, vol. 13, no. 6, pp. 868-876, 2005.
- [18] M. Boukhnifer and A. Ferreira, "H_∞ loop shaping bilateral controller for a two-fingered tele-manipulation system," *Control Systems Technology, IEEE Transactions on*, vol. 15, no. 5, pp. 891-905, 2007.
- [19] W. Sun, H. Gao, and O. Kaynak, "Finite frequency control for vehicle active suspension systems," *Control Systems Technology, IEEE Transactions on*, vol. 19, no. 2, pp. 416-422, 2011.
- [20] J. Dong, S. M. Salapaka, and P. M. Ferreira, "Robust control of a parallel-kinematic nanopositioner," *Journal of dynamic systems, measurement, and control*, vol. 130, no. 4, 2008.

- [21] T. W. Seo, H. S. Kim, D. S. Kang, and J. Kim, "Gain-scheduled robust control of a novel 3-dof micro parallel positioning platform via a dual stage servo system," *Mechatronics*, vol. 18, no. 9, pp. 495-505, 2008.
- [22] H. C. Liaw, B. Shirinzadeh, and J. Smith, "Enhanced sliding mode motion tracking control of piezoelectric actuators," *Sensors and Actuators A: Physical*, vol. 138, no. 1, pp. 194-202, 2007.
- [23] —, "Sliding-mode enhanced adaptive motion tracking control of piezoelectric actuation systems for micro/nano manipulation," *Control Systems Technology, IEEE Transactions on*, vol. 16, no. 4, pp. 826-833, 2008.
- [24] J. -C. Shen, W. -Y. Jywe, C. -H. Liu, Y. -T. Jian, and J. Yang, "Sliding-mode control of a three-degrees-of-freedom nanopositioner," *Asian Journal of Control*, vol. 10, no. 3, pp. 267-276, 2008.
- [25] C. -Y. Lin and P. -Y. Chen, "Precision tracking control of a biaxial piezo stage using repetitive control and double-feedforward compensation," *Mechatronics*, vol. 21, no. 1, pp. 239-249, 2011.
- [26] J. Yao, Z. Jiao, D. Ma, and L. Yan, "High-accuracy tracking control of hydraulic rotary actuators with modeling uncertainties," *Mechatronics, IEEE/ASME Transactions on*, vol. 19, no. 2, pp. 633-641, 2014.
- [27] Z. Li, H. Liu, B. Zhu, H. Gao, and O. Kaynak, "Nonlinear robust attitude tracking control of a table-mount experimental helicopter using output-feedback," *Industrial Electronics, IEEE Transactions on*, 2015, doi 10. 1109/TIE. 2015. 2414396.
- [28] J. Yao, Z. Jiao, and D. Ma, "Extended-state-observer-based output feed-back nonlinear robust control of hydraulic systems with backstepping," *Industrial Electronics, IEEE Transactions on*, vol. 61, no. 11, pp. 6285-6293, 2014.
- [29] Z. Li, H. Liu, B. Zhu, and H. Gao, "Robust second-order consensus tracking of multiple 3-dof laboratory helicopters via output feedback," *Mechatronics, IEEE/ASME Transactions on*, 2015, doi 10. 1109/TMECH. 2014. 2388240.
- [30] W. Sun, Y. Zhao, J. Li, L. Zhang, and H. Gao, "Active suspension control with frequency band constraints and actuator input delay," *Industrial Electronics, IEEE Transactions on*, vol. 59, no. 1, pp. 530-537, 2012.
- [31] J. Douglas and M. Athans, "Robust linear quadratic designs with real parameter uncertainty," *Automatic Control, IEEE Transactions on*, vol. 39, no. 1, pp. 107-111, 1994.
- [32] H. -J. Shieh, Y. -J. Chiu, and Y. -T. Chen, "Optimal pid control system of a piezoelectric micro-positioner," in *System Integration, 2008 IEEE/SICE International Symposium on*. IEEE, 2008, pp. 1-5.
- [33] S. Aggarwal, M. Garg, and A. Swarup, "Comparative analysis of traditional and modern controller for piezoelectric actuated nanopositioner," *International Journal of Nano Devices, Sensors and Systems (IJ-Nano)*, vol. 1, no. 2, pp. 53-64, 2012.
- [34] Y. Li and Q. Xu, "Adaptive sliding mode control with perturbation estimation and pid sliding surface for motion tracking of a piezo-driven micromanipulator," *Control Systems Technology, IEEE Transactions on*, vol. 18, no. 4, pp. 798-810, 2010.
- [35] J. J. Gorman, N. G. Dagalakis, and B. G. Boone, "Multiloop control of a nanopositioning mechanism for ultraprecision beam steering," in *Optical Science and Technology, SPIE's 48th Annual Meeting*. International Society for Optics and Photonics, 2004, pp. 170-181.
- [36] T. Low and W. Guo, "Modeling of a three-layer piezoelectric bimorph beam with hysteresis," *Microelectromechanical Systems, Journal of*, vol. 4, no. 4, pp. 230-237, 1995.
- [37] J. Li and L. Yang, "Adaptive pi-based sliding mode control for nanopositioning of piezoelectric actuators," *Mathematical Problems in Engineering*, vol. 2014, 2014.
- [38] Q. Xu and Y. Li, "Micro-/nanopositioning using model predictive output integral discrete sliding mode control," *Industrial Electronics, IEEE Transactions on*, vol. 59, no. 2, pp. 1161-1170, 2012.

Image process using Raspberry Pi mini computer

Tumurchudur Lkham^{#1}, U ranchimeg Tudevda gva^{#2}

[#]Mongolian University of Science and Technology (MUST)

Ulaanbaatar, Mongolia

¹chuugii@must.edu.mn

²uranchimeg@must.edu.mn

Abstract— Image processing is used in various fields, this paper presents the implementation of image processing operations on Raspberry Pi. Raspberry Pi is a powerful and affordable small computer. Hence, it consists of high functional features and is sold at low cost for easy access to everybody. The Raspberry Pi board is single, powerful computer having 85.6×56 mm Dimension of business card, and it is used to reducing the complexity of system in real time application. This platform is mainly based on C/C++. Raspberry Pi consist of Camera slot Interface (CSI) to interface the Raspberry Pi camera. Here, the dark and low contrast images are captured by a camera and then enhanced in order to identify the particular region of image. Algorithms for edge, corner and line detection have been implemented by using Simulink with the Computer Vision System Toolbox and other in-build tools.

Keywords— C/C++, Raspberry Pi, image enhancement algorithm, hardware design, software design, image processing

I. INTRODUCTION

The very first computer had a size as large as a room and consumed high amount of power. However, things had been changed since semiconductor technology was introduced. The size of computer became rapidly shrinking while the computing power was growing. The advent of integrated circuits (ICs) or chips had speeded up the miniaturization of computer technology where smaller and smaller elements are integrated to build a chip [1]. This is described by so called Moore's law which says that the number of transistors on chip is doubling every two years [2]. As a result, personal computers (PCs) and other devices such as laptops, mobile phones, tablets have been developed and using in daily life. Moreover, the process of miniaturization of computers continues leading to an emergence of very small computers which can be used in many applications covering every aspect of life. Ideas for such new applications could be developed by anybody interested if they have knowledge about computer operation, programming and access to a very small experimental platform. Raspberry Pi for technology enthusiasts, hobbyists, engineers who are using it in commercial applications as well as in universities both in education and research projects. This success of Raspberry Pi is due to the unique combination of proper price, hardware, software and the cre-

ation of global community of users and developers sharing information.

The interaction between the PC and Raspberry Pi is handled by Matlab and Simulink software where Simulink makes possible porting of the Matlab software to wide variety of devices and platforms. Matlab in Raspberry Pi can run both in a simulation mode where the board is connected to a PC and in a standalone mode where a software is downloaded onto the board and runs independently from a PC. Experience gained by using Raspberry Pi with Simulink has universal applications.

General information of Raspberry Pi such as released models and components including the details of its architecture; Central Processing Unit (CPU), Graphic Processing Unit (GPU), Random Access Memory (RAM) and general purpose input and output (GPIO) pins with an emphasis on the recently released model of Raspberry Pi, namely Raspberry Pi 2 model B. Basic Operating System (OS) and setup instruction including a secure digital (SD) card preparation, connection and configuration instructions are presented.

II. IMAGE PROCESSING

In general, any digital image processing algorithm consists of three stages:

- ① Input Image (Original Image);
- ② Image Processor (Processed Image);
- ③ Output Image (Output Image).

In the input stage image is captured by a camera. It sent particular system to focus on Pixel of image that gives its output as a processed image.

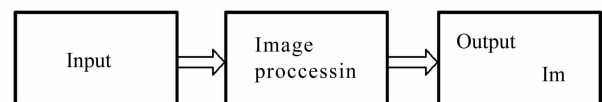


Fig. 1 General Block diagram of image processing

III. RASPBERRY PI HARDWARE DESIGN

All Provides a background information of Raspberry Pi starting with general information such as released models and basic components including GPIO opera-

tions. It also provides the information regarding the history of this single board computer as well as a detail of its system on a chip (SoC) including CPU, GPU and RAM.

The Raspberry Pi board is the central module of the whole embedded image capturing and processing system as given in Fig. 2. Its main parts include: main process-

ing chip unit, memory, power supply HDMI Output, VGA display, Ethernet port, and USB ports.

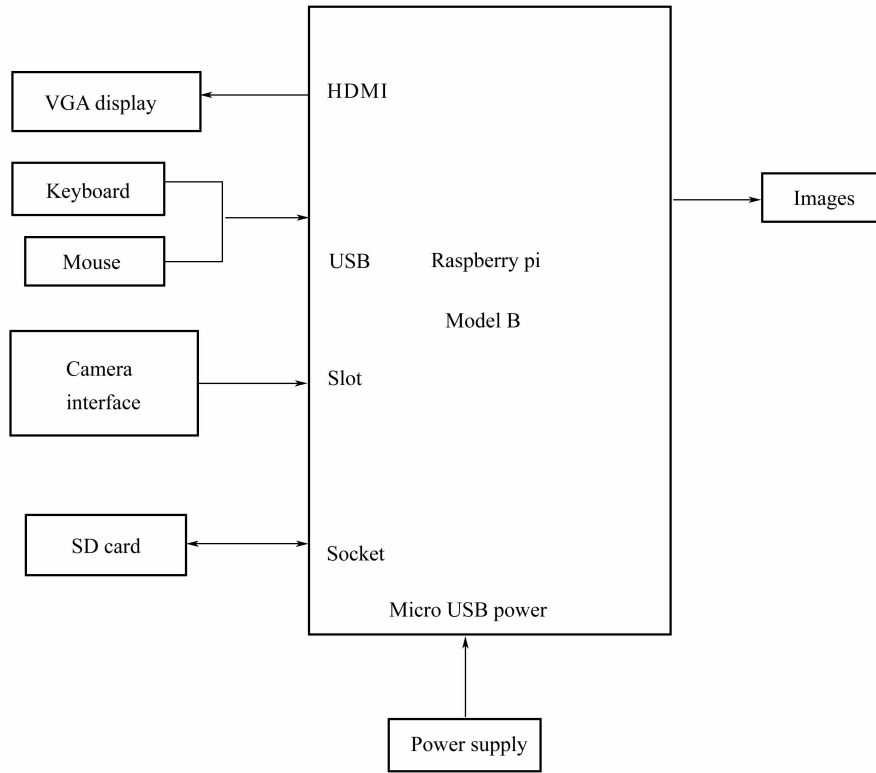


Fig. 2 Block Diagram

A. Raspberry Pi board

Raspberry Pi, as shown in Fig. 3, is a single-board computer having a size as small as a credit card from the Raspberry Pi foundation. An idea of producing Raspberry Pi began in 2006 with a realization that young generation is missing knowledges about computer operation. A group of academics and engineers from University of Cambridge. Therefore, decided to develop a very small computer which everyone could afford to buy to create learning environment in programming[3]. The Raspberry Pi project became promising with the appearance of cheap and powerful mobile processors with many advanced features allowing a possible development of Raspberry Pi which was continued under specially created Raspberry Pi foundation with the first product launched in 2012.

Development of Raspberry Pi computer is a continuing process. Until spring of 2015, Raspberry Pi foundation has released five models in two generations of the computer. The first generation consists of four models as shown in Fig. 4. First released was model B followed by

model A, model B + and model A + respectively. The two latter models are upgraded versions of their previous releases to make the computer more efficient and convenient to users especially by having lower power consumption and more (Universal Serial Bus) USB ports. The second generation consists of only one model called Raspberry Pi 2 model B whose specification is based on Raspberry Pi model B + but with faster CPU and more memory [17].

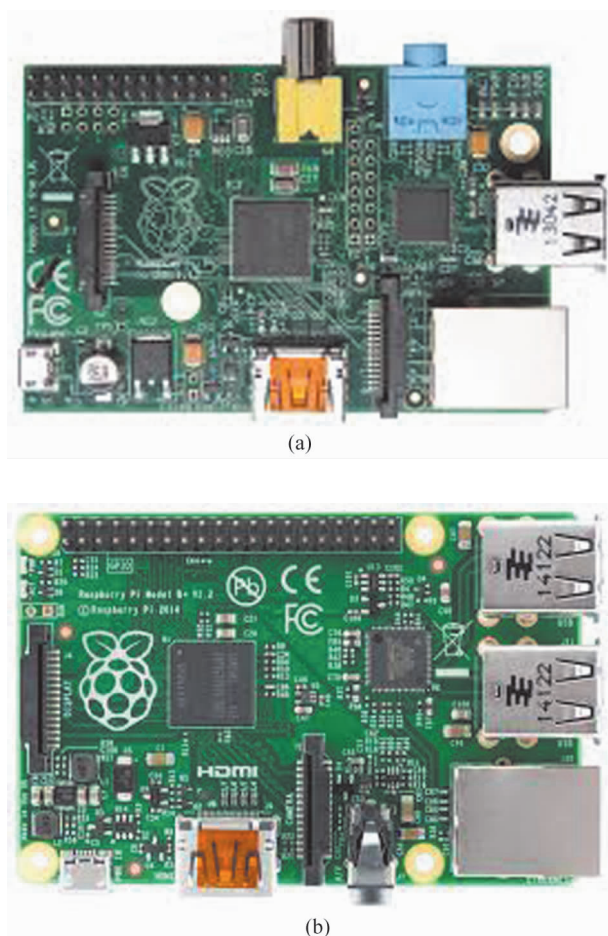


Fig. 3 Raspberry Pi's generation 1 and generation 2
(a)Raspberry Pi model B[15];(b)Raspberry Pi 2 model B[16]

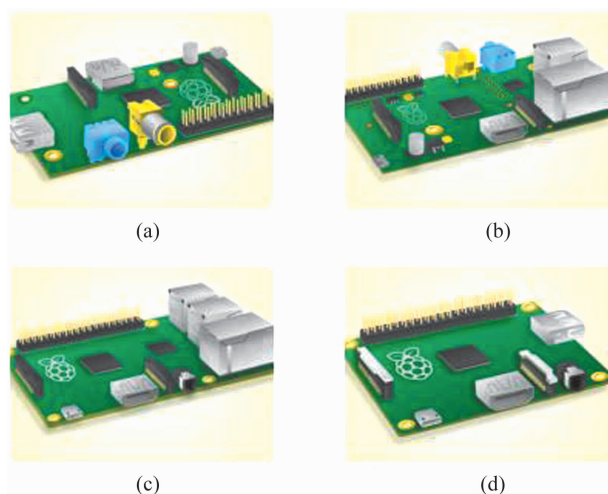


Fig. 4 Raspberry Pi's released models
(a)Raspberry Pi model A;(b)Raspberry Pi model B;
(c)Raspberry Pi model B+;(d)Raspberry Pi model A+

Raspberry Pi model B, as shown in Fig. 4 (b), was released in the early 2012 with the specification of 256 MB RAM, two USB ports and one Ethernet port. Later in the same year, in a new release, the amount of RAM was increased to 512 MB followed by the release of the lower-spec board, model A as shown in Fig. 4 (a). Model A was released with same amount of RAM of the older model B at 256 MB but with one USB port and no Ethernet port. These first generation computers use BroadcomSoC, BCM2835, which integrates 700 MHz single-core ARM1176JZF-S CPU, VideoCore IV GPU and variety of peripherals. While model B can be used in any applications, the cheaper model A is useful in specific applications that require light-weight and low power consumption such as robotics or any portable media services.

In early 2015, the next generation of Raspberry Pi computer called Raspberry Pi 2 model B was released. The model represents significant upgrade of the Raspberry Pi capabilities and was enthusiastically welcomed by the community. Raspberry Pi 2 model B uses a much more powerful BCM2836 which offers 900 MHz quad-core ARM Cortex-A7 CPU, 1 GB RAM and the same specification of graphics processor as previously making it run approximately six times faster comparing to its predecessors at an unchanged price. Moreover, the supports of Windows 10 Internet of Things (IoT) version is available for Raspberry Pi 2 model B for free of charge providing a great potential for future usage [4][5].

Table I summarizes major features of the Raspberry Pi models. Further details are explained in the next section [18].

Table I MAJOR FEATURES OF THE RELEASED MODELS OF RASPBERRY PI

Raspberry Pi	Generation 1				Generation 2
Specifications	Model A	Model B	Model A +	Model B +	Model B
Power	300 mA	700 mA	200 mA	600 mA	900 mA
Ethernet Port	No	Yes	No	Yes	Yes
USB Port	1	2	1	4	4
GPIO	26	26	40	40	40
SD Card Slot	SD	SD	microSD	microSD	microSD
SoC	BCM2835	BCM2835	BCM2835	BCM2835	BCM2836
CPU	700 MHz	700 MHz	700 MHz	700 MHz	900 MHz
	ARM11	ARM11	ARM11	ARM11	ARM Cortex-A7
	single-core	single-core	single-core	single-core	quad-core
RAM	256 MB	512 MB	256 MB	512 MB	1 GB

B. Power supply

Power Connector is to power the board with a 5 V micro USB port. The need of the power consumption required depends on the peripheral devices attached to the board, for example, model B requires around 700 mA at the idle state but can consume up to 1 000 mA if connects to other devices. In model A + and B +, the linear regulator was replaced by a switching regulator causing lower power consumption at 200 mA and 600 mA for each model respectively. More power is required for Raspberry Pi 2 model B of at least 900 mA since it has more advanced four core processors. It is recommended to use a power adapter that can produce at least 1,200 mA, however, a normal mobile phone charger with a micro USB head can be used as a power supply. Higher power is required for more stable operation [6].

C. HDMI

High-Definition Multimedia Interface (HDMI) Port is a connection for High Definition (HD) resolution display device such as a computer screen or a smart television. All models provide a resolution from 640×350 to $1,920 \times 1,200$ including Phase Alternating Line (PAL) and National Television System Committee (NTSC) standards.

D. Camera Interface

Camera Serial Interface (CSI) Connector is a connector for a camera module, as shown in Fig. 5. The camera is specifically designed for Raspberry Pi and can be connected with a ribbon cable providing a support for both photography and HD video [7].



Fig. 5 Camera Pi

E. Audio Connector

Audio Connector is an audio output through a 3.5 mm headphone jack which also supports an analogue video output for model A +, B + and Raspberry Pi 2 model B. For earlier models, A and B, the connector is only for an audio output.

F. Ethernet Port

Ethernet Port is a network connector providing a speed of 10/100 Mbit/s through an RJ45 Local Area Network (LAN) cable. Ethernet Port is available in model B, B + and Raspberry Pi 2 model B.

G. USB Port

USB Port supports any USB devices such as a keyboard, a mouse, a Wi-Fi dongle or a webcam. The USB port provided is a 2.0 version. Both model A and A + have one USB port while model B has two USB ports. There are four USB ports in model B + and Raspberry Pi 2 model B.

H. GPIO

GPIO is a set of universal input and output connector pins for general purposes such as connecting expansion boards or devices in order to control CPU or checking power consumption. There are 26 pins in model A and B and 40 pins in model A + , B + and Raspberry Pi 2 model B.

I. Light-Emitting Diodes (LEDs)

Light-Emitting Diodes (LEDs) indicate the board status. For example, when it is powered, the LED appears red, and when it connects to the network, the LED appears yellow. Model A and B provide 5 LEDs in which the first two are for indicating an SD card status and a power status and the other three are for a network status. Model A + , B + and Raspberry Pi 2 model B, provide only 2 LEDs to indicate SD card and power status. The LEDs for network status are built-in at Ethernet socket.

J. Display Serial Interface (DSI) Connector

Display Serial Interface (DSI) Connector is for connecting the board with a display module such as PiTFT, which is a Thin Film Transistor liquid crystal display shown in Fig. 5 (b). The connection is done with a flexible flat cable.

K. Display Serial Interface (DSI) Connector

Display Serial Interface (DSI) Connector is for connecting the board with a display module such as PiTFT, which is a Thin Film Transistor liquid crystal display shown in Fig. 5 (b). The connection is done with a flexible flat cable.

L. SoC

SoC is a Broadcom chip containing CPU, GPU and RAM. The first generation of Raspberry Pi, namely model A, B, A + and B + , uses BCM2835 chip providing 700 MHz ARM1176JZF-S single-core CPU, VideoCore IV GPU and 256 MB or 512 MB RAM depending on the model. For the second generation, Raspberry Pi 2 model B, the chip was upgraded to BCM2836 which provides 900 MHz ARM Cortex-A7 quad-core CPU and 1 GB RAM with the same GPU [12].

SOFTWARE DESIGN

In late 2012, Raspberry Pi's GPU driver code running on the ARM was released as open source and available for download [8]. Then later, in 2014, the full register-level documentation including a graphics driver stack of the chip were released publicly [9]. From the diagram shown in Fig. 6, the part covered with ARM was the first release followed by the part called VideoCore IV GPU. The parts, colored in orange, indicate the closed source such that the driver code running on the ARM was available, however, the graphic libraries provided by out-sourced suppliers are closed source [10]. VideoCore IV

GPU was a binary blob, which is a closed source binary driver, prior the latter announcement making it difficult for users to fully take advantage of the chip. Although the release from Broadcom was meant for a mobile SoC called BCM21553, it is compatible with BCM2835, a SoC used in Raspberry Pi. This release allows users to fully understand its internal operation as well as develop own open source drivers or write codes for General Purpose computing on Graphic Processing Unit (GPGPU). Hence, making it attractive to academics, engineers and hobbyists who are interested in exploiting the capability of the chip.

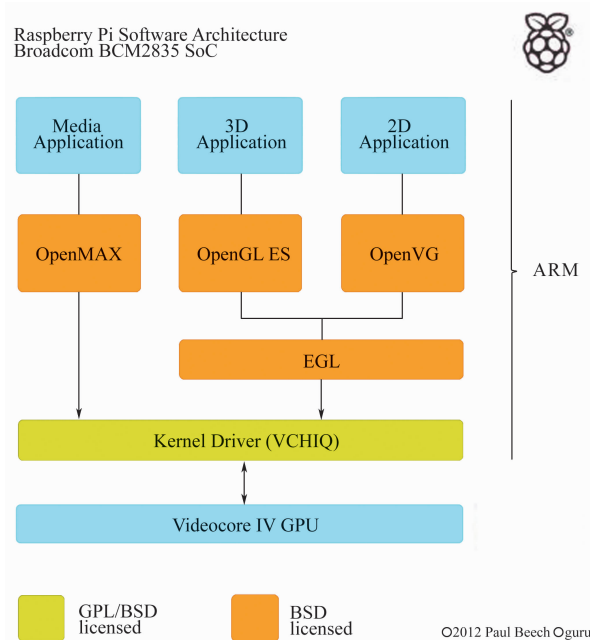


Fig. 6 Raspberry Pi's software architecture

A. OpenCV

Open Source Computer Vision contains a library programming functions mainly aimed at real-time applications. Open CV supports a lot of programming languages such as C ++ , python, Java, etc. , and is available on different platforms including Windows, Linux, Android, and OS. Here C/C ++ is used as a programming language. Multiple language bindings are available for Open CV, such as Open CV Dot Net and Emgu CV [13].

V. RESULT AND TEST

Matlab, integrated with Simulink, is a programming language software developed by MathWorks. It provides numerical computation tools and models used by academics and engineers for various systems design and development [11]. It is also widely used in university education. Matlab and Simulink support package for Raspberry Pi is composed of two parts of which one is for Matlab and another one is for Simulink. Matlab support enables development software for algorithms which can run in Raspberry Pi [14].

VI. CONCLUSION

Previously, image processing was usually made via python programming language on the Raspberry Pi, but the separation of the image process was very difficult. We have studied the basic possibilities of Raspberry Pi and tested its development with the help of Matlab software on the C/C++ programming language. Its advantage is the possibility to use commands for distinguishing the image easily. In addition, it is limited with the device and testing Matlab software simulator. In the future, we have been working to review and improve algorithms which is determining the image. See Fig. 7 to Fig. 9.



Fig. 7 Original Image

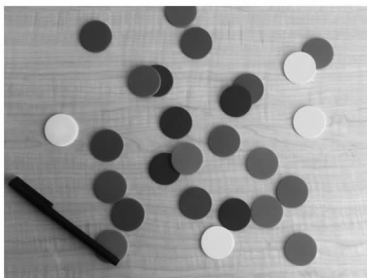


Fig. 8 Gray image



Figure 9 Dark and Bright image

ACKNOWLEDGMENT

I implemented the algorithm to enhance an image in different enhancement degree using the raspberry Pi. It was found that the algorithm developed for the Raspberry Pi executes successfully and gives a very colorful image.

REFERENCES

- [1] "Computer," Internet: <http://en.wikipedia.org/wiki/Computer> [accessed on Apr 24, 2016].
- [2] "Moore's law," Internet: http://en.wikipedia.org/wiki/Moore%27s_law [accessed on Apr 27, 2016].
- [3] "The making of pi," Internet: <https://www.raspberrypi.org/about/> [accessed on Apr 20, 2016].
- [4] "Raspberry pi 2 model b," Internet: <https://www.raspberrypi.org/products/raspberry-pi-2-model-b/> [accessed on May 1, 2016].
- [5] "Raspberry pi 2 on sale now at 35," Internet: <https://www.raspberrypi.org/raspberry-pi-2-on-sale/> [accessed on Apr 25, 2016].
- [6] "Power," Internet: <https://www.raspberrypi.org/help/faqs/#power> [accessed on Apr 27, 2016].
- [7] "Camera module," Internet: http://www.geeetech.com/wiki/index.php/Raspberry_Pi_Camera_Module [accessed on Apr 27, 2016].
- [8] Bradbury, "Open source armuserland," Internet: <https://www.raspberrypi.org/open-source-arm-userspace/> [accessed on Apr 27, 2016].
- [9] "Android for all; Broadcom gives developers keys to thevideocore R kingdom," Internet: <http://blog.broadcom.com/chip-design/android-for-all-broadcom-gives-developers-keys-to-the-videocore-kingdom/> [accessed on Apr 28, 2016].
- [10] L. Upton, "Libraries, codecs, oss," Internet: <https://www.raspberrypi.org/libraries-codecs-oss/> [accessed on Apr 28, 2016].
- [11] "Matlab," Internet: <http://en.wikipedia.org/wiki/MATLAB> [accessed on Apr 30, 2016].
- [12] Siriphat pomyem signal and image processing with matlab on Raspberry pi platform, Examiner: Prof. Irek Defee. May 2015 pp 3-8.
- [13] Introduction to Computer Vision in Python, Brian Thorne HitLabNZ, University of Canterbury, 2009, p. 2.
- [15] "Example", Internet: <http://www.mathworks.com/help/images/examples/detect-and-measure-circular-objects-in-an-image.html>. <http://www.nexuscyber.com/raspberry-pi-model-b-rev-20-512mb>.
- [16] "Raspberry pi image", Internet: https://en.wikipedia.org/wiki/Raspberry_Pi.
- [17] "Raspberry pi," Internet: http://en.wikipedia.org/wiki/Raspberry_Pi [accessed on Feb 16, 2015].
- [18] Siriphat pomyem signal and image processing with matlab on Raspberry pi platform, Examiner: Prof. Irek Defee. May 2015 p. 5.

Image Restoration via nonlocal supervised coding

Ao Li^{#1}, Deyun Chen^{#2}, Guanglu Sun^{#3}, Kezheng Lin^{#4}

[#]Postdoctoral research station of Computer Science and Technology,

Harbin University of Science and Technology

No. 52 XueFu Road, Nangang Dist, Harbin, China

¹dargonboy@126.com

²chendeyun@hrbust.edu.cn

³sunguanglu@hrbust.edu.cn

⁴link@hrbust.edu.cn

Abstract— Sparse representation (SR) and nonlocal technique (NLT) have shown great potential in computer vision and low-level image processing. However, due to the degradation of the observed image, SR and NLT may not be accurate enough to obtain a faithful restoration results when they are used independently. To improve the performance, in this paper, a novel nonlocal supervised coding strategy based NLT for image restoration is proposed. The novel method has three main contributions. First, to exploit the useful nonlocal patches, a nonnegative sparse representation vector is introduced, whose coefficients can be taken as the supervised weights among the patches. Second, a new objective function is proposed, which integrated the supervised weights learning and the nonlocal sparse coding to guarantee a more promising solution. Finally, to make the minimization tractable and convergence, a numerical scheme based on iterative shrinkage thresholding is developed to solve the above underdetermined inverse problem. The extensive experiments validate the effectiveness of the proposed method.

Keywords— Image restoration, sparse coding, nonlocal technique, nonnegative supervised weights

I. INTRODUCTION

As a fundamental problem in image processing, image restoration has been extensively studied in the past decades [1] to [4]. It also has various useful applications, such as medical imaging, remote sensing, astronomy, and so on. Image restoration aims to recovery the original image from its degradation observation. The challenge of many image restoration tasks is that they are ill-posed. To cope with the ill-posed inverse problem, lots of the techniques has been developed, most of them are under the regularization. For an observed imagey, the degradation can be formulated as the following linear model

$$y = Hx + n \quad (1)$$

where x is the clean original image, H is the degradation operator, and n is generally the additive gaussian white noise with covariance σ^2 .

By regularization technique, the ill-posed model in (1) can be solved by

$$\min_x \{ \|y - Hx\|_2^2 + \lambda J(x) \} \quad (2)$$

where $\|\cdot\|_2$ denotes the l_2 -norm, (\cdot) is a function to constrain the prior knowledge of the original image,

and λ is a positive scalar parameter to trade off the two terms.

In fact, the formulation in (2) can be also viewed as a variant of Bayesian framework when $J(x)$ is taken as the prior distribution of natural image. To obtain more better performance, many regularization forms have been studied [5] to [8].

It is aware that the prior knowledge plays a significant role in restoration performance, so how to design the effective regularization to constrain the prior knowledge of natural image is at the core of image inverse problem.

Classical regularization is studied with the assumption that the images are local smooth except its edges [9], [10]. So, by this viewpoint, a widely used regularization called total variation (TV) is proposed [11]. It compute the norm of first-order difference of natural image to constrain the local smooth property. Since the total variation favors piecewise constant and only exploit the local statistical model, it tends to smear out some details in image and failed to recover fine structures. To better preserve the image details, some extensive TV-based models are developed subsequently to improve the performance [12], [13], [14].

Recently, the sparse representation is proposed and widely used in image restoration problems. Sparse representation (SR) assumes that the image can be presented as a linear combination of a few basis from a set called dictionary, i. e. $x \approx D\alpha$, where D is the dictionary and α is the coefficients vector. With the sparse assumption, SR constructs the regularization by constrain the l_0 -norm to α , which counts the number of nonzero elements in a vector [15]. As the l_0 -norm is a NP-hard problem, which can be approximately solved with some greedy algorithms [16]. To obtain the more stable solution, l_0 -norm is often relaxed to its convex version by l_1 -norm, which has been successfully used in various image restoration applications [17], [18]. Though the current SR-based methods has obtained the promising restoration results, there exist two main drawbacks can be deeply developed to improve the performance. First, dictionary learning requires more adaptive representation ability to enhance the sparsity. Meanwhile, the

learning process commonly is accompanied by high computational complexity. Second, the sparse coding can be designed more accurately to preserve the fine structures in image. So, many literatures about the dictionary learning and coding are studied to make some progress.

In recent years, the nonlocal prior, as one of the most significant properties of natural images, has been explored in image restoration. Generally, the nonlocal prior is depicted by nonlocal self-similarity in global image, which is firstly utilized to synthesis textures. Then, Buades [19] introduces the nonlocal technique to image denoising, and proposed a novel weighted-based denoising approach, which is quite effective in sharpening image and preserving fine structure by utilizing the nonlocal statistics. Inspired by the nonlocal techniques (NLT), some researchers exploit embedding the nonlocal self-similarity to other regularization to obtain better performance for image restoration. Zhang, et al [20] integrated the nonlocal technique and total variation to form the nonlocal TV regularization, which has been successfully used in image deconvolution and compressive sensing. In addition, two effective numerical algorithm based on bregman iteration are also proposed to guarantee the convergence. Another patch-based SR method with NLT is proposed by Dong, et al in [21]. They introduced the concept of coding noise and centralized the sparse coding with nonlocal self-similarity patches to suppress the coding noise. The paper shows that the proposed method can obtain a good estimation of sparse coding of original image. However, the NLT is employed by the Euclidean distance which is not effective to reflect the inherent similarity among the patches. So, in this paper, different from the Euclidean distance adopted in conventional NLT, we employ a new way to exploit the nonlocal self-similarity among the patches as the supervised weights and propose a so-called supervised coding strategy for image restoration, which integrates the NLT-based weighted exploitation and sparse coding in the same objective function to make them promote each other during the iteration.

The rest of the paper is organized as follows. In Section 2, the way to exploit the supervised weights is elaborated. Section 3 presents the proposed objective function and the numerical scheme is given yet. Extensive experimental results are reported in section 4. Section 5 concludes the paper.

II. LEARNING THE SUPERVISED WEIGHTS

SR has been widely used in image inverse problem, such as imagedeblurring, inpainting and compressive sensing. The conventional SR-based methods aim to find some atoms to represent the image more sparsity. However, Yu, et al [22] pointed out that the locality is the more significant property than conventional sparsity and proposed the local coordinate coding (LCC) method, by which the signal can be represented via a linear combination of several neighbor atoms in given dictionary. Given a signal $x \in R^m$ and dictionary $D \in R^{m \times N}$, LCC can

be modeled by the following minimization

$$\hat{\alpha} = \operatorname{argmin} \left\{ \|x - D\alpha\|_2^2 + \gamma \sum_{i=1}^N \|x - D_i\|_2^2 |\alpha_i| \right\} \quad (3)$$

where D_i denotes the i -th atom of dictionary D and α_i denotes the i -th element of the coding vector α . The second term in (3) can be seen as a reweighted coding, which take the distance between the signal and atom as the constrain weight. With the coding process in (3), the signal can be encoded by a more discriminative way which also can guarantee the sparsity naturally. In other words, the signal will be represented by several neighbor atoms associate with itself.

For NLT-based image restoration with SR, it generally has the assumption that the patch and its neighbor ones have similar coding with same dictionary. That is, these self-similarity patches has great contribution to obtain a good estimation of the current encoded patch. Nevertheless, the contribution of neighbor patches is conventionally measured by computing the Euclidean distance, which is not effective enough to mining the inherent correlation among the patches. Motivated by LCC, we can take the similar scheme in (3) to learn the correlation coefficient vector among the self-similarity patches, which can assign the contribution according to the similar structure. In other words, we can take those coefficients as the weights to supervised the coding approximation, which is belief to obtain better performance.

Let $x_i \in R^{m^2}$ be a vectorized patch centralized around i -th pixel in image, and S_{x_i} is a set, which take the similar patches of x_i as its columns. In light of the scheme in (3), we adopt the S_{x_i} as the dictionary alternatively and rewritten the minimization by the following form

$$\min \left\{ \|x_i - S_{x_i} z_i\|_2^2 + \gamma p_i \otimes z_i \right\} \quad (4)$$

$$z_i^k \geq 0, k = 1, 2, \dots, \text{length}(z_i)$$

where p_i is the a constrain vector, whose k -th element is distance between x_i and the k -th atom of S_{x_i} , denoted by $\|x_i - S_{x_i}^k\|_2^2$. \otimes denotes the Hadamard product operator. It is worth noting that correlation coefficient of z_i in (3) can not be used as the supervised weight in objective function since it may be negative. Thus, we add the extra nonnegative constraint as $z_i^k \geq 0$, where z_i^k is the k -th element in z_i . Here, the first term in (4) is used to minimize the reconstruction error among the similar patches. With the constraint vector p_i , the second regularization term can assign the large weight to the similar patch whose linear reconstruction contribution is large yet. Thus, the regularization in (4) can be viewed as a practical weights learning mechanism to reflect the correlation effectively between the encoded patch and its self-similarity one.

III. NLT-BASED SUPERVISED SPARSE CODING

In this study, to overcome the drawback of conventional sparse representation model, we proposed a novel

supervised coding strategy based on NLT, aiming to obtain a better accurate coding by embedding the NLT to SR model. The comprehensive objective function for supervised coding is presented as

$$\min_{\alpha, z} \left\{ \begin{aligned} & \frac{1}{2} \| y - HD\alpha \|_2^2 + \lambda \| \alpha \|_1 \\ & + \eta \sum_i \sum_k z_i^k \| \alpha_i - \hat{\alpha}_{i,k} \|_2^2 \\ & + \sum_i (\| x_i - S_{x_i} z_i \|_2^2 + \gamma p_i \otimes z_i) \end{aligned} \right\} \quad (5)$$

s. tx = Doα, z_i^k ≥ 0, k = 1, ..., length(z_i) for ∀ i

where $Do\alpha = (\sum_i R_i^T R_i)^{-1} \sum_i R_i^T D\alpha_i$, $\hat{\alpha}_{i,k}$ is the good estimated coding of the k -th similar patch of x_i . $\| \cdot \|_1$ denotes the l_1 norm, and λ, η, γ are all the positive scalar parameters to trade-off among the regularizations. The first term, called data fidelity, is to evaluate the linear reconstruction error. The second term is to constrain the sparsity of coding to adaptive the sparse prior of natural image. The third term is to constrain the coding approximation between the similar patches. It is should noted that the correlation coefficient is also utilized in the third term as the supervised weight to assign the contributions of those similar patches. Moreover, it also links the correlation exploration and supervised coding, which can make them to promote each other in the iteration to obtain an overall optimal solution. The last two terms in (5) is the weight learning regularization, aiming to embed the NLT to SR model.

The problem of (5) is a constrain minimization problem with multi-variables, which can be solved with alternating optimization technique. So, in this paper, a iterative numerical scheme based on alternating technique is adopt to obtain the approximate solution of optimization problem in (5). Firstly, for in j -th iterative step p , α is computed while other variables is fixed and the minimization can be compacted as

$$\hat{\alpha}^j = \operatorname{argmin} \left\{ \begin{aligned} & \frac{1}{2} \| y - HD\alpha \|_2^2 \\ & + \eta \sum_i \sum_k z_i^k \| \alpha_i - \hat{\alpha}_{i,k}^{j-1} \|_2^2 \\ & + \lambda \| \alpha \|_1 \end{aligned} \right\} \quad (6)$$

It is an unconstraint problem, which can be solved by the classical iterative shrinkagethresholding (ISTA) method. We derived the first two terms with respect to α as

$$(HD)^T (y - HD\alpha) + 2\eta \sum_i \sum_k z_i^k (\alpha_i - \hat{\alpha}_{i,k}^{j-1}) \quad (7)$$

And then, ISTA can be implemented by an iterativethresholding operation as follows

$$\alpha^{temp} = \alpha^{j-1} + (HD)^T (y - HD\alpha^{j-1}) + 2\eta \sum_i \sum_k z_i^k (\alpha_i^{j-1} - \hat{\alpha}_{i,k}^{j-1}) \quad (8)$$

$$\alpha^j = \operatorname{shrinkage}(\alpha^{temp}, \lambda) \quad (9)$$

where $\operatorname{shrinkage}(x, \tau) = \max(|x - \tau|, 0) \operatorname{sgn}(x)$, and

$\operatorname{sgn}(\cdot)$ is the sign function. And then, the high quality estimation image x^j can be recovered by computing $x^j = Do\alpha^j$.

Secondly, correlation vector z_i is solved when α is fixed. The minimization defined in (5) is written as

$$\min_z \left\{ \sum_i \left(\| x_i - S_{x_i} z_i \|_2^2 + \gamma p_i \otimes z_i \right) + \sum_k \eta z_i^k \| \alpha_i - \hat{\alpha}_{i,k} \|_2^2 \right\} \quad (10)$$

Based on the separable form in (10), it can be independently further formulated as

$$z_i = \operatorname{argmin} \{ \| x_i - S_{x_i} z \|_2^2 + W \otimes z_i \} \quad (11)$$

s. t z_i^k ≥ 0, k = 1, 2, ..., length(z_i) ∀ i

where $W = \gamma p_i + \eta m_i$, and $m_i = \eta \| \alpha_i - \hat{\alpha}_{i,k} \|_2^2$ is the k -th element of m_i . The optimization in (11) can be seen as the reweighted sparse coding problem. In this study, the method in [23], which is known to be a good solution to reweight nonnegative coding, is used to solve the minimization problem of (11).

So far, we have acquired the efficient and effective solutions to the comprehensive objective function in (5), and the overall numerical scheme is described in the following Algorithm 1.

Algorithm 1 Numerical Scheme for Objective Function in(5)

Input: The observed image y and degradation operator H ; Sparse dictionary D ; Scalar parameters λ, η , and γ .

Output: estimated original image \hat{x} .

Set $t = 0, x^{(0)} = y, \alpha^{(0)} = 0$.

Repeat for t

1. Divide the image $x^{(t)}$ into overlapping patches $\{x_i^{(t)}\}_{i=1}^N$;
2. Searching similar patches for each $x_i^{(t)}$ to construct $S_{x_i^{(t)}}$;
3. Compute each z_i with $S_{x_i^{(t)}}$ and $\alpha^{(t)}$ by solving (11);
4. Update the $\alpha^{(t+1)}$ by computing (7) and (8) iteratively;
5. Estimate original image by $x^{(t+1)} = Do\alpha^{(t+1)}$;
6. Update $t = t + 1$;
7. Then, if $\operatorname{mod}(t, T) = 0$, jump to step 1.

Other, jump to step 2.

Until The criterion is satisfied.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In our experiments, several benchmark images are used to verify the performance of the proposed NLT-based supervised coding method for image restoration (NLTCIR). The benchmark images are shown in Fig. 1. The parameter setting of NLTCIR is as follows: the size of patch is 5×5 with overlapping 2-pixel-width. Because dictionary learning is not the issue of our paper, so we adopt the KSVD in [24] to learn an effective dictionary D directly to present the patches sparsely. The maximum iteration number for t is set to be 30 and the inner loop for supervised coding with re-

spect to (7) and (8) is set to be 20 empirically. The NLTCIR is compared to several state-of-the-art methods, including a fast algorithm for total variation method called SALSA [14], nonlocal total variation (NLTV) in [20], and the NCSR method by [21]. We apply these methods to image deblurring and a Gaussian deblurring kernel, whose scale is 1.5, is exploited for simulation. Moreover, the additive Gaussian noise standard is set to be 0.5. Due to the limited pages, only part of the comparison visual experimental results are shown in Fig. 2 and Fig. 3. To further evaluate the quality of the results, the Peak Signal to Noisy Rate (PSNR) is also taken as the objective metric, whose values are presented in Fig. 4.



Fig. 1 Benchmark images

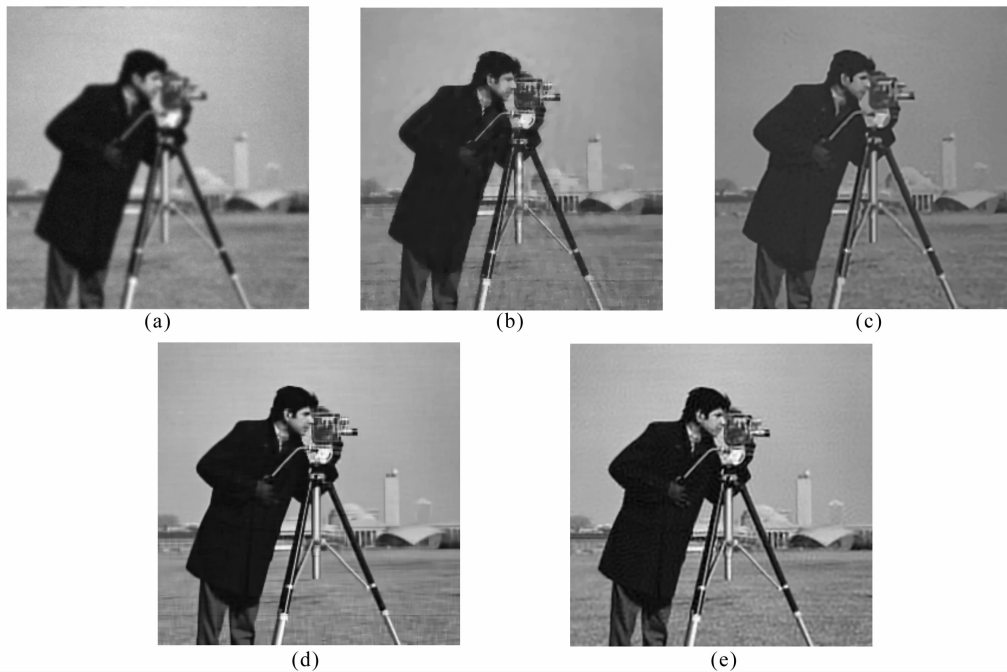


Fig. 2 Restoration results of Cameraman
 (a) Blurry Image; (b) SALAS; (c) NLTV; (d) NCSR; (e) NLTCIR

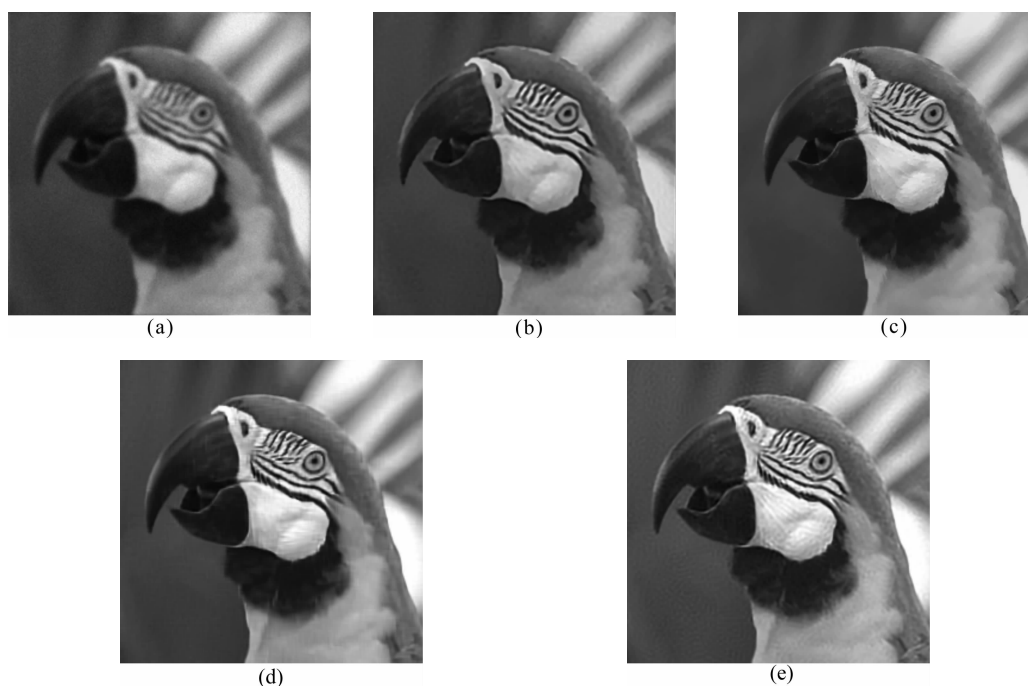
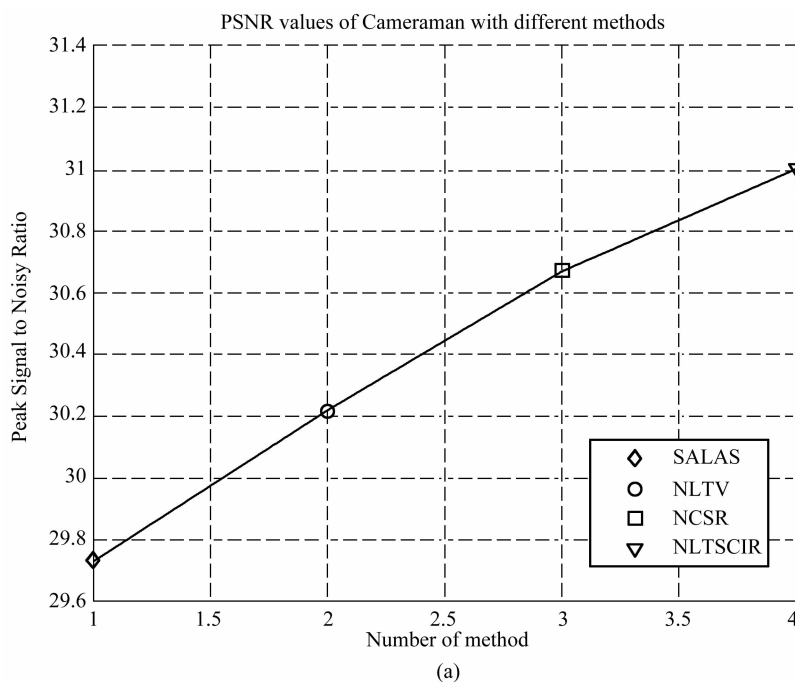


Fig. 3 Restoration results of Parrot
(a) Blurry Image; (b) SALAS; (c) NLTV; (d) NCSR; (e) NLTCIR

From Fig. 4, we can conduct that the proposed NLTCIR significantly outperforms other comparison excellent methods on all the test images. From the visual comparison results, we can observe that the NLTCIR achieves competitive performance compared with other leading methods. The SALSA acquires the inferior results and smears out the image with some mottles. NLTV is better than SALSA by exploiting the nonlocal character in image, but still fails to recover part of fine structures. NCSR and proposed NLTCIR produce similar

visual results and obtain significant improvement over other competing methods. However, the NLTCIR shows better performance on preserving more sharper edges and texture details than NCSR. Furthermore, the highest PSNR value also verifies the competitive performance of proposed methods, compared with other state-of-the-art ones. The good performance of NLTCIR is attributed to the employment of the powerful correlation exploring and the NLT-based supervised coding.



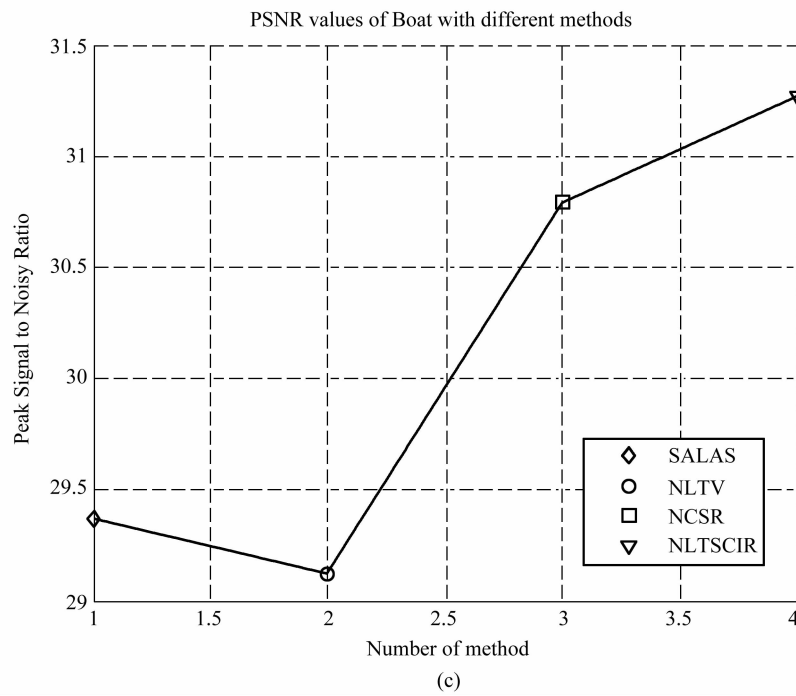
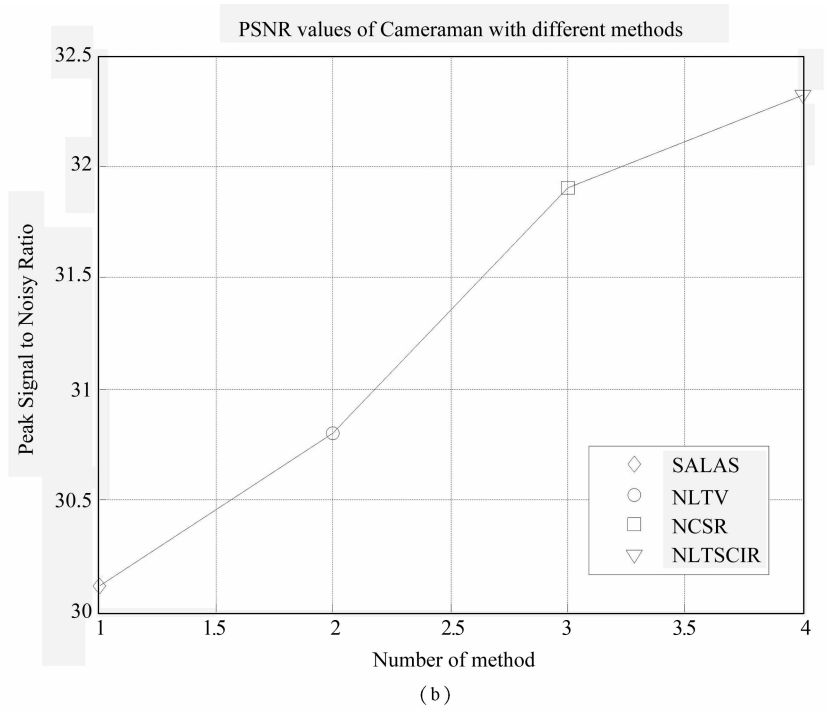


Fig. 4 PSNR values of benchmark images with different methods
 (a) PSNR for Cameraman; (b) PSNR for Parrot; (c) PSNR for Boat

V. CONCLUSION

In this paper, a novel NLT-based supervised coding strategy for image restoration is presented, which explores the correlation between self-similarity patches and proposes a supervised coding by using the correlation as the constrain weights. Also, a new regularization for supervised coding is utilized in comprehensive objective function, and links the supervised coding and nonnegative weights learning to make them promote each other iteratively. The experiments on several benchmark images demonstrate that the proposed NLTSICR outperforms existing excellent methods.

ACKNOWLEDGMENT

This paper was supported by National Natural Science Foundation of China (Grant 61501147), Natural Science Foundation of Heilongjiang Province (Grant F2015040) and Postdoctoral Science Foundation of Heilongjiang Province (Grant 2501051410).

REFERENCES

- [1] H. Takeda, S. Farsiu, and P. Milanfar. Kernel regression for image processing and reconstruction. [J] *IEEE Transactions on Image Processing*, 2007, vol. 16, no. 2, pp. 349-366.
- [2] W. Dong, G. Shi, and X. Li, "Nonlocal image restoration with bilateral variance estimation; a low-rank approach. [J] *IEEE Transactions on Image Processing*, 2013, vol. 22, no. 2, pp. 700-711.
- [3] S. Roth and M. J. Black. Fields of experts. [J] *International Journal of Computer Vision*, 2009, vol. 82, no. 2, pp. 205-229.
- [4] X. Li. Image recovery from hybrid sparse representation; a deterministic annealing approach. [J] *IEEE Journal of Selected Topics in Signal Processing*, 2011, vol. 5, no. 5, pp. 953-962.
- [5] T. Goldstein, S. Osher. The split Bregman algorithm for L1 regularized problems. [J] *SIAM Journal of Imaging Science*, 2009, vol. 2, pp. 323-343.
- [6] J. F. Cai, S. Osher, Z. W. Shen. Split Bregman methods and frame based image restoration. [J] *Multi-scale Modeling and Simulation*, 2009, vol. 8, pp. 5057-5071.
- [7] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. [J] *Society for Industrial and Applied Mathematics*, 2009, vol. 51, no. 1, pp. 34-81.
- [8] Weisheng Dong, Lei Zhang, Guangming Shi, et al. Image Deblurring and Super Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization. [J] *IEEE Transactions on Image Processing*, 2011, vol. 20, no. 7, pp. 1838-1857.
- [9] D. Geman, G. Reynolds. Constrained restoration and the recovery of discontinuities. [J] *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1992, vol. 14, no. 3, pp. 367-383.
- [10] D. Mumford, J. Shah. Optimal approximation by piecewise smooth functions and associated variational problems. [J] *Comm. on Pure and Appl. Math.*, 1989, vol. 42, pp. 577-685.
- [11] L. Rudin, S. Osher, E. Fatemi. Nonlinear total variation based noise removal algorithms. [J] *Physical D, Nonlinear Phenomena*, 1992, vol. 60, nos. 1-4, pp. 259-268.
- [12] J. Oliveira, J. M. Bioucas-Dias, M. Figueiredo. Adaptive total variation image deblurring: A majorization-minimization approach. [J] *Signal Processing*, 2009, vol. 89, no. 9, pp. 1683-1693.
- [13] G. Chantas, N. P. Galatsanos, R. Molina, et al. Variational Bayesian image restoration with a product of spatially weighted total variation image priors. [J] *IEEE Transactions on Image Processing*, 2010, vol. 19, no. 2, pp. 351-362.
- [14] M. Afonso, J. Bioucas-Dias, M. Figueiredo. Fast image recovery using variable splitting and constrained optimization. [J] *IEEE Transactions on Image Process.*, 2010, vol. 19, no. 9, pp. 2345-2356.
- [15] J. Portilla. Image restoration through l_0 analysis-based sparse optimization in tight frames. [C] in *Proceeding of IEEE International Conference on Image Processing*, 2009, pp. 3909-3912.
- [16] J. A. Tropp and S. J. Wright. Computational methods for sparse solution of linear inverse problems. [C] *Proceedings of IEEE*, 2010, vol. 98, no. 6, pp. 948-958.
- [17] M. Elad, M. A. T. Figueiredo, Y. Ma. On the role of sparse and redundant representations in image processing. [C] *Proceedings of IEEE*, 2010, vol. 98, no. 6, pp. 972-982.
- [18] M. J. Fadili, J. L. Starck. Sparse representation-based image deconvolution by iterative thresholding. [C] *Astronomical Data Analysis. IV*, 2006, pp:1-10.
- [19] A. Buades, B. Coll, J. M. Morel. A non-local algorithm for image denoising. [C] *Int. Conf. on Computer Vision and Pattern Recognition*, 2005, pp. 60-65.
- [20] X. Q Zhang, Martin Burger, Xavier Bresson et al. Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction. [J] *SIAM Journal of Imaging Science*, 2010, vol. 3, Issue 3, pp. 253-276.
- [21] Weisheng Dong, Lei Zhang, Guangming Shi, et al. Nonlocal centralized sparse representation for image restoration. [J] *IEEE Transactions on Image Processing*, 2013, vol. 22, no. 4, pp. 1620-1630.
- [22] K. Yu, T. Zhang, Y. Gong. Nonlinear learning using local coordinate coding. [C] *Conference on Neural Information Processing Systems*, 2009.
- [23] J. Yang, Y. Zhang. Alternating direction algorithms for L1-problems in compressive sensing. [J] *SIAM Journal on Scientific Computing*, 2011, vol.

33, no. 1, pp. 250-278.

[24] M. Aharon, M. Elad, A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries

for sparse representation. [J] IEEE Transactions on Signal Processing, 2006, vol. 54, no. 11, pp. 4311-4322.

Online Activities Club of Music Education and Music Group

Oyuntuya G. ^{#1}

[#] School of Music Education, Department of Piano, Mongolian State University of Art and Culture
Ulaanbaatar, Mongolia

¹oyuntuya_g0518@yahoo.com

Abstract— Who created humanity around the world, and many scientific disciplines, while self-awareness. One of academics, its one branch of music education is based on similar principles guide the formation of skills based on students behavior, needs and interests listen to music, as well the development of instinct and talent to understand, evaluate, and create. Education environment for rapid growth in information and communications technology, young people learn to demand began to change with learning objectives, content and methods in line with the growing. For example: training activities were based on new methodologies study skills in a practice expertise based only knowledge. Knowledge and musical training can be taught online. For example: student correspondent held music education more accessible training online. Because: correspondence course students will stay in rural areas studied by 1 ~ 2 times a vocational teacher came to the school year and to reinforce lessons online training period to the next to come, go to the course to learn a lesson repetition can make it survive.

Keywords— Musical training course, knowledge of music, students skill, behavior, talent, online course

INTRODUCTION

Teacher training it is a means of social and creative student cooperation with the aim to student needs.

Online music lessons to teach the followings.

① Can be expected to listen to developing children's thinking and music vocal melody and express the tone patterns, musical melody describe the feeling.

② To provide basic knowledge of music theory and replace key learning note reading and writing, musical composition, such friends as will pave the way for deepening the content.

Thus, online learning music first element of the new ownership will also deepen reinforce a lesson form of organizing musical training has felt sidelined, only music classes to online training sessions circle outside group. In this regard, the core curriculum is arranged below and development of all children under the program.

II. ONLINE COURSE OF MUSIC EDUCATION

To organize a music education by online course.

① Student's knowledge and skills, based on an interest in music and dance, including folk art types. Children often something more than just interested, successfully performed the item. Therefore, interest is one of the

grounds to be a baseline for the development of discovered gifted children.

② Exercise long-term exercise, each student and develop a prosperous course and learning outcomes, we need to provide training should be content and ways to overcome difficulties appropriate, require a certain amount of force child labor in their levels from simple tasks. It is also desirable to develop a training plan for children, according to their respective potential and included folk art elephants legacy train.

③ Each type of selected training as appropriate. / Individual, group, small groups, etc. / It features a music training.

④ Teaching methods and optimize decided to training materials, organized plan.

Teach online classes should be used in combination with creative students to enable exploration gas and pedagogical methods and custom conversion, authoring and researchers study from the experience of other teachers.

Talent development speed of the main conditions of the system, ownership consecutive conducted exercises work, education and training activity and the knowledge, skills and practices depends on an individual.

III. ONLINE TRAINING OF PUBLIC WORKS

To carry out public works by online training:

① Song to play a diversity of music, folk songs;

② Dance was dominated by dance, social dance festival;

③ Reading literacy was associated with less routine work.

Manipulating games to play, which means people were paying attention to culture, improve social activity to involve the whole community work. But now will be taught through songs, dances, poems and songs online training.

Give music group online training for club:

① The development of art talent;

② To promote Aesthetic;

③ Free time to spend properly;

④ Prepared for Celebrations.

Art direction to prepare for examination, the competition order will be held in long and short-term.

These activities are necessary to conduct quality pedagogical students considered as important as a school because the build teaching certain things.

Arole in the music group overall development of the

child: " Pupils developed by the Ministry of Education to promote all things ,a vision and a way to learn about art and the development of each child that will earn their talent and beauty individual preferences willing to express creativity through artistic feelings, development and arts and culture" ,core curriculum is included in " music education through a child was included clear that to achieve this result/attachments / content of this program is the second innovation of education standards.

IV. CONCLUSIONS

" Music education and music teaching a group of club activities online" , the subject of Mongolian country music training system to ensure the certainty of the XXI century Education Sector policy as part of reforms " core" program, examples of parallel view as possible of the implementation of training into online forms we get to the final conclusion of our work.

REFERENCES

- [1] Amgalan. Kh. School of art sophisticated Course-UB. ,2008.
- [2] Batsukh. Ch. Learning technology innovation for music's aesthetic education UB. ,2012.
- [3] Batsukh. Ch. Training concept developed skills's musical talents and concepts, methods and tools-UB. ,2004.
- [4] Journal of Educational Development theory methods-UB. ,2007.
- [5] Budkhuu. Kh. Study integrate students enjoy beauty musical training UB. ,2004.
- [6] Ichinkhorloo. Sh. Education philosophy and ideas: before and now /Editor. JamtsL. /UB. ,2008.
- [7] Ichinkhorloo. Sh. Training /Editor. Gyatso L. /UB. ,2008.
- [8] All public education in the Mongolian-UB. ,2003.
- [9] Otgonchimeg. Z. Mongolia to study the implementation of the music artist-UB. ,2011.
- [10] Oyunbileg. Choi. Music education methodology 1-3 class-UB. ,2006.
- [11] Purevsuren. D. Mongolian musical traditions of education reform-UB. ,2007.
- [12] Education legal basis for raising /Editor. Ichinkhorloo. N/ UB. ,2000.
- [13] Education legal basis for raising /Editor Ichinkhorloo. N / UB. ,2006.
- [14] The basic education core curriculum-UB. ,2015.
- [15] Musical education standards-UB. ,2004.
- [16] Gantomor. J. Modern Mongolian Music History / Editor Oyunbileg. Choi/ UB. ,2011.

On Designing Reliable and Energy-Efficient MACs for Unidirectional Nodes

Philip Parsch^{#1}, Alejandro Masrur^{#2}

[#]Computer Science Department, Technische Universität Chemnitz D-09107 Chemnitz, Germany

¹philip.parsch@cs.tu-chemnitz.de

²a.masrur@cs.tu-chemnitz.de

Abstract— Wireless sensor networks (WSNs) are gaining in importance with an increasing need for interconnectivity in the advent of Internet of Things. In some application scenarios, such as building and home automation, WSNs need to comply with deadlines and guarantee a reliable communication, for which a suitable MAC (Medium Access Control) is of paramount importance. Most available MACs are based on bidirectional devices or nodes with the capability of acknowledging packets and performing retransmissions if necessary. However, in many application scenarios, nodes merely report data to a sink and do not need any feedback. Hence, unidirectional nodes can be used instead reducing energy consumption considerably. To this end, two major approaches have been presented to guarantee reliability requirements in unidirectional networks: First, nodes send a number of redundant data packets with fixed inter-packet times to guarantee deterministic delay and packet loss behavior. Second, inter-packet times are not fixed but randomly selected between specified boundaries to allow for lower delays. In this paper, we summarize both approaches and compare them in simulation using OMNeT++. Finally, we provide a guide to select the most suitable MAC depending on a set of different application scenarios.

Keywords— Sensor/actuator networks, cost efficiency, reliability, single-hop communication, unidirectional nodes

I. INTRODUCTION

Wireless sensor networks (WSNs) have attracted much attention both in literature and industry during the last years. With the upcoming trend of Internet of Things and its increasing need of interconnectivity of electric devices, WSNs are expected to even continue growing in importance. In general, WSNs can be found in wide range of application domains such as home automation, body area networks, environmental monitoring, surveillance, etc. These normally substitute wired solutions such as field buses wherever the latter are not viable due to technical or economic limitations.

WSNs use different communication technologies, which can generally be classified into unidirectional or bidirectional nodes. Unidirectional nodes can only transmit or receive data, while bidirectional nodes are capable of both transmitting and receiving. Clearly, in a unidirectional WSN, data packets cannot be acknowledged or retransmitted making it difficult to implement reliable communication. Bidirectional nodes, on the other hand, can implement these features making more reliable, multi-hop communication possible. However, this also

leads to a higher computational cost (to implement the more sophisticated communication protocols), which, together with a more complex transceiver circuitry, makes bidirectional WSNs be considerably more expensive than uni-directional ones, in particular, as the number of nodes in the network increases.

In contrast, unidirectional nodes are of low-cost and have a higher energy efficiency compared to bidirectional nodes [1], since they do not need to power a receiver and monitor the communication channel. Considering that the overall costs as well as energy consumption are key factors when designing WSNs, unidirectional nodes have been used in various scenarios in the past: home automation [2], environmental monitoring [3], RFID [4], etc., where potentially hundreds of sensor nodes report data to a sink (either in a periodic fashion or upon the occurrence of specific events) without need for external control [5].

As a result, if the application under consideration tolerates it, unidirectional nodes allow for a more cost-effective WSN. However, since a unidirectional WSN may incur in packet loss, special medium access control (MAC) protocols have to be used to improve reliability. Unfortunately, existing solutions like CSMA (Carrier Sense Multiple Access), TDMA (Time Division Multiple Access) or slotted Aloha [6] cannot be applied, since they rely on carrier sensing or synchronization (which unidirectional nodes are unable to perform).

A number of approaches have been proposed to improve the reliability of unidirectional communication. In [7], for example, a method is presented to recover the packet with the highest signal strength when packets collide at the communication channel. In [1][5], a mix of unidirectional and bidirectional nodes is used instead. Although these methods improve the average performance of a (mixed or fully) uni-directional WSN, they do not allow for guarantees on the resulting communication reliability.

Since, it is important for a network designer to know the performance of a network at design time, rather than finding the optimum in a long try and error process during deployment, other approaches have been presented that allow quantifying reliability and guaranteeing a bounded delay. In general, these can be divided in two main categories: first, fully reliable MACs that use fixed inter-packet times [2][8] and second, probabilistic approaches that randomize these timings [9]. To this

end, this paper aims to provide an overview and help selecting the right method for a specific case. Our contributions can be summarized as follows.

- ① We examine three MAC protocols [2][9][8] that allow quantifying specific reliability requirements and a bounded delay in a fully unidirectional network. To this end, we summarize their working principles and assumptions and compare them analytically.
- ② We perform simulations based on OMNet ++ in order to assess performance characteristics, such as average delay, packet loss, etc. These results are then compared with each other to finally help selecting the right MAC for a set of different application scenarios.

The rest of this paper is structured as follows. Related work is discussed in Section II. Next, Section III explains our system model and assumptions. Section IV and Section V introduce the different MAC protocols for unidirectional networks. Section VI presents our experimental evaluation based on simulation and Section VII concludes the paper.

II. RELATED WORK

Saving costs and energy by using unidirectional nodes is an idea that has been used multiple times in the past: RFID Systems [10][4], long range outdoor networks [3], wireless body area networks [11] and indoor networks [12]. The challenge hereby is to find communication schemes to allow for reliability, energy efficiency, and a bounded delay. As already mentioned, unidirectional nodes cannot perform carrier sensing nor synchronization, therefore, existing algorithms such as CSMA and Aloha cannot be used.

This behavior poses a problem for communication reliability, since there is no feedback for lost or corrupt packets. In this context, Cardell-Oliver et al. [13] identified three *error-contention* strategies: temporal diversity, spatial diversity and code-based methods. In temporal diversity, data packets are, for example, transmitted repeatedly at different times, while spatial diversity aims to separate devices geographically such that they do not fall within each other's range and, hence, cannot interfere with one another. Code-based methods add redundant data, which can be used by the receiver to correct damaged packets. However, the increased packet length leads to less energy efficiency and higher collision probability due to longer transmission times. Cardell-Oliver et al. further concluded that temporal and spatial diversity yield better results for indoor scenarios than code-based strategies [13].

An approach based on time diversity, called Timing Channel Aloha (TC-Aloha), has been proposed by Galluccio et al. [14]. This scheme encodes parts of the information in the inter-packet separation of the different nodes. As a result, the packet length and transmission time can be reduced, thus, decreasing energy consumption and the probability of packet collision. In order to

improve reliability, packets are transmitted multiple times. Clearly, to recover the information embedded in the inter-packet separation, at least two packets must be received. By an analytical framework and experimental results, Galluccio et al. [14] prove that TC-Aloha increases data throughput; however, no guarantees can be given on whether data always reaches the corresponding receivers on time.

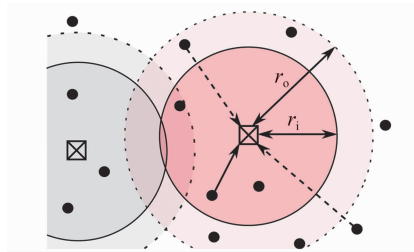


Fig. 1 Example of a unidirectional WSN with transmit-only (solid circles) and receive-only nodes (checked boxes); r_i represents the range within which the receiver collects packets, while r_o indicates the range in which transmitters can interfere with each other

Another approach presented in [1] consists in using two different types of transmitters. Simple unidirectional transmitters (forming clusters) for cost reduction and so called *cluster heads* with receiving capability. The cluster heads collect the packets from their corresponding unidirectional clusters and forward them to receivers. Since cluster heads can perform carrier sensing and acknowledge packets, more sophisticated communication schemes can be implemented upon them. However, if many cluster heads are necessary for a good coverage in a home-automation scenario, costs increase rapidly. Moreover, this method cannot guarantee a fully reliable communication and packets may potentially never reach their receivers due to collisions, in particular, within one cluster.

Although these previously mentioned approaches improve the overall communication reliability, they either do not allow giving precise guarantees on delay and reliability or they use hybrid nodes, which increase costs. In this paper, however, we consider exclusively designs for unidirectional networks that can provide these guarantees. To this end, the following section summarizes the system models as well as the assumptions made by the presented MAC protocols.

III. SYSTEM MODEL AND ASSUMPTIONS

We consider a WSN consisting of n simple transmit-only nodes that are spatially distributed and transmit data upon activation. A transmitter node is activated by an event, e. g., a motion detector gets triggered, the temperature rises to a certain level, etc. Upon activation, the corresponding transmitter broadcasts its data within a certain range and every receiver within that range will receive this data-see Fig. 1. If a receiver is connected to

a transmitter¹, it will process data; otherwise this is discarded. Since transmit-only nodes are unable to detect whether data has reached its corresponding receivers, they transmit a sequence of $k \in \mathbb{N} > 0$ redundant packets — to increase chances of a successful reception — where k is a natural number greater than zero and a design parameter common to all nodes in the system.

Packets transmitted by nodes consist of an identifier, a data, and a check sum or CRC (Cyclic Redundancy Check) field. Usually, these have a relatively constant size consisting of a few bytes. Transmitting one such packet takes an amount of time that depends on the number of bits to be transmitted and the bandwidth of the communication channel. We refer to this time to as packet length. In this paper, we denote by $l_{\max} \in \mathbb{R} > 0$ the maximum length of any packet in the network. We assume that even the smallest overlapping between any two packets on the communication channel produces interference and, hence, packet loss.

We further assume that each receive-only node constantly monitors the communication channel and, hence, no special measures have to be taken before sending data. For example, no preamble needs to be sent to wake up a receiver, etc. In many applications, this assumption does not pose any additional restrictions, since receive-only nodes are usually attached to unrestricted power supply, e. g. , a lamp in a home automation setting is attached to the electric network. In other cases, where power supply is restricted, the presented method can be extended to account for preambles, e. g. , by adjusting the packet length or increasing the number of packets sent. However, this is out of scope in this paper.

For each transmit-only node in the network, it must be guaranteed that one packet arrives within a deadline measured from the node's activation time. In the context of home automation, typical deadlines are around 500 ms. This is, for example, the time by which a wireless light switch should turn on the light, or a motion sensor should detect the presence of a person. A greater delay is often unacceptable, since it negatively impacts the quality of the system. Similar deadlines are also common in body area networks. In this paper, we consider that this deadline is the same for all nodes and denoted by $d_{\max} \in \mathbb{R} > 0$ — this is typical from single-hop WSNs where data needs to be conveyed in one direction and within a given time upper bound.

Finally, each transmit-only node is assumed to be activated only once within a time interval of length d_{\max} . In our previous example, this means that the wireless light switch sends only one sequence of packets every 500 ms. This is a logical design assumption, since multiple activations of the transmitters lead to unnecessary interference and do not help achieving the design goal of the system.

IV. RELIABLE MAC PROTOCOLS

In this section, we introduce two MAC protocols [2][8] for designing fully reliable networks based on unidirectional devices. Fully reliable means that it is guaranteed that at least one packet of a sequence of redundant packets reaches the corresponding receiver within a deadline d_{\max} in the worst case. However, this reliability can only be guaranteed if interference from outside the network is negligible and interference solely originates from simultaneous transmissions within the network itself. Although this assumption strongly limits the area of applications, e. g. being impractical for large WSNs or in noisy environments, there still exists numerous application scenarios. For example, the authors [2][8] mention indoor networks as a possible application area, since walls usually shield environmental noise to a great extent. Or they suggest using different frequency bands, such as 868 MHz to avoid interference with common devices using 2.4 GHz, such as WLAN routers, microwaves, etc.

A. Periodic MAC

In our previous work in [2], we presented an approach for fully reliable communication based on unidirectional devices. In summary, this consists in each node sending a sequence of redundant packets upon activation with constant inter-packet time. For each node in the network, its inter-packet time has to be selected to ensure that at least one of its packet arrives in the worst case. After each sequence, there is a transmission pause in which transmitter cannot be triggered anew.

In more detail, the authors [2] first start by deriving the minimum number of packets k per sequence. Due to the asynchronous nature of unidirectional WSNs, any node could possibly start transmitting at any time instance, hence, if there is a total number of n nodes in the system, each node could suffer at least $n - 1$ collisions in the worst case. Since for successful communication at least one packet of a sequence must reach the receiver, the minimum number of packets per sequence k must be consequently set to at least $k = n$, i. e. , each node sends at least as many packets per sequence as there are nodes in the system. Although this number of k is relatively high, it is the necessary minimum for full reliable communication in unidirectional WSNs. This also holds for the *optimized periodic MAC*[8], as shown later.

Next, methods for finding suitable inter-packet times (periods) are introduced. To this end, the authors fix $k = n$, i. e. , k is set to the minimum number of packets per sequence possible. This reduces both energy usage and interference, since there are no more transmissions than necessary. However, this also restricts the selection of inter-packet times, as there can now be at most 1 collision between any two nodes in the sequence. In order to prevent multiple collisions between any two nodes i and j , the following condition must hold for $1 \leq i \leq n, 1 \leq j \leq n, 1 \leq k_i \leq n - 1$, and $i \neq j$

¹An identifier is sent in each packet by the transmitter, which is then recognized by its corresponding receiver.

$$\text{mod}\left(\frac{k_i \times p_i}{p_j}\right) \geq 2 \times l_{\max} \quad (1)$$

where $\text{mod}(\cdot)$ is the modulo operation, l_{\max} is the maximum length of a packet, while p_i and p_j are the (constant) inter-packet times of the i -th and the j -th node respectively.

To better understand (1), Fig. 2 displays exemplary sequences of three nodes in the case of a simultaneous activation, i. e., all nodes start transmitting at the same point of time. We can see that beside the first unavoidable collision, there are no further collisions between any two nodes. In addition, (1) states that there must be at least a space of $2 \cdot l_{\max}$ between any period and its combinations. This ensures that collisions, in which packets do not overlap perfectly as in Fig. 2, but are shifted relative to each other, do not cause another collisions within that sequence. For example, if node 2 starts transmitting $l_{\max} - \epsilon$ before node 1, where ϵ is an infinitesimal small number, both first packet are still lost, but the second packets do not interfere. If node 2 starts transmitting even earlier, e. g. $2 \cdot l_{\max}$ before node 1, packet 2 will be lost, but no other packets collide with 1.

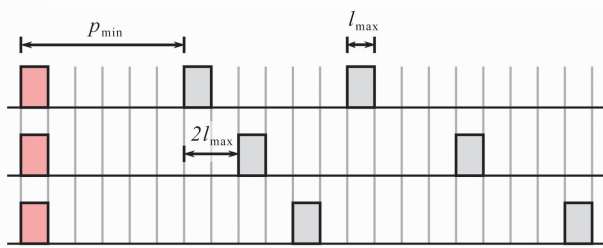


Fig. 2 Illustration of Equation 1 for the case of three nodes, i. e., $n = 3$. In the case of a simultaneous activation, the first packet of each node will get lost (reddish shading). However, there will be no further packet losses for the next two packets (i. e., $n - 1$) of each node (grayish shading)

Finally, (1) is used to find the following formula to calculate inter-packet times

$$p_{\min} \geq 2 \times (n - 2) \times (n - 1) \times l_{\max} + 2 \times l_{\max} \quad (2)$$

where p_{\min} or p_0 is the smallest inter-packet times of all nodes and all further periods are exactly $2 \cdot l_{\max}$ longer, i. e., $p_i + 1 = p_i + 2 \cdot l_{\max}$ — see (1).

We can see that period lengths increase quadratically with n and the overall maximum delay, which is $(n - 1) \cdot p_i$ of a node i for sending n packets, rises with the power of three for an increasing network size. This means that the overall maximum delay is very high for higher n making the *periodic MAC* generally be suitable for only small networks.

In addition to the long inter-packet times, there must also be a transmission pause implemented after each sequence. By this, double collisions with two subsequent sequences are prevented, as shown in Fig. 3. This pause must have an overall length of the longest sequence of all nodes, hence, it doubles the maximum delay. For example, let us consider a wireless light switch in an home automation network with a typical deadline of 500 ms. If the

user accidentally switches the light on, he has now to wait for the data transmission for up to 500 ms and an additional 500 ms to be able to trigger the switch anew. Although this might not be a problem, it generally decreases the quality of service of the system.

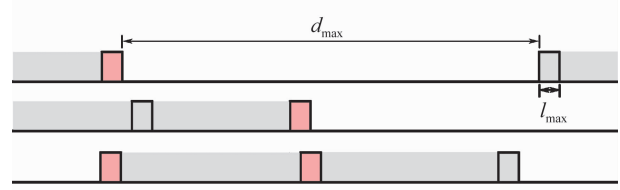


Fig. 3 Illustration of the transmission pause after each sequence for the case of three nodes. The last packet of the top node interferes with the first packet of bottom node (reddish shading).

Given that the second packet of the bottom node might also be lost, its third (and last) packet can only be guaranteed to reach its receiver (grayish shading), if the top node is forced to wait for at least d_{\max} time before transmitting anew

B. Optimized periodic MAC

As we previously discussed the *periodic MAC*, we could see that making unidirectional communication fully reliable incurs in great pessimism; A high packet number of at least n packets is required for each node per sequence, where n is the total number of transmitters in the system, as well as long inter-packet times lead to a high overall delay. In order to improve performance, efforts have been made to reduce inter-packet times (as we already know that the packet number cannot be further decreased).

The *optimized periodic MAC* [8] follows a similar principle as the *periodic MAC*. Each transmitting node sends a sequence of n packets upon activation with constant inter-packet times, followed by a transmission pause. However, in contrast to the *periodic MAC*, nodes can have multiple different inter-packets times instead of just one—see Fig. 4. These lead to non-symmetric periodic patterns that were found by using an ILP (Integer Linear Programming) solver and optimized for the shortest total transmission duration, however, these patterns greatly differ in length, meaning that some transmitters are very fast at sending their packets whereas others have total transmission durations which are multiple times longer. This leads to big differences in terms of the average and total delay and hinders the system analysis.

Further, the ILP solving requires high computational power. Hence, the authors of the *optimized periodic MAC* were only able to generate periods for a very low $n \leq 6$. For higher n , another paper [15] was proposed offering two different algorithms of heuristic nature for finding inter-packet times. The first one uses prime numbers to generate unique period times, but as prime numbers also strongly increase in its value as more are needed for higher n , these result in even longer inter-packet times as for the *periodic MAC*. The second algorithms first generates a large set of possible periods and then picks the shortest ones by performing a check that is similar to

Equ. (1) of the *periodic MAC*. By doing so, the second algorithm is able to find periods that are approximately 30% shorter than those of the *periodic MAC*. However, due to its still fairly high computational complexity, generating periods requires a much higher processing time. For example, finding all inter-packet times for $n = 30$ requires a time of 72 hours to generate, whereas the *periodic MAC* takes less than 1ms².

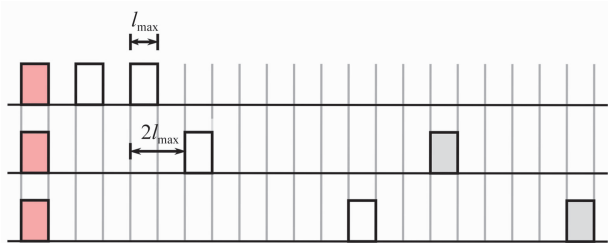


Fig. 4 Exemplary inter-packet times for the *optimized periodic MAC* in the case of three nodes, i. e. , $n = 3$

In summary, this means the *optimized periodic MAC* allows for a shorter maximum delay, but requires more memory on the nodes, as multiple inter-packet times must be saved and higher computation time for calculating these.

V. PROBABILISTIC MAC PROTOCOLS

The previously discussed MAC protocols have shown that making a network fully reliable results in high costs, i. e. , a high number of packets must be sent separated by long inter-packet times. In addition, to achieve this reliability, external noise must be neglected, which is often not the case, for example, in dense networks with a high number of nodes or in noisy environments. In order to reduce this pessimism and to be able to account for external interference (which is of stochastic nature), probabilistic MAC protocols have been presented.

In our previous work in [9], we proposed a probabilistic MAC protocol for guaranteeing a bounded delay and specific reliability for unidirectional networks. It consists in each node sending a sequence of k redundant packets within a maximum deadline d_{\max} with inter-packet times that are randomly selected between a lower boundary t_{\min} and upper boundary t_{\max} . In contrast to the previously mentioned approaches, no transmission pause is required after a sequence. In addition, after triggering a node it will not start transmitting immediately, but first waits a random period within $[t_{\min}, t_{\max}]$. Hence, there will be a greater delay on average, which, however, can be neglected in most applications, since delay is not critical as long as packets arrive before the deadline.

Given the packet size l_{\max} , the deadline d_{\max} , the number of nodes n and the required communication reliability

p , the *probabilistic MAC* [9] allows computing the missing variables t_{\min} and t_{\max} and k . To this end, let us first consider t_{\max}

$$t_{\max} \leq \frac{d_{\max} - l_{\max}}{k} \quad (3)$$

Since packets are sent in intervals that are randomly selected between t_{\min} and t_{\max} , increasing this interval results in a higher number of choices selecting periods, consequently reducing the collision rate. Hence, in order to increase the interval $t_{\max} - t_{\min}$, t_{\max} should be as high as possible. This is done in (3), as in the worst case, in which a node randomly picks k times t_{\max} as its inter-packet time, starts transmitting its last packet at $d_{\max} - l_{\max}$ and finishes it exactly at d_{\max} . Next, let us consider the lower interval boundary t_{\min}

$$t_{\min} \leq t_{\max} - \frac{2(n-1)l_{\max}}{k\sqrt{1-p}} \quad (4)$$

On the other hand, t_{\min} is a subtraction of t_{\max} and a second term that depends on all remaining parameters. This second term must not be too large in order that t_{\min} does not get negative and the system becomes not feasible anymore. This means that, as expected, both the network and packet sizes are limited as they directly increase the term's size. Further, we can see that increasing reliability p decreases t_{\min} as well. This is a logical conclusion since for a higher reliability, the interval $t_{\max} - t_{\min}$ must be larger. We can also conclude with (4) that a fully reliable communication with $p = 1$ is not possible with the *probabilistic MAC*, as the denominator would become 0. This means that the *probabilistic MAC* tolerates packet loss to some extent in exchange for a lower delay and a lower packet count ($k \ll n$).

The last missing parameter k is more complex to determine, since both (3) and (4) depend on it. Therefore, the authors [9] performed a set of experiments to numerically find its optimum. In general, k should be chosen as small as possible, as it directly affects the systems energy consumption.

VI. EXPERIMENTAL RESULTS

In this section, we present our experimental results comparing the three proposed techniques. To this end, we have performed a simulation based on the OMNeT++ network simulation framework [16] and an extension for mobile and wireless networks named MiXiM [17]. This allows us to effectively simulate our network with different physical parameters and to record statistical values for very large numbers of transmissions.

We assume that the network has been set up correctly in a way that physical effects, such as fading, shading and reflection of radio waves do not cause packet loss and can therefore be neglected. The data rate of transmission has been fixed to 128 kbit/s and the packet size is 3 bytes (8 bits for identifier, 8 bits for data, and 8 bits for check sum). The transmission time of a single packet takes consequently 187.5 μ s, i. e. , this is the value of

²All calculations were performed on an AMD A8-6500 processor at 3.5 GHz and with 8 GB of memory)

l_{\max} . The deadline d_{\max} has been set to 500 ms and in case of the *probabilistic MAC*, we chose $k = 2$ (required minimum) and a reliability of $p = 0.95$.

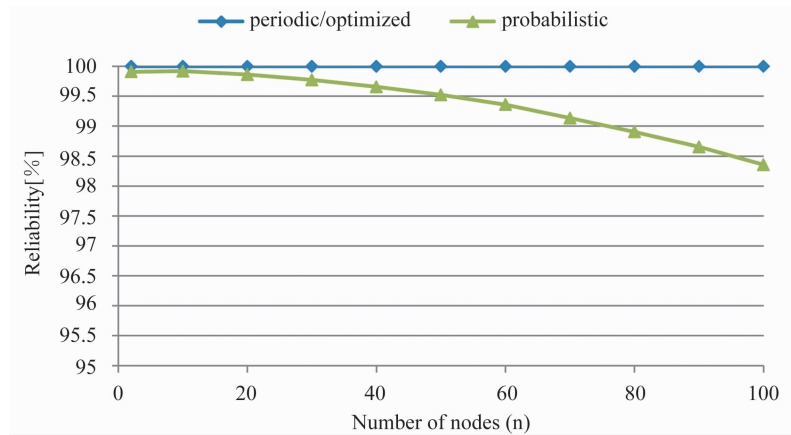


Fig. 5 The measured reliability of the three compared MAC protocols is depicted. Each algorithm was executed for different numbers of transmitter nodes (n); for each value of n , 100,000 different packet sequences have been simulated

The simulated network consists of one receiver and a selectable number of n transmitters that are all within range of one another and, hence, interfere with one another. The receiver node is a simple data sink, whereas transmitter nodes are data sources that transmit packets with a certain pattern according to the compared algorithms.

All transmitter nodes run independently of one another and are triggered by random time events to ensure that different possible combinations of packet transmissions are considered. Recording and processing of simulation data is done by the framework at runtime. In particular, the time stamps of the different packets sent are compared to determine whether packets overlap and, hence, get lost.

Fig. 5. shows the reliability of the different MAC protocols as n increases. As expected, both the *periodic* and *optimized periodic* protocols do not lose any packets as

they were specifically designed for fully reliable communication. The *probabilistic MAC*, however, incurs in slight packet loss with further increases with a rising n . This is due the fact that a higher n increases the channel load and consequently the collision probability. Still, its reliability stays well over the calculated 95%.

Let us now consider the average transmission delay, i. e., the delay from triggering the node to the successful reception of its data. As can be seen in Fig. 6, both *optimized* and *periodic* protocols show a low delay for small n , which rises quadratically for higher n . This is because for a low number of transmitters, the collision count is low and typically the first packet of a sequence succeeds. For higher numbers, the collision count rises and the overall delay increases as the periods increase as well. Since the periodic protocol typically has longer periods than the *periodico ptimized MAC*, its average delay is higher.

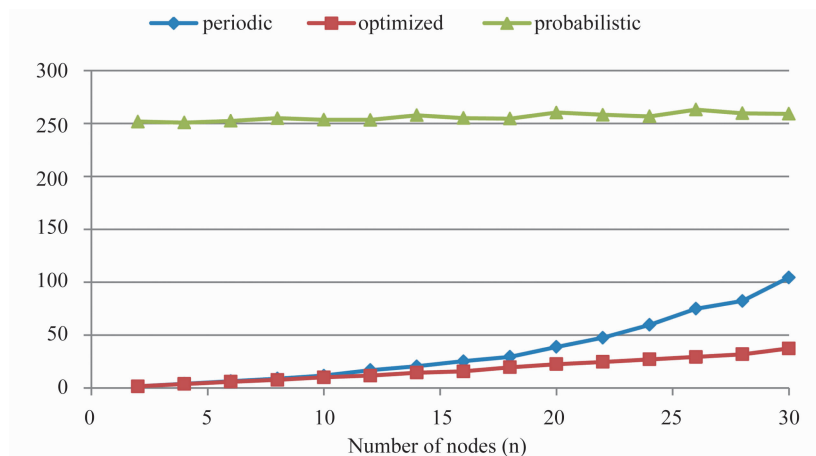


Fig. 6 The average delay for receiving a packet in numbers of l_{\max} is depicted for the *random* and our *proposed* algorithm. Each algorithm was executed for different numbers of transmitters (n); each time 100,000 packet sequences have been simulated

In contrast, the *probabilistic* MAC starts with a relatively high delay due to its random waiting time before the first packet of each sequence. The delay then slowly increases for higher n as the channel load is greater and the loss rate increases.

Although the average delay is higher for the *probabilistic* MAC for $n \leq 30$, its worst case delay is much lower than for the fully reliable protocols. For example, let us consider a network with 30 nodes. For the *probabilistic* MAC, the worst case delay will only be 500 ms, whereas the *periodic* and the *optimized periodic* have a delay of 1,500 ms and 4,000 ms respectively.

Finally, let us consider the typical energy consumption

of the three presented MACs, as shown in Fig. 7. To this end, each node was triggered between 6 to 10 times a day (typical value in home automation scenarios) and the total energy consumption used for transmission was recorded. We can see that the *probabilistic* MAC consumed fairly low energy, as always $k=2$ packets have been sent independent of n . The two reliable protocols, however, consumed a large amount of energy which rises quadratically with n . This means that in order to be still able to operate nodes with batteries, both fully reliable MAC protocols should only be considered for smaller networks $n \leq 30$, for example, home automation networks.

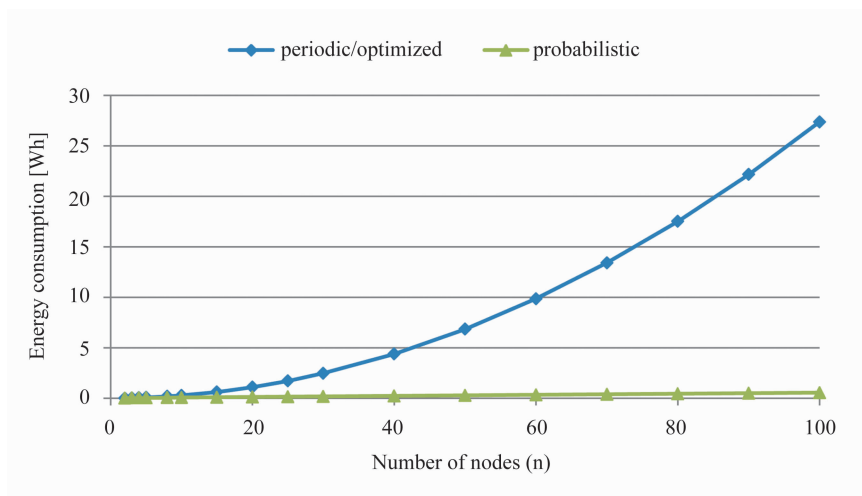


Fig. 7 The yearly energy consumption used for packet transmission by the different MAC protocols is shown. On average, each node is triggered a number from 6 to 10 times per day, which corresponds to the common activation rates in home automation networks. The energy consumption during transmission is 5 mW

VII. CONCLUSIONS

In this paper, we presented different MAC protocols for purely unidirectional networks, i. e., networks of nodes with either the capability of transmitting or receiving data packets. These protocols can be further divided into fully reliable and probabilistic approaches.

The presented reliable approaches consist in sending a sequence of redundant packets with carefully selected inter-packets times. These avoid multiple collisions between any two nodes and ensure that at least one packet reaches the corresponding receiver in the worst case. The reliability, however, comes at the costs of increased delay and energy consumption. In addition, these MAC protocols only work in a noise free environment, where no external interference is present. As a consequence, those protocols are best suited for smaller indoor networks with less than 30 nodes. Examples can be home automation, Internet of Things (IoT), environmental monitoring and control loops, where data loss cannot be tolerated.

On the other hand, probabilistic approaches try to improve the delay and energy consumption by allowing packet loss to some extent. In addition, they allow considering external interference, since it shares the same prob-

abilistic nature as the protocol itself. This makes them the ideal choice for a broad variety of applications including difficult settings, such as networks with a high number of nodes or noisy environments. Examples can be environmental monitoring, IoT, sensor clusters, etc.

Beside doing a summary and analysis of the presented approaches, we performed a set of experimental simulations based on the OMNeT ++ framework. Our experiments show important criteria such as energy consumption, packet loss, etc. and validate the presented analysis.

ACKNOWLEDGMENT

The authors would like to thank the professorship Computer Engineering of TU Chemnitz for their support as well as the anonymous reviewers for their valuable comments and suggestions.

REFERENCES

- [1] B. Blaszczyszyn and B. Radunovic, "Using Transmit-only Sensors to Reduce Deployment Cost of Wireless Sensor Networks," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, 2008.

- [2] P. Parsch, A. Masrur, and W. Hardt, "Designing Reliable Home-Automation Networks based on Unidirectional Nodes," in *Proceedings of the IEEE International Symposium on Industrial Embedded Systems (SIES)*, 2014.
- [3] C. Huebner, S. Hanelt, T. Wagenknecht, R. Cardell-Oliver, and A. Mon-salve, "Long Range Wireless Sensor Networks Using Transmit-only Nodes," in *Proceedings of the ACM Conference on Embedded Networked Sensor Systems (SenSys)*, 2010.
- [4] G. Mazurek, "Collision-Resistant Transmission Scheme for Active RFID Systems," in *Proceedings of the International Conference on Computer as a Tool (EUROCON)*, 2007.
- [5] J. Zhao, C. Qiao, R. S. Sudhaakar, and S. Yoon, "Improve Efficiency and Reliability in Single-Hop WSNs with Transmit-Only Nodes," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 3, pp. 520-534, 2013.
- [6] L. G. Roberts, "Aloha Packet System With and Without Slots and Capture," *Computer Communication Review (SIGCOMM)*, vol. 5, pp. 28-42, 1975.
- [7] B. Radunovic, H. L. Truong, and M. Weisenhorn, "Receiver Architectures for UWB-Based Transmit-Only Sensor Networks," in *Proceedings of the IEEE International Conference on Ultra-Wideband (ICU)*. IEEE, 2005, pp. 379-384.
- [8] B. Andersson, N. Pereira, and E. Tovar, "Delay-Bounded Medium Access for Unidirectional Wireless Links," in *Proceedings of the International Conference on Real-Time Networks and Systems (RTNS)*, 2007.
- [9] P. Parsch and A. Masrur, "A Reliability-Aware Medium Access Control for Unidirectional Time-Constrained WSNs," in *Proceedings of the International Conference on Real Time and Networks Systems (RTNS)*, 2015.
- [10] B. Zhen, M. Kobayashi, and M. Shimizu, "To Read Transmitter-only RFID Tags with Confidence," in *Proceedings of the IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, 2004.
- [11] H. Keong, K. Thotahewa, and M. Yuce, "Transmit-only Ultra Wide Band Body Sensors and Collision Analysis," *IEEE Sensors Journal*, vol. 13, pp. 1949-1958, 2013.
- [12] B. Tas and A. Tosun, "Data Collection Using Transmit-only Sensors and a Mobile Robot in Wireless Sensor Networks," in *Proceedings of the International Conference on Computer Communications and Networks (ICCCN)*, 2012.
- [13] R. Cardell-Oliver, A. Willig, C. Huebner, T. Buehring, and A. Mon-salve, "Error Control Strategies for Transmit-only Sensor Networks: a Case Study," in *Proceedings of the IEEE International Conference on Networks (ICON)*, 2012.
- [14] L. Galluccio, G. Morabito, and S. Palazzo, "TC-Aloha: A Novel Access Scheme for Wireless Networks with Transmit-only Nodes," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 3696-3709, 2013.
- [15] B. Andersson, N. Pereira, and E. Tovar, "Delay-Bounded Medium Access for Unidirectional Wireless Links," CISTER-Research Centre in Real-Time and Embedded Computing Systems, Tech. Rep., 2007.
- [16] A. Varga, "The OMNeT++ Discrete Event Simulation System," in *Proceedings of the European Simulation Multiconference (ESM)*, 2001.
- [17] A. Köpke, M. Swigulski, K. Wessel, D. Willkomm, P. T. K. Haneveld, T. E. V. Parker, O. W. Visser, H. S. Lichte, and S. Valentin, "Simulating Wireless and Mobile Networks in OMNeT++: The MiXiM Vision," in *Proceedings of the International Conference on Simulation Tools and Techniques for Communications, Networks and Systems (SIMU-Tools)*, 2008.

Optimization of Composite Winding Strategy Based on Robot

Xu Jiazhong^{#1}, Yang Hai^{*2}, Liu Meijun^{#3}

[#] School of Automation, Harbin University of Science and Technology
No. 52 XueFu Road, Nangang Dist, Harbin, China

¹xujiazhong@126.com

^{*} School of Mechanical Engineering, Harbin University of Science and Technology,
Harbin 150080, China

Abstract— Composite winding workstation of industrial robot with 6 degree of freedom was studied, in which individually analyze the influence of the variation parameters of enveloping form, length of the hanging filament, geodesic and non-geodesic to winding craft and the robot kinetic stability. MATLAB was used to calculate robot winding track in different winding strategies and determine winding track after projecting. Joint simulation of robot winding dynamics based on ADAMS and MATLAB was conducted to obtain curve of each joint torque of the robot, then the influence of the winding track after projecting on kinetic stability of robot was analyzed. The experiment of dry fiber winding of composite products was carried out, which indicates the optimized winding method adopted to stabilize the winding pattern and no slip yarn, overlapping, overhead and other phenomena in the winding process, and the winding precision completely meets the design requirements.

Keywords— Composite, robot, winding strategy, united simulation, optimization

I. INTRODUCTION

In the composite forming technologies, winding is an important technology of the production of composite products. Winding machine is the core of the equipment of winding products. At present, the composite molding equipment and process control research are more mature [1], [2], realized the axisymmetric composite products automatic winding. But for a non axisymmetric composite such as elbow, tee etc. production, they still adopt backward mechanical winding machine and the manual winding, or used 2 ~ 3 degrees of freedom CNC winding machine for semi-automatic winding of artificial auxiliary. For mechanical and CNC winding machine, whose adaptability and flexibility is poor, there are low degree of freedom, unable to solve the problem, such as complex shape of composite products (elbow, tee, etc.) automatic winding [3], [4]. Industrial robot has the advantages of good general, more degrees of freedom, high precision and strong extensibility, so it is suitable for complex high precision automatic winding shape of composite products.

Many scholars abroad have carried out research, involving composite products molding with robot and robot track. Polini and Sorrentino [5], [6] has studied the effect of the safe distance of workpiece and winding ten-

sion, using robot to composite winding, on the properties of products; Aized and Shirinzadeh [7] studied the analysis and optimization of robot fiber laying process using response surface method; Tian and Collins [8] have studied using genetic algorithm for robot track planning, able to quickly get the optimized track; Toussaint [9] has studied robot track optimization by the method of approximate reasoning, and applied to non-LQG system. In addition, the abroad has carried out the research of robot winding equipment, France MF Tech company developed Pitbull and Fox multi-axis robot winding control system [10]; The Netherlands Taniq company independently research and development the Scorpo robot, used for the winding of fiber reinforced rubber products, Scorpo robot can respectively winding rubber lining layer, fiber reinforced layer and the protective layer by replacing the robot end device [11]; Canada Compositum company independently researched and developed a variety of models of the robot and the numerical control system of the automatic control system, with CNC system, robot and winding machine to complete the composite container (natural gas, hydrogen storage tank) production.

Currently there are less domestic organizations to study the robot winding, for complex composite products of the same winding process, there are many robots winding strategy, different winding strategy which effect on winding process and robot stability and mechanical properties, such as tremor, jitter and overload phenomenon appeared in the process of robot winding. However, enveloping form, length of the hanging filament and geodesic winding and non-geodesic winding are a key parameter to influence winding strategy, and there are correlation and influence between each parameters. Therefore, the winding strategy, to make the robot movement stability best, needs a strategy optimization to influence the parameters of the winding. In this paper aiming at composite winding robot, designs a winding pattern of solid of revolution suitable for the robot winding, analyzes winding model of the geodesic of rotors mandrel, and get the doffing point trajectory; for doffing points trajectory, we respectively, in the MATLAB platform, calculate the robot end winding path using different forms of enveloping, length of the hanging filament and geodesic

winding and non-geodesic winding strategy, analyze the winding strategy that affects the winding process and the performance of robotics, and determine the planning of the robot track; we using the ADAMS and MATLAB simulation, in the end of the robot on filament winding tension, verify the stability when the robot catches the mandrel winding according to the planning track; we conduct the shape product dry filament winding experiment according to the planning track and the winding method, verify the feasibility of the optimized winding strategy feasibility to make products.

II. SOLID OF REVOLUTION LINES DESIGN AND TRAJECTORY PLANNING

Winding lines design is the basis of the winding process, it is closely related to the quality of the winding products and the realization of the molding process. For complex composite products, we designed the linear design of cylinder, cone and ellipsoid composites of rotors mandrel in a way of the geodesic winding, and obtained the doffing point trajectory of the mandrel surface; for doffing points trajectory, we respectively, in the MATLAB platform, calculate the robot end winding track using different forms of enveloping, length of the hanging filament and geodesic winding and non-geodesic winding strategy, and plan the motion trajectory of robot winding.

A. Solid of revolution winding lines design

(1) Cylindrical geodesic winding

Cylinder and conical section of geodesic winding for composite products mandrel as shown in Fig. 1. Cylinder mathematical model as shown in Fig. 1 (a), the cylinder adopt stable geodesic winding, if the doffing points in the cylindrical section where the fiber fall off are regarded as the assumption point M in the process of winding, the equation of point M , suitable for the surface of the cylinder to meet the surface, is $r(R\cos\theta, R\sin\theta, z)$. z is the axisymmetrical z axis coordinates, r is the radius of each point along the z axis direction, R is the radius of the cylinder, Set up a $Oxyz$ coordinate system, the x axis goes through L to intersect at the rectangular bottom edge. M on the xOy projection is M_1 , the Angle from the x axis forward to OM_1 is θ . The filament winding Angle for cylinder section is ψ . Point located on the cylinder and xOy surface in winding angles around the z axis to rotate 360° from xOy distance below is h . There are

$$\begin{cases} z = M_1M = \frac{\theta}{2\pi} \cdot h = (R \cdot \tan\psi) \cdot \theta \\ x = R\cos\theta \\ y = R\sin\theta \end{cases} \quad (1)$$

Get on the surface of the cylinder for the geodesic equation is

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = r\theta\cos\psi \end{cases} \quad (2)$$

(2) Conical geodesic winding

Cone mathematical model as shown in Fig. 1 (b),

conical half-apex angle is β . The geometric characteristics of rotating cone head shows that the shape is formative by $y = r(z)$ generatrix around the rotating axis z . In polar coordinates of the cone surface the expression is

$$r(\theta_0, z) = \{r(z)\cos\theta_0, r(z)\sin\theta_0, z\} \quad (3)$$

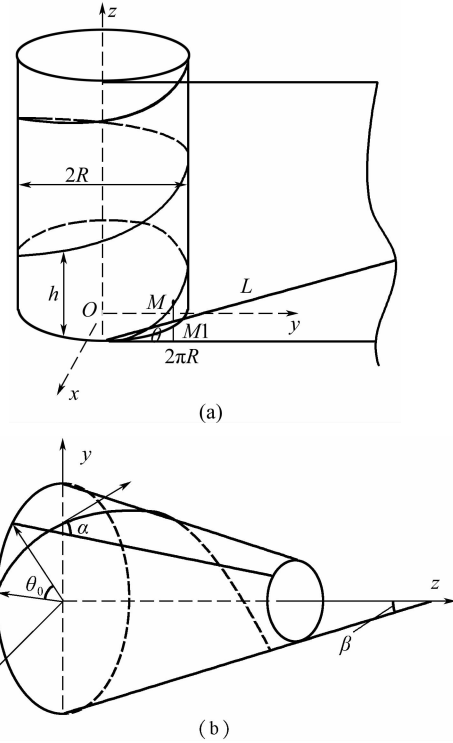


Fig. 1 Mathematical model of cylinder and conical section of geodesic winding for composite products mandrel
(a) Mathematical model of cylinder;
(b) Mathematical model of conical section

In expression, θ_0 is the corner of mandrel center. The expression (3) is carried out calculations in the way of $\theta'_0, \theta''_0, z', z''$. According to the differential geometry theorem, the curvature of the cone can be obtained by Gaussian curvature equation

$$k_g = \frac{d\theta_0}{ds} = -\frac{d\alpha}{ds} - \frac{r'\sin\alpha}{r\sqrt{1+r'(z)^2}} \quad (4)$$

In expression, k_g is the geodesic curvature, s is the filament winding parameters of arc length, α is the winding angle. Based on differential geometry, surface mean curvature, gauss theorem, differential geometry theorem and Euler formula (5), it can be seen

$$k_n = \frac{r'}{(1+r'^2)^{\frac{3}{2}}} \cos^2\alpha - \frac{1}{r\sqrt{1+r'^2}} \sin^2\alpha \quad (5)$$

In expression, k_n is the normal curvature, k_1 is the meridian curvature, k_2 is the ring to curvature, Set $k_g/k_n = \lambda$, λ is the geodesic normal curvature ratio, then the trajectory differential equations as follows

$$\frac{d\alpha}{dz} = -\frac{r'\tan\alpha}{r} - \lambda \left(\frac{r''}{1+r'^2} \cdot \cos\alpha - \frac{\sin\alpha\tan\alpha}{r} \right) \quad (6)$$

When λ is 0, differential equation is the differential equation of geodesic trajectory.

(3) *Ellipsoid geodesic winding*

Models of ellipsoid and geodesic on surface of ellipsoidal dome for composite products mandrel as shown in Fig. 2. To solve the geodesics on the surface of the solid of revolution is equivalent to solve the geodesics on the ellipsoid ΔNAB [12] [13]. The method that the ellipsoid is seen as the ball is adopted to solve the geodesics on the surface of the elliptical head, then establish the geodesic coordinate system (O, B, H). Said ellipsoid Fig. 2 (a) mathematical model of the semi-major axis length is a , the length of short half axis is b , the flat rate of ellipsoid is f , the first eccentricity is e , and the second eccentricity is e' . According to the formula of Clairaut here are

$$\begin{aligned} \sin\gamma_0 &= \sin\gamma_1 \cdot \cos\beta_1 \\ &= \sin\gamma_2 \cdot \cos\beta_2 \end{aligned} \quad (7)$$

Inexpression, γ is the azimuth angle, β is called as parameter latitude, β is to satisfy the following formula

$$\tan\beta = (1 - f) \cdot \tan\phi \quad (8)$$

Inexpression, ϕ is the latitude angle. Fig. 2 (b) is a round sphere, according to Clairaut formula: $\sin\gamma_0 = \sin\gamma \cdot \cos\beta$. In the NEP triangle, the length of the edge EP and the edge of the opposite Angle ω are corresponding to the length s_n of the geodesic line AB side and the edge angle λ_n of AB to the other side in the Fig. 2(a) ellipsoid. Satisfy the following equation (10), supposing $k = e' \cdot \cos\gamma_0$

$$\frac{s_n}{b} = \int_0^\sigma \sqrt{1 + k^2 \cdot \sin^2\sigma'} d\sigma' \quad (9)$$

$$\lambda_n = \omega - f \cdot$$

$$\sin\gamma_0 \int_0^\sigma \frac{2 - f}{1 + (1 - f) \sqrt{1 + k^2 \times \sin^2\sigma'}} d\sigma' \quad (10)$$

Inexpression, σ is the ellipsoid arc length, σ' is the derivative of ellipsoid arc length.

As shown in Fig. 2 (b), geodesic sphere surface as starting point E, with initial azimuth γ_0 , the geodesic line is across the equator; Point P can represent A and B any point of geodesic line AB, and each point is with parameter $\beta, \gamma, \sigma, \omega, s_n$ and λ_n , where $s_{12} = s_1 - s_2$ means the length of AB. s_1 is the A coordinates, s_2 is the B coordinates; $\lambda_{12}, \sigma_{12}$ and ω_{12} are the same definition, γ_2 express the forward direction of the point B.

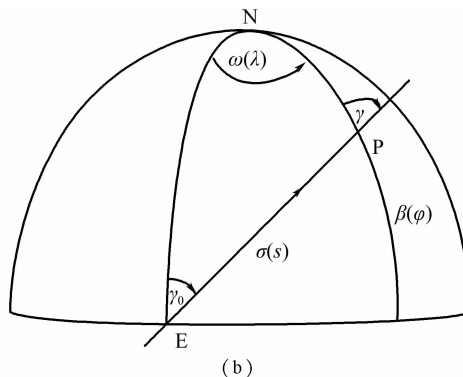
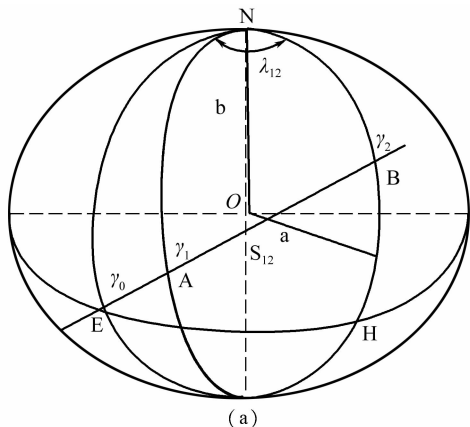


Fig. 2 Models of ellipsoid and geodesic on surface of ellipsoidal dome for composite products mandrel (a) Model of ellipsoid; (b) The geodesic on the surface of the ellipsoidal dome

B. The winding trajectory planning

According to the solid of revolution winding equation, we gained the doffing point trajectory of the mandrel surface, using robot movement in the form, including the envelope form of pay-out movement around the work-piece, length of the hanging filament and geodesic winding and non-geodesic winding etc., and we analyzed the winding process and robot performance.

(1) *The winding trajectory of the envelope form*

When it is winding, the pay-out moves around the workpiece, its line formed the envelope line relative to the workpiece. According to the shape of the envelope, it can be divided into open enveloping cylinder, closed enveloping cylinder, entire enveloping cylinder and constant free fiber length, 4 kinds of forms, schematic diagrams of 4 types of enveloping forms as shown in Fig. 3. We planed the axis trajectory of the robot in the four kinds of forms, then by using the Matlab calculated the axis trajectory, finally got its motion characteristic curve, Movement relationship curves between carriage, cross carriage, pay-out and mandrel under the condition of 4 types of enveloping patterns as shown in Fig. 4.

The open enveloping cylinder, when the end reverse, the characteristic curve is choppy. The sawtooth can lead to continuous reversing, the acceleration and deceleration of the end terminal is too large, leading to different degrees of tightness of the fiber. The fiber resin concentration and winding stress are uneven, which affect products quality and performance. From the point of the slope of the curve, the end is almost constant motion; when it is reversing, there is the curve of discontinuity. There are vibration problems for large inertia at the end of the actuator, then the robot body movement process is not stable. When the closed enveloping cylinder is in reversing the sawtooth phenomenon of the motion characteristic curve is improved, it still exists repeated start and stop action in the mid-piece, and it still exists the mutation rate when it is reversing. The entire enveloping cylinder in the process of marching in the middle is very smooth, but vibration problem is in the reversing area.

Constant free fiber length avoids the above problems, and its characteristic curve is smooth, Except on both ends of the reversing there is no other directional movement in the end, explaining its movement is stable; It

won't cause the problem of tremors and the too large deceleration, so it can be used as the preferred method in this paper.

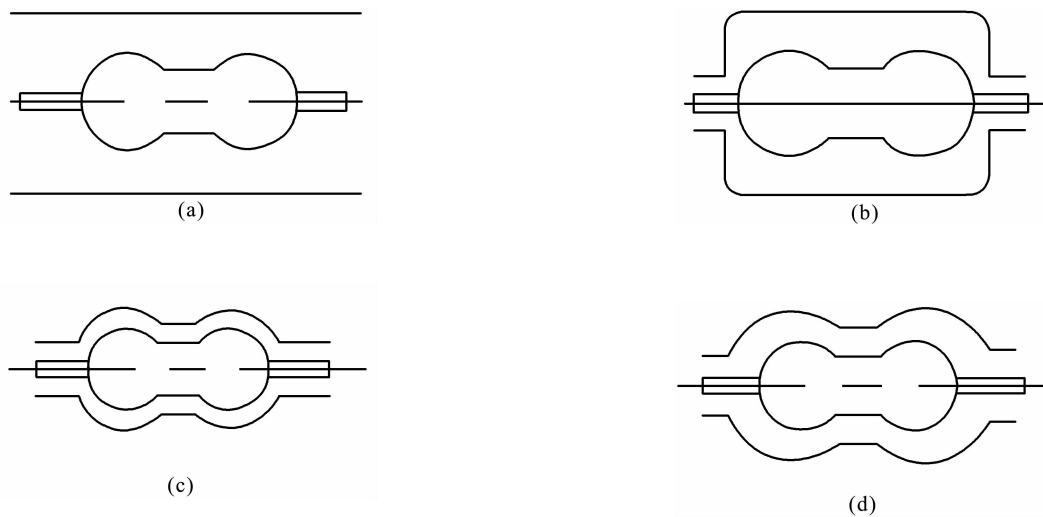
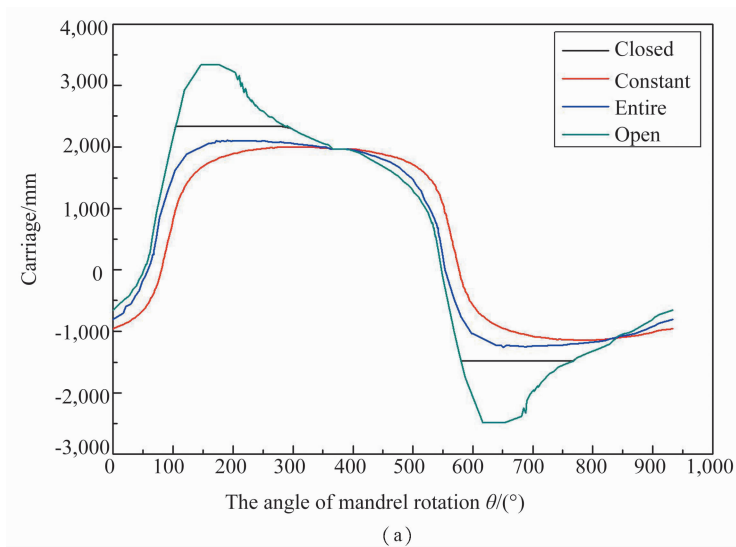


Fig. 3 Schematic diagrams of 4 types of enveloping forms ; open enveloping cylinder ,closed enveloping cylinder ,entire enveloping cylinder and constant free fiber length
 (a) Open enveloping cylinder ; (b) Closed enveloping cylinder ;
 (c) Entire enveloping cylinder ; (d) Constant free fiber length



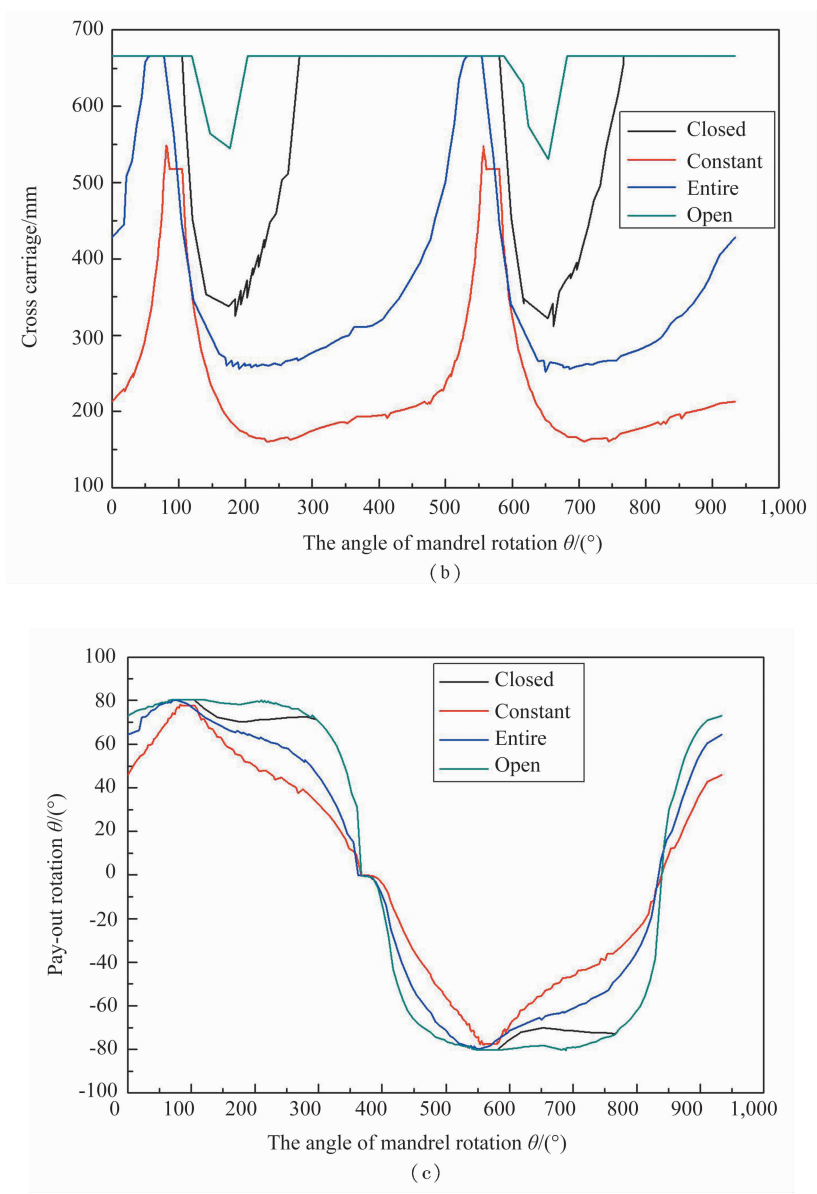


Fig. 4 Movement relationship curves between carriage, cross carriage, pay-out and mandrel under the condition of 4 types of enveloping patterns: open enveloping cylinder, closed enveloping cylinder, entire enveloping cylinder and constant free fiber length
 (a) Movement relationship between carriage and mandrel;
 (b) Movement relationship between cross carriage and mandrel;
 (c) Movement relationship between pay-out rotation and mandrel

(2) The length of hanging filament is fixed length winding trajectory

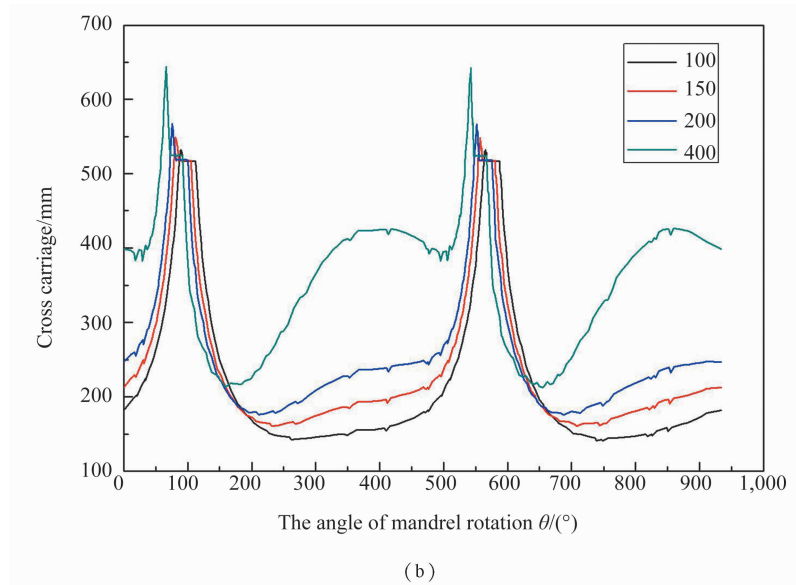
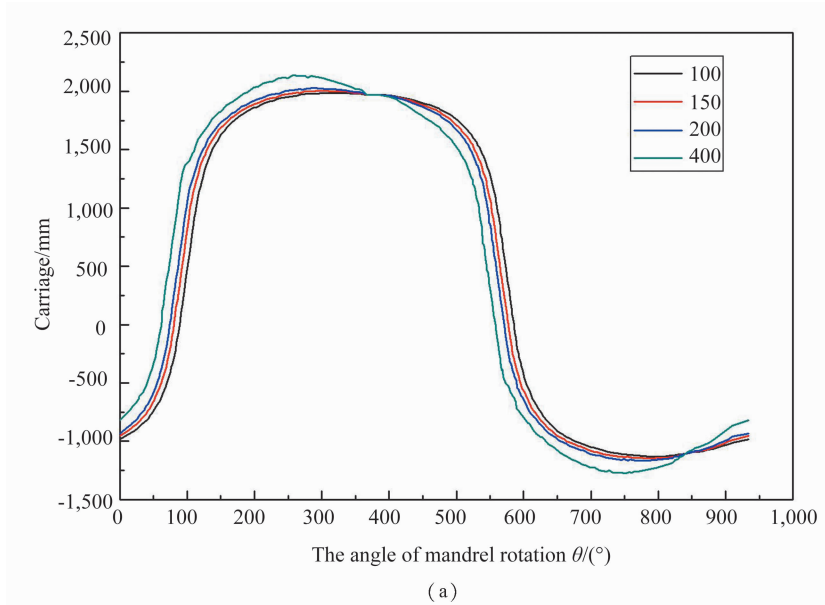
The theory algorithm shows that when the hanging filament λ is the fixed length, the equations of motion are with finite number of solutions. $\lambda = 100$ mm, 150 mm, 200 mm, 400 mm are taken to calculate. When the fiber finishes a complete winding, movement relationship curves between mandrel and carriage, cross carriage, pay-out while the length of hanging filament is equal to 100 mm, 150 mm, 200 mm and 400 mm is shown in Fig. 5.

With the increase of the value λ , the movement curve

raised, and the carriage and the cross carriage movement increases. Winding equivalent radius of pipe the robot work radius also needs to increase; the larger λ is, the greater the each axis of motion curve is shaking. It is easy to cause fiber tension in winding process can't keep constant, lead to the fiber from gaps or covers each other. This is not conducive to increase the rate of filament winding products. In addition, the λ is greater, in the conical section of small diameter directional area, the faster the reversing speed. The robot will produce loss, leading to reduce the service life of the robot. When λ is smaller, movement relationship curves between mandrel

and carriage, cross carriage more gentle, more stable movement, but when the width of the pay-out is greater than 2λ , the guide pay-out during winding will collide with the surface of the mandrel and therefore we need to

ensure the robot is with adequate moving space of the case, and the length of hanging filament is as short as possible.



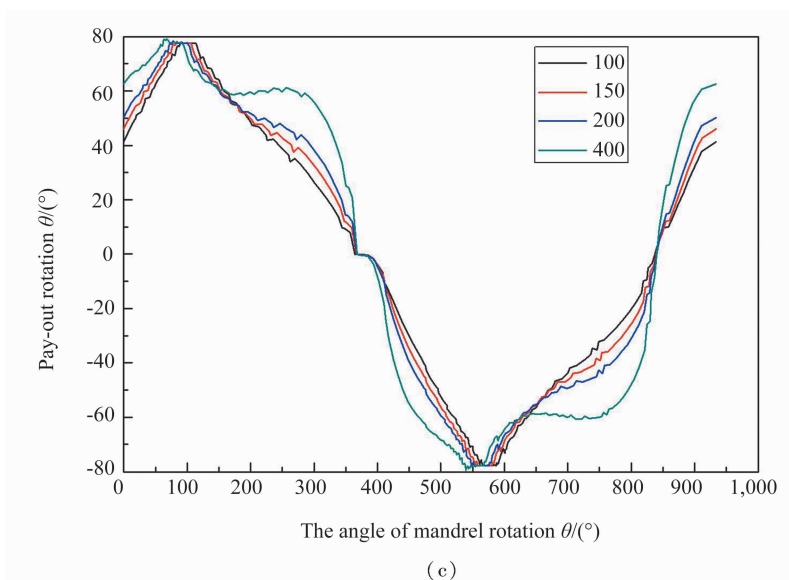
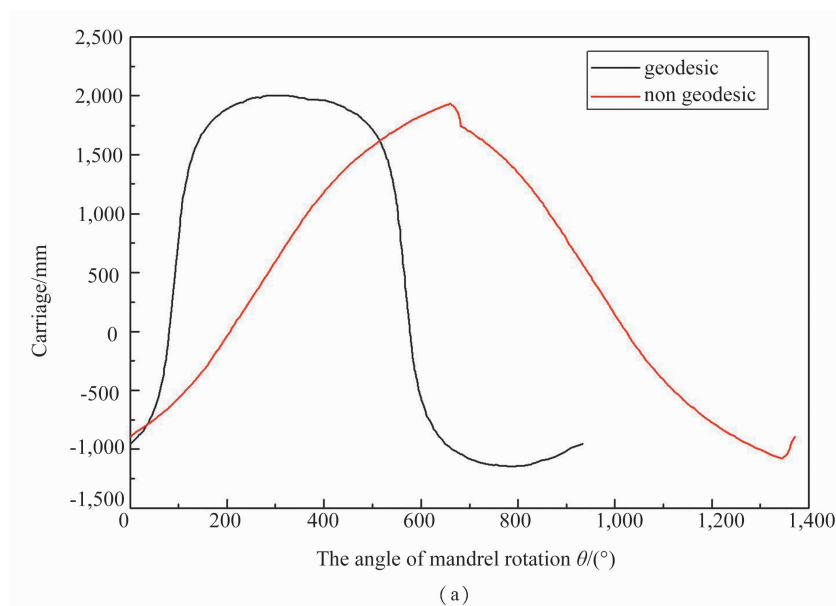


Fig. 5 Movement relationship curves between mandrel and carriage, cross carriage, pay-out while the length of hanging filament is equal to 100, 150, 200 and 400 mm (a) Movement relationship between carriage and mandrel; (b) Movement relationship between cross carriage and mandrel; (c) Movement relationship between pay-out and mandrel

(3) Geodesic and non-geodesic winding trajectory

Whether to adopt the geodesic winding not only determined by the shape of the workpiece, but also on the performance of the molded article and molding robot capable of stable operation. While the geodesic (helical) winding has the advantage that the linear is stable and there are non-slip filament, but it is not suitable for all

the complex material products' winding. Hence, we should analyze the motion characteristics for each axis which of the geodesic (helical) and non-geodesic winding. The movement relationship curves between mandrel and carriage, cross carriage, pay-out for geodesic / non geodesic is shown in Fig. 6.



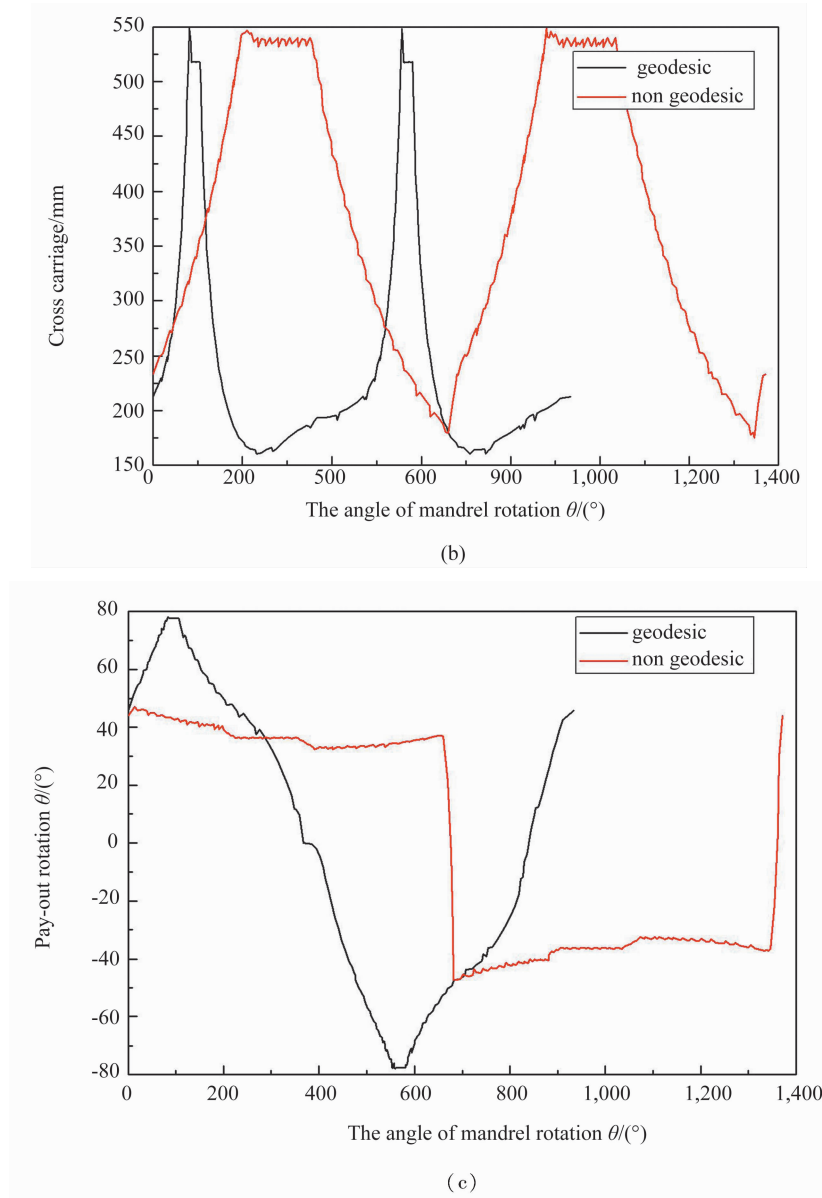


Fig. 6 Movement relationship curves between mandrel and carriage, cross carriage, pay-out for geodesic/non geodesic
 (a) Movement relationship between carriage and mandrel; (b) Movement relationship between pay-out and mandrel;
 (c) Movement relationship between cross carriage and mandrel

When it helical winding, the carriage curve is not jagged or mutations, being continuous and no abnormal points, indicating that the carriage more smoothly, less prone to tremors or shaking. Compared with geodesic winding, there are significantly abnormal points during the carriage is reversing in the non-geodesic winding way. When it helical winding, cross carriage reversing also appeared mutation rate, and the pay-out motion curves have jagged individually, but considering the non-geodesic winding, the cross carriage and the pay-out have significant jitter when they reverse; therefore the helical winding way is better. But it needs further optimization and is eliminated rate mutation.

Through analyzing the above three forms of motion, we can conclude that the robot trajectory curve, adopting a way of the constant free fiber length and smaller

length of the hanging filament, is relatively smooth. It can be used as the ideal path of the robot winding, but we still need robot winding trajectory simulation to verify the stability of the robot winding.

III. ROBOT MOTION SIMULATIONS

According to the winding of six degrees of freedom robot mechanical structure, we build the virtual model of the robot, whose base is fixed on the ground. The six degrees of freedom winding robots' posture means the position and posture at which the robot movement of each connecting rod and each joint. In the process of filament winding, when the pay-out moves along the preset winding process, robots can have a lot of kinds of path and at the same time. This paper selects the dumbbell core

mandrel, using the optimized robot winding path. Namely fixed length fiber free length is 150 mm, point of winding is $-5/3$, constant mandrel speed, the trajectory of the geodesic winding, which is 50 mm wide. As shown in Fig. 7, the robot winding position while the starting angle of six joints are respectively 0° , -90° , 90° , -90° , -90° and 0° .



Fig. 7 The robot winding position while the starting angle of six joints are respectively 0° , -90° , 90° , -90° , -90° , 0°

From the base to the end of the robot, the robot joints numerals are joint 1~6, the rotation angle of each joint respectively is $\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6$, so we can obtain the size of the joint angle by the pose forward kinematics equation, that is the inverse kinematics of the robot equation

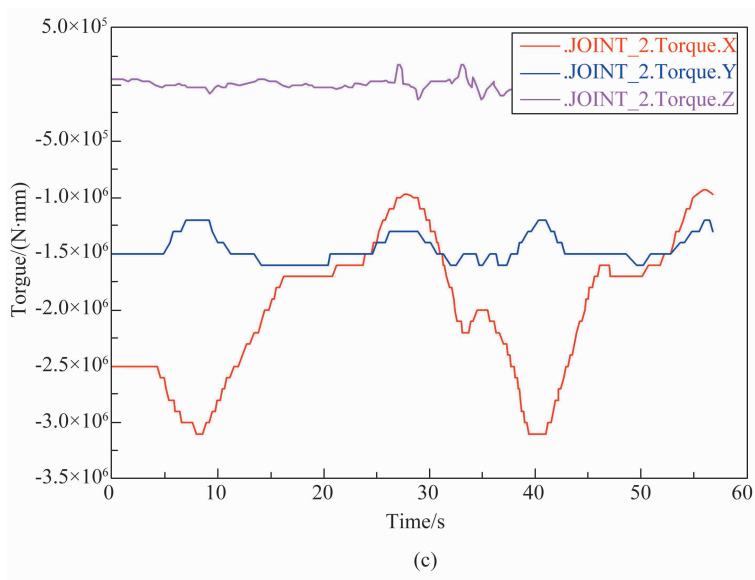
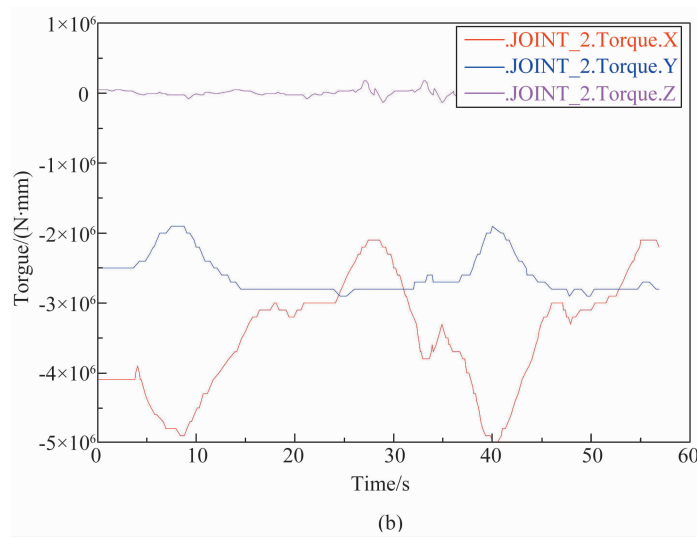
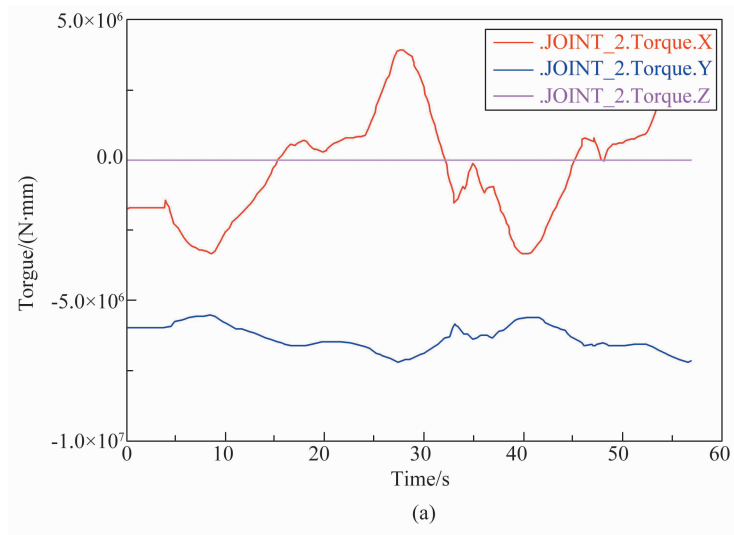
$$\begin{cases} \theta_1 = \tan^{-1}\left(\frac{X_F}{Y_F}\right) \\ \theta_4 = \tan^{-1}\frac{K}{H - Z_F} - \sin^{-1}\left(\frac{\sqrt{(H - Z_F)^2 + K^2}}{2L_2} + \frac{L_3^2 - L_2^2}{2L_2\sqrt{(H - Z_F)^2 + K^2}}\right) \\ \theta_6 = \pi - \sin^{-1}\left(\frac{\sqrt{(H - Z_F)^2 + K^2}}{2L_3} + \frac{L_3^2 - L_2^2}{2L_3\sqrt{(H - Z_F)^2 + K^2}}\right) - \sin^{-1}\left(\frac{\sqrt{(H - Z_F)^2 + K^2}}{2L_2} - \frac{L_3^2 - L_2^2}{2L_2\sqrt{(H - Z_F)^2 + K^2}}\right) \end{cases}$$

In expression, X_F, Y_F and Z_F are the coordinates of the robot end in the robot base coordinate system, H is the origin distance from ground robot joints 1, L_1, L_2 and L_3 are respectively 2, 4, 6, arm length of robot joints. The Pose take

$$K = \frac{Y_F}{\cos\theta_1} - L_1\cos\theta_2$$

In the condition that the track of the robot's terminal in this position and orientation is exactly the same, due to the different angles of each joint and connecting rod position in the winding course of the winding, when the end of the robot by fiber tension, each joint bear component are different, thereby affecting to each joint driving force and driving torque, the robot power consumption during the winding process of a difference. Therefore, the combination of winding dumbbell products as simulation object, consider the process of winding speed, acceleration and other parameters of the dumbbell product during the winding process the fiber tension by analysis and simulation.

Using Matlab the pose inverse kinematics equation is solved to obtain the curves of the joint angles and the time t , and outputs the data to the ADAMS, to control each robot model axis motor runs at a specified track, and the end of the robot is applied 50 N tension load. Get the joint drive torque curve through the simulate function and post-processing functions of ADAMS software, Driving torque simulation curves of six joints while end of robot is applied tension loading of 50 N as shown in Fig. 8.



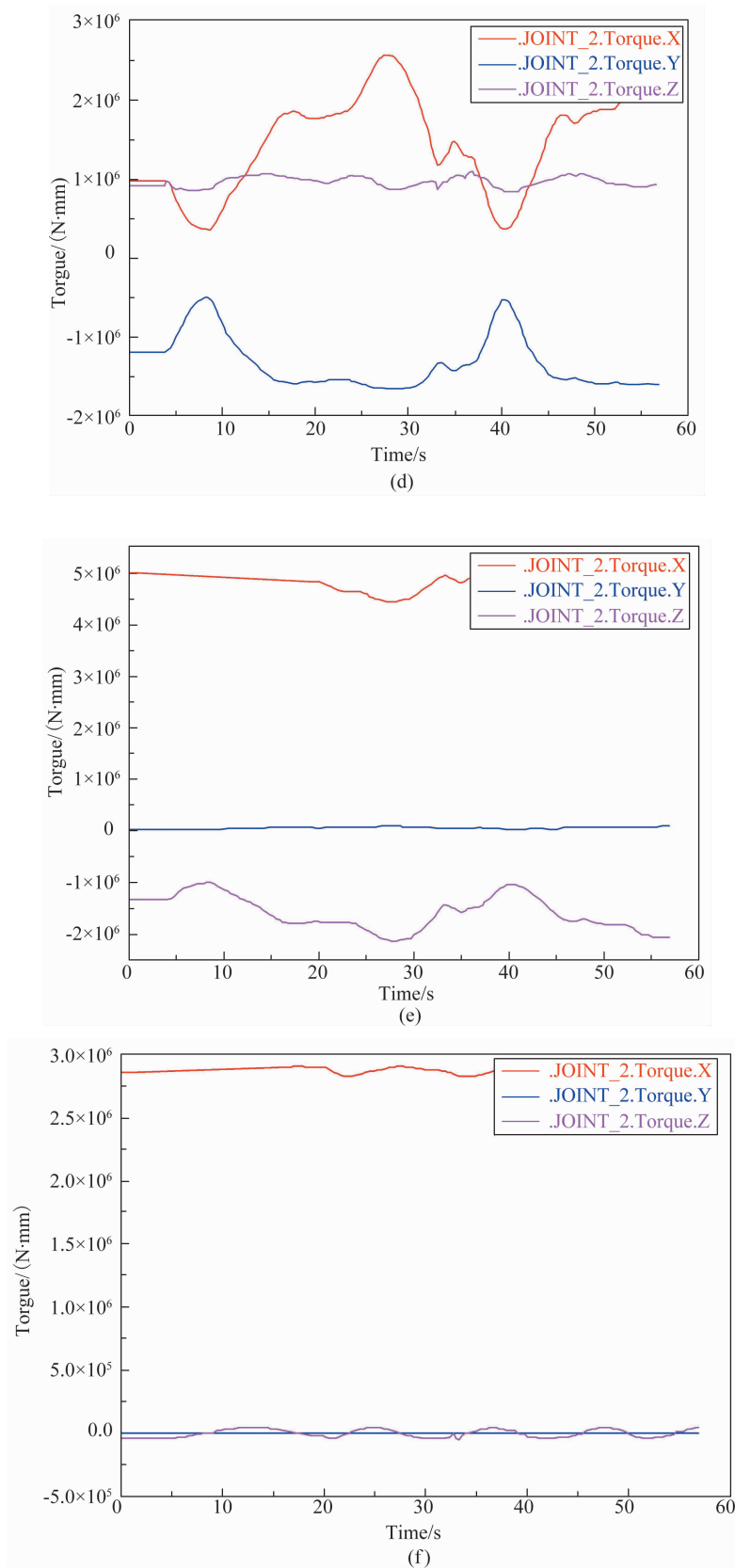


Fig. 8 Driving torque simulation curves of six joints while end of robot is applied tension loading of 50 N (a) Driving torque simulation curves of robot joint_1; (b) Driving torque simulation curves of robot joint_2; (c) Driving torque simulation curves of robot joint_3; (d) Driving torque simulation curves of robot joint_4; (e) Driving torque simulation curves of robot joint_5; (f) Driving torque simulation curves of robot joint_6;

In the process of robots winding, each robot joint torque curve is relatively stable. The torque curve slope of the x axis direction appears too large commutation place, due to the mutation of movement speed of the mandrel and moving in the opposite direction in the commutation process, But the x axis points torque is small, negligible effecting on robot; There is little change on the torque curve of the y direction, it is in a stable state. Because, in the winding process, the motion range of the robot in the z axis direction was small, the torque on the z axis direction was close to zero and the curve is smooth. Through the curve analysis of each robot joint torque, it was found that, under the joint action by the tension of the fibers, the torque curve of each robot joint is steady and the joint torque is small. The energy consumed less during exercise, so the robot with the optimized winding path which robot hand grasping mandrel smooth movement in the process of winding way, the way and the winding path can be used for the robot.

IV. ROBOT WINDING EXPERIMENTS

The experiments equipment use the filament winding workstation based on KUKA robot, the equipment parameters are as follows. Maximum of yarn speed: 50 m/min; robot winding range: 2,000 mm; winding trajectory accuracy: ± 0.3 mm; robot load: 210 kg; the winding experiment employed dumbbell core mandrel. The specific dimensions are as follows: pole diameter 52 mm, cylinder section diameter 190 mm, cylinder length 150 mm, ball segment radius 145 mm. In this study, The length of hanging filament is 150 mm, winding pattern number is $-5/3$, mandrel speed is constant, and 50 mm wide geodesic yarn were winding yarn. We adopted the way of grasping mandrel winding by robot hand to produce, robot hand using mandrel winding process winding way, as shown in Fig. 9. Winding pattern of geodesic for a cycle of dumbbell core mandrel using a single yarn while length of free fiber is 150 mm, winding pattern is $-5/3$ and width of yarn is 50 mm as shown in Fig. 10, and winding lines are wound as shown in line measurement, Measure cycle data between adjacent yarn of winding pattern for a cycle of dumbbell core mandrel shown in Fig. 11.

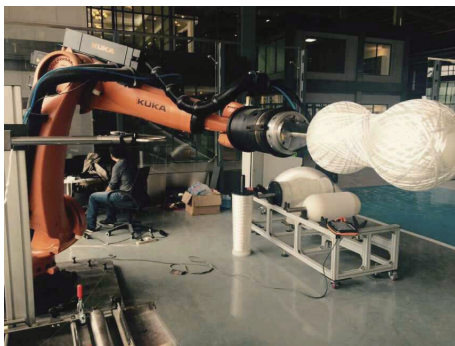


Fig. 9 Winding way of robot hands catch mandrel

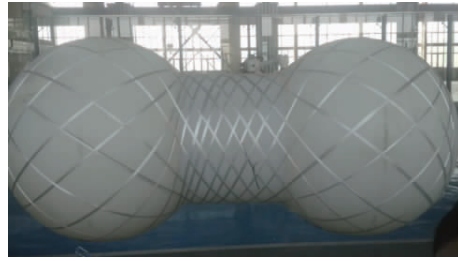
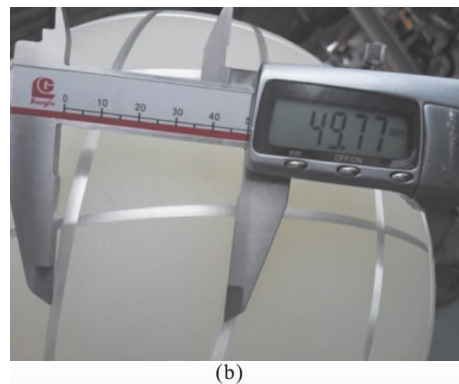


Fig. 10 Winding pattern of geodesic for a cycle of dumbbell core mandrel using a single yarn while length of free fiber is 150 mm, winding pattern is $-5/3$ and width of yarn is 50 mm



(a)



(b)

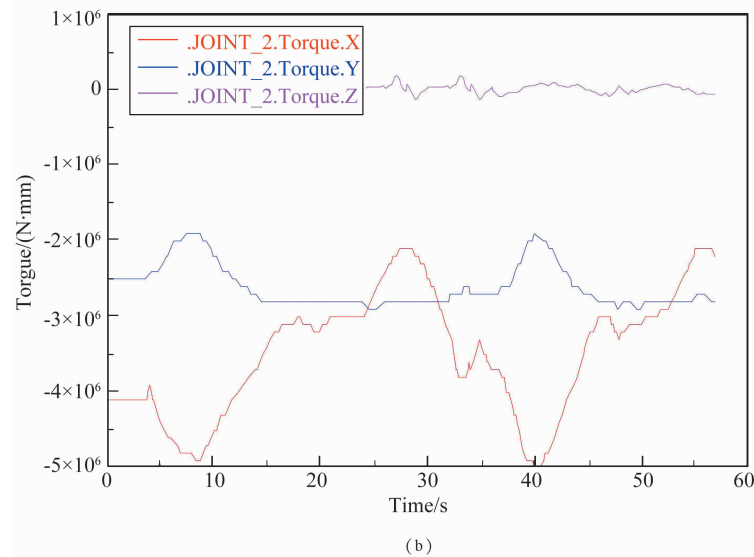
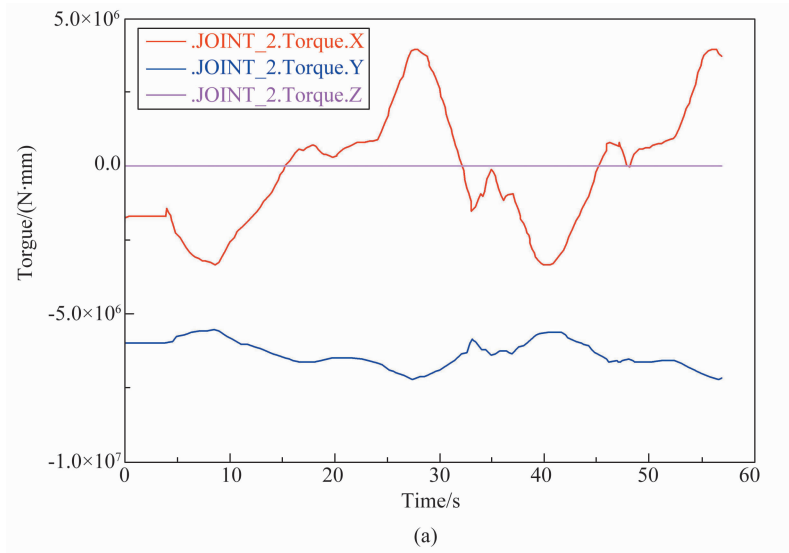
Fig. 11 Measure cycle data between adjacent yarn of winding pattern for a cycle of dumbbell core mandrel
(a) Pattern measurement of cylindrical section;
(b) Pattern measurement of sphere section

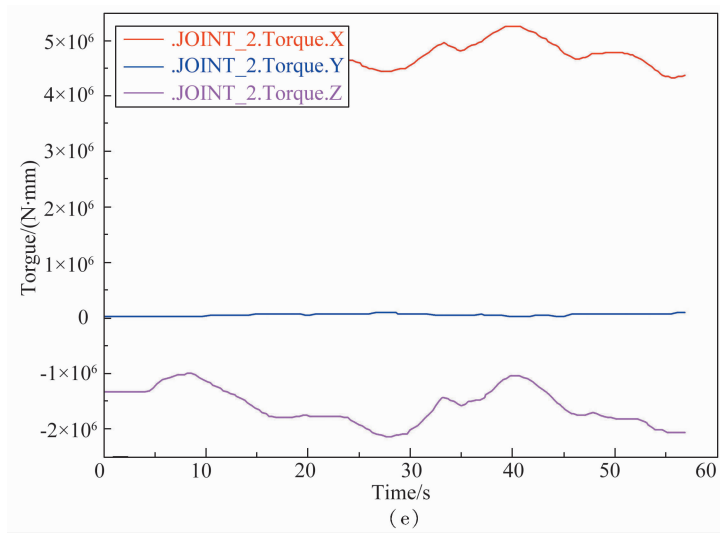
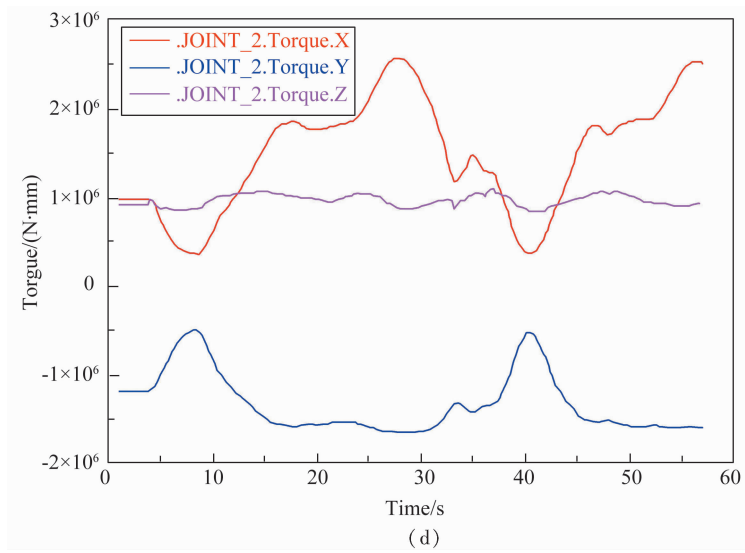
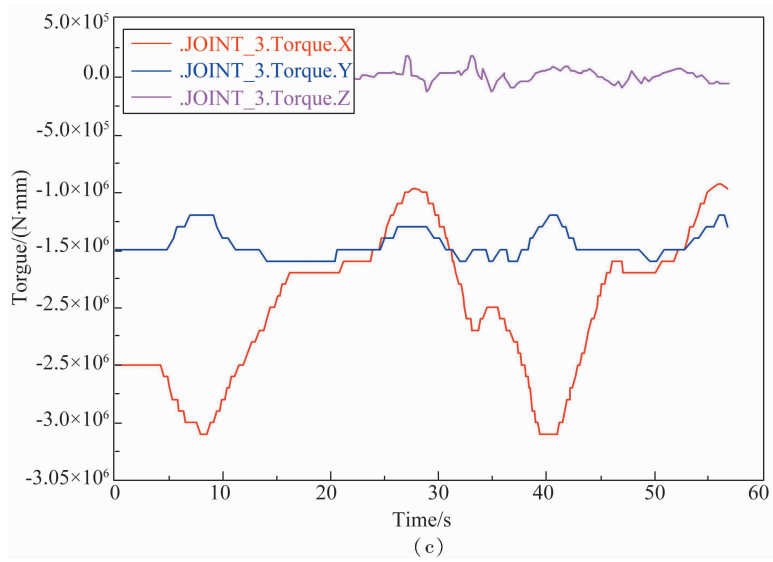
During the filament winding workstation, based on KUKA robot, complete dumbbell products winding process, each robot axis moves smooth, and there is not slipping yarn, overlap, overhead and so on. Through measuring the winding line, the distance between adjacent yarn sheet is 49.77 mm, and the width of the yarn used in this experiment is 50 mm, thus the winding traj-

jectory accuracy is less than 0.3 mm and the winding precision of the workstation meets the design requirements in order to achieve the composite products stable winding, which proves the feasibility of the linear design and the robot motion trajectory.

In robot grabbing dumbbell core mandrel winding process, the winding mandrel is applied 50 N tension; Then take the value of the angle of each joint real-time motion of the robot, and get each joint drive torque curve using ADAMS, driving torque active curves of six

joints while end of robot is applied tension loading of 50 N as shown in Fig. 12. During the winding process, the robot running smoothly, and there is not skew or dumping phenomenon. According to Fig. 12, each robot joint torques, that adopts the way of catching the mandrel by robot hand, is smaller, and the energy consumed during exercise is less. In addition, the curve and robot motion simulation resulting torque curve are consistent, and then verify the reasonableness of robot motion simulation.





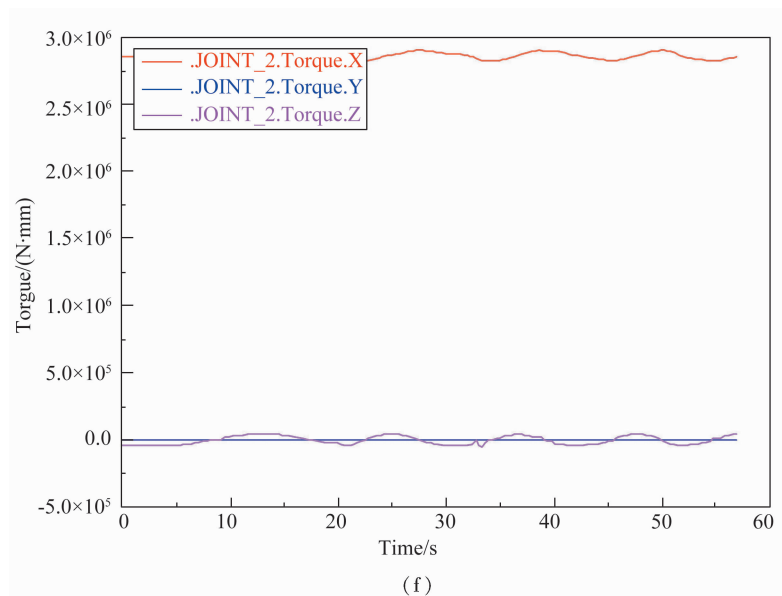


Fig. 12 Driving torque active curves of six joints while end of robot is applied tension loading of 50 N
 (a)Driving torque active curves of robot joint_1;(b)Driving torque active curves of robot joint_2;
 (c)Driving torque active curves of robot joint_3;(d)Driving torque active curves of robot joint_4;
 (e)Driving torque active curves of robot joint_5;(f)Driving torque active curves of robot joint_6

V. CONCLUSIONS

① Through the analysis of such geodesic winding model of the cylinders, cones and ellipsoid composites products, we calculated each geodesic equation, and obtained the model the doffing point trajectory of the robot winding.

② Various parameters influence wound strategies to optimize the constant free fiber length and smaller length of the hanging filament geodesic winding track is a typical product winding track robot stable operation.

③ Through the ADAMS and Matlab we carried on the collaborative simulation of the combination of winding movement adopting the way that robots catch winding mandrel by robot hand. When the winding movement on the basis of the robot, under the action of fiber tension, each robot joint torque is small and its movement is smooth.

④ Through the experiment, the line of products is stable; there are not the phenomena of slipping yarn, overlap, overhead. The winding precision meets the design requirements; robot moves smoothly, each joint torque curve of the motion simulation and is almost the same as the size and trend of the joint torque curve in the winding experiments. It verifies the feasibility of linear design and motion path, and the correctness of the simulation results.

REFERENCES

- [1] SHI YY, YU T, HE X D, et al. Mechanism and optimization of process parameters coupling for composite tape winding [J]. *Acta Materiae Compositae Sinica*, 2015, 32(3): 831-839 (in Chinese).
- [2] SHIRINZADEH B, CASSIDY G, OETOMO D, et al. Trajectory generation for open-contoured structures in robotic fiber placement [J]. *Robotics and Computer-Integrated Manufacturing*, 2007, 23(4): 380-394.
- [3] RAHMAN H, JAMSHED R, KHAN A, et al. Design of tape wound composite cylindrical shells incorporating different failure criteria and winding kinematics [J]. *Advanced Materials Research*, 2012, 570(9): 53-62.
- [4] KOUSSIOS S, ZU L, WENTZEL CM. Filament winding: Process overview & novel developments [C] // *International SAMPE Technical Conference*. Baltimore, MD: Society for the Advancement of Material and Process Engineering, 2012: 465-466.
- [5] POLINI W, SORRENTINOL. Actual safety distance and winding tension to manufacture full section parts by robotized filament winding [J]. *Journal of Engineering Materials and Technology-Transactions of the ASME*, 2006, 128(3): 393-400.
- [6] POLINI W, SORRENTINO L. AR models to forecast roving tension trend in a robotized filament winding cell [J]. *Materials and Manufacturing Processes*, 2006, 21(8): 870-876.
- [7] AIZED T, SHIRINZADEH B. Robotic fiber placement process analysis and optimization using response surface method [J]. *International Journal of Advanced Manufacturing Technology*, 2011, 55(1-4): 393-404.
- [8] TIAN L F, COLLINSC. An effective robot trajectory planning method using a genetic algorithm [J]. *Mechatronics*, 2004, 12(5): 455-470.
- [9] TOUSSAINT M. Robot trajectory optimization using approximate inference [C] // *Proceedings of the*

26th International Conference On Machine Learning. Montreal, QC: Omnipress, 2009: 1049-1056.

- [10] WANG G Y, CAO J, LI C W. Winding technology of composite elbow based on dual-port RAM of PMAC [C] // 2011 International Conference on Computer Science and Service System. Northeast Forestry University: IEEE Computer Society, 2011: 2842-2845.
- [11] RATERINK J C, NOOIJ S M, KOUSSIOS S. Improving the performance of fiber reinforced presurizable products [D]. Delft: Delft University of Technology, 2009.
- [12] W C, WANG R G, LIU W B, et al. Measurement of slippage coefficient between fiber and mandrel surface for non-geodesic filament winding [J]. Acta Materiae Compositae Sinica, 2012, 29(3): 191-196(in Chinese).
- [13] CHARLESF F K. Algorithms for geodesics [J]. Journal of Geodesy, 2013, 87(1): 43-55.

Path Planning of Mobile Robot Using Virtual Plane Approach in Dynamic Environment and Its Applications

T. Khurelbaatar^{#1}, P. Enkhtsoyt^{#2}, B. Zorig^{#3}

[#]Department of Electronic, School of Information and Communications Technology,
Mongolian University of Science and Technology, Mongolia

¹khurelbaatar@must.edu.mn

Abstract— In this paper, we present sensor fusion approach for path planning of mobile robot in dynamic environment. This sensor fusion approach is based on the position and orientation estimation using combination of camera, encoder signals to improve and correct the measurement of moving object's position and velocity. We design our mobile robot whose name is LION with the function of the remote control system to navigate the mobile robot in indoor environment using a wireless network. We use Matlab software for the path planning of a mobile robot based on virtual plane approach and combine this simulation with real mobile robot. For this simulation uses image processing for observe view of the surroundings, one of the main issues is robot navigation and collision detection of moving bodies using the web camera. Based on collision detection, the robot will start avoiding from obstacles.

The image data of web camera is processed using Matlab software for path planning algorithm at the user PC and the navigation data is transmitted to mobile robot from user PC through the Bluetooth for remote controlling the mobile robot.

Keywords— Path planning algorithm, service robot, robot designing and navigation, control law

I. INTRODUCTION

For autonomous robot systems which work without human operators, one of the important things is path planning of mobile robot that is necessary to plan a collision free path minimizing a cost such as time, energy and time.

This paper uses a path planning algorithm based on virtual plane approach to plan an optimal or feasible path avoiding obstacles when a mobile robot moves from a start point to goal point in dynamic environment [1]. Virtual plane approach is an invertible transformation equivalent to the workspace which is constructed by using a local observer speed of the mobile robot and orientation angle are independently controlled using simple collision cones and collision windows constructed from the virtual plane. Therefore, based on the virtual plane, it is possible to determine the intervals of the linear velocity and the paths that lead to collisions with moving obstacles [2].

We present fast measuring method using a sensor's fusion to correct position errors in robust localization scheme. This sensor fusion approach is based on the position and orientation estimation using combination of camera, encoder signals to improve and correct the

measurement of moving object's position and velocity. We use the Kalman filter algorithm that is used for estimating position and orientation of the robot to combine sensors information [3].

The performance of the proposed method is demonstrated by simulation movement results using experiment with moving obstacles.

We design our training robot whose name is LION robot and combine this robot's navigation system with path planning simulation based on concept of virtual space using Matlab.

This paper is organized as follows, see Fig. 1. First, we introduce modelling and path planning of the mobile robot with two wheels. In section III, we introduce the implementation of Kalman filter to combine a camera data with encoder signal. The implementation of the developed system is presented in section IV. The section V is the conclusion of the work.



Fig. 1 "LION" robot model

II. MODELING AND PATH PLANNING OF THE MOBILE ROBOT

A. Mobile Robot Kinematics

The Fig. 2 shows the navigation for the mobile robot by the orientation and the speed. The line of sight of the robot l_r is the imaginary straight line that starts from O and it is directed toward the reference point of robot R and the line of sight angle, α_r which is the angle made by l_r . The line of sight of the robot l_r is the imaginary straight line that starts from the origin and is directed toward the reference center point of the robot R . The line-of-sight angle, α_r is the angle made by the sight l_r . The dis-

tance l_{gr} between robot R and the goal G is calculated by

$$l_{gr} = \sqrt{(y_g - y_r)^2 + (x_g - x_r)^2} \quad (1)$$

where (x_g, y_g) is the coordinates of the final goal point and (y_r, x_r) is the state of the robot in $\{W\}$. The mobile robot has a differential driving mechanism using two wheels and the kinematic equation of the wheeled mobile robot can be given by

$$x_r = v_r \cos \theta_r \quad (2)$$

$$y_r = v_r \sin \theta_r \quad (3)$$

$$v_r = \alpha_r \quad (4)$$

$$\theta_r = w_r \quad (5)$$

where α_r is the robot's linear acceleration and v_r, w_r are the linear and angular velocities, respectively. θ_r and v_r are the control inputs of the mobile robot. The line-of-sight angle φ_{ir} which is obtained from the angle made by the line of sight l_{gr} is given by the following equations

$$\cos \varphi_{ir} = \frac{|x_g - x_r|}{\sqrt{(x_g - x_r)^2 + (y_g - y_r)^2}} \quad (6)$$

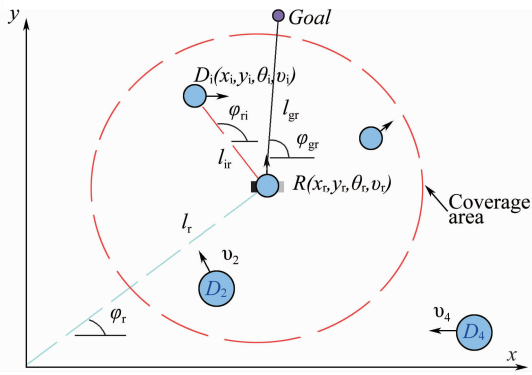


Fig.2 Geometry of the navigation problem.

In this paper, we suggest virtual planning method which is developed [4] for path planning for service mobile robot in library. This method is derived directly from the relative equations of motion of robot and dynamic obstacle.

The evolution of the range between the robot and obstacle for virtual planning method is given by following equations

$$l_{ir} = v_i \cos(\theta_i - \theta_{ir}) - v_r \cos(\theta_r - \varphi_{ir}) \quad (7)$$

$$l_{ir} \varphi_{ir} = v_i \sin(\theta_i - \theta_{ir}) - v_r \sin(\theta_r - \varphi_{ir}) \quad (8)$$

Above equation shows the tangential component of the relative velocity. We can see the proof of the equations for the tangential and the normal components of the relative velocity from thereference [5]. A negative sign of l_{ir} indicates that the robot is approaching from obstacle D . If a zero rate, range implies constant distance between the robot and obstacle. The system presents a nice and simple model that allows real time representation of the relative motion between robot and obstacle.

B. Mobile Robot Kinematics

The kinematic configurations of the mobile obstacle are given by

$$x_i = v_i \cos \theta_i \quad (9)$$

$$y_i = v_i \sin \theta_i \quad (10)$$

$$\theta_i = w_i \quad (11)$$

The distance between the robot R and the obstacle D is given by

$$l_{ir} = \sqrt{(y_i - y_r)^2 + (x_i - x_r)^2} \quad (12)$$

where l_{ir} is distance between robot R and obstacle D . The line-of-sight angle φ_{ir} is the angle which is made by the line of sight l_{ir} and it is given by

$$\cos \varphi_{ir} = \frac{|D_x - R_x|}{\sqrt{(D_x - R_x)^2 + (D_y - R_y)^2}} \quad (13)$$

$$\sin \varphi_{ir} = \frac{|D_y - R_y|}{\sqrt{(D_x - R_x)^2 + (D_y - R_y)^2}} \quad (14)$$

C. Navigation Process

Kinematic-based linear navigation laws are used to navigate the robot toward the final goal [7]. A linear navigation law is given by

$$\theta_r = M \gamma_{gr} + c_1 + c_0 e^{-at} \quad (15)$$

where θ_r is direction of mobile robot, M is navigation parameter, γ_{gr} is angle of line of sight, c_0 and c_1 are direction terms and a is given constant for heading regulation.

Fig. 3 shows simulation result of trajectory planning for the mobile robot motion from a start point to the final point. We can see changing navigation parameters give us different trajectories to follow for mobile robot.

Fig. 4 shows simulation results for instant path planning while robot moves in a dynamic environment with moving obstacles.

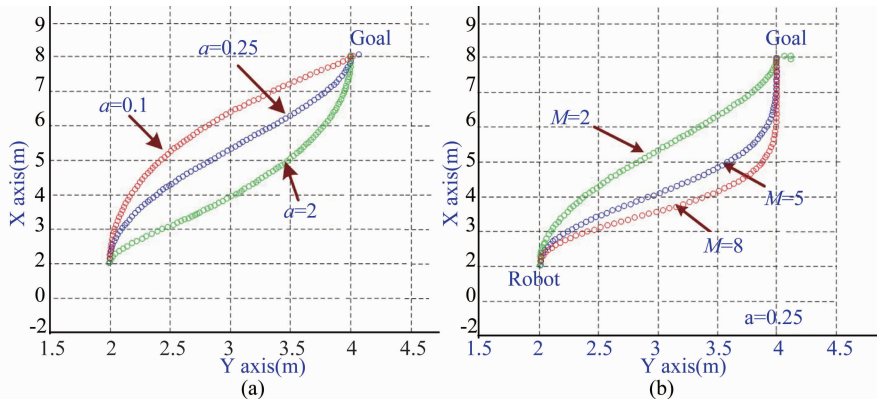


Fig. 3 An illustration of the used approach, where the initial configuration is satisfied by the choice of the control law parameters

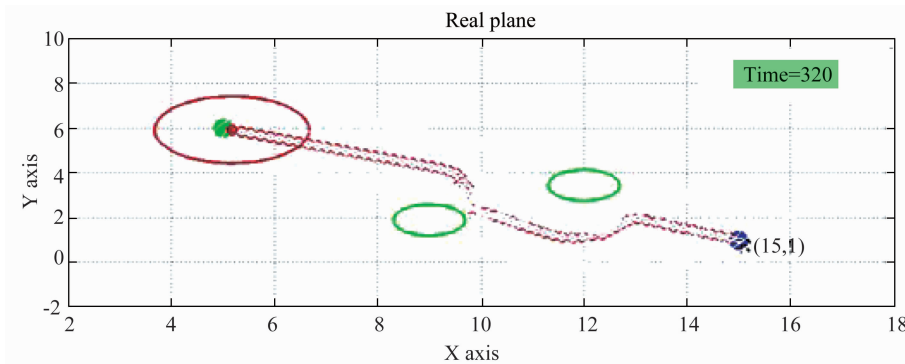


Fig. 4 Collision avoidance using the robot-transformed approach versus obstacle-transformed approach

III. IMPLEMENTATION OF THE KALMAN FILTERING

We use the Kalman filter that is used for estimating position and orientation of the robot to combine sensors information. In our used method, we combine the rotation encoder data with camera data, using the error model method to estimate more reliable position. Fusion of the odometer sensor and camera information, and more accurate position estimation can be acquired. Fig. 5 shows the position (x_c, y_c) estimated by camera combined with position (x_e, y_e) and angle (θ_e) measured by odometer.

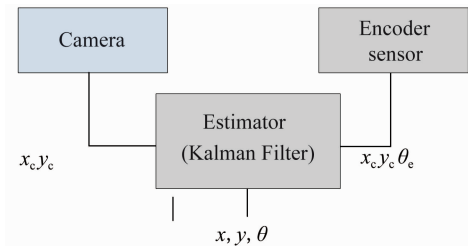


Fig. 5 Implementation of Kalman filter

From the position estimator, we get the new estimated position for the mobile robot after fusion of the camera and encoder's information through the Kalman filter, as x, y, θ . Using the above error models, we design the indirect feedback Kalman filter as the following state equations of the system

$$x(k+1) = A(k)x(k) + w(k) \quad (16)$$

where $v(k)$ is measurement noise. It is assumed that $w(k)$ and $v(k)$ are zero-mean Gaussian white noise sequences. $A(k)$ are system matrices [3].

IV. ROBOT DESIGN AND ITS APPLICATION

We design the training robot that is a mobile mini-robot, completely programmable in C and especially developed for educational purpose at the electronics department of school of information and communications technology, Mongolian university of science and technology (MUST), Mongolia. Assembly is simple for experienced

electronic technicians and feasible for a novice. As for assembly, only our prepared printed circuit boards (PCB) and standard parts used as shown in Fig. 1. For programming, we use freeware only. Therefore, training robot is suitable tool or product for the introduction of processor-controlled amateur electronics, projects in schools and universities, studies and adult education centres. Special tools, which are freeware for private users, have been used for all electronic development phases and software design, proving how robots can be designed without using expensive tools or machines Fig. 6.



Fig. 6 Robot implementation

Training robot is equipped with a RISC-processor and two independently controlled motors, four status LED, four photo resistor sensors for line tracer, two collision-detector switches, two odometer-sensors, keyboard controller and a wireless communication set for programming and remote control by a PC, see Fig. 7.

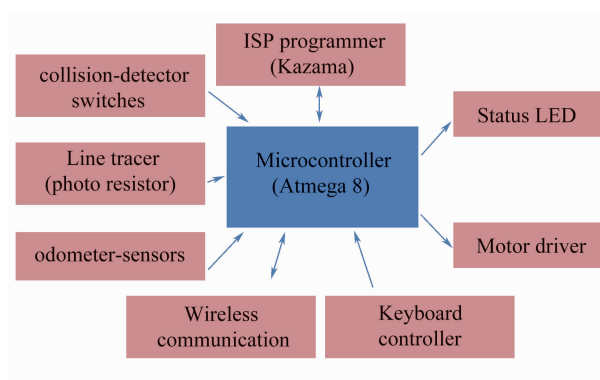


Fig. 7 Control system's block diagram of training robot

Fig. 8 shows the real simulation results that combine our robot navigation system with path planning simulation based on concept of virtual space using Matlab.

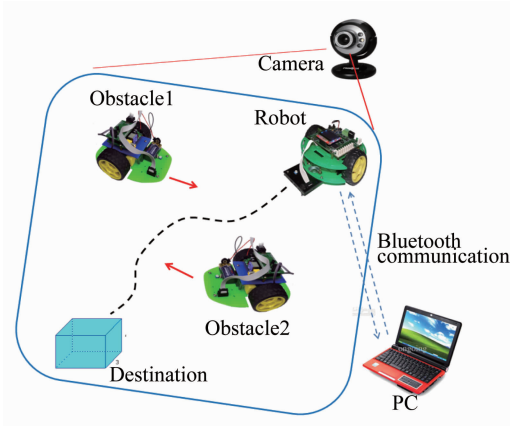


Fig. 8. Implementation of control system

V. CONCLUSIONS

In this paper, we suggested the path planning algorithm based on concept of virtual space for collision detect and avoid obstacles in dynamic environment. The notion of the virtual plane can be combined with various classical methods for path planning and navigation in dynamic environments. We fast measuring method using a sensor's fusion to correct position errors in robust localization scheme and found the sensor fusion approach is better than single sensor approach. Also, we designed our training robot whose name is LION robot and combine this robot navigation system with path planning simulation based on concept of virtual space using Matlab. Then we obtained good results of real experiments.

VI. ACKNOWLEDGMENT

This work was supported by project of school of information and communications technology, Mongolian university of science and technology.

REFERENCES

- [1] O. Hachour, "Path planning of autonomous mobile robot," *International Journal of Systems Applications, Engineering & Development.*, vol. 2, pp. 178-190, Apr. 2008.
- [2] Z. Liu, J. Zhao, L. Zhang, G. Chen, and D. Li, "Realization of mobile robot trajectory tracking control based on interpolation," *2009 IEEE ISIE*, Seoul, 2009, pp. 648-651.
- [3] I. Zunaidi, N. Kato, Y. Nomura, and H. Matsui, "Positioning system for 4-wheel mobile robot: encoder, gyro and accelerometer data fusion with error model method," *Chiang Mai Univ (CMU) Int J.*, vol. 5, pp. 1-14, 2006.
- [4] F. Belkhouche, "Reactive path planning in a dynamic environment," *IEEE T Robot.*, vol. 25, pp. 902-911, Aug. 2009.
- [5] F. Belkhouche, and B. Belkhouche, "Wheeled mobile robot navigation using proportional navigation" *Adv Robotics.*, vol. 21, pp. 395-420 Apr. 2007.
- [6] F. Belkhouche, T. Jin, and C. H. Sung, "Plane collision course detection between moving objects by transformation of coordinates" *Control Applications, 2008, CCA 2008 IEEE International Conference on.*, pp. 1013-1018, Sep. 2008.
- [7] T. Khurelbaatar, T. Amartuvshin, and D. Enkhzul, "Control design and implementation of mobile robot in the dynamic environment" *Journal of Advanced Information Technology.*, vol. 2, pp. 35-43. Apr. 2012.

Reform of Cultural studies curriculum in MUST

Narantsatsral. D^{#1}

[#]*Department of Humanity, School of Business administration and Humanity,
Mongolian University of Science and Technology Ulaanbaatar, Mongolia*

¹*narantsatsral5@yahoo.com*

Abstract— This article shows the results of analysis of cultural studies academic program taught Mongolian Universities. We are introduced to innovate and improve outcomes about curriculum of Introduction to Cultural Studies. Before other Mongolian Universities need to change the scientific study of Introduction to Cultural studies program in the current conditions of globalization is just one of problems. We assume articles introduced gateway postmodernist, de constructionist theories and approaches to innovate a curriculum to publish textbook conform to the needs and handbook. And we viewed with the same update gateway that inside the many countries in US, in Europe, Asia and Africa, as a definition of Cultural studies.

Keywords— Cultural studies, postmodernist, new theories, Mongolian Universities, analyze

I . INTRODUCTION

The notion of culture, with a wide range of academic study, because surely it should be studied and appropriate position and field of the social sciences. Our researchers understand the culture represented by the culture. " Britannica " dictionary; culture is defined as is wisdom in considering the role of social institutions. This intelligent concept is similar to literary criticism, philosophy, sociology, anthropology, history records, and value. Culturology is the Eastern European, cultural studies can be a term used in Western Europe. It shows that these countries every culture qualification prepares professionals understand that the study narrowed. Development and research of cultural studies in the Mongolian and developing experienced more than less time. But what are the major objects of cultural researchers? Where the theoretical and methodological sources? What is a study of the angle of view, our research feature? As such issues can still be sufficiently well defined. We found each other that three different " cultural studies " in the development of the program. These are the first of the Russian and then culturology in our culture emerging in modern-Soviet Russian science of translating that study. Researchers often have to translate that science of culturology (see the Radugin 1999, Tumurbaatar and Tumorhuyag 2001, Gurevich 2004 and Dorjdagva 2010). Secondly, the US anthropologist Leslie White, a trend of anthropological research preached about studying any culture called culturology. According to read the works he published in 1949, Culture, Science; human culture and civilization studies, (see the Science of Culture; A Study of Man and Civilization) he mentioned that some studies are made on the

basis of a science. His culturology as the name has been studied as a methodology and theoretical approach of the culture (see the White, 1949). Thirdly, clultural studies in America and Europe. Origin of the Sciences in 1964, counting from the established Center for Contemporary Cultural Studies University of Birmingham. Some authors have recently become a popular addition to the use of the the term rather culturology translated into English in their professional and research areas (see the Monk-Erdene, 2014). These conceptions are clearly many for something different, such as the origin of history, theory, methodology, and angle to see the objects and issues. Great high school curriculum and Russian researchers are the same with relative culturology but are used in any other science the term of Cultural studies. And the origin appears not believe that in uncoordinated US anthropologist Leslie White basis set research trends and the scientific basis for any such item is the origin of the Russian-culturology study of Culture. Which of the three studies on culture what we have taught university students? Any consideration of what are we doing? Which direction would be closer to the world standards and hold what? Will that should answer the first question.

II . CULTURAL STUDIES IN MUST

" Cultural Studies " major opens the 1998 – 1999 of the school year and are taught the basic education and specialized courses. Over 16 years, " Cultural Studies " 10 times prepared in bachelor's degree 146 specialists, master's degree 40, doctor's degree 12 career specialists graduated. Employment of graduates are working 100% , is a graduate of 44. 4% in the profession of which, 55. 6% in other sectors. Now bachelor 25 students and 1 M. Sc and 2 Ph. D study.

Cultural studies academic programs specialized units on MUST Engineering undergraduate studies and was selected since 1998. Cultural Studies academic programs conducted 16 years experience in engineering schools. Every semester about 500 bachelor students selected this courses. But the Cultural Studies academic program was defined just general overview on culturology in recent years, cultural history seems program is taught for a long time. Therefore, the upgrade program has encountered a problem we need. To develop from government common requirements adopted policy, the Ministry of Education towards the higher education program, MUST's mission and policies, international engineering education CDIO system, liberal arts education as the quality of edu-

education and programs in accordance with the international standard, such for Cultural Studies academic program is to develop curricula. Social and cultural rights to develop the current program, a common form MUST and Mongolian all Universities not only in freedom, we need publish sample textbook, it's Cultural studies new theory of post-modernist, de construction and approach. The key to initiate the introduction of research and training because we cultural studies program is to develop a critical need, validity and importance to society.

III. ANALYZE OF CULTURAL STUDIES ACADEMIC PROGRAMS

The United States, such as the Stanford academic program called Cultural Studies in Stanford University, Victoria University, Georgia State University and the Association of Cultural Studies, Canada, England explained quite a few university cultural studies comparative research focus of academic programs as well as research organizations and professional associations. We are selected Mongolian some universities:

- ① National University of Mongolia;
- ② Mongolian University of Science and Technology;
- ③ University of Humanities;
- ④ University of Law enforcement;
- ⑤ University of Culture and Art;
- ⑥ Institute of Chingis Khan.

Cultural Studies Introduction course credit hours and tutorials taught in schools is shown by Table I below. Many schools have time for 2 considered more credit 16/32 lectures.

Table I
ACADEMIC CREDIT FOR EACH SCHOOL

No	Universities	Credit	Lecture / Seminar Hour
1	National University of Mongolia	2	Lecture 16 hour, Seminar 32 цаг
2	University of Humanities	2	Lecture 30 hour
3	MUST	2	Lecture 16 hour, Seminar 32 hour
4	University of Law enforcement	2	Lecture 32 hour, Seminar 32 hour Practice and Laboratory 32 hour
5	Institute of Chingis Khan	3	Lecture 32 hour, Seminar 32 hour
6	University of Culture and Art	2	Lecture 32 hour

However, Table II by following, shown by comparing the difference and the same aspects of high education

goals. It was in some schools aim to compare developed more because of the state party slogans like, abstract propaganda as a given school and a mind completely understood the program, rather than academic high school goals.

Table II
DIFFERENCES AND SIMILARITIES IN ACADEMIC PURPOSES

No	Universities	Differences	Similarities
1	NUM	People and social, environmental, scientific, ethnic and language related factors of culture	1. The importance of cultural research 2. Problems of human and social cultural relations 3. The national culture and development 4. Universal culture and anthropology 5. Function of culture 6. Mongolian tangible and intangible cultural heritage
2	University of Humanities	Phenomenon and cultural definition, relations between speech and Culture	
3	MUST	People and social, environmental, scientific, ethnic and language related factors of culture	
4	University of Law enforcement	Mongolian tradition, the ethnic characteristics, people prefer the wicked and the concept of decoration, climate and traditional knowledge	
5	Institute of Chingis Khan	Civilization and cultural anthropology	
6	University of Culture and Art	Modern culture, influencing factors and cultural heritage of tradition and innovation	

Table III by the content is shown when comparing the academic content of different or the same side with some of the top universities Mongolian culture rather than science content of academic studies or ethnic studies like this are beyond the content of the lessons that we tell the world and a too-grade localized and isolated.

Table III
DIFFERENCES AND SIMILARITIES IN SUBJECT MATTER

No	Universities	Differences	Similarities
1	NUM	Cultural history, the famous concept of cultural agents are written and communication skills	Basic understanding of the culture and the provision of culture and nature, structure, material, and intellectual cultural commonalities and unique situation, development patterns
2	University of Humanities	Explore the importance of the concept of culture in the social and cultural development	
3	MUST	Provide comprehensive knowledge of the culture	
4	University of Law enforcement	Nomadic culture, civilization and Mongolian image and hold national trends	
5	Institute of Chingis Khan	Independent study of scientifically concept of culture	
6	University of Culture and Art	Humanities education to cultivate the human personality	

Table IV shows that compared different and similar aspects of student awareness shown is the national culture, cultural theory and social relations and cultural awareness in addition to aesthetics, ethics, religion eclectic mix science and ethnography science. Because professional and non-professional lecturers teaching their point of view.

Table IV
SIMILARITIES OF KNOWLEDGE FOR STUDENTS

No	Universities	Differences	Similarities
1	NUM	General theory of knowledge and cultural history	1. Possession cultural functions of Public Relations 2. To express own freely position 3. Human development consists of Culture and How to join and build
2	University of Humanities	Aesthetics, ethics and the legal sense, orientation	
3	MUST	Human nature, society, technology, religion, and identity practices have a communicator	
4	University of Law enforcement	Access to a national concept of racial justice any issue and submit order	

Table IV Continued

No	Universities	Differences	Similarities
5	Institute of Chingis Khan	To study the presentations and in litigation with knowledge of the culture	1. Possession cultural functions of Public Relations 2. To express own freely position 3. Human development consists of Culture and How to join and build
6	University of Culture and Art	The cultural concept of ethnic nationalities	

Similarities to provide academic skills:

- ① working with people;
- ② Human Relations;
- ③ society to behave properly;
- ④ Differences of academic skills;
- ⑤ University of Culture and Art-to understand and learn the basic steps of the world for the many country's cultural development;
- ⑥ University of Law Enforcement-Mongolian culture and recognize the historical conditions developed and be proud of him.

But Table V is Introduction to Cultural Studies academic programs indicate Mongolian universities that summarizes the US University of Georgia State University's academic program of Introduction to Cultural Studies. Such comparison, the difference came in many other things, how to solve this difference? To see that respondents in the opinion questions the article.

Table V
COMPARISON OF CURRICULUM "INTRODUCTION TO CULTURAL STUDIES"

Mongolian Universities	Georgia State University
Lecture I; The concept of cultural understanding, development and significance	Lecture I ;Introduction
Lecture II : Cultural consideration attitudes and cultural model, cultural functions and universal culture and cultural anthropology	Lecture II : The Politics of CultureL (Barbie Nation Film) Lecture III : Culture and Power Graeme Tuner, British Cultural Studies: An Introduction; Chapter 1 Karl Marx
Lecture III : Primitive culture and ancient religion forms	Lecture IV ; Hegemony and Resistance Antonio Gramsci, " Hegemony, Intellectuals and the State ", and Stuart Hall, " Encoding/Decoding" Lecture V ; Cynicism and Utopia Fredric Jameson, " Reification and Utopia in Mass Culture " and Thomas Frank, " New Consensus for Old "

Table V Continued

Mongolian Universities	Georgia State University
Lecture IV: Medieval culture Meso America	Lecture VI: Audience and Gender Janice Radway, Reading the Romance
	Lecture VII: Postfeminisms Susan Douglas, The Rise of Enlightened Sexism
Lecture V: XVIII century Western European sociopolitical and religious life, new ideas	Lecture VIII: Communicative Capitalism Jodi Dean, Democracy and Other Neoliberal Fantasies
	Lecture IX: Fantasy Stephen Duncombe, Dream: Reimagining Progressive Politics in an Age of Fantasy
	Lecture X: Play Jane Mc Gonigal, Reality Is Broken: Why Games Make Us Better and How They Can Change the World
Lecture VI: Byzantine culture	Lecture XI: Science Bruno Latour, On the Modern Cult of the Factish Gods
Lecture VII: The XXth century theoretical disputes arising among the Western European culture and cultural crisis	Lecture XII: Myth Janice Hocker Rushing & Thomas S. Frenzt, Projecting the Shadow: The Cyborg Hero in American Film
	Lecture XIII: Mysticism Jeffrey Kripal, Mutants and Mystics: Science Fiction, Superhero Comics and the Paranormal
Lecture VIII: XX century Mongolian culture	Lecture XIV: Research report
Lecture IX: Human and social problems of cultural relations	Lecture XV: Research report

Although Mongolian universities lecturers teach each with their own different curriculum, they are talking about how to submit a standard curriculum taught science courses, aligned to different content is to change the academic program teams. Some lecturers teach specify the form of classes, such as historical and cultural philosophy and ethnography. 9 topics shown in the table above can represent all the universities in Mongolia. Table 5 only apply to universities in the 15 topics of the University of Georgia on the other side would like to emphasize that Canada, end duplication of some US and British universities. But the difference between "Introduction of cultural studies" course topics. University as the main patron students must be taught by a good teacher quality class should be no difference exercise topics and books. The university that offers a course to students and other teachers, different and interesting programs either because they are taught, and same topics, and lecturers are compete each other. But this does

not mean to teach anything to anyone to teach within the already published and researched the topic.

Main problem is inconsistent with grade teaching academic content of the Mongolian state university by Western universities. Why should comply with it? Almost as popular science concept should serve the mind of the world's universities, research institutions and science. One of the science should be our main concepts specified class universities of the world agree. If the core content, our concepts is dangerous little to establish the extent of the inconsistency, social and humanities as opposed to matter each other. Specific social and humanitarian sciences that are able to discover the common law has served frequently as science for centuries, probably like. So this feature to be carefully pondered, the world will not delay feet walking together fun social theory, will be misleading. In particular, researchers studying a mission, like the collapse of Mongolian socialist system following social and humanitarian sciences theoretical yield, new gates, current conditions and culture in search of development consistent with the world-class science of mind, and they will need to focus on that face us.

This perspective took considers students enrolled in the Western Cultural Studies founders as Richard Hoggart, creation of Raymond Williams, Stuart Hall has graduated from school, not a treatise. Some may be that our teachers are half-baked knowledge based on studies of their cultural studies. A simple example, when someone Mongolian culture are both studying in a university research career a student was moved to the school career as a foreign recognized a few of the lessons of a student, considered a credit. And what exactly does the opposite when we were walking a student at the University of Europe, to learn this profession? So if we Curriculum of cultural studies to introduce the combination of more efficient with every new legal adoption of emerging research topics in the industry, in addition to side change compared with Western cultural studies and their stars scientific and methodological and theoretical approach.

IV. CONCLUSIONS

Mongolian national universities with the same theory by already logged culturology of concern, which is revised curriculum of cultural studies? Because we already have mentioned, L. White's culturology studies of XIX century social evolutionary tendency of the theory itself from its strong critical, since published many research articles about the lack of any errors of modern social and humanitarian studies so that is already insignificant. Soviet countries has also improved the discretion side replaced by soviet-style culturology-western cultural studies and compensate for your shortcomings. We have reason that these standards do improve their research approaches and theoretical methods to follow.

Once we have developed, "Introduction to Cultural Studies" program for Mongolian universities of liberal arts education system log. In addition, books translated textbooks and compromise Mongolian soil and knowl-

edge that students give our students even go landing elsewhere. Refer to Table VI of the revision, the proposed program.

Table VI
PROPOSALS TO CHANGE THE ACADEMIC OF "INTRODUCTION TO CULTURAL STUDIE"

Week	Topic	Lecture hour	Seminar hour
1	The objective of the cultural studies, research object, research methods and approaches.	2	2
2	The concept of culture, cultural characteristics and importance.	2	2
3	Theories about the origin of culture and cultural roles and functions. Cultural anthropology.	2	2
4	Ancient Western and Oriental cultures: (Egypt, Messopotami, China, India, Meso America, Grece and Roman culture).	2	2
5	Medieval culture: (Byzantine, Slav) Christian influence in European culture.	2	2
6	Renaissance culture Western global cultural new stage(a new flow of ideas)	2	2
7	Cultural crisis and trends in contemporary cultural theory	2	2
8	The Mongolian cultural policy: (1921-1940) , (1940-1990) (Cultural Rights in the Constitution)	2	2
9	The tangible and intangible cultural heritage	2	2
10	Communication skills and culture (Independent culture of human relations) , (individuals social cultural roles)	2	2
11	Popular culture and sub-culture, youth culture Multiculturalism	2	2
12	Urban Culture (urbanization and culture shock)	2	2
13	Daily life culture Leisure time	2	2
14	Media, representation and the cyber culture	2	2
15	Specific professional culture	2	2
16	Cultural freedom and globalization	2	2

And professional lecturers is the need for efforts to teach Introduction to Culture Studies are taught the academic quality and the latest knowledge and information theory to law-free entry to foreign countries Mongolian and Mongolian researchers have inhibited the specific conditions of outer space. Consequently, the situation is already to create the conditions foundations of Cultural Studies science discipline or order, established research sector has scientific fields likely, but not certain order and chaos, who teaches an amateur field. Any sane stone three pillars necessary for an independent science sector priorities of the scientific theory of history, objects, professional experts to find among its sciences building.

REFERENCES

- [1] Berry, Ellen E. болон Epstein, Mikhail N. " *Transcultural Experiments: Russian and American Models of Creative Communication*". New York: St. Martin's Press. 1999.
- [2] D. Bum-Ochir, G. Monk-Erdene. " *Theory of social and cultural anthropology*". Ulaanbaatar: Admon. 2012.
- [3] P. S. Guryevich. " *Culturology*". Ulaanbaatar: Admon. 2004.
- [4] T. Dorjdagva, N. Sarantuya, J. Dolgorsuren. " *Basic of culturology*". Ulaanbaatar: NUM Press. 1998.
- [5] T. Dorjdagva. " *Culturology*". Ulaanbaatar: Admon. 2010, 2012.
- [6] G. Monk-Erdene. " *Which cultural studies?: Needed to determine the field*" " Cultural studies sciences development trends: Mongolian and international practice" report. Ulaanbaatar: Mongolian University of Culture and Arts.
- [7] D. Narantsatsral, G. Monk-Erdene " *Curriculum of cultural studies*", National IX seminar of cultural studies. Mongolian association of cultural studies, NUM, MUST, Education University. 2015.
- [8] D. Narantsatsral, G. Monk-Erdene " *Change the curriculum of cultural studies*", Scientific meeting of the 2015. Professor compilation of scientific works. 2015.
- [9] D. Narantsatsral. " *Liberal art programs in the field of innovation: Cultural Studies*" and " Science and education reform and innovation-women" research conference. Freidrich Ebert Stiftung Foundation of Germany, Federation MWF, democracy, justice, the Mongolian Women's Federation. 2015.
- [10] A. A. Radugin. (translated by Kh. Lkhamdorj. P. Chojil). " *Culturology*". Ulaanbaatar: Mongolian University of Culture and Arts. 1999.
- [11] J. Tomorbaatar, N. Tomorkhuyag. " *Basic of culturology*". Ulaanbaatar: Bit service. 2001.
- [12] White, Leslie A. *The Science of Culture: A Study of Man and Civilization*. New York: Farrar, Straus and Cudahy. 1949.

Research on Dynamic Network Self-Configuration toward SDN

Zhong-Nan Zhao^{#1}, Jian Wang^{#2}

[#]*School of Computer Science and Technology, Harbin University of Science and Technology,
NO. 52 XueFu Road, Nangang Dist, Harbin, China*

¹*piconet@126.com*

²*wangjianlydia@163.com*

Abstract— Software Defined Network (SDN) is a kind of novel style network different from legacy OSI structure which makes the transmission and control logic decoupled, namely, control plane as well as data plane. With the development of SDN, variety of network function virtualization (NFV) appears, and in the same time brings an opportunity to survivability enhancement to study. In this paper, a technology of autonomic growth for survivability is proposed, and meanwhile the feasibility of self-configuration for cognitive network under SDN is discussed.

Keywords— SDN, NFV, survivability enhancement, self-configuration

I. INTRODUCTION

In recent years, Internet has become the current generic real-time computer network. With the popularity of computer network, various types of mission-critical system are gradually becoming more and more networked in order to achieve remote operation and enlarge application scale. However, there exists a considerable risk in the mission-critical system due to the openness and complexity of the network, the design defect of hardware and software systems and the inevitability of human errors and other factors. These systems, once destroyed, are sure to cause failure, and it will have a huge impact on the national economy and people's lives.

It is hoped that the software system, even if encountering an accident, an attack or a system failure, can continuously and stably provide critical services, for it is the only way to improve the survivability of the system. To meet the requirements of the growth of system survivability, the system needs to implement dynamic allocation in the structure, behavior, properties and other aspects of self-sensing module according to the perception of the external environment and the results of the feedback event handler during the operation, and autonomic regulation. As the new network architecture, SDN (Software Defined Networking) not only separates the controlling and forwarding, but also achieves a programmable centralized control. The core idea is to separate the control plane of network and the forwarding plane of data. That is to say, control power can be separated from network devices. By doing so, some related software platforms centralized in the controller layer can be utilized to flexibly distribute network resources according to the

needs by using the underlying hardware which is programmable. Besides, it can also turn resources into an IT service, which is able to be provided to customers by automated processes and software. So far, it has become the most popular research directions in the field of network in the world.

II. SOFTWARE DEFINED NETWORK

SDN was originated from the "Clean Slate" subject in Stanford University (one top American university). The subject was funded by the US project—GENI, aiming to improve the existing traditional network architecture. As soon as this new network architecture appeared, academic and business communities showed their enthusiasms. There is no doubt that SDN is one of the hottest technologies currently, for it has broad prospects and great research value. Besides, as a new implementation of network architecture, it has brought new vitality and challenges to traditional network.

A. The Definitions of SDN

Currently, SDN is divided into about two kinds. [1] Generalized SDN: referring to those open interfaces of application resource, and network architecture which can be implemented in software programming control. [2] Narrow SDN: referring to those that meet the requirements of ONF (Open Networking Foundation), and software-defined network which is standards-based OpenFlow protocol.

The basic idea of SDN is the traditional network architecture and network control network, forwarding tightly coupled relationship decoupling. As a result, they build the open, programmable network architecture. In simple terms, SDN can be separated from the control of the traditional network equipment out, and then passes control to the centralized controller management, without relying on the underlying network equipment (switches, routers, firewalls, etc.). It not only shielded from the bottom differences in a variety of network devices, and because of foreign control is completely open, so users can follow their own ideas and needs to define any network routing and transport policy rule you want to achieve, making the system more flexible and able to meet the intelligent user it needs.

B. The Profile of SDN

SDN is a kind of separation of data and control of software programming to achieve the new network architecture. Shown in Fig. 1, SDN is divided into three planes and two interfaces.

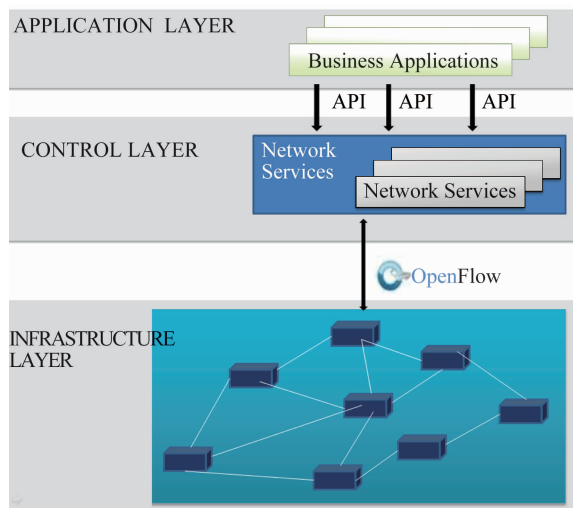


Fig. 1 The architecture of SDN

Infrastructure Layer: which includes some network elements, each network element can provide network traffic. Wherein the network device can be either a hardware switch, the virtual switch may also be, e. g. , OVS, of course, can be other physical devices, such as routers, firewalls and so on. All forwarding entries are stored in the network device. The user data packets are processed in this layer, forwarding.

Southbound interface: infrastructure located between plane and control plane, responsible for data exchange and interoperability SDN controller and network elements. OpenFlow is the most famous work in the south to the protocol interface.

Control Layer: this layer comprises a controller and a network operating system in two parts, of which the most important is the SDN controller. SDN network controller is the central component of the control network traffic. As an important task, it is responsible for processing the data plane resources, maintaining global view of the network and network devices in the network information collection. They form the strategies according to the demand generation strategy, send a stream under the policy table and then control the behavior of the network^[1]. A SDN network controller may have one or more, between multiple controllers can be master-slave relationship, it can be a peer relationship. One controller can control multiple devices. Of course, a piece of equipment can also be controlled under multiple controllers. Under normal circumstances, the controller is running on a separate server.

NBI: located between the control plane and application plane, the upper application gets the underlying interfaces to network resources through the north, and by

NBI to send data to the underlying network. The simplest, most traditional NBI is CLI, SNMP; the most popular NBI is REST API interface.

Application Layer: includes a variety of applications, open programming interfaces and network view provided by the control layer, allowing user-defined network control and network services logically by software. Application layer is to provide users with services, including load balancing, security, network performance monitoring, including congestion, delay and other network performance testing and management, topology discovery, and many other services. These services will eventually software applications presented in a way out, and they may be located on the same server with controller, it can also run on other servers to communicate between them with the controller via the communication protocol.

The basic characteristics of SDN include: standardization for centralized control and forwarding control after the separation, open standardized interfaces between the control planes and forwarding plane supports flexible software programming. Centralized network controller in a centralized manner in the various devices on the network control and management, always monitor the status of the entire network, making it possible to reduce the consumption of resources and the exchange of information transmission, and can be optimized for global control. Flexible software programmability, making automated management and control capacity of the network has been improved, can effectively solve the current scale of resources faced by network system expansion is limited, network flexibility is poor, it is difficult to quickly meet the demand for services and other issues.

C. Related Technologies

The core of SDN technology is OpenFlow. OpenFlow was first put forward by Professor Nick McKeown and other researchers at Stanford University. It was initially unable to overcome the innovation on Internet, the real network is simple, easy to achieve and produce, which SDN has a common goal, and thus it has been widely attention. OpenFlow is a network device specification by itself, this specification includes the network infrastructure layer forwarding device OpenFlow switch the functional requirements and the basic components, and it is responsible for the switch controller or remote control to a network in OpenFlow. OpenFlow controller cluster OpenFlow agreement. OpenFlow agreement refers to the switch controller and the communication between OpenFlow followed protocol. In OpenFlow network, the main task of a switch is to receive various instructions delivered by the controller, and then following the instructions to switch the incoming data stream which is processed and forwarded. Therefore, OpenFlow switch function can be divided into two parts, one is the data plane which is responsible for transmitting and dealing with the data stream, the other part of the control plane is responsible for communication, the reception controller send control messages to the controller, as convection table to be processed, etc., but also its status will switch

to upload information to the controller. OpenFlow controller is responsible for maintenance and monitoring of the entire network information, such as network topology, network resources, so as to analyze the processing and forwarding for a variety of data streams, and then send to the switch, according to analysis by the switch out of the policy processing and forwarding data stream.

III. SURVIVABILITY ENHANCEMENTS

Survivability enhancement provides the important research content as survivability, and it aims at surviving ability about changing system for exploration and research technology. Throughout the history of the development about computer security technology, the first-generation security technology mainly studies the method to prevent the main attack, such as firewalls, encryption technology. The second-generation security technology mainly studies how to successfully find technical intrusions, such as various intrusion detection technologies. With the expanding scope of the application of computer, all kinds of application system came into being, at the same time, the outside world for the attacks and intelligence systems are also increasingly diverse, which leads to the first-generation and second-generation security technology want to shut out the intruder security technology completely is no longer feasible, so the third-generation security technology appeared, it realized if computers want to avoid attacks, the invasion and failure is not completely possible, so professors do not pursue the absolute security system, they put focus on research how to ensure the completion of the core mandate of the system to minimize intrusion and malicious attacks on the system of negative damage. Survivability enhancement technology is generally taken as the third-generation security technology of security systems.

At present the most common enhancement technology is supported by diversity redundancy, threshold mode and isolation repair technology. And diversity redundancy includes redundancy resources, the redundancy method, temporal redundancy and redundancy. But the cost of redundancy is often too high, so the redundancy technology has not been widely used. Threshold mode has been widely recognized certification system survivability, enhancing storage system design, but the viability of the system is affected because asynchronous environments share secrets. Although the cost of isolation and remediation pay few, it cannot tolerate large-scale physical failures and natural disasters, which undoubtedly will affect the survival of the system enhancing. Depending on the different applications, a number of security technologies can also be suitably applied to enhance survivability studies, such as encryption, digital signatures, authentication and other traditional security technologies.

IV. SELF-CONFIGURATION TECHNOLOGIES FOR SUR ENHANCEMENT

To meet the system requirements for enhanced survivability paper borrowed from the ideological configuration, the system has the ability to self-configuration, and the system can ensure that in the absence of human intervention in case of intervention or less under good run. On the basis of a comprehensive study on the SDN and survivability enhancement technology, it is proposed based SDN survivability enhancement self-configuration technology, and there are two key research directions: network resource virtualization and system self-configuration technology.

A. SDN/NFV

To meet the system requirements for enhanced survivability paper borrowed from the ideological configuration, the system has the ability to self-configuration which can ensure that in the absence of human intervention in case of intervention or less under good run. On the basis of a comprehensive study on the SDN and survivability enhancement technology, it is proposed based SDN survivability enhancement self-configuration technology. There are the two key research directions: network resource virtualization and system self-configuration technology.

Network virtualization technology is similar to storage virtualization and server virtualization, for the underlying physical network resources abstracted into a resource pool to allow multiple virtual networks to dynamically access network resources, control network devices, network traffic management so that they can be used transparently between multiple virtual networks^[2]. In the SDN architecture, state of the switch, link information and network topology information can be obtained through a centralized controller, which is well supported virtualization technology in the development of abstract network resources network. Network virtualization technology can shield the underlying physical devices heterogeneous differences, so that a plurality of logical isolated virtual network can coexist on the same physical network, each virtual network can use your own protocol architecture, and not each other influence, and capable of rational allocation of the entire network node and link resources based on dynamically changing user needs, support a variety of network architectures, network protocols, network systems to run simultaneously, effectively improve the utilization of resources.

Virtualization of network resources is done through a corresponding virtualization technology. The first valid network resource should be virtualized, so that resources can be reasonably allocated. The critical core issue of Virtual network mapping is to set up a reasonable correspondence between the underlying physical topology node resources and physical paths and virtual network virtual nodes and virtual link.

Traditional network virtualization solutions are on the physical network by adding multiple virtual nodes and virtual links to either build an entire virtual network topology, under this scheme, network management need to manage each network node, which makes network management has become extremely complex. If the entire network as a router can reduce this complexity, this virtual network service providers will no longer need to manage the entire network will be able to deploy the services it provides, but also to make the physical network platform provides a more user-friendly use. SDN virtualization technology is mainly used to obtain information and the state of the entire network through a centralized controller, and then use this information to network resources abstracted, and then can abstract over resources division and isolation, and can finally take advantage of the unique SDN programmatic definition of the control logic, control logic defined in accordance with the use of virtual networking resources. SDN-based network resource virtualization first needed to solve the network resource virtualization, network virtualization, followed by the distribution of resources and, finally, monitoring and adjustment of network resources.

Since SDN network equipment has the good network programmability, network management and network studies the human eye can easily control network devices, deploy new network protocol. SDN network control plane and data plane are separated, allowing users to define their own virtual network, define their own rules and network control strategy, network service providers to provide users with end-controlled network services, even in the hardware adding new applications directly on the device. This programmable network platform not only to unlock the software and network link between a specific hardware, but also the intelligent network software and hardware are fully integrated high speed, making the network smarter and flexible.

B. Self-Configuration on SDN

Because of the complexity of software systems goes on, its function is also becoming stronger and stronger. However, the system is also more complex to software configuration, and traditional software configuration is done manually, which is time-consuming and laborious. Besides, the system's credibility is also not high. By giving the software the ability to self-configure, the computer may be enabled to independently allocate the resource configuration and components, improving the credibility of the system, which can solve the traditional software configuration flaws.

(1) Autonomic Computing and Cognitive Network

With the development of technological advancements, it not only makes computer network become more open, but the scope of its application is also increasingly expanding. Gradually penetrated into people's daily life, people are more and more dependent on software systems continue to improve, with the number of users increases, customer service and safety needs of the application on top of software systems is also raising. Users

want their software used by the system in a timely manner to perceive the software system, the external environmental change, and can be configured to automatically adjust itself by components and resource allocation, allows the system to be able to provide normal service, in any cases necessary, to meet the users high demand for trusted systems. However, the traditional network architecture can only rely on manual configuration of the system to adjust, and the current software does not take the initiative and its own operating environment perceived behavioral state, so the traditional network architecture and cannot meet the high users' of the software trusted demand. In order to meet users' demands for high-trusted software systems, we need to improve the current traditional network architecture. The introduction of self-configuring autonomic computing ideology will add some attributes to the current self-regulatory system in the network, which makes the software system can perceive the system internal and external environment, and internal systems in real-time dynamic configuration, and "with technical management technology" smart idea

IBM's autonomic computing thought was first proposed in 2001, and the idea was inspired by the main body of complex autonomic nervous system. Autonomic computing refers to the system of the entire network system hardware resources dynamically and proactively adjusted according to the needs of the dynamic changes within and outside the system, in order to manage and improve its own. That is autonomic computing is an emerging approach to systems management, implementation of the system can be realized even in the case of a small amount without human intervention, "self-configuring" "self-optimization" "self-healing" "self-protection", and ultimately "with technology management", the idea.

Cognitive network is a natural extension of autonomic computing from single or cluster level to the level of the network, which will be introduced into the network of biological self-regulatory level. Furthermore, the system is designed to achieve a high level of brain cognitive intervention.

(2) Status of Researches

Self-configuring is one of the most important attributes in autonomic computing the main goal of which is to be able to customize the system to configure the system to achieve autonomic computing in autonomy, that is to say, it also requires the system to be able to in the case of little or no intervention, automatically complete a variety of such repair, optimization, protection and other activities. For the current study self-configuring autonomic computing has begun to take shape. For example, literature [3] proposed a dynamic self-configuration framework to address Qos traditional network users on the system cannot be directly applied to Qos such as integrated services and differentiated services, etc. Next generation network (NGN) in this issue, drawing on the concept of self-configuration, and this framework into the utility function and interrupt mechanism, which used to represent the user utility function Qos priority inter-

rupt mechanism is used to dynamically correct transmitted data packet process priority. As a result of active self-configuration feature for users, Qos issues falling under network congestion situation has also been improved accordingly. Literature [4] introduces a highly flexible component architecture, this architecture are mainly used in automotive control systems, architecture is a real-time self-configuration, it can be configured to support distributed dynamic context-aware behavior. The highly scalable architecture, through the appropriate parts of the software at the same time embed a plurality of different dynamic decision points, utilization of dynamic decision point rating generated automatically take the appropriate configuration actions, and can realize the current software operating status were evaluated.

Nizar Msadek, et al. presents an organic computing system with a reliable, scalable, fault-tolerant and self-configuring algorithm [5], which aims on the one hand in order to load the service on average a distribution node as a typical load balancing scheme, on the other hand, in order to add services with different levels of importance to the node, making it more important services on a more trusted nodes can be assigned to. Moreover, the algorithm further comprises a failure handling mechanism, making the system even in the case of a fault in the system can still continue to host service which gives a perception can dynamically load changes^[6], and dynamically customize the system to adjust the configuration parameters of the adaptive middleware configuration framework that is layered queuing network performance prediction model based on the guidance of search optimal resource allocation, allows the system to have the ability to self-regulate, able to adapt to changes in the environment, in order to meet user Qos. Literature [7] and P2P Mobile Agent technology combined while adding self-configuration feature, on this basis, the proposes of a system based on self-configuring architecture policy, the establishment of a BestPeer P2P system, the system can be based on policies independent configuration of various components, assemblies, and a set of self-configuring system assessment methods, examples of the evolution of the system can be configured to analyze the nature of the system.

(3) Feasibility Study

A self-configuring software system includes five aspects: external environmental factors, the managed resource, context-aware, self-configuring, and self-configuring policy managers. Among them, the managed resource by the passive configuration, self-configuring and self-configuration manager policy belongs to the configuration of the main activities of the initiator and executor, environmental awareness is a bridge connection between them. The main task of environmental perception is responsible for the underlying environmental information forwarded to the upper application, and it provides configuration motivation. Since the configuration manager to be responsible for collecting environmental information for analysis, planning, decision making, self-configuring finalize behavior based on self-configuring poli-

cies dynamically. Although currently self-configuration technology has several ways to achieve, all relevant fields are relative to them. The traditional self-configuring software technology has been unable to adapt to the current complex computing environments.

SDN will be removed from the control of network devices out, and can be centrally controlled by a user-defined program, not only can mask differences in the underlying physical device, and the controller according to the underlying network device status of resources in a timely manner and configuration management, and allocation of resources with the concept of system administrators will.

Based on the above discussion, SDN platform based network self-configuration is entirely feasible.

V. CONCLUSIONS

With the current rapidly development of network technology and the continuous expansion of the network, the computer network has been widely used in many areas of government, military, education, scientific research, commerce, etc., which the security of all kinds of application systems also face a very big risk, the traditional network architecture can not meet the current security needs and survivability requirements of the system. SDN appears just to meet the needs of users to adapt to the current development of computer networks, SDN simplify current network devices to optimize network structure, making the network more and more flexibility, speed of response has been great improve. These advantages SDN-based SDN people's attention more and more, making the SDN has been rapid development, but SDN is still in the early stages of development, its development process, there have been many challenges, so it restricted to a certain extent, SDN development of. SDN-depth study on the current development of the computer network will have a profound impact.

VI. ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (61403109) and the Scientific Research Fund of Heilongjiang Provincial Education Department (12541169).

REFERENCES

- [1] McKeown N, Anderson T, Balakrishnan H. OpenFlow: enabling innovation in campus networks. ACM SIGCOMM Computer Communication Review, 2008, 38(2): 69-74.
- [2] Chowdhury N M M K, Boutaba R. A survey of network virtualization. Computer Networks, 2010, 54(5): 862-876.
- [3] Feng G, Wang H, Li B, et al. Dynamic self-configuration of user QoS for next generation network. Proceedings of the 2009 Sixth IFIP International Con-

- ference on Network and Parallel Computing. Gold Coast, Australia, 2009 :80-85.
- [4] Anthony R, Pelc M, Ward P, et al. A run-time configurable software architecture for self-managing systems. Proceedings of the 2008 International Conference on Autonomic Computing. Chicago, IL, USA, 2008 :207-208.
- [5] Nizar Msadek, Rolf Kiefhaber, Theo Ungerer. A trustworthy, fault-tolerant and scalable self-configuration algorithm for Organic Computing systems. Journal of Systems Architecture, 2015, 61 :511-519.
- [6] Hu Jianjun, the official Dutch Shuqing. Based on the performance of the model self-configuring middleware framework. Software, 2007, 9 :2117-2129.
- [7] LinYuming, Caiguo Yong, Li gifted. Based on Modeling and Analysis form self-configuring system policy. Computer and Modernization, 2008(2) :63-66.

Research on Spherical Beamforming Algorithm for Sound Source Localization

Yuechan Liu^{#1}, Zhen Zhou^{#2}, Chao Sun^{#3}

[#]The Higher Educational Key Laboratory for Measuring & Control Technology and Instrumentations of Heilongjiang, Harbin University of Science And Technology

No. 52, XueFu Road Nangang Dist, Harbin, China

¹liyuechan@hrbust.edu.cn

²zhzh49@126.com

³sc13579@126.com

Abstract— A high-resolution and high precision spherical array beamforming method for sound sources localization based on virtual sources algorithm is proposed. By means of the spherical array beam scanning, the focused spatial spectrum of actual sound sources can be obtained and each scanning point is assumed to be the virtual sound source. The actual sources focused spectrum is the result of focused beam output by the entire virtual sources, and then high precision sound source localization can be realized by calculating the contribution of each virtual source to the sound field. Compared with the conventional methods, the performance of the proposed method under some influence factors, such as frequency, array aperture, sound field mode orders are analyzed. The simulation results show that this method is not limited by frequency and aperture of array, effectively reducing the spatial aliasing. The focused spatial spectrum indicates the higher resolution sound sources location and the stronger ability to suppress the fluctuation of the background noise.

Keywords— Spherical array beamforming, high resolution, sound source localization, virtual sound source

I. INTRODUCTION

Focused beamforming array signal processing technique is widely applied to the radiation noise detection and positioning of the aircrafts, automobiles and ships due to its good tolerance and simplicity. However, regular array is tended to achieve the blind scanning area. The spherical array can be widely used for sound field analysis and sound sources localization in three dimensions space for its symmetry and rotation.

The spherical array beamforming was first presented by Meyer and Elko [1] to [3], and the sound field is decomposed into spherical harmonic function including the scattering sound field. And then, the sound sources can be located by using the orthogonality of the spherical harmonic function in three-dimensional space. Li, et al studied the design and optimization of array distributions based on the genetic algorithm [4], [5], which make the reduction of the side lobe in focussing spectrum. the undistorted steady focussing spectrum can be obtained through quadratic constraint on the weighted coefficient by using the least mean square algorithm (LMS). However, the amount of calculation is increased

for the iterative optimization algorithm, and its performance is influenced by the number of array element, array aperture, sound field decomposition mode order, and the source frequency, resulting in the serious confusion, low resolution and unachievable accurate positioning for multiple sound sources [6].

A high-resolution and high precision spherical array beamforming method for sound sources localization based on virtual sources algorithm is proposed, which utilize the theory of conventional spherical beamforming and the structure and sound properties of rigid spherical array. The performance of array aperture, sound frequency and sound field modal order is analyzed. Results of simulation show that the method proposed can effectively suppress the confusing and obtain higher resolution, positioning accuracy and noise suppression capability. The positioning performance is not affected by the frequency and the influence of array aperture, which can be applied to detect the underwater low-frequency noise sources.

II. VIRTUAL SOURCES AOSHERICAL ARRAY FOCUSED BEAMFORMING

In the direction (θ, φ) , the weight coefficient can be represented as

$$W = [w_1, w_2, \dots, w_l, \dots, w_L]^T \quad (1)$$

In [4], $w_l(\theta, \varphi)$ is derived as following

$$w_l(\theta, \varphi) = \sum_{n=0}^N \frac{1}{2i^n b_n(ka)} \sum_{m=-n}^n Y_n^m(\theta, \varphi) Y_n^{m*}(\theta_l, \varphi_l) \quad (2)$$

After completescanning, the output of spherical Fourier transform focused beamforming (SFTFB) can be written as

$$F_{SFTFB} = W^H P \quad (3)$$

The maximum value from the focussing spectral means the direction of the sound source. The SFTFB algorithm has the advantages of easier implement and simple calculation process, but has the drawbacks such as low spatial resolution, bigger sidelobe, false source in multiple sound source localization and sensitivity to environment. In order to overcome these problems, one need to utilize the correction factor and optimize the microphones layout [4], which increase the complexity of calculation. In view of this,

Rafaely and Li, et al applied distortionless constraints and robustness constraints to the conventional spherical beamforming[3],[8], transforming the SFTFB problem into a quadratic constraint spherical focused beamforming(QCSFB) optimization problem.

Minimum

$$W^H P W \quad (4)$$

Distortionless constraints

$$W^H V(\theta, \varphi) = 1 \quad (5)$$

Robust constraint conditions

$$\|W\|^2 \leq T_0 \quad (6)$$

The virtual sources method assumes that the sound sources exist in the scanning direction except the real sound source. The vector $X = [x_1, x_2, \dots, x_i, \dots, x_M]^T$ is set to the sound source vector in the scanning plane, where x_i is the amplitude of the i th virtual sound source, r_i is the distance to the array for each scan point, θ_i is pitching angle, φ_i is azimuth angle. The received sound field of l th microphone can be written as

$$\begin{aligned} p_l(ka, \theta_l, \varphi_l) &= \sum_{i=1}^M x_i \sum_{n=0}^N 4\pi i^n b_n(ka) \\ &\quad \sum_{m=-n}^{m=n} Y_n^{m*}(\theta_i, \varphi_i) Y_n^m(\theta_l, \varphi_l) \\ &= \sum_{i=1}^M x_i B_{il} \end{aligned} \quad (7)$$

$$B_{il} = \sum_{n=0}^N 4\pi i^n b_n(ka) \sum_{m=-n}^{m=n} Y_n^{m*}(\theta_i, \varphi_i) Y_n^m(\theta_l, \varphi_l) \quad (8)$$

So we have focusing output in the scanning direction (θ, φ) from(3) and(4)

$$\begin{aligned} F(\theta, \varphi) &= w_1(\theta, \varphi)p_1 + w_2(\theta, \varphi)p_2 + \dots + w_L(\theta, \varphi)p_L \\ &= \sum_{i=1}^M x_i (B_{i1}w_1(\theta, \varphi) + B_{i2}w_2(\theta, \varphi) + \dots + \\ &\quad B_{iL}w_L(\theta, \varphi)) \end{aligned} \quad (9)$$

Because of the virtual sound sources is no contribution to the sound field, the scanning results should be the same as the SFTFB result. The output at each scan point is equal to the sum of all virtual sources beamforming. The output matrix can be written as

$$F_{\text{SFTFB}} = D X \quad (10)$$

where M is the number of scanning points, and D is represented as

$$D = \begin{bmatrix} B_{11}\omega_1(\theta_1, \varphi_1) & B_{22}\omega_2(\theta_1, \varphi_1) & \dots & B_{ML}\omega_L(\theta_1, \varphi_1) \\ B_{11}\omega_1(\theta_2, \varphi_2) & B_{22}\omega_2(\theta_2, \varphi_2) & \dots & B_{ML}\omega_L(\theta_2, \varphi_2) \\ \vdots & \vdots & \ddots & \vdots \\ B_{11}\omega_1(\theta_M, \varphi_M) & B_{22}\omega_2(\theta_M, \varphi_M) & \dots & B_{ML}\omega_L(\theta_M, \varphi_M) \end{bmatrix} \quad (11)$$

So the amplitude of each virtual source is

$$X = D^{-1} F_{\text{SFTFB}} \quad (12)$$

The real sound source can be located by the scanning point which has the largest amplitude. This method is named virtual source spherical array focused beamforming method(VSSFB). The basic idea of this method is that the actual beamformer result is the contribution of all the virtual sound sources, and the amplitude of the sound sources

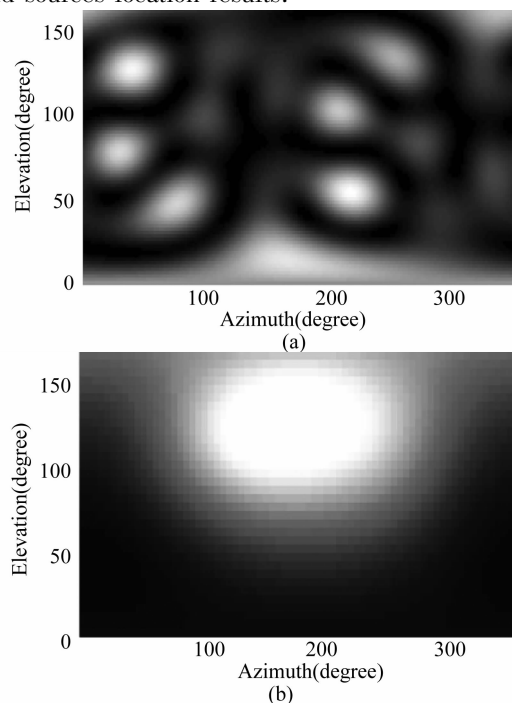
shows the contribution of the virtual point source to the sound field, which determines the position of the sound source. The method proposed not only has advantages of avoiding the complex process in solving the optimal weighting, simplifying the operation, but also improving the precision of positioning and eliminating the confusion due to the sidelobe interference and low resolution at low frequency through the conventional beamforming.

III. THE SIMULATION ANALYSIS OF VSSFB

The virtual source method of sound source localization algorithm brings the acoustic inverse problem and the ill-posed nature of solving process due to the inevitable measurement error within the sound pressure data, making the serious deviation from the true solution. Therefore, regularization filter technology must be used for suppression and filtering the smaller singular value. In the paper, the truncated singular value filtering method (TSVD) is applied to removing the impacts.

The target sound source localization through the method is validated with the setup of the spherical microphone array is described as earlier in this paper. The sound sources are located on the positions $(90^\circ, 140^\circ)$ and $(140^\circ, 250^\circ)$. In order to compare the low-frequency performance, the dimensionless frequency $ka = 1$, and the search step is 0.5° , also the $SNR = 20$ dB.

The sound source localization results of three methods are shown in Fig. 1. At the low frequency, many false sound sources are appeared by using the SFTFB. The sidelobe can be eliminated with the implementation of QCSFB, but the spatial resolution is not high for the location accuracy of multiple sound sources. As shown in Fig. 1(c), the resolution can be greatly improved without sidelobe by the method proposed, and the sound sources are clearly visible, which obtain high resolution sound sources location results.



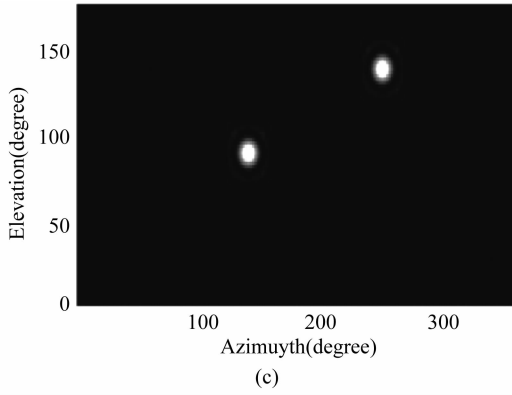


Fig. 1 Sound sources localization simulation results (a) SFTFB; (b) QCSFB; (c) VSSFB

In order to compare the performance of three algorithms, the localization error curves, maximum sidelobe level and main beamwidth are given respectively in Fig. 2.

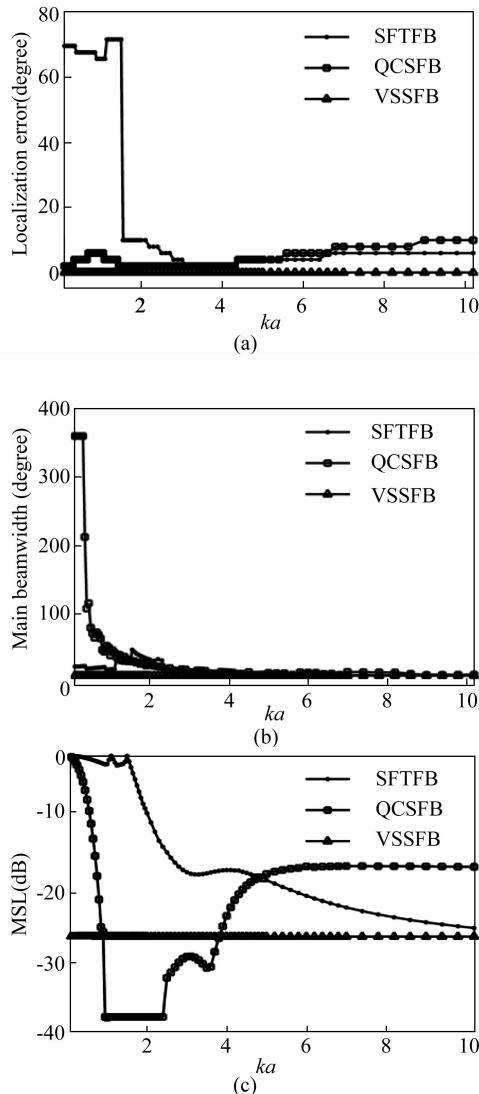
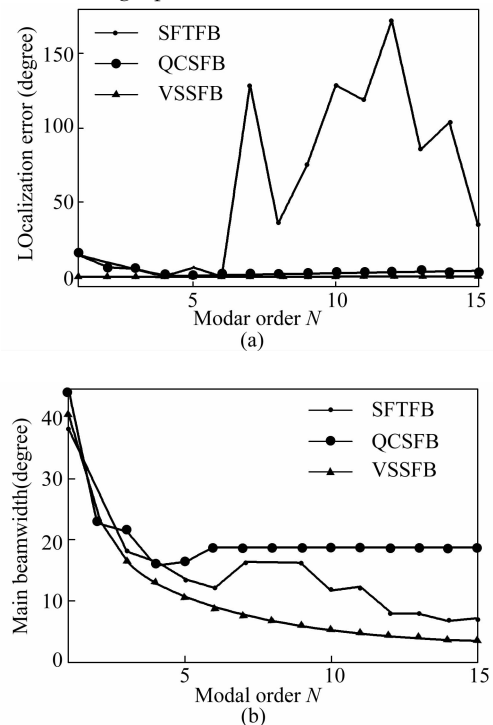


Fig. 2 The performance of three methods with ka (a) Location error; (b) Main beamwidth; (c) Maximum sidelobe level

As shown in Fig. 2, the localization performance has been demonstrated by the three methods with different

dimensionless parameters ka . Though the SFTFB in $ka < 1.5$, the MSL is almost the same with main lobe level, resulting in the presence of large amounts of false peak and incorrect results, just as shown in Fig. 1 (a). After quadratic constrained optimization, the MSL is greatly reduced, but the main lobe width is too wider to achieve high spatial resolution and multiple sound sources location, just as shown in Fig. 1 (b). However, the sound source localization error by the proposed method is not affected by the parameter ka , achieving significantly higher accuracy of sound source localization than other two methods. The multiple sound sources can be distinguished by high enough resolution and low sidelobe level. This method proposed can be used for high precision beamforming at full frequency, just as shown in Fig. 1 (c).

Fig. 3 shows the influence of modal order to localization performance. Because of the three methods all have good performance in $ka = 3.4$, which is shown in simulation results in the previous, we choose $ka = 3.6$, and 64 microphones as simulate parameter. When $N > 6$, the error and MSL are increased obviously by SFTFB, and the sidelobe interference will appear because the modal order is limited by the number of microphones, which is given in literature [7] $L \geq (N + 1)^2$. When $N < 6$, the performance of QCSFB is improved with the increasing of modal order. Its performance is no longer influenced with $N > 6$, and the localization error is not zero. However, the method proposed can have very high spatial resolution and obtain invariable positioning error, which is constant zero. The main beamwidth with satisfied MSL is decreased with the increasing of modal order and is consistently lower than other two methods. Therefore, in practical application, we select the appropriate modal order according to the resolution requirement by the VSSFB to achieve high precision sound source localization.



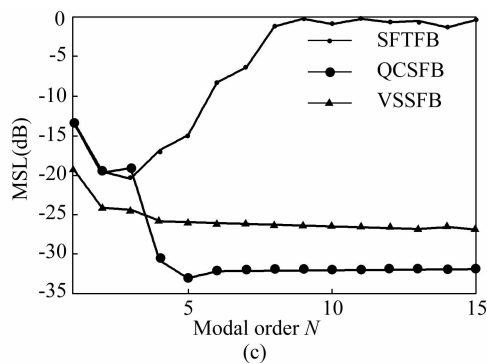


Fig. 3 The performance of three methods with modal order
(a) Location error; (b) Main beamwidth;
(c) Maximum sidelobe level

IV. CONCLUSIONS

Utilized the spherical array beamforming for three-dimensional multiple sound source location, A high precision spherical array beamforming method for sound sources localization based on virtual sources algorithm is proposed. By means of the spherical array beam scanning, the focused spatial spectrum of actual sound sources can be obtained and each scanning point is assumed to be the virtual sound source. The actual sources focused spectrum is the result of focused beam output by the entire virtual sources, and then high precision sound source localization can be realized by calculating the contribution of each virtual source to the sound field. The simulation results show that the method proposed can effectively eliminate low resolution, greater lobe effect and poor performance at low frequency by using the traditional algorithm. Further, the method has the advantages of background noise suppression and good stability.

ACKNOWLEDGMENT

We gratefully acknowledge the valuable cooperation of Dr. Zhou and the members of his laboratory in preparing and discussing this thesis.

REFERENCES

- [1] Meyer J, Elko G W. *A highly scalable spherical microphone array based on an orthonormal decomposition of the sound-field*, In: *Prco. ICASSP.*, 2002; vol. 2, pp 1781-1784.
- [2] Abhayapala T D, Ward D B. *Theory and design of high order field microphones using spherical microphone array*, In: *Prco. ICASSP.*, 2002; vol. 2, pp. 1949-1952.
- [3] Ward D B, Abhayapala T D. *Reproduction of a plane wave sound field using an array of loudspeakers*. *IEEE Transactions on Speech and Audio Processing*. 2001; vol. 9, pp. 697-707.

- [4] Z Li, Ramani D, Elena G et al. *Flexible layout and optimal cancellation of the orthonormality error for spherical microphone arrays*. In: *Prco. ICASSP.*, 2004; vol. 1, pp. 41-44.
- [5] Z Li, Ramani D. *A robust self-reconfigurable design of spherical microphone array for multi-resolution beamforming*. In: *Prco. ICASSP.*, 2005; vol. 2, pp. 1137-1140.
- [6] Boaz R. *Spatial aliasing in spherical microphone arrays*. *IEEE Transtions on Signal Processing*. 2007; vol. 55, pp. 1003-1010.
- [7] Qian Chen, Yang Yixing, Guo Guoqiang. *On robust supergain beamforming in mode space for a circular array mounted on a sphere*. *ACTA Acoustica (In Chinese)*, 2010; vol. 36, pp. 623-633.
- [8] Boaz R. *Spherical microphone array with multiple nulls for analysis of directional room impulse response*. In: *Prco. ICASSP.*, 2008; vol. 1, pp. 281-284.
- [9] YanShefeng, Hou Chaohuan, Ma Xiaochuan. *From element-space to modal array signal processing*. *ACTA Acoustica (In Chinese)*, 2011; vol. 36, pp. 461-468.

Resource-based Application Clustering in Wireless Sensor/Actuator Networks

Mirko Lippmann^{#1}, Mirko Caspar^{#2}, Wolfram Hardt^{#3}

[#]Computer Science Department, Technische Universität Chemnitz

D-09107 Chemnitz, Germany

¹mirko.lippmann@informatik.tu-chemnitz.de

²mirko.caspar@informatik.tu-chemnitz.de

³wolfram.hardt@informatik.tu-chemnitz.de

In the last years, sensor networks have increasingly evolved from the field of research in the direction of the practical use. They are primarily used in the field of home automation and environmental observations. In order to achieve a high duration of the sensor node, it is important to minimize power consumption. Many optimization techniques considered with the routing of data packets and the transmitting and receiving process [1]. Another approach is the optimization of the applications of each sensor node. In traditional sensor networks, data sets are acquired through the application and transmitted to a data sink, where they may have a lot of unneeded measured values. Each of these values is attended with an amount of energy for the acquisition and the transport. In order to avert this dissipation of energy, a dynamic application scheduler can be used.

In this paper, we present the resource-based application clustering approach (RACA). RACA could be used for the application clustering layer of the multidimensional clustering concept [2]. A major challenge is the independence from the operating system of the sensor nodes and their hardware. A heterogeneous sensor node environment builds the initial situation for this approach.

RACA distributes the applications in a wireless sensor network dynamically while taking the resources of the individual sensor nodes into account. Modifications on the network or in the resources of a sensor node could lead to changes in the participation of the node to the application. Each sensor node requires a runtime environment. In this environment, all resources, states and properties of the node are composited in a homogeneous model.

The execution model is parted in two sections. The first part is the hardware-dependent section. This section represents the hardware abstraction layer that contains all the drivers for the sensors and actuators, timers and the communication modules. This represents the hardware abstraction layer (HAL). A defined interface between this HAL and the application layer can be used for data exchange. The second part is the application layer.

In RACA we use a set of different three roles for ad-

ministrative procedures. These roles are the application node, application master and the coordinator. Except the coordinator role, a sensor node can hold multiple roles at the same time if the node controls different application or tasks. The first role, the application node, represents a local executing part of an application. This means, a local task of an application is running on this sensor node, for example the reading of a temperature sensor. These tasks are managed by the application master. The application master is responsible for one global application, like the measurement of a temperature of a room. It manages application task and the execution of selected nodes. If an application node is incapable to execute the part of the application, for example in the case of insufficient resources, the application manager searches and chooses other nodes for the execution. The last role is the coordinator role. This is a unique role in the wireless network. All events and requests for applications are received by this node. With this information, the coordinator starts a resource request to all sensor nodes in the network. Each node that satisfies the requirements, sends a reply. The coordinator selects the best node and appoint it as the new application master for this application. All collected data are transmitted to the application master. From this time, the application master is responsible for the task execution. After the transmission is complete, the coordinator can observe the new application master node. If there is a failure, a new application master will be searched by the coordinator.

RACA was used in some OMNeT++ simulations. In the comparison with classical approaches, a better energy performance could be ascertained. In further works, a complete scenario will be rebuilt in real hardware. If these components are similar to the simulated environment, a comparison with the simulation results will be possible.

ACKNOWLEDGEMENT

We gratefully acknowledge the cooperation of our project partners and the financial support of the DFG

(Deutsche Forschungsgemeinschaft) with in the Federal Cluster of Excellence EXC 1075 "MERGE" and it is supported by the nanett project in cooperation with the Bundesministeriums für Bildung und Forschung (BMBF) FKZ 03IS2011A.

REFERENCES

[1] M. Patil and R. C. Biradar. A survey on routing

protocols in wireless sensor networks. In Networks (ICON), 2012 18th IEEE International Conference on, pages 86-91, Dec 2012.

[2] Mirko Lippmann, Mirko Caspar, and Wolfram Hardt. Mehrdimensionales Clustering in drahtlosen Sensor- und Aktuatornetzwerken. In Dresdner Arbeitstagung Schaltungs- und Systementwurf (DASS2016). ISBN: 978-3-8396-1046-6, pp. 27-30

Service Control Algorithm of Providing Efficient Video-on-Demand Service using Hybrid Mechanism

Otgonbayar Bataa^{#1}, Erdenebayar Lamjav^{#2}, Purevtseren Bayarsaikhan^{#3},
Uuganbayar Purevdorj^{#4}, Chuluunbandi Naimannaran^{#5}, Young-il Kim^{*6},
Khishigjargal Gonchigsumlaa^{#7}

[#]*School of Information and Communication Technology (SICT) of Mongolian University of Science and Technology (MUST), acronyms acceptable, Ulaanbaatar, Mongolia*

¹b_otgonbayar2002@yahoo.com

^{*}*Electronics and Telecommunications Research Institute (ETRI), 138 Gaejeongno, Yuseong-gu, Daejeon, Korea*

⁶sosa247@gmail.com

Abstract— According to 3GPP/LTE standard service control is one of the key problems in eMBMS service in broadband wireless access network. The service control algorithm has to be performed on Fast, Pyramid and Reverse Fast broadcasting schemes. In this paper, we have provided performance analysis of mobile IPTV service control algorithm for 3GPP/LTE. This paper is proposed new algorithm to providing hybrid mechanism for efficient video-on-demand service.

Keywords— Component, Multicast efficient factor, VOD, segment, resource allocation, broadcasting scheme

I. INTRODUCTION

LTE technology is able to provide wide range coverage, high data rates, secured transmission, and mobility supported at vehicular speeds. It should have high QoS to transmit audio, video, voice, data services such like video gaming, mobile IPTV. IPTV is originally toward to fixed terminals such as set top box. However, it is developed for fixed and mobile convergence according to the future approach and requirements for mobility support. Video on Demand (VoD) service, which presents the requested video can be watched anytime anyway, will be major portion rather than following fixed schedule. VoD service is classified in Unicast and Multicast VoD. VoD service is based on an unicast transmission so it will not so impact on the network side when there is enough capacity and service request rate will be moderately low. Multicast VoD is based on broadcast, multicast transmission and system bottleneck will be observed in bandwidth, downlink resources of wireless access network rather than content server. Because of it's possible to improve server performance by doing HW and SW extension, Multicast VoD requests like the top 10 ~ 20 videos

are known to comprise 60% ~ 80% of the total demand. So it's assumed to improve service efficiency of popular videos. Unlike traditional TV broadcasting, mobile IPTV stream can't be broadcasted in all over the internet, so we used multi-channel multicast concept for Multicast VoD within limited number of wireless channels. So, to increase system capability we have to define any service control algorithm. Proposed Service Control Algorithm is performed on following situations.

- ① More numbers of users access at same time when transmit only one content.
- ② Allocate more different types of contents at least channels.

II. PERFORMANCE METRICS FOR MOBILE IPTV SERVICE CONTROL ALGORITHM

At First, we have decided server-side algorithm before transmission, it's going to.

Step 1: Proceed the number of sessions n_i with required video V_i , include with length L_i that meets the requirement for the VoD connection time less than $ct < 10$ s

$$n_i = \text{Log}_2\left(\frac{L_i}{sd} + 1\right) \quad (1)$$

Step 2: Segmenting the video V_i equally to $2^{n_i} - 1$ segments, where n_i is available channels.

Step 3: Distribute the segments to probability connections. The segments are allocated continuously S_1 to S_{n_i} , by normal series $2^{(n_i-1)}$.

Step 4: All videos allocated for the multi-session multicasting since segmentation processing.

Next, we decide server-side Algorithm for transmission. See Fig. 1.

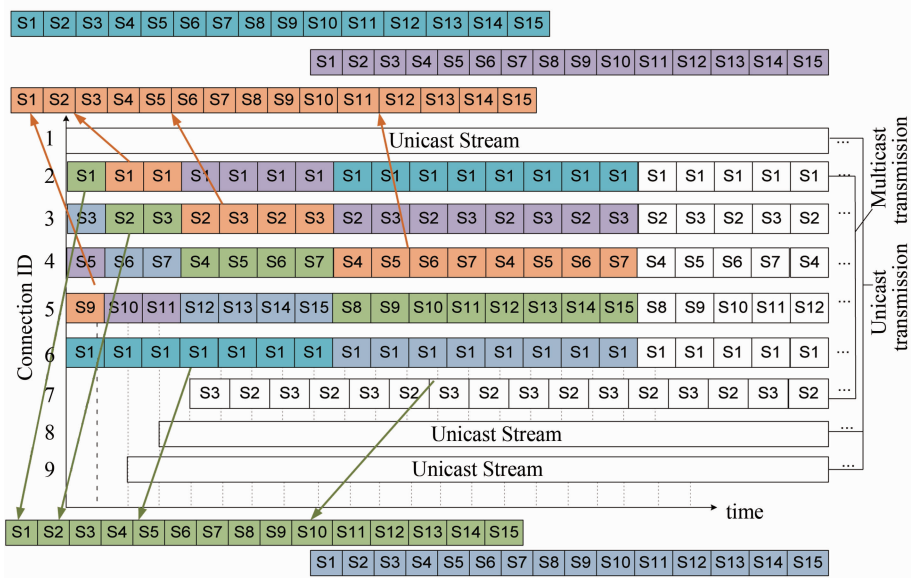


Fig. 3 Content allocation step 2

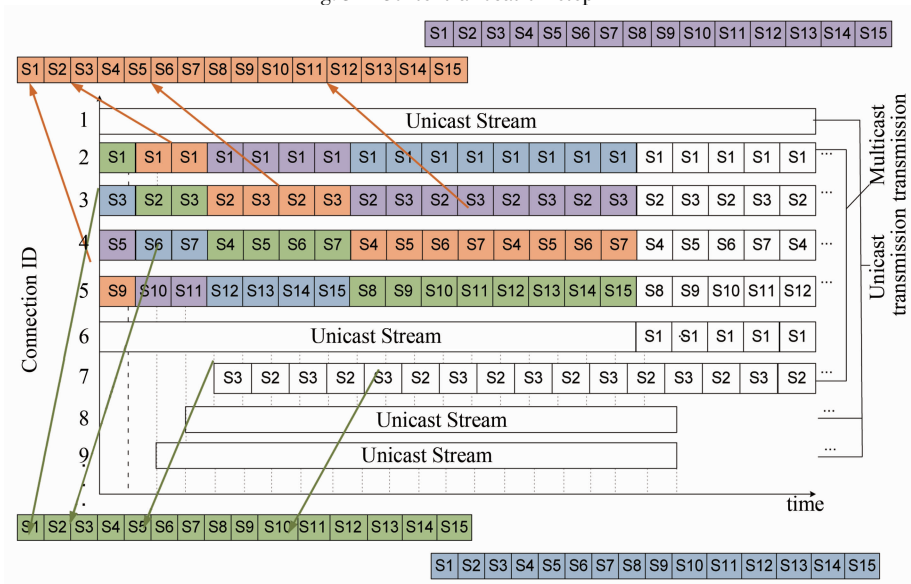


Fig. 4 Content allocation step 3

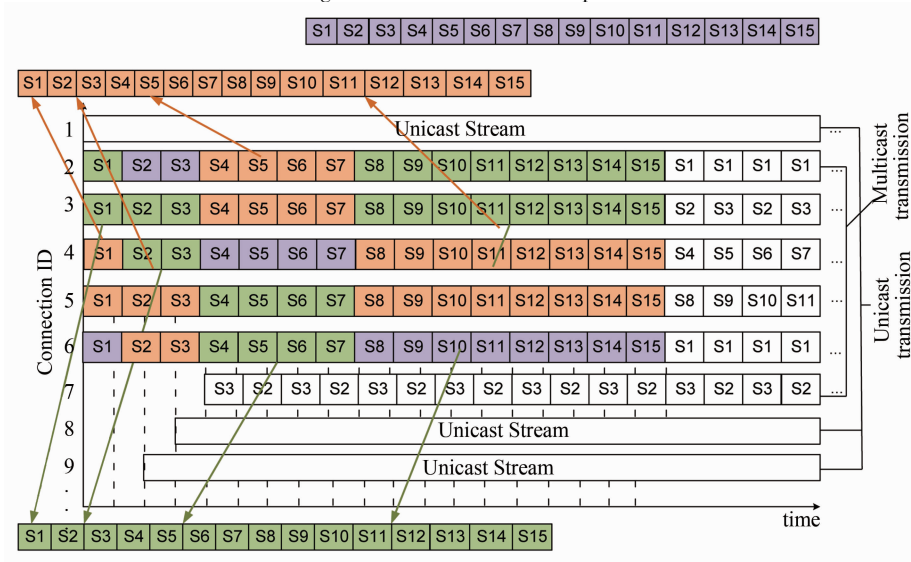


Fig. 5 Content allocation step 4

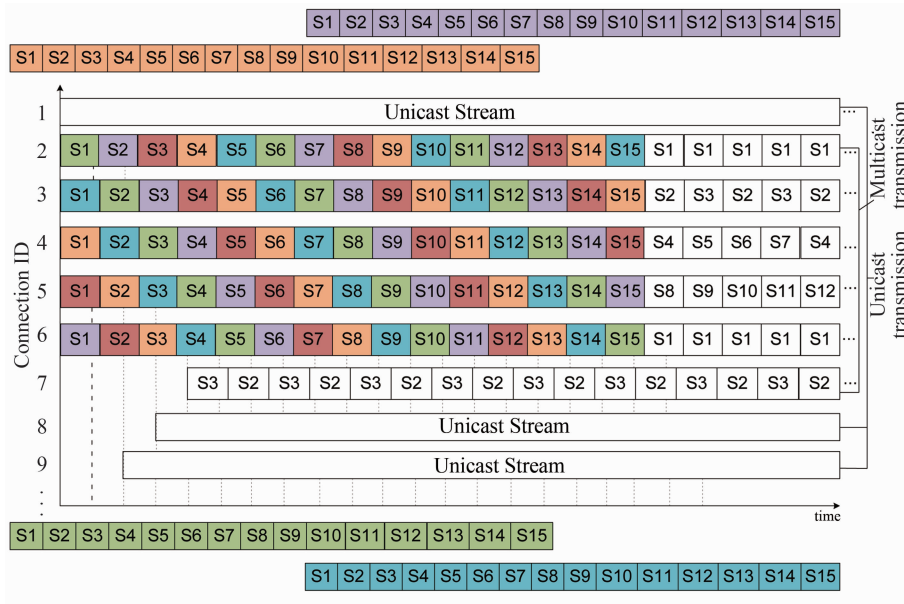


Fig. 6 Content allocation step 5

In Table III, we show the comparison of those three schemes. Our proposed scheme: Fast Broadcasting make a better performance than Reverse Fast and Pyramid [3]. After then, we set up the value for the parameters: number of Fast channels, Pyramid channels, Reverse Fast channels for observing the performance of the waiting time as the Fig. 7.

Table III
THE PERFORMANCE COMPARISON

Aspect of each scheme approaching	Maximum waiting	Buffer requirement	Maximum client handle bandwidth
Fast broadcasting	$sd < 10$ s	$(2^{n-2}/N)^*$ 10%	$\sum_{i=1}^k n_i - 1$
Pyramid broadcasting	$sd < 10$ s $sd < 0.1$ s	$(2^{n-2}/N)^*$ 57%	$\sum_{i=1}^k n_i / N$
Reverse Fast broadcasting	$sd < 0.1$ s	$(2^{n-2}/N)^*$ 94%	$\sum_{i=1}^k n_i$

Fig. 8 shows a hybrid method for allocate content to available free channel. First, we decide VoD request, so MEF value determined for multisession multicasting or single session unicasting service, then our segmentation rule has being work.

Our research result is analyzed and illustrated by 2 graphs which have best performance. It shows us advantage of our algorithm (Fig. 8 and Fig. 9). The MBMS server can access over 300 serving channels, and there are over 100 contents, their lengths are average 90 min

(5,400 s). The request arrival ratio varies from 5/60 to 20/60, and is calculated by Zipf distribution with skew factor 0.271.

We create the adaptive resource allocation method to choose the best resource allocation that minimizes service blocking probability. Our graph result depicts the blocking probability with different request rates, there may be the number of multicasted contents varies from 1 to 25 that mean the total number of multi-channel multicasted contents (m) in previous analysis model is changed 1 to 25. The results of our environment calculated that served 5 videos with multi-session multicasting and 1 video with unicasting, then we can have the best performance. This result shows how many contents delivered by multi-channel multicast serving to achieve the best performance. In general, the VoD service blocking probability is increased when the total service request rate is increased. If service request rate is lower than 3/min, then customers can't recognize the difference. With request arrival ratio $\lambda = 6/\text{min}$, we applied our adaptive video allocation method to find the best allocation that minimizes service blocking probability. The result depicts the blocking probability with fixed video length (i. e., 50 min) where the number of multicasted video varies from 1 to 10[1].

We using multi-session multicast takes unless than unicast so that enables better performance for overall VoD services in blocking probability point of view, under the same condition.

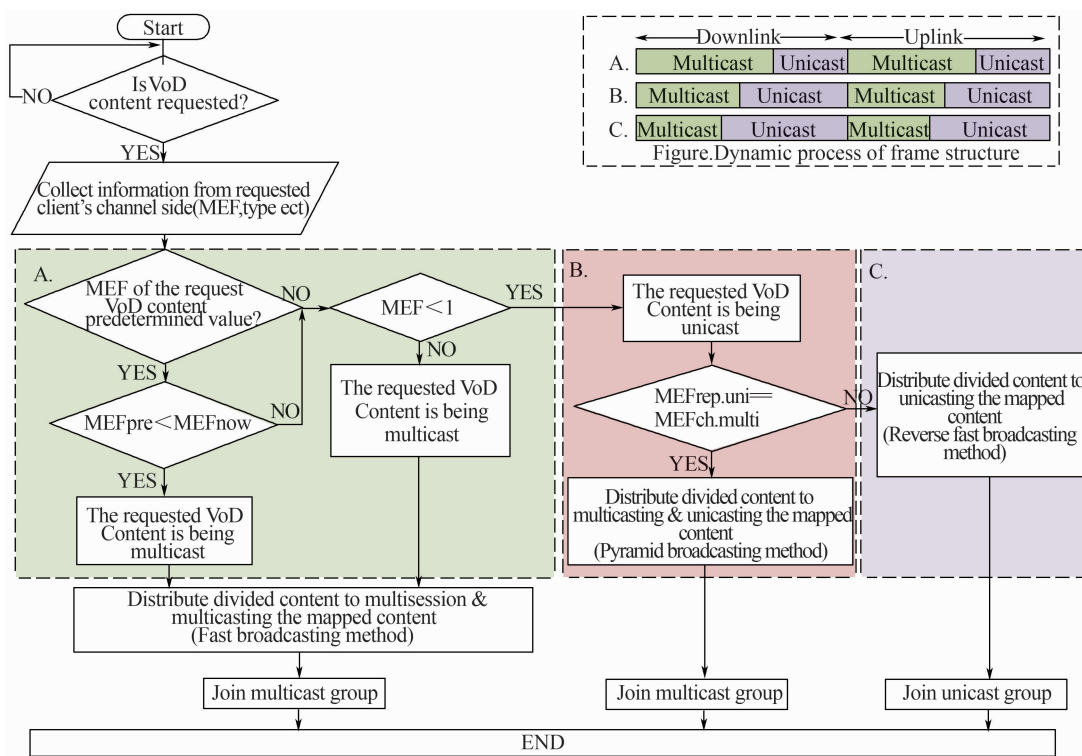


Fig. 7 Proposed Service control algorithm

Finally, based on the above calculated video allocation, we show the performance of blocking probability on the proposed hybrid mechanism compare to unicast mechanism. Service request rate rises, the proposed hybrid transmission scheme shows more robustness.

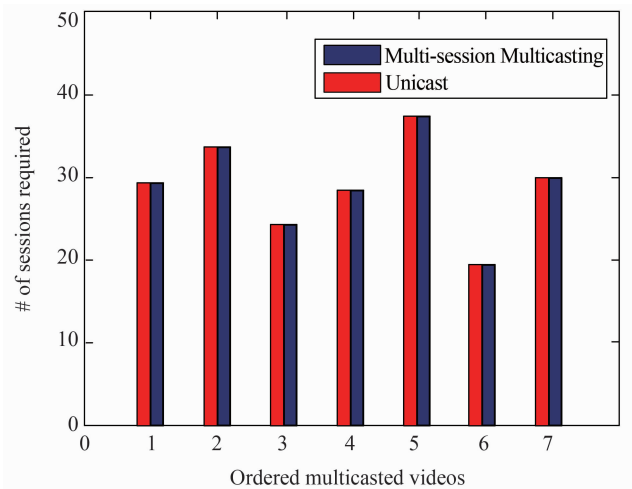


Fig. 8 Videos allocated for Pyramid broadcasting method

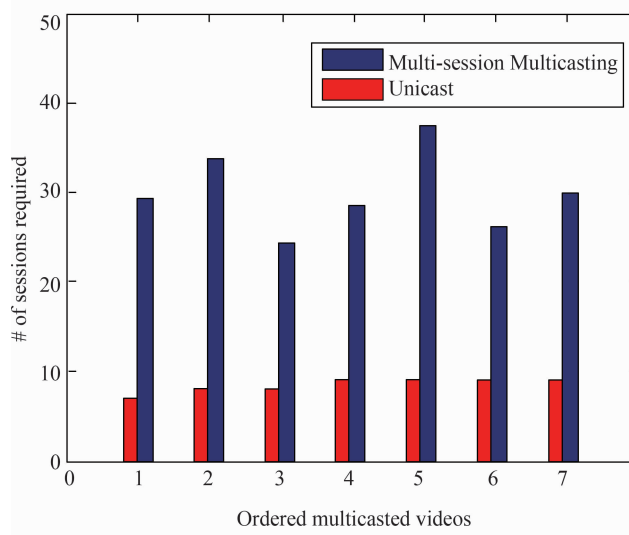


Fig. 9 Videos allocated for Fast broadcasting method

ACKNOWLEDGMENT

This work was supported by the IT R&D program of KCC/MKE[2010-10035206, The Development of IMT-Advanced Mobile IPTV core technology], Electronics and Telecommunications Research Institute of Korea and Mobicom corporation of Mongolia.

CONCLUSION

In this paper, we present an effective service control algorithm that may be reduce the blocking probability and increase overall throughput. Client request is dependent from video on demand service control algorithm. Our proposed algorithm is based on combined unicast and multichannel multicasting mechanism which is reduced overall bandwidth consumption of the system (Fig. 7). The research results of the proposed VoD

service control algorithm are illustrated in Fig. 10 and Fig. 11. Simulation results indicate that our algorithm may offer a satisfactory performance for VoD service.

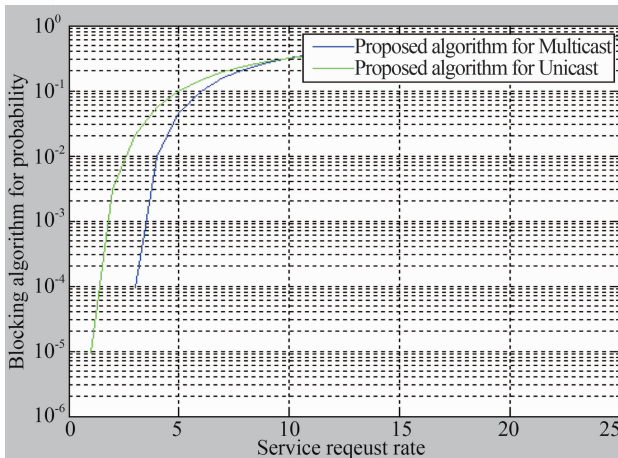


Fig. 10 Blocking probability of VOD services

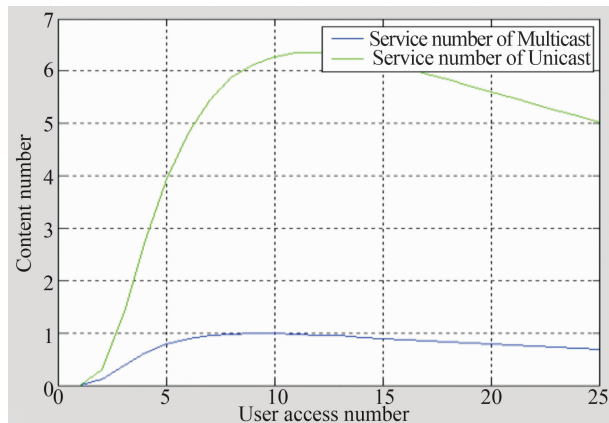


Fig. 11 User access number of per content

tions," Master's thesis, School of Electrical Engineering and Computer Science University of Central Florida, 2011.

- [6] Jong Min Lee, Hyojin Park, Kyun Choi, " System and method of providing efficient video-on-demand service using unicast/multicast in internet protocol network" Korea advanced institute of science and technology, Feb 2010.

REFERENCES

- [1] Dai-Boong Lee, Wan Kim, and Hwangjun Song, " An Effective Mobile IPTV Channel Control Algorithm over WiMAX Network" IEEE Trans. Consumer Electronics, Pohang, Korea. 2010.
- [2] Chultemdarjaa Khaltmaa, Takuro Sato " Study on high reliable IPTV transmission system over cellular network" , Graduate School of Global Information and Telecommunications Studies, WASEDA UNIVERSITY, June 2010.
- [3] Po-Tsun Liu and Yu-Wei Chen, " A Hybrid Broadcasting Method for Video-on-Demand Services" National Taipei University of Technology, Graduate Institute of Information and Logistics Management, Taipei, Taiwan, 2011.
- [4] Ailan Hu Ioanis Nikolaidis Peter van Beek, " On the Design of Efficient Video-on-Demand Broadcast Schedules" Oct. 2010.
- [5] James Z, Wang Ratan Guha, " Efficiently Allocating Video Data in Distributed Multimedia Applica-

Study on Process of Metamodel-based Atmosphere Pressure Modeling Method

Guobing Sun^{#1}

[#]*School of Automation, Harbin University of Science and Technology*

No. 52 XueFu Road, NanGang Dist Harbin 150080, China

¹*sunguobing@live.com*

Abstract— In this era of natural environment simulation, atmosphere are considered as an important factor effecting the other natural environment factors. Here, a method based on metamodel is used, this kind of method has not been proposed in my related literature, but it has not been implemented in detail, here the process is proposed to implement the modeling atmosphere pressure.

Keywords— Metamodel-based, modeling process, atmosphere pressure, simulation

I. INTRODUCTION

In Contemporary modeling and simulation society, there are lots of methods proposed for modeling natural environment, such as EDM (Environmental Data Model), TCDM (Terrain Common Data Model), MEL (Master Environment Library), and so on [1] to [6]. These methods are efficient in some specific domain, but sometimes they are complex or not sufficient when used in common Natural Environment (NE) modeling processed. Due to this, we have ever proposed a simple meta-model based method some literature [7] to [10], but the detail implementation of this approach is absent. So it is not so easy to get a natural environment model. Here, the process is proposed to implement the metamodel-based atmosphere pressure.

At first, an idea should be hold that model is an instance of meta-model and meta-model is an abstract of model. In this paper, a detail process is proposed to build atmosphere pressure models and meta-models. There are six sections in this paper. Section II provides brief description of meta-modeling common process. Section III focuses on the requirement of atmosphere pressure. Section IV gives the metamodels of atmosphere pressure. In section V, the model is built and an example is taken.

II. COMMON PROCESS FOR META-MODELING

In this section, the concept of meta-model is presented, and a common process is proposed for BE meta-modeling. 0.63 cm. Please do not re-adjust these margins.

A. BE Meta-model and Meta-modeling

In knowledge engineering, if a concept itself is repeatedly used, meta is used as a prefix to the concept. Meta-x means x of x, so meta-model is a model of model [5].

There are two types of meta-model, one is used in software engineering, and the other is used in the operational research. Here the first meta-model is studied, it defines the syntax and semantics in certain domain, it can be used to describe all systems in the specific domain. This type of meta-model can be reused in related process of modeling and simulation [6].

B. Common Process for BE Meta-modeling

According to the deep research to the process for BE model, a common process is proposed for meta-modeling, see Fig. 1. The process is composed of 6 steps [7].

(1) Confirming the simulation requirement

Here we should confirm whether the BE simulation is needed in complex military simulation systems, if be, the next step is on.

(2) Picking up the BE entities in Requirement

Here three parts of BE entities should be picked up.

① BE data requirement;

② BE entities which effect the military entities;

③ BE entities which are impacted by the military entities.

(3) Confirming the military entities

Here the military entities are meant to be those which interact with the BE entities, this step makes the models simply. These military entities can be divided to five categories, such as trafficability, sensor, weapon, infrastructure and personnel, and one entity should not be included in different categories.

(4) Analyzing the internal dynamics of BE Entities

Here the BE internal dynamics is analyzed, and its attributes, behavior and constraint should be confirmed, then BE internal dynamics model will be built.

(5) Analyzing the effects and impacts of BE factors

Here the two kinds of models should be built separately if any one is needed, and the attributes, behavior, interaction and constraint should also be confirmed. In this step the BE effect model and BE impact model are modeled.

(6) Analyzing the BE model

Here all three kinds of models mentioned above should be analyzed carefully, and all the classes, attributes, behavior, interaction and constraint in them should be abstracted to meta, then the meta-model is built, including meta-class, meta-attribute, meta-behavior, meta-interaction and meta-constraint, which will be discussed later.

This process shows a common reference steps for

building BE meta-model, and these steps can be chosen in specific simulation.

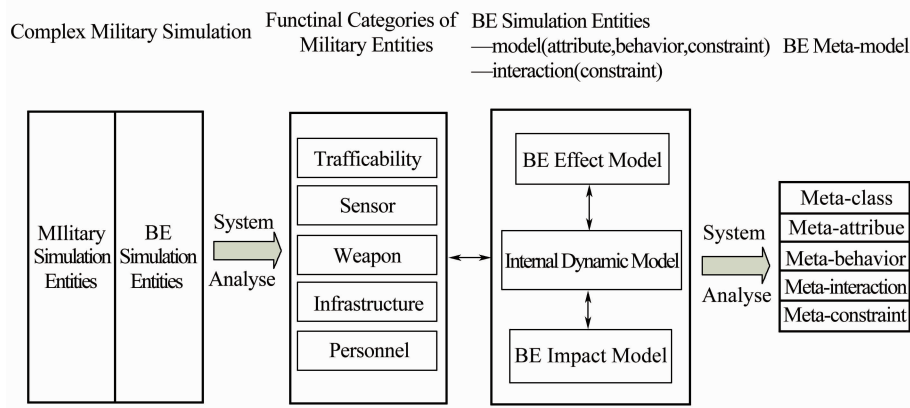


Fig. 1 Common process of BE meta-modeling

III. CONFIRMING ATMOSPHERE PRESSURE REQUIREMENT

In this section, the process is proposed to implement the atmosphere pressure meta-model.

A simple requirement is proposed, just an aircraft to get the real-time elevation through the pressure sensor on it. A model should be built to get this relationship between elevation and atmosphere pressure. According to the model, the related elevation can be confirmed. So the detail requirement can be listed as Table I to Table IV.

Table I
FACTORS OF BE

Factor	Identifier	Type
Elevation	AirElevation	double
Atmosphere pressure	AirPre	double

Table II
FACTOR OF AIRCRAFT

Factor	Identifier	Type
Atmosphere pressure	CraftPre	double

Tabel III
FACTORS FOR INTERACTION

Factor	Identifier	Type
Atmosphere pressure	AirPre	double

Tabе IV
RELATED FACTOR

Factor	Identifier	Type
Elevation	intElevation	double

IV. META-MODELING OF ATMOSPHERE PRESSURE

In "A meta-model is a model of models" [5] and "A model is an instance of meta-model" [6] imply that a meta-model is a model of another model.

According to [7] to [9], the contents of BE meta-model should include five factors.

A. Meta-class

Meta-class is a prototype to define the variables and methods of certain BE objects, and it includes following information, such as names, methods, attributes and events. To BE, different BE objects and any kinds of interaction should be listed. Here the `mmodel_airPreName;name` is used to present the atmosphere pressure meta-class.

B. Meta-attribute

Meta-attribute presented here focuses on the specific attributes of BE meta-object, and could be abstracted from model attributes. Here the `mmodel_airPreAttr;string` is used to present the atmosphere pressure meta-attribute.

C. Meta-behavior

Meta-behavior is an abstract of behavior in BE internal dynamic models. Here the `mmodel_airPreBehav;string` is used to present the atmosphere pressure meta-behavior.

D. Meta-interaction

Meta-interaction is an abstract of interaction between BE entities and military entities. Here the `mmodel_airPreInter;string` is used to present the atmosphere pressure meta-interaction.

E. Meta-constraint

Meta-constraint is an abstract of constraint the BE model should keep to. Currently, the constraint of models is becoming a popular problem in modeling and sim-

ulation research. When space environment is modeled, some constraint should be included. Here the `mmodel_airPreCons`; string is used to present the atmosphere pressure meta-behavior.

Here two values are presented such as potential value and space value.

Potential value means how many times the meta-models can be instanced. The value subtract one while an instance is performed, and no more instance can be performed when it is zero.

Space value presents whether the meta-models can be instanced to the level not the next neighbor, "1" means it can, "0" means it can not. The space value can only be "1" or "0".

In this paper, the superscript is used to present space value, and the subscript is used to present potential value. These two values constrain each other, and the instance process should meet the two values.

In order to present this method easily, potential value and space value are used to present air density meta-model in NE as shown in Fig. 2. Here an `instance_of` relationship is used to present the linguistic instance, it means that the atmosphere pressure meta-model is a linguistic instance of meta-model, and the atmosphere pressure meta-model is linguistic instance of air meta-model, this type of instance process doesn't make potential value changed. Here all the contents of atmosphere pressure meta-models are listed in the Fig. 2.

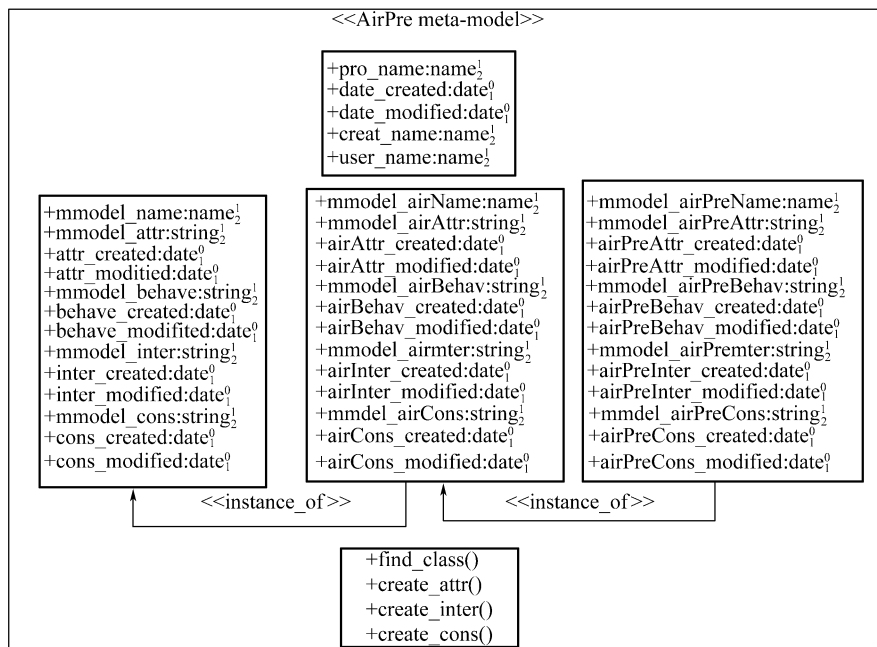


Fig. 2 Meta-models of air pressure in atmosphere pressure

V. MODEING OF ATMOSPHERE PRESSURE

In this section the process is proposed to implement the atmosphere pressure model and the simulation result is given.

In order to build the atmosphere pressure model, the simple relationship show in equation(1) is got from Fig. 2.

$$P = P_0(t/t_0)^{-g/LR} \quad (1)$$

Here acceleration of gravity $g = 9.806,65 \text{ m} \cdot \text{s}^{-2}$, common air constant $R = 287.053 \text{ m}^2 \cdot \text{K}^{-1} \cdot \text{s}^{-2}$. The other related parameters are shown in Table V.

Elevation <i>H</i> /km	Initial Atmosphere Pressure P_0 /Pa
0 ~ 11	101,325
11 ~ 20	22,632

In this process, the resolution requirement is 1 km, then the result can be got as shown in Table VI.

If the resolution is small, a different result can be got. In order to present the simulation result directly, Fig. 3 is drawn, here the resolution is 0.1 km. When the pressure is $4.986 \times 10^4 \text{ Pa}$, its related elevation is $57 \times 0.1 \text{ km} = 5.7 \text{ km}$.

Height/km	Pressure/Pa
0	101,325
1	89,876.285,187,271,2
2	79,501.424,641,667,0
3	70,121.162,236,430,0
4	61,660.444,130,465,0
5	54,048.286,145,761,4

Table VI Continued

Height/km	Pressure/Pa
6	47,217.642,475,895,0
7	41,105.275,732,340,5
8	35,651.628,335,825,0
9	30,800.695,260,495,4
10	26,499.898,139,253,3
11	22,699.960,739,233,4
12	19,399.393,646,323,9
13	16,579.578,567,877,5
14	14,170.338,784,134,0
15	12,111.791,461,904,7
16	10,352.802,535,044,1
17	8,849.706,368,285,17
18	7,565.213,106,910,44
19	6,467.476,153,031,74
20	5,529.296,262,737,94

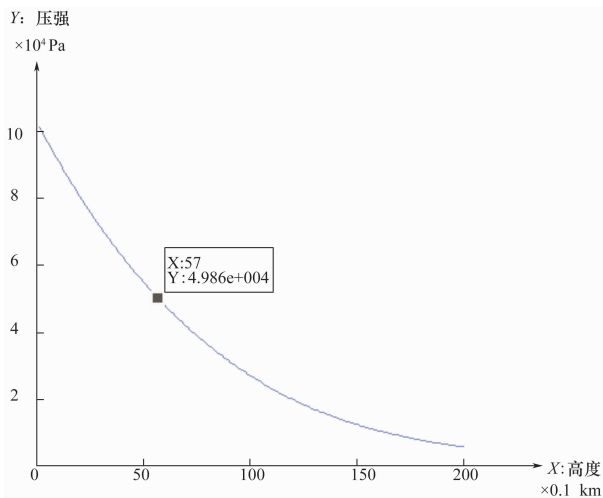


Fig. 3 Relationship between pressure and elevation

VI. CONCLUSION

The method based on meta-model has been studied for a long time, and it has been proposed to NE domain by our team, but the detail process was not presented clearly. Here, atmosphere pressure was used to

give the total modeling process for the meta-model-based method.

Of course, there are some insufficiencies. The relationship between the requirement and the natural environment is not shown directly, so for the further, we would like to focus on the knowledge of natural environment domain. In addition, the related knowledge should put into the NEMMS[10] more and more.

VII. ACKNOWLEDGMENT

I would like to express my gratitude to all those who have helped me during the writing of this thesis. I gratefully acknowledge the help of my supervisor Professor Huang Jinjie. I do appreciate his patience, encouragement, and professional instructions during my thesis writing.

Last but not the least, my gratitude also extends to my wife, Che Xiaodan who have been assisting, supporting and caring for me all of my life. kindly gave me a hand among my work.

REFERENCES

- [1] D. M. Dale, C. J. Annette, E. N. Melissa. Environmental Data Modeling: REDM, DREDM, Ontology and Metrics. Proc. of the Spring 2003 Simulation Interoperability Workshop. 2003;03S-SIW-132.
- [2] J. L. Michael, D. Virginia. SEDRIS: A Tool for the Management and Exchange of Multi-domain Environmental Data. Proc. of the Fall 2006 Simulation Interoperability Workshop. 2006;06F-SIW-095.
- [3] R. Gitzel. The Role of Metamodeling in Model-driven Development. Proceeding of the 8th World Multi-Conference on Systemics, Cybernetics and Informatics (SCI2004). 2004;19-23.
- [4] J. D. Bruce. Standards and the Master Environmental Library. Proc. of the Spring 1999 Simulation Interoperability Workshop. 1999;99S-SIW-042.
- [5] A. Talevski, E. Chang, T. S. Dillon. Meta Model Driven Framework for the Integration and Extension of Application Components. *Proceedings of the Ninth IEEE International Workshop on Object-Oriented Real-Time Dependable Systems (WORDS'03)*. 2003.
- [6] Thomas Kühne. Matters of (meta-) modeling. Software System Model. DOI 10.1007/s10270-006-0017-9.
- [7] SunGuobing, Gao Yajuan, Du Hongyue, Huang Jinjie. Natural Environmental (NE) Modeling Process Based on Ontology and Meta-model. International Forum on Strategic Technologies 2010. Paper ID:3089.
- [8] Sun Guobing, Huang Jinjie, Gao Yajuan. Research on Battlefield Environment Simulation Meta-model. 2010 ETP/IITA Conference on System Science and Simulation in Engineering. 2010;295-298.
- [9] Sun Guobing, Huang Jinjie, Gao Yajuan. Research on Battlefield Environment Meta-model Hierarchy. Computer Science. 2010,38(3), pp. 295-298.

[10] Sun Guobing, Xiang Quan, Huang Jinjie, Cheng Guangming. Design of Natural Environment Modeling Management System Based on Meta-model. In-

ternational Forum on Strategic Technologies 2011. 2011 ;800-803.

Towards Impact Detection and Localization on a Piezo Metal Composite

Frank Ullmann^{#1}, Wolfram Hardt^{#2}

[#]*Computer Science Department, Technische Universität Chemnitz
D – 09107 Chemnitz, Germany*

¹*frank.ullmann@informatik.tu-chemnitz.de*

²*wolfram.hardt@informatik.tu-chemnitz.de*

The integration of piezoelectric elements into metal sheets enables the manufacturing of smart metal components. Such components can be used, for example as smart input device in the automotive industry. The presented research is embedded into the research on mass production enabled manufacturing processes for functional integration. In this process, the piezo compound is extruded and joined with a structured metal surface and some copper electrodes to obtain a hybrid assembly compound consisting of two metal sheets with piezo core. After a polarization step to amplify the resulting voltage from the piezo core, the metal sheet can be formed into a shape that is needed. The last part of the process is the signal processing and classification with an embedded system and optional communication. Further information about the mechanical engineering part are discussed in [1].

While in on-going mechanical engineering research the mass production enabled large scale production of these components is addressed, the problem to create a mass production enabled embedded processing system for that use case is nearly untouched. In this paper, a hardware/software co-design approach is presented that is used to process the input and in turn predicts the impact position.

Especially in health-monitoring systems the detection and localization of impacts are important because impacts can easily lead to damages that cannot so easily be located while maintenance. Therefore, different approaches with embedded piezoelectric sensors were developed that enable damage detection while in use. To be able to classify the impact's location without knowing the concrete structure of the test object, just by training the system, machine learning techniques are often used. In this paper, a neural network is evaluated. Using a machine learning approach is useful especially in the targeted mass production enabled use case because without a learning approach every change in the production process results in different behavior. As a consequence, the localization algorithm have to be tuned every time, i. e. by time-consuming simulation. This step is not necessary when using an adaptive algorithm based on machine learning techniques.

The aim is to develop a low-cost embedded system for localization which has very limited calculation resources. To avoid excessive and complex computations

only time domain features are used. In the following, the developed system is described.

At first, the voltage from the piezo sensors have to be pre-processed to match the requirements of the following processing elements. This is done by using an electronic circuit based on interconnected operational amplifiers. The voltage is increased by 0.5 V to shift the voltage signals to a positive only range of between 0 V and 1 V. To ensure that the voltage does not exceed or go below the maximum rating of the embedded processing system, it is limited by an upper and lower threshold. The resulting voltage is then sampled by a high-speed analog-digital converter (ADC) in order to digitalize it. If this step is done, the values are forwarded to the Piezo Integration Interface Board which is capable of processing the input. Depending on the needs of the localization procedure some features are required as input to the algorithm. These features have to be extracted from the voltage history before. With suitable features impacts can be detected and subsequently be localized.

If an impact occurs the generated voltage will start to oscillate around its idle level. This results in a changing of the gradient of the data if the according voltage is imagined over time. The higher the frequency or amplitude of the voltage curve the higher the gradient. The single absolute values of it are then summed to obtain a value independent measure of the gradient. If this gradient exceeds a moving average threshold an impact is detected and the location can be estimated.

While the just mentioned procedure is implemented using an FPGA, the localization is currently implemented using a desktop PC. To monitor and test different algorithmic localization approaches a desktop application was developed that is based on the programming language C++ and uses the Qt framework [2] in combination with the QCustomPlot library [3] to create the graphical user interface. In a first test, an artificial neural network was implemented to estimate the impact location. Based on the Fast Artificial Neural Network library [4] (libfann) an MLP (multilayer perceptron) with backpropagation and one bias neuron per layer was modeled.

To obtain data to train and test the network, a U-shaped profile with a length of 500 mm, a height of 105 mm and a thickness of around 1.5 mm was used. Three electrodes are asymmetrically mounted at the bottom.

The profile was excited at 20 points that are scattered over its surface and the resulting voltage curves were recorded.

The neural network was trained and tested with data from the time domain only. As input to the net the time differences between the instants of time are used, that are the moments when the impact induced mechanical wave arrives at the electrodes. Based on this the network predicts the position of the impact on the surface that is represented by two coordinates which are the longitudinal position and the transversal position along the surface.

Different network configurations were testing, varying in the number of hidden neurons, activation function, learning rate and dimensionality to be predicted. The best results concerning localization accuracy were achieved with a sigmoid activation function and a learning rate of 0.4.

In the initialization phase of the network training step, all weights are randomly calculated. As a consequence, the results can only be evaluated statistically. Every configuration was tested 1000 times to achieve a good representativeness.

From the results could be seen, that three layers are sufficient to predict the longitudinal position of the impact on the formed profile. With nine hidden neurons the net performs best, a median position deviation of 22 mm was achieved. In contrast, a transversal position estimation was not possible due to the symmetrical orde-

ring of the electrodes in transversal perspective. This problem will be solved in future work, some experiments are pending that concern the optimal placement of the electrodes. In another on-going investigation the evaluation of other machine learning algorithms are carried out.

ACKNOWLEDGMENT

We gratefully acknowledge the cooperation of our project partners and the financial support of the DFG (Deutsche Forschungsgemeinschaft) within the Federal Cluster of Excellence EXC 1075 "MERGE".

REFERENCES

- [1] V. Kräusel, A. Graf, M. Heinrich, R. Decker, M. Caspar, L. Kroll, W. Hardt, and A. Göschel, "Development of hybrid assembled composites with sensory function," *CIRP Annals-Manufacturing Technology*, vol. 64, no. 1, pp. 25-28, 2015.
- [2] The Qt Company, "Qt Framework Documentation," <http://doc.qt.io/qt-5/index.html>, Accessed: 30.04.2016.
- [3] Emanuel Eichhammer, "QCustomPlot," <http://www.qcustomplot.com/>, Accessed: 30.04.2016.
- [4] S. Nissen, "Implementation of a fast artificial neural network library (fann)," 2003.

The Innovation Process Making Education and Knowledge into Work Creation

M. Byambasuren ^{#1}, Ch. Oyunjargal ^{*2}

[#] General technical department, Transportation Institute of Mongolia Ulaanbaatar, Mongolia

¹byambaamyagmar@yahoo.com

^{*} Civil and architecture engineering institute, Mongolian University of Science and Technology Ulaanbaatar, Mongolia

²oyunaaj2013@yahoo.com

Abstract— The current trend of education is making knowledge into work creation. It is much more practicable/suitable to specify the quality changes in higher education as an innovation reform. Within the innovation of education reform, it is necessary to remodel the curriculum according to the current requirements, to give students the knowledge of new techniques of 21st century such as creative and system thoughts and communication skills/ethics and first of all, it is important to focus on the result of training processes. Therefore, this paper offers the opportunities to direct students for developing themselves in whole life on the basis of good formation of character by assessing themselves by 5 potentialities and planning their further development.

Keywords— Innovation of education reform, 5-sided principle

I . INTRODUCTION

In the processes of the globalization and establishment of knowledge-based community, there are different types of universities such as open, on-line profit, corporate and world standard. These universities have been competing for the world market of education by their training quality, the level of research work, new knowledge and potentialities. Acquiring higher education is massive in our country. Although this desire/enthusiasm is appraisable, the training quality and supplying cannot satisfy the requirements of learners, employers and the

public/society.

The 21st century is a century with different manners and a lot of values coexist and the social development is speeding up. Employers and public demand the equal formation including intelligence, mental strength, healthy body-build, accountabilities and skills working in a team. In order to form these natures, the roles and participation of educational institutions, universities, colleges and schools are very important.

Education system concentrates more on the improvement of intelligence and its assessment. Thus, the process from secondary school leavers to higher-educated or professional person could not become a mechanism for developing the equal formation of a person. Employers need the information about behavior and attitude of a person except the diploma for choosing a proper employee.

II . METHODS

The objective of education ensures that to carry on the issue of person formation equally and to develop the previous system until now. In the system of Mongolian education, the current strategy of education is leading up to make the knowledge into the work creation for the analysis of innovation processes. The process of education innovation is given in Table I .

Table I
THE PROCESS OF EDUCATION INNOVATION

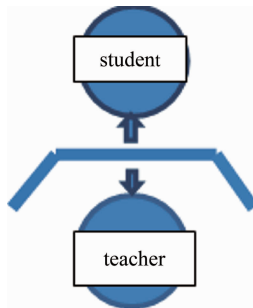
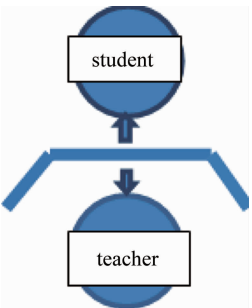
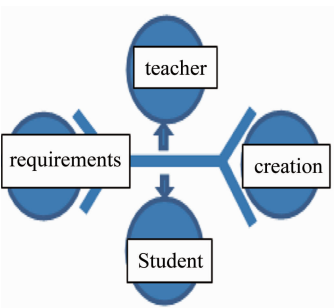
Correlation	Before 1990	1990 – 2010 Credit was introduced	After 2010
Formulation of knowledge and potentiality relations			

Table I Continued

Correlation	Before 1990	1990 – 2010 Credit was introduced	After 2010
Tendencies	Teacher-knowledge distributor Student-knowledge collector	Teacher-knowledge server Student-customer	Teacher-knowledge source, student-knowledge searcher, knowledge co-inventors based on the usage
Tendencies to get knowledge	Students were active to get knowledge	Students were weak to get knowledge	Students are active, potential, cooperative and they are application /usage/ determiners and creators
Expenses	It is free to get knowledge and skills	It is fee-paying to get knowledge and skills	Students have to pay for getting knowledge and skills
Profession choice	To choose an occupation due to his/her knowledge and skills	On the requests of parents, a few of them choose according to their knowledge and skills	To choose according to their own interests, knowledge, skills and social needs
Training methodology	Traditional training methodology	Electronic, on-line and distance training methodology	Electronic training methodology which tends to traditional one
Tendencies to get knowledge	Students were active to get knowledge	Students were weak to get knowledge	Students are active, potential, cooperative and they are application/usage/ determiners and creators
Expenses	It is free to get knowledge and skills	It is fee-paying to get knowledge and skills	Students have to pay for getting knowledge and skills
Profession choice	To choose an occupation due to his/her knowledge and skills	On the requests of parents, a few of them choose according to their knowledge and skills	To choose according to their own interests, knowledge, skills and social needs

Within the innovation of education reform, it is important to develop the interests and needs for the self assessment, the feeling of advance and lifelong development.

We recently carried out a survey of students on the specific questionnaire and interviewed in a focus group.

III. RESULTS

The results demonstrate are the followings.

When they defined their future goals, 70% of the students defined them in one or two years. This means that they are short-sighted. So they can't be successful, cause they don't know where do they get.

After assessing their 5-sided potentialities themselves, when we correlated their marks/grades, the correlation between their communication skills and their marks was more great.

And we came to the conclusion that this cannot demonstrate the students' 4 years investment.

① When we analysed on the correlation between intelligence, basic knowledge of general education and mental strength of a student, there was a straight correlation.

② When we carried out the intelligence survey of the students from 1st to 4th year on same questionnaire, the

survey suggested that their capacity to receive information increased, they noted everyday common news rather than receiving scandal news, the understanding of scientific terms improved and converting abilities into formula and graphic display language also increased.

③ Within the scope of study, we conducted training on the evaluation of ability to work in groups, the introduction of its significance, the integrated potentialities considered as social needs and the communication skill for the first time and every student made the starting assessment.

IV. CONCLUSIONS

It has been seen from the above experiment and survey that in the process of education innovation, making knowledge into skillful work creation is very important. Possessing knowledge and capability is an education which creates from people's usages and it is a work creation which can satisfy the consumer management. Thus, there is a tendency as application-training and research-work creation of current education.

REFERENCES

[1] The urgent issues of higher education reforms.

Brochure of research papers,2015.

- [2] The 2nd forum for well-educated women-scientists of Mongolia,2014.
- [3] L. Suvd, M. Byambasuren, Ch. Oyunjargal A Mongolian teacher of a new century,2015.
- [4] D. Badarch Reform innovation of higher education.
- [5] Von Dun Yon"5-sided principle" ,2007.

The Integrated Flexible Scheduling Algorithm of Complex Product based on the Shortest Processing Time Scheduling Strategy

Xie Zhi-qiang^{#1}, Xin Yu^{#2}

[#] School of Computer Science and Technology, Harbin University of Science and Technology NO. 52 XueFu Road, Nangang Dist, Harbin, China

¹xiezhiqiang@hrbust.edu.cn

²xinyhrb@qq.com

Abstract— Constraints in the problem of integration and flexible scheduling, complex products with no waiting between processes, which combines program, no-wait constraint, formerly a virtual node algorithm set against proposed. For standard flexible procedures, critical path method is to take the Allies to confirm scheduling command and adaptive flexible scheduling strategy is to take handler. For the flexible procedure, no-waiting constraints use the shortest processing time to make a schedule. For better flexible scheduling program structure, expansion flexibility flexible manufacturing and virtual tree trees also been proposed. The examples show that the algorithm can solve problems in meaningful and easy to implement.

Keywords— Flexible procedure, no-wait constraint conditions, expanded flexible processing tree

I . INSTROCIION

The integrated flexible scheduling problem(FSP) is to expand the scheduling optimization problem. In previous studies, researchers question as a certainty. But in reality, there are two problems: a flexible scheduler means can be arranged on different devices; and no-wait constraint means that one or several groups of programs need to be done continuously at the same time.

Compared with the traditional job shop scheduling, the integrated flexible scheduling problem of complex product with no-wait constraint between procedures is a NP-hard problem and it is more complicated[1], [2]. Because of the procedures have several available processing devices. And the equipment constraint is reduced, but the search scope of the best equipment is expanded. At the same time, the no-wait constraint between procedures is under consideration, the difficulty of solving problem is further increased [3].

So far, the research for FSP is mainly solve FSP of batch manufacturing[4] to [7] and FSP of small volume production for multi-varieties[8] to [10]. However, these algorithms did not take the constraint relationship between procedures into consideration.

Xie and his groups put forward a series of solutions to the integrated scheduling problem of complex products such as subsection idea [11] and layer priority algo-

rithm [12]. The scheduling problem of complex products with delay procedures was solved by converting the delay procedures into virtual procedures. And the integrated flexible scheduling problem of complex products was solved by using the method of selecting the shortest processing time [13], etc.

The literature [14] simplified the integrated flexible scheduling problem by choosing the equipment according to the shortest processing time, but the algorithm did not consider that the flexible scheduling has the characters of multiple equipment and multiple processing time with no-wait constraint between procedures. Therefore, it is necessary to study the integrated flexible scheduling problem of complex product with no-wait constraint between procedures.

II . THE MATHEMATICS MODEL OF FLEXIBLE SCHEDULING

In order to solve complex products processing and assembly, FSP processing equipment and assembly equipment is uniformly defined as a processing device, and therefore, will not be processed and assembled uniformly defined process. Mathematical model is then as follows.

The procedures number is n , and the equipment number is m . Procedure x on equipment y is the predecessor of procedure i on equipment k . S_{ik} is the starting time of procedure i processed on equipment k , and 8 is the finishing time. C_i is the earliest finishing times of procedure i on its equipment subset. The aim is make the processing time shortest.

$$T = \text{Min} \{ \max(C_i) \}, \quad i = 1, 2, \dots, n$$

$$S. t: \quad C_i = \min(E_{ik}) \quad (1)$$

$$\min(S_{ik}) \quad (2)$$

$$S_{ik} - E_{xy} \geq 0 \quad (3)$$

$$S_{ik} - E_{(i-1)k} \geq 0 \quad (4)$$

$$S_{ak} - E_{bk'} = 0 \quad (5)$$

Equation (1) denotes the earliest finishing time of procedure i on equipment subset. Equation (2) denotes the procedure is processed as early as possible. Equation (3) denotes procedure i must start processing after its

predecessor finished processing. Equation (4) denotes procedure i must be processed after procedure $i-1$ finished processing. Equation (5) denotes procedure a and b have the no-wait constraints.

In order to simplify the complex FSP products, we proposed the following definition.

Standard flexible procedure: it does not have no-wait constraint generally flexible procedures.

Flexible program with no-wait constraint: Flexible program completion time must be its predecessor start time; flexible program start time must be its successor is complete.

No waiting for a flexible team: an assembly made of a plurality of consecutive no-wait process constraints.

EFP-tree (expanded flexible processing tree): a tree shows the relationship of the flexible product with No-Wait procedure.

VFP-tree (virtual flexible processing tree): expanded flexible processing tree with processing time equipment set.

Real node: the flexible procedure node with no-wait constraint in the VFP-tree, of which the processing time can't be changed dynamically.

Forged node: a predecessor real node of the real node procedure in the VFP-tree, of which the processing time can't be changed dynamically.

Virtual node: In VFP tree, its processing time can be dynamically changed, because the device is busy set of flexible procedures without a no-wait constraint node.

Set of nodes: node through a real, many fake nodes and virtual nodes predecessor assembly thereof.

Predecessor procedures: Do not wait predecessor without constraint without waiting for a flexible team.

III. THE DESIGNING OF FLEXIBLE SCHEDULING ALGORITHM

There are two procedures in a complex product with FSP no-wait constraint between programs; it can not wait for the limiting process which may be carried out in the same device in different devices and flexible processes and standard flexible process to be processed.

A. The scheduling priority for procedures in No-Wait flexible workgroup

Without waiting for a flexible program, that may affect the scheduling team formerly known as the relevant procedures in the program of the Working Group. Such a program and its related procedures constitute a set of nodes based on VFP tree functions. This means that the set of nodes, including real-time nodes, nodes and virtual nodes forgery. There are four cases to be introduced as follows.

If the node set has more than one virtual node, determined by the Allied critical path method (ACPM) scheduling order, and adaptive flexible scheduling strategy scheduler virtual program. If the actual node does not satisfy the constraints, the process will be delayed at

the start time of the working group without waiting for the program until the restrictions are satisfied.

If the working group is not the only no-wait, which means that when there is VFP EFP-trees and tree several working groups need to determine scheduling command. According tree traversal process, if the program has a more realistic path to a node on the presence will be higher, so the first path, this arrangement can reduce the requirements in the path occupied by other device, then the path in the working group can early arrangements.

If the path has the same number of real nodes, schedule the workgroup with less virtual nodes firstly, which could reduce the effect of workgroup scheduling. It is because that the workgroups having less relative nodes will be affected by the relative procedures.

If the virtual node number is still equal, schedule the path with long processing time firstly according to the allied critical path strategy. The procedures on the path have obviously effect on the total processing time of product, processing them earlier, and the total processing time will be shorten.

After scheduling all workgroups, the allied critical path strategy and adaptive flexible scheduling strategy are adopted to schedule the remaining virtual nodes procedures.

B. Use the allied critical path strategy to determine the scheduling order of standard flexible procedures

Critical path is the maximum path from the node to the root node of the processing time of the process. The flexible scheduling problem, the process may be treated in different devices, and the processing time may be different. In order to make all programs as short as possible the total processing time to a minimum processing time of the device it is most likely to be selected. References [15] and [16] calculated for each program by the minimum processing time, which means that the initial processing time virtual node tree VFP confirmed the critical path, and then the program is determined by the ACPM virtual scheduling command.

If the path maximum processing time is greater than 1 in the group number of the device can affect the selection process of the device, in order to narrow the range of options, the first scheduling device number is programmed device settings in the minimum processing time minimum path.

C. Process the standard flexible procedures by using the adaptive flexible scheduling strategy

In order to make procedures finish earlier, begin processing time and end processing time of procedures need to be considered. The procedure can complete processing earlier if the begin time is earlier and the processing time is shorter.

If there is equipment in the virtual node equipment set which has one idle time to make the procedure begin

processing earliest, then it is selected. If there exist several selectable equipment, select the proper one by equipment balanced strategy. In this way, the procedure can be processed at the earliest time and the processing time is shortest, so the end processing time of the procedure is earliest.

If the earliest start time of the processing procedure of the device is occupied, the start processing time may be delayed, so the end of the processing time of the program will be late in the process on the device having a long processing time. In this case, the device settings and VFP tree longer processing time will be selected, and put it into the correct virtual node. The search appliance has a longer processing time, but the end of the processing time earlier, until at the earliest time possible program termination Discovery.

If the first end of the processing time of the equipment is not the only choose a short processing time, to reduce the elapsed time of the device. If the short processing time the device is not the only device in accordance with the appropriate balance a strategic choice.

If the program can not set the device from all devices at the earliest time to begin processing, the program scheduling optimization strategy for each device set of completion of the comparison between a predetermined time. Select the device can make the process at the earliest time the process is terminated, if the proper equipment is not the only choose a short processing time.

Adaptive and flexible scheduling policies take full advantage of these features. There are many practical devices and choose from a variety of time and equipment is flexible. According to part of this strategy, the device is sequentially selected by the processing time of the short length of the program. The final processing time may be less than earlier end processing time from the processing equipment in the shortest time on the device. So the purpose of short complex product flexible scheduling total working time can be achieved.

D. Use the equipment balanced strategy to select the same equipment

Since there are multiple equipment with the same begin processing time and end processing time, the method that calculate the total working hours of all scheduled procedures on each optional equipment separately is adopted. Then the equipment with minimal total working hours is selected. This strategy takes full account of the processing load for equipment, make every equipment using balanced. Thereby the parallel working hours of whole processing systems can be enhanced.

IV. Scheduling algorithm

Step 1: Build the EFP-tree and VFP-tree according to the constraint conditions, and create the corresponding lists named ListO for EFP-tree and Listo for VFP-tree.

Step 2: Calculate the number of real nodes in each path of VFP-tree, and create its list which only include real nodes according to the real node number from more

to less (if there are several paths with the same number of real nodes, select the path with less virtual nodes procedure, if still same, select the path by ACPM), and then assign the link list to $List1... ListR (r < n)$.

Step 3: Select the minimum of $Listx (x \text{ belong to } R)$.

Step 4: Traverse the path from began to end, search for the first no-scheduled real node which is closest to leaf nodes.

Step 5: If there is no no-scheduled forged node in the set of this real node, turn to Step 6, otherwise, turn to Step 10.

Step 6: If there is no no-scheduled virtual node in the set of this real node, turn to Step 7, otherwise, turn to Step 11.

Step 7: If the real node can be scheduled in the equipment set to begin processing at the earliest time, then select the equipment by equipment balanced strategy and schedule the procedure, turn to Step 8, otherwise, search for the equipment in this set which can make the procedure begin processing as early as possible, and then schedule the procedure by equipment balanced strategy, turn to Step 8.

Step 8: If the real node satisfy the constraint in the No-Wait workgroup, turn to Step 9, otherwise, turn to Step 17.

Step 9: If subsequent real node exists in the real node list, select the subsequent node, and turn to Step 5, otherwise, turn to Step 18.

Step 10: Find the path which has the largest number of forged nodes, and select it as scheduling path, turn to Step 4.

Step 11: Confirm the scheduling order of virtual nodes according to using the ACPM, and select the virtual node which needs to be processed firstly, turn to Step 13.

Step 12: Select the next virtual node to be scheduled according to ACPM.

Step 13: If the virtual node could be processed at the earliest begin processing time, then process it, turn to Step 15, otherwise, turn to Step 14.

Step 14: Select the equipment with earliest end processing time according to the adaptive flexible scheduling strategy, then process the procedure on it, turn to Step 15.

Step 15: If there are no-scheduling virtual nodes in the set, turn to Step 12, otherwise, turn to Step 16.

Step 16: schedule the real nodes in the set, turn to Step 8.

Step 17: Judge whether the begin processing time of the node could be delayed, if so, put off processing the procedure, otherwise, all of the procedures in this no-wait workgroup are processed in the end, turn to Step 9.

Step 18: If there are no-scheduling paths in $Listx$, select the path with the minimum x , turn to Step 4; if not, the rest of standard flexible procedures are scheduled according to ACPM and adaptive flexible scheduling strategy, turn to Step 19.

Step 19: The end.

V. Conclusions

Comprehensive and flexible scheduling algorithm complex product without waiting for the proposed linkage constrained scheduling method used, in line with no-wait constraint between programs. Because it takes full advantage of the process is more flexible scheduling of equipment selection function, the program is scheduled to achieve an early end of the processing time of the device.

As a result, the algorithm provides a reference to resolve particularly flexible scheduling problem, which has certain theoretical and practical value of the synthesis problem.

ACKNOWLEDGMENT

This project is supported by National Natural Science Foundation of China (No. 61370086), Science and Technology Research Foundation of Heilongjiang Province Education Department (No. 12531105), Heilongjiang Province Postdoctoral Science Foundation (LBH-Q13092), Heilongjiang Province Postdoctoral Science Foundation (LBH-Z15096), and National University of Computer Education Research Association (No. ER2014018).

REFERENCES

- [1] Samarati P, Sweeney L. *Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression*. In: Proceedings of the 1998 IEEE Symposium on Research in Security and Privacy. Paloalto, CA: IEEE, 1998. 1-19.
- [2] Q. C. Huang, Q. Chen. *A Fast Job Oriented Scheduling Algorithm*. Chinese Journal of Software, vol. 10, n. 10, pp. 1073-1077, 1999.
- [3] K. Jansen, M. Mastrolilli, R. Solis Oba. *Approximation Algorithms for Flexible Job Shop Problems*. International Journal of Foundations of Computer Science, vol. 16, n. 2, pp. 361-379, 2005.
- [4] Abul O, Bonchi F, Nanni M. *Never walk alone: uncertainty for anonymity in moving objects databases*. In: Proceedings of the 24th International Conference on Data Engineering. Cancun; IEEE, 2008. 376-385
- [5] Yang Y. Z. *Cultural-based genetic tabu algorithm for multi-objective job shop scheduling*. Mathematical Problems in Engineering, Volume 2014, Article ID 230719, 14 pages, 2014.
- [6] Yu-Yan Hana, Dun-wei Gong, Xiao-Yan Suna & Quan-Ke Panb. *An improved NSGA-II algorithm for multi-objective lot-streaming flow shop scheduling problem*. International Journal of Production Research, vol. 52, n. 8, pp. 2211-2231, 2014.
- [7] Huo Z, Huang Y, Meng X F. *History trajectory privacy preserving through graph partition*. In: Proceedings of the 1st International Workshop on Mobile Location-based Service. Beijing, China: ACM, 2011. 71-78.
- [8] Shao, XY; Liu, WQ; Liu, Q; Zhang, CY. *Hybrid discrete particle swarm optimization for multi-objective flexible job-shop scheduling problem*. International Journal of Advanced Manufacturing Technology, vol. 67, n. 9-12, pp. 2885-2901, 2013.
- [9] Gao S, Ma J F, Sun C, Li X H. *Balancing trajectory privacy and data utility using a personalized anonymization model*. Journal of Network and Computer Applications, 2014, 38(1):125-134.
- [10] D. X. Chen, H. X. Wang. *Reverse Algorithm for Solving Nonstandard Job-shop Scheduling Problem Based on Redundancy*. Computer Integrated Manufacturing Systems, vol. 10, n. 10, pp. 1238-1241, 2004.
- [11] Z. Q. Xie, G. J. Ye, D. L. Zhang, G. Y. Tan. *Study of a New Nonstandard Job Shop Scheduling Algorithm*. Chinese Journal of Mechanical Engineering, vol. 21, n. 4, pp. 97-100, 2008.
- [12] Z. Q. Xie, J. Yang, G. Yang, G. Y. Tan. *Dynamic Job-Shop Scheduling Algorithm with Dynamic Set of Operation Having Priority*. Chinese Journal of Computers, vol. 31, n. 3, pp. 502-508, 2008.
- [13] Z. Q. Xie, T. Mo, G. Y. Tan. *Dynamic Job-shop Scheduling Algorithm of the Non-close-joining Operations*. Chinese Journal of Mechanical Engineering, vol. 44, n. 1, pp. 155-160, 2008.
- [14] Z. Q. Xie, S. Z. Hao, Guangjie Ye, Guangyu Tan. *A new Algorithm for Flexible Job-shop scheduling with Constraints between Jobs*. Computers & Industrial Engineering, vol. 57, n. 3, pp. 766-772, 2009.
- [15] Z. Q. Xie, S. H. Liu, P. L. Qiao. *Dynamic Job-shop Scheduling Algorithm Based on ACPM and BFSM*. Journal of Computer Research and Development, vol. 40, n. 17, pp. 977-983, 2003.

FPGA-based Cluster Platform for Reconfigurable Distributed Embedded Systems

Michael Nagler and^{#1} Stephan Blokzyl^{#2} Wolfram Hardt^{#3}

*Technische Universit at Chemnitz Faculty of Computer Science
Professorship Computer Engineering D-09107 Chemnitz, Germany
{nagm,sblo,hardt}@informatik.tu-chemnitz.de*

Abstract— The functional requirements of embedded systems in automotive engineering, aircraft construction, and space industry are continuously rising. Nowadays, the individual tasks of these systems are very complex, so they have to be separated into different subsystems. Also the field of application puts high requirements to these systems with respect to reliability and real-time capabilities. To provide a high performance system, that satisfies conservative cost restrictions, this paper proposes a cluster platform for distributed embedded systems. The communication infrastructure of the cluster is based on a recon. gurable bus architecture for on-chip communication and Ethernet for interchip communication. The proposed solution shows, that a cluster infrastructure can also be realised on cost-optimised FPGAs with less logic resources.

I. INTRODUCTION

The functional requirements of embedded systems in automotive engineering, aircraft construction, and space industry are continuously rising. Nowadays, the individual tasks of these systems are very complex, so they have to be separated into different subsystems[9]. Also the growing degree of automation forces the system designers to build up increasingly powerful systems with a high reliability and real-time capabilities. Next to this high performance, these systems, especially in the fields of autonomous driving and unmanned aerial vehicles, have to satisfy conservative cost restrictions like size, weight, energy consumption, and price[2].

An important aspect in the mentioned field of application is the localisation of the vehicle in the real world and the estimation of its surroundings. For example an autonomous car on a multi-lane motorway needs information about the lane it is driving on and about the other cars around it. An unmanned aerial vehicle during the landing approach needs information about its relative position to the runway and obstacles in front of it. A variety of sensors like ultrasonic sensors, laser sensors, and electro optical sensors are used in today's applications to gain knowledge about the environment. Especially the latter ones provide a huge amount of data, that has to be processed in a short amount of time.

Programmable hardware modules like Field Program-

mable Gate Arrays (FPGA), as they provide a good trade-off between the speed of circuit level specialised ASICs and the programming flexibility of general purpose processors [6], are commonly used to accelerate such image processing applications. In [2] an FPGA-based runway detection based on colour images is proposed. Currently the working software prototype is mapped to the hardware platform presented in [1]. The image processing algorithms shall be realised on one of the biggest FPGAs from the *Xilinx Virtex-6 DSP* family. Although this chip is very powerful and provides a lot of logical resources, it is on the one hand very expensive and on the other hand it can be a single point of failure. To counter this problem, a cluster out of FPGAs can increase the reliability of such a system. The proposed data and task parallelisation concept in [2] will also fit to such distributed architectures. With the use of several FPGAs, smaller and more cost-optimised chips can be used.

In this paper an FPGA-based cluster platform for reconfigurable distributed embedded systems is presented. The nodes of the cluster will be realised with cost-optimised *Xilinx Spartan-6* FPGAs. In section II other FPGA-based cluster systems and systems used for real-time image processing are discussed. Section III presents the system setup. Section IV focuses on the communication infrastructure. In section V the results in the form of occupied resources are presented. Finally, section VI draws a conclusion and gives an outlook about the future work.

II. STATE OF THE ART

FPGA-based cluster systems are commonly used for high performance applications. There are several architectures in the academic field like the *Novo-G* supercomputer, developed at the *US National Science Foundation's Center for High-Performance Reconfigurable Computing at the University of Florida* [4], or commercial solutions like the *DNBF_I2_I2 Cluster*, developed by the *DINI Group* [5]. As both are used for high performance computing, they will not fit to the field of application that is considered in this work. Nevertheless, these supercomputing architectures form the base for hardware acceleration architectures in the embedded field.

* We gratefully acknowledge the cooperation of our project partners and the financial support of the DFG (Deutsche Forschungsgemeinschaft) within the Federal Cluster of Excellence EXC 1075 "MERGE".

One of this architectures is the *CUBE FPGA Cluster*, that consists of up to 512 *Xilinx Spartan-3* FPGAs [6]. It provides a cost effective cluster framework to solve research tasks and real-world applications. Its symmetric design and fixed communication infrastructure ease the application design, but make it not flexible enough for the applications that are considered in this work.

Other architectures, like the one described in [1], realise the complete application on a huge single chip with a lot of logic resources. A failure in this chip will directly result in a failure of the complete system. In terms of reliability such single point of failures should be avoided. This can be done, using several redundant processing units [8]. In the case of a failure, another chip will continue the execution of the application. On the one hand such a setup will be a waste of resources as the second chip is only needed, if there is a failure in the first. On the other hand it will have a negative impact on the strict cost restrictions of such embedded systems in terms of size, weight, and price.

Next to these FPGA-based solutions, a lot of image processing application are realised with GPUs. Parallel computing architectures like the *CUDA* framework from the *NVIDIA Corporation* [3] combine the flexibility of software programming with the acceleration capabilities of parallel execution on hardware. The only drawback of these architectures is the still high energy consumption such systems have compared to FPGA-based solutions. To combine these two worlds, there is also research going on to compile *CUDA* kernels onto FPGAs [7].

III. SYSTEM SETUP

Since all presented solutions have their pros and cons, as they are specialised to a dedicated problem, this paper proposes an architecture, that is placed between the fixed architecture of the *CUBE FPGA Cluster* and the single chip solution from [1]. The evaluation platform is realised on *Digilent Nexys 3* boards with a *Xilinx Spartan-6 XC6LX16-CS324* FPGA on them. The communication between the cluster nodes is realised over Ethernet. The topology of the cluster is flexible and can be adapted to the dedicated application.

The different tasks of the application are realised as modules on the FPGA. Each node in the cluster provides several slots, where these modules can be placed. Each slot has the same amount of logical resources, so the modules can be distributed arbitrary on the chip. The amount of slots per FPGA is also flexible to fit the task granularity of the application. Although all slots have the same size, the different modules can exceed this. In this case the module will occupy several slots. Nevertheless, it is not possible to split a slot between two modules. Fig. 1 illustrates this exemplarily for an FPGA with 9 slots and 5 tasks. Task A needs 3 slots, Task B and Task E two, and Task C and Task D one. As depicted, the slots of such bigger tasks have to be next to each other. This requires a mapping strategy to find a good distribution of the modules to the slots. This is an independent

topic and not considered in this work.

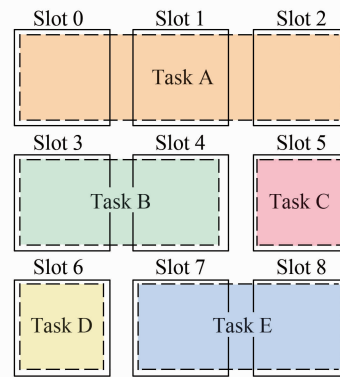


Fig. 1 Exemplary mapping of tasks to the module slots

IV. COMMUNICATION INFRASTRUCTURE

The communication in the cluster is divided into two different parts: the on-chip communication and the inter-chip communication. The first one is realised by a specially developed bus architecture, the *Reconf Bus (RCB)*. The communication between the cluster nodes is realised with a bus bridge, that connects two RCB instances on different FPGAs with each other. The data between two FPGAs is transmitted via Ethernet.

A. On-Chip Communication: *Reconf Bus*

The RCB is a bus architecture specially developed for distributed FPGA applications. Next to the support for partial reconfiguration by the possibility to disable and reset single modules, it also considers the resource restrictions of FPGAs. As several modules are placed on a single FPGA, the resources occupied by the bus, have to be minimised. To achieve this, the bus can be parameterised in terms of data width, the maximum number of connectable modules, and the number of possible parallel transmissions.

The requirement for real-time capabilities leads to important characteristics of the bus: The communication is packet oriented, so the sender does not have to wait for the establishment of the communication channel. This is important especially for inter-chip communication to avoid the blocking of the complete communication line. Also the maximum packet length and waiting time are limited, so the bus will not block in case of an error inside a module. Each module has a priority value assigned to it. When two modules want to send a package at the same time, conflicts can be resolved and no module specific behaviour has an impact on this. To avoid a permanent blocking of a module with a low priority, the values are dynamic. Each time a module successfully starts a transmission, it gets the lowest priority on the bus, while all other modules increase their priority value.

Fig. 2 shows a schematic representation of the bus architecture. It consists of a *Port Connector*, one or more

Transfer Lanes, a *Switch Matrix*, and the central *Bus Controller*. The modules are connected to the *Port Connector*, that recognises request to the bus and forwards them to the central controller. It also contains the logic to disable and reset the modules. The *Transfer Lanes* handle the package transmission between two connected modules. If more than one lane is available, accordingly more packets can be transmitted in parallel. The *Switch Matrix* can establish arbitrary connections between the

modules and the *Transfer Lanes*. During a transmission a *Transfer Lane* is connected to exactly two modules, the sender and the receiver. The central *Bus Controller* manages the requests of the *Port Connector* and the *Transfer Lanes*, resolves the conflicts based on the priorities, and provides the control signals for all bus components. The logic for disabling and resetting the modules can be accessed via a *Control Interface*.

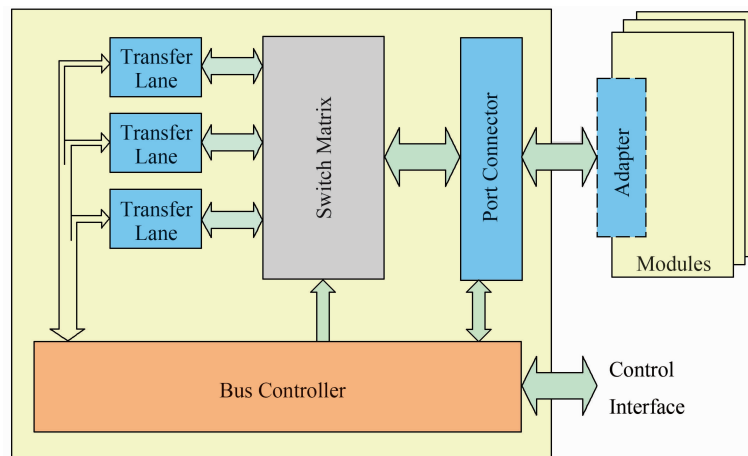


Fig. 2 Architecture of the Reconf Bus

The depicted *Adapter* is not directly a component of the bus. It is used to connect the modules with the bus. An easy interface supports the module implementation, as the complex RCB protocol is realised within the *Adapter*. Furthermore it has routing capabilities to enable a routing of the packets between the modules. To realise this, it provides internal ports to the module for incoming and outgoing packets. Depending on which port a packet is received, the module can handle the data differently. The outgoing ports are used to address an entry in a routing table inside the *Adapter*. Depending on the module's needs, the *Adapter* can be parameterised by the number of incoming and outgoing ports and the data width of the routing table. The routing information can be configured by sending a packet to the module, so there has to be no memory interface to configure the adapters.

B. Inter-Chip Communication :RCB-Ethernet-Bridge

The *RCB-Ethernet-Bridge* is realised as a module for the RCB and can be directly connected to its module ports. For the Ethernet communication the *Tri-Mode Ethernet MAC* IP core from XILINX is used. As the Nexys 3 boards only have one Ethernet port, this port has to be used to connect it to all neighbouring nodes. Commonly such infrastructures are realised with high-speed serial interfaces like the *GTP Transceivers* [11], that are also available on some *Spartan-6* FPGAs. The bridge is designed to model these point-to-point-communications, so the transmission medium (Ethernet) can be changed with little effort.

The structure of a *RCB-Ethernet-Bridge* module is de-

picted in Fig. 3. To model the point-to-point-communications, the module has several connections to the RCB (*Adapters*). Each adapter realises a connection to a different neighbouring node. Received data and data to transmit is buffered in First-In-First-Out (FIFO) memories. To synchronise between the 100 MHz clock frequency of the RCB and the 25 MHz of the Ethernet transceiver, these FIFOs provide an independent read and write clock. Separate Finite-State-Machines (FSM) control the sending and receiving of data packets over the Ethernet core. This core provides an easy to implement LocalLink [10] interface. As the IP cores for the *GTP Transceivers* provide a similar interface, they can be exchanged easily by adapting the FSMs.

Data, that is received over the RCB, is stored in the TX FIFO. The Ethernet packet is sent using the User Datagram Protocol (UDP) to minimise the communication overhead, as all adapters have to share one port. The destination IP and MAC address as well as the destination port in the UDP frame can be configured for each adapter separately. The destination UDP port is used to select the adapter on the receiving side. The Ethernet packets are generated by the TX FSM. It checks with a round robin scheduling strategy the transmission memories of all adapters to treat the different logical connections equally.

Data, that is received over the Ethernet interface, is processed by the RX FSM. The destination port of the UDP frame is used to select the adapter to which the data is forwarded. The adapter is then responsible to send the data, using the configured routing information, over the RCB.

tion with the available adapter simplify the integration of new modules. It could also be shown, that the needed resources are very low, so the proposed cluster platform can be realised even on very small FPGAs.

In the next step, the communication overhead shall be considered. An easy image processing, using the example of a JPEG compression, will be realised with the cluster platform. The results will then be compared with a single FPGA solution on a *XILINX Virtex-6* FPGA.

REFERENCES

- [1] S. Blokzyl and W. Hardt. Generisches Framework zur Parallelisierung von Echtzeitfähiger Bilddatenauswertung für Rekon. gurierbare Integrierte Schaltkreise. In *Studierendensymposium Informatik 2016 der TU Chemnitz*, pages 145-148, Apr 2016.
- [2] S. Blokzyl, M. Vodel, and W. Hardt. Fpga-based approach for run-way boundary detection in high-resolution colour images. In *Sensors Applications Symposium (SAS), 2014 IEEE*, pages 59-64, Feb 2014.
- [3] NVIDIA Corporation. CUDA Parallel Computing — What is CUDA? <http://www.nvidia.co.uk/object/cuda-parallel-computing-uk.html>. Accessed: 2016-05-04.
- [4] A. George, H. Lam, and G. Stitt. Novo-g: At the forefront of scalable recongurable supercomputing. *Computing in Science Engineering*, 13 (1):82-86, Jan 2011.
- [5] DINI Group. Product Brief: DNBFC S12 12 Cluster. http://www.dinigroup.com/product/data/DNBFC_S12_PCIe/.les/DNBFC_S12_12_Cluster_v10_hi.pdf, Jun 2010.
- [6] O. Mencer, Kuen Hung Tsoi, S. Cramer, T. Todman, W. Luk, Ming Yee Wong, and Philip Leong. Cube: A 512-FPGA cluster. In *5th Southern Conference on Programmable Logic, 2009. SPL*, pages 51-57, April 2009.
- [7] A. Papakonstantinou, K. Gururaj, J. A. Stratton, D. Chen, J. Cong, and W. M. W. Hwu. Fcuda: Enabling efficient compilation of cuda kernels onto fpgas. In *Application Specific Processors, 2009. SASP '09. IEEE 7th Symposium on*, pages 35-42, July 2009.
- [8] Thilo Streichert, Dirk Koch, Christian Haubelt, and Jürgen Teich. Modeling and design of fault-tolerant and self-adaptive recon. gurable networked embedded systems. *EURASIP J. Embedded Syst.*, 2006 (1):9-9, January 2006.
- [9] C. B. Watkins and R. Walter. Transitioning from federated avionics architectures to integrated modular avionics. In *2007 IEEE/AIAA 26th Digital Avionics Systems Conference*, pages 2. A. 1-1-2. A. 1-10, Oct 2007.
- [10] XILINX. LocalLink Interface Specification. http://www.xilinx.com/aurora/aurora_member/sp006.pdf, Jul 2005.
- [11] XILINX. Spartan-6 FPGA GTP Transceivers-Advance Product Specification (UG386 v2.2). http://www.xilinx.com/support/documentation/user_guides/ug386.pdf, Apr 2010.